

Leonie Wiemeyer

# **Intertextuality in foreign-language academic writing in English**

A mixed-methods study of university students'  
writing products and processes in source-based  
disciplinary assignments

A doctoral thesis submitted for the degree of Doctor of Philosophy (Dr. phil.) in English linguistics  
at the University of Bremen, 9 October 2020

under the supervision of

Prof. Dr. Marcus Callies, University of Bremen, Germany

Prof. Casey Iezzi, Ph.D., Boise State University, USA

Third evaluator: Prof. Diane Pecorari, Ph.D., City University of Hong Kong, China

Date of Defence: 26 August 2021



# Contents

Abstract	iii
Zusammenfassung auf Deutsch	iv
Danksagung – Acknowledgements	vii
<b>1 Introduction</b>	<b>1</b>
<b>2 Intertextuality in academic writing</b>	<b>7</b>
2.1 Forms and functions of citations	10
2.2 Types of intertextual strategies	14
2.2.1 Direct quotation	14
2.2.2 Paraphrasing and summarisation	17
2.2.3 Generalisation	20
2.3 Signalling intertextuality: attribution, documentation, reporting structures, and evaluation	23
2.3.1 Attribution, averral, and documentation: Distinguishing voices in the text	24
2.3.2 Reporting structures and reporting verbs	26
2.3.3 Evaluation and writer’s stance	28
<b>3 Intertextuality in L2 academic writing</b>	<b>31</b>
3.1 The complex issue of apparent plagiarism: Explaining transgressive intertextuality	32
3.1.1 Intertextuality, cultural background, and language proficiency	33
3.1.2 Intertextuality and academic writing expertise	36
3.2 Beyond plagiarism: Unintentional textual borrowing and learning processes	37
3.3 Manifestations of intertextuality in L2 academic writing	38
3.3.1 Direct quotation in L2 student writing	38
3.3.2 Paraphrasing and summarisation in L2 student writing	42
3.3.3 Attribution, reporting structures, and documentation in L2 student writing	45
3.4 Towards proficiency-oriented approaches: The source text as a language resource in L2 student writing	49
<b>4 Writing processes of source-based academic writing</b>	<b>52</b>
4.1 Models of L1 and L2 writing processes	52
4.2 The influence of source use on the writing processes	54
4.3 Process-oriented perspectives on source-based L2 student writing	56
<b>5 A mixed-methods study of intertextual strategies in L2 academic writing</b>	<b>64</b>
5.1 Rationale for this study	64
5.2 Research questions	66
5.3 The text type ‘reading report’	68
5.4 Conception of the study design	70
<b>6 Corpus study</b>	<b>72</b>
6.1 The corpus	72
6.2 Data annotation and analysis	74
6.3 Results and discussion	82
6.3.1 Strategies of source text material selection and structuring	83

6.3.2	Intertextual strategies	87
6.3.2.1	Paraphrasing	87
6.3.2.2	Summarisation	97
6.3.2.3	Direct quotation	98
6.3.2.4	Meta-level observations	102
6.3.2.5	New content introduced by the students	104
6.3.3	Documentation, attribution, and reporting structures	105
6.3.4	Textual borrowing, patchwriting, and strategic re-use of source text vocabulary	110
6.3.4.1	Copying from the source text and patchwriting	110
6.3.4.2	The source text as a language resource	115
6.4	Interim summary: Corpus study	116
<b>7</b>	<b>Process study</b>	<b>119</b>
7.1	Study design and methodological decisions	119
7.2	Pilot study	124
7.3	Study participants	128
7.4	Data collection procedure	129
7.4.1	Screen recording	129
7.4.2	Retrospective interviews with stimulated recall	131
7.5	Coding and analysis of process data	133
7.6	Transcription, coding, and analysis of stimulated recall and interview data	144
7.7	Results and discussion	145
7.7.1	The pre-writing stage	147
7.7.2	The writing stage	149
7.7.2.1	Linearity of the writing stage	149
7.7.2.2	Activities constituting the writing stage	151
7.7.2.3	Writing acts and non-writing acts in the writing stage	152
7.7.2.4	Intertextual writing processes	158
7.7.3	The post-writing stage	176
7.8	Interim summary: Process study	177
<b>8</b>	<b>Discussion</b>	<b>180</b>
8.1	Implications	191
8.1.1	Benefits of mixing methods in research into L2 source-based writing	192
8.1.2	Pedagogical implications	193
8.2	Limitations	199
8.3	Directions for future research	200
<b>9</b>	<b>Conclusion and outlook</b>	<b>203</b>
	<b>References</b>	<b>207</b>
	<b>List of figures</b>	<b>217</b>
	<b>List of tables</b>	<b>218</b>
	<b>Appendices</b>	<b>219</b>

# Abstract

This dissertation explores intertextuality in the writing processes and products of English-language academic writing. Using a mixed-methods approach, it investigates how German university students use the source text when writing in their foreign language (L2) English. Intertextual strategies are analysed in a corpus of reading reports, i.e. disciplinary assignments based on a single source text. Furthermore, six L2 writers' processes of writing a reading report are studied using screen recordings and stimulated recall data. Intertextual links such as paraphrases as well as strategies of attribution, referencing, reporting, and (strategic) textual borrowing are analysed in order to shed light on how students create and combine intertextual manifestations in L2 academic writing. By combining a corpus and a process study in an innovative study design, this dissertation contributes to our understanding of how novice writers create and conceptualise intertextuality. The study adopts a positive perspective by focusing on proficient source use. It examines students' strategies and identifies patterns of intertextuality in the writing process and the product. Its results bear important implications for foreign-language writing research and the teaching of source use in academic contexts.

**Keywords:** second language writing; academic writing; intertextuality; writing process; source use; paraphrasing; learner corpus research; mixed methods; stimulated recall; foreign language writing; source-based writing; integrated writing; learner writing; novice writers; corpus linguistics; English linguistics; writing research;

# Zusammenfassung auf Deutsch

Intertextualität ist ein charakteristisches Merkmal des akademischen Registers. Sie beschreibt die Bezugnahme auf die Literatur des Forschungsgebiets, auf die Ideen, Theorien und Forschungsergebnisse anderer Autorinnen und Autoren. Verweise auf die zitierte Literatur sind in wissenschaftlichen Texten omnipräsent. In allen akademischen Disziplinen integrieren Autor\*innen Wörter, Ideen und Konzepte aus Quellen auf vielfältige Weise in ihre Texte. Dabei entstehen durch die Verknüpfung verschiedener Wissensstränge neue Bedeutungen und Argumente. Die formal korrekte und sorgfältige Kennzeichnung der Zitation ist ein zentraler Bestandteil guter wissenschaftlicher Praxis und eine grundlegende Anforderung an wissenschaftliche Publikationen. Die Missachtung wissenschaftlicher Konventionen kann zu Plagiatsvorwürfen führen und sich erheblich auf den wissenschaftlichen Erfolg einer Wissenschaftlerin oder eines Wissenschaftlers auswirken.

Wissenschaftliche Autor\*innen bedienen sich unter anderem direkter Zitate, Paraphrasen und zusammenfassender Sätzen (*summaries*), um Informationen aus Quelltexten direkt oder indirekt in den eigenen Text einzuarbeiten. Diese intertextuellen Strategien beruhen oft auf der strategischen Entlehnung von Textmaterial aus dem Ausgangstext, dem sogenannten *textual borrowing*. Autor\*innen können darüber hinaus aus einer Vielzahl von Mitteln wählen, um die Intertextualität im Text sichtbar zu machen, beispielsweise indem sie ihre Quellen durch Quellenangaben im Text und in der Bibliographie dokumentieren, den Ursprung des Gesagten im Text erwähnen und Zitatverben sowie indirekte Rede verwenden, um zitierte und paraphrasierte Passagen als solche zu kennzeichnen. Auf diese Weise machen Autor\*innen nicht nur deutlich, inwieweit sich ihr Text auf relevante wissenschaftliche Literatur stützt, sondern werden zugleich wissenschaftlichen Standards in Bezug auf die Verwendung und Kennzeichnung von Quellen gerecht. Die Formen der Kennzeichnung von Quellen sind in hohem Maße abhängig von der jeweiligen wissenschaftlichen (Teil-)Disziplin. Die Sichtbarmachung von Quellen unterliegt dennoch nicht nur zugrunde liegenden wissenschaftlichen Konventionen, sondern oft auch individuellen Vorlieben.

Wissenschaftliches Schreiben ist sowohl sprachlich als auch rhetorisch anspruchsvoll und erfordert umfangreiche Lese- und Schreibkompetenzen (Grabe & Zhang 2013; Samraj 2013). Dies gilt im besonderen Maße für das Schreiben in einer Fremdsprache, vor allem wenn die Lernenden sich beim Schreiben auf Quelltexte beziehen müssen (Abasi & Akbari 2008, C. Thompson, Morton & Storch 2013; Davis 2013). Dazu kommt, dass nicht jede Art von Intertextualität positiv bewertet wird und die unangemessene Übernahme von Passagen aus Quellen zu Plagiatsvorwürfen führen kann. In der Schreibforschung werden die Probleme von Studierenden beim Schreiben in der Fremdsprache in der Vergangenheit häufig im Kontext von Plagiarismus diskutiert (siehe Überblick in Pecorari & Petrić 2014). Obwohl Studierende unabhängig von ihrer Muttersprache ähnlichen Herausforderungen beim quellenbasierten Schreiben begegnen, beispielsweise bei der Integration von Quelltextmaterial in ihre Texte und deren angemessene Zitierung, stützen sich Autor\*innen in der Fremdsprache stärker auf Quelltextmaterial und neigen zu sogenanntem *Patchwriting*, bedienen sich also ausgiebiger am Quelltext (z.B. Campbell 1990; Keck 2006; Shi 2004). Studien zeigen eine Vielzahl an Schwierigkeiten beim Schreiben von fremdsprachlichen wissenschaftlichen Texten auf, u.a. sich in akademischer Sprache auszudrücken, mangelnde Quellenbelege und -verweise, ein begrenztes Repertoire an Zitatverben, unangemessene direkte Kritik, kopierte Passagen ohne Anführungszeichen, umfangreiche direkte Zitate ohne offensichtliche Funktion, fehlende Kennzeichnung von Paraphrasen und unzureichende Abgrenzung eigener Ideen von denen der zitierten Autor\*innen (Borg 2000; Cheng 2006; Dong 1996; Groom 2000; Petrić 2012; Verheijen 2015).

In den letzten Jahren sind Schreibforscher\*innen verstärkt der Frage nachgegangen, warum Schreibende in der L2 vermeintlich plagiierten und ob der sprachliche Hintergrund ausschlaggebend für die unangemessene Verwendung von Quellen ist. Entsprechende Studien zeigen auf, dass problematische Quellenverwendung oft weder auf kulturelle Unterschiede noch auf Täuschungsabsichten oder Faulheit zurückzuführen ist, sondern vielmehr im mangelnden Wissen der Studierenden zu wissenschaftlichen Schreibkonventionen begründet liegt (z. B. Pecorari 2006). Aufgrund ihrer geringen Erfahrungen mit wissenschaftlichen Texten verfügen diese häufig über ein noch wenig ausgeprägtes wissenschaftliches Vokabular und müssen ihre intertextuellen Kompetenzen erst noch aufbauen. Diese Erkenntnisse

fürten in der Folge zu einer Neuausrichtung der Forschung im Bereich des quellengestützten Schreibens, die betrügerische von entwicklungsbedingten Formen unangemessener Intertextualität abgrenzt (vgl. Chandrasoma, Thompson & Pennycook 2004). Entsprechend geht man in der L2-Schreibforschung heute davon aus, dass von den Konventionen abweichende Quellenverwendung zumeist ein Ausdruck fehlender Schreiberfahrung ist. Patchwriting dient zudem als Lernstrategie beim Erwerb wissenschaftlicher Lese- und Schreibfertigkeit und ist folglich keine Form des Plagiarismus im engeren Sinne (vgl. Campbell 1990; Currie 1998; Keck 2006; Flowerdew & Li 2007). In der jüngeren Forschung wurden daher unter anderem die Strategien von fremdsprachlich Schreibenden bei der Quellenverwendung sowie die Herangehensweise erfolgreicher Studierender untersucht (z.B. Ädel & Garretson, 2006; Harwood & Petrić 2012; Hirvela & Du 2013; Li 2013; Petrić 2007, 2012; Petrić & Harwood 2013). In diesem kompetenzorientierten Diskurs zum quellengestützten fremdsprachlichen Schreiben auf Englisch ist die hier vorgestellte Studie angesiedelt.

Die vorliegende Arbeit vereint eine Korpus- mit einer Prozessstudie zur Untersuchung von Intertextualität im wissenschaftlichen Schreiben in der Fremdsprache Englisch. In den beiden Teilstudien werden sowohl die Formen der Integration von Wörtern und Ideen aus dem Quelltext im Schreibprodukt untersucht als auch die Strategien der Studierenden bei der Strukturierung ihres Schreibprozesses und ihres Textes unter Verwendung des Quelltextes betrachtet. Die in der Korpusstudie analysierten Texte, sogenannte Reading Reports<sup>1</sup>, wurden von deutschsprachigen Studierenden in ihrer Fremdsprache Englisch im Rahmen von sprachwissenschaftlichen Einführungskursen verfasst. Für die Prozessstudie verfassten Studierende Reading Reports in einem experimentellen Setting. Dabei wurde der Schreibprozess mithilfe einer Bildschirmaufnahmesoftware im Hintergrund aufgezeichnet. Anschließend wurden die Studierenden – ebenfalls im 2. Semester desselben Bachelorstudiengangs – im Rahmen einer Stimulated-Recall-Prozedur zu ihrem Schreibprozess befragt, wobei die Aufzeichnung des Schreibprozesses als Stimulus diente.

Die Mixed-Methods-Studie geht auf das Desiderat ein, sowohl den Schreibprozess als auch das Produkt quellenbasierter Schreibaufgaben zu berücksichtigen, um fundierte Empfehlungen für Lehrer\*innen zu entwickeln. Die Studie trägt der Vielschichtigkeit der Intertextualität Rechnung, indem sie eine Vielzahl intertextueller Phänomene wie Paraphrasieren, direktes Zitieren, Zusammenfassen, Quellenverweise und textuelle Entlehnung (*textual borrowing*) sowie die strategische Wiederverwendung des Quelltextvokabulars untersucht. Die manuelle Annotation sowie quantitative und qualitative Analysen ermöglichen dabei eine fundierte und detaillierte Beschreibung der tatsächlichen intertextuellen Schreibpraxis von L2-Autor\*innen. Die folgenden Forschungsfragen liegen der vorliegenden Arbeit zugrunde:

1. Welche intertextuellen Strategien wenden Studierende in der L2 an und wie kombinieren sie diese, sowohl im Schreibprozess als auch im fertigen Reading Report?
2. Wie dokumentieren und attribuieren L2-Autor\*innen Quelltextinformationen und -material; wie verwenden sie Zitatverben und *reporting structures*?
3. Wie kopieren und verwenden Schreibende in der L2 Quelltextmaterial, sowohl im Schreibprozess als auch im fertigen Reading Report? Wird das Entlehnen von Ausdrücken strategisch eingesetzt, und wenn ja, wie und warum?

Die vorliegende Studie zeigt, dass die studentischen Autor\*innen in der L2 Englisch im Schreibprozess vielfältige intertextuelle Strategien anwenden und auf individuelle Weise kombinieren. Der Schreibprozess der Studierenden wird durch die Präsenz des Quelltextes beeinflusst, auf den sie während des Schreibens ständig zugreifen. Die unterschiedlichen Herangehensweisen an die Arbeit mit dem Quelltext zeigen sich in der Prozess- wie auch in der Korpusstudie. Die wichtigsten intertextuellen Strategien sind Paraphrasieren und direktes Zitieren, die bei allen Studienteilnehmer\*innen in ähnlichen Prozessen entstehen. Dies umfasst unter anderem sogenannte D/A/S-Strategien, also das Löschen, Hinzufügen von Material und das Ersetzen von Quelltextmaterial. Diese sind in den schriftlichen Produkten erkennbar und wurden auch in den Schreibprozessen auf der Mikroebene beobachtet. Die Studierenden verwenden *meta-level observations*, d.h. Kommentare auf der Metaebene, um ihre Reading Reports zu strukturieren und ihnen zum Teil neue Bedeutungen hinzuzufügen. Eine wichtige Erkenntnis der Korpusstudie ist, dass Zusammenfassungen zum Teil auf einzelnen Sätzen des Quelltextes basieren, die bereits eine zusammenfassende

<sup>1</sup> Ein Reading Report ist eine Zusammenfassung eines wissenschaftlichen Artikels mit einem Fokus auf die verwendeten Methoden sowie die Ergebnisse und Schlussfolgerungen der Autor\*innen. Die Studierenden erhielten vorab eine Aufgabenbeschreibung, s. Anhang 11.

Funktion haben. Diese *summary paraphrases* unterscheiden sich somit funktional von *summaries* und *paraphrases*. Im Gegensatz zu früheren Studien sind kopierte Textpassagen (*exact copies*) in beiden Teilstudien äußerst selten, was die fortgeschrittene wissenschaftliche Schreibkompetenz der Studierenden unterstreicht.

Es wird in beiden Teilstudien deutlich, dass sich die Studierenden der Notwendigkeit bewusst sind, die intertextuelle Natur ihrer Sätze herauszustellen, und die verschiedenen Möglichkeiten der Kennzeichnung von Intertextualität auch stilistisch einzusetzen wissen. So wird die Zuschreibung von Quelltextmaterial (*attribution*) in allen Studierendentexten auf vielfältige Weise realisiert. Eine große Anzahl von Sätzen enthält eine Form der Zuschreibung; meist werden die Autor\*innen des Ausgangstextes referiert. Dies geschieht beispielsweise in Kombination mit Zitatverben und *reporting structures*, von denen die Studierenden ein breites Spektrum einsetzen, einschließlich einiger kreativer und ungewöhnlicher Verwendungen. Es finden sich – im Gegensatz zu früheren Studien – hingegen keine Belege dafür, dass einzelne Studierende wiederholt auf dieselben Verben zurückgreifen. Vielmehr zeigen vor allem die Interviews, dass die Studierenden mit Bedacht darauf achten, Zitatverben im Text zu variieren. Jedoch verwenden einige Studierende grammatische Strukturen der indirekten Rede sowie Zuschreibungen nur aus stilistischen Gründen und sind sich ihrer intertextuellen Funktionen offenbar nicht bewusst, was die Notwendigkeit sensibilisierender Übungen in Universitätsseminaren hervorhebt.

Quellenangaben werden von den Studierenden kaum verwendet, was wahrscheinlich eine Folge der Interpretation der Aufgabenstellung durch die Studierenden ist. In beiden Teilstudien zeigen sich eindeutige Unsicherheiten hinsichtlich der korrekten Formatierung von Quellenangaben und ob und an welchen Stellen diese erforderlich sind. Insgesamt verweisen die Studierenden in ihren Reading Reports auf unterschiedliche Weise und in unterschiedlichem Maße auf den Quelltext. Die meisten verfügen am Ende des ersten Studienjahrs demzufolge sowohl über ein Verständnis der Notwendigkeit, auf den Quelltext zu referieren, als auch über ein ausreichendes Repertoire entsprechender sprachlicher Mittel. Nichtsdestotrotz sind die genauen Anforderungen wissenschaftlicher Texte und intertextueller Phänomene manchmal noch von Unsicherheiten behaftet.

Die Studienergebnisse unterstreichen, dass das Kopieren von Quelltextmaterial im Schreibprozess entweder mit großer Sorgfalt und strategisch eingesetzt oder gänzlich vermieden wird. Die Studierenden in dieser Studie treffen weitgehend bewusste Entscheidungen bei der Übernahme von Textstellen aus dem Ausgangstext, die sich an ihrem Wissen zu wissenschaftlichen Schreibkonventionen orientieren. Die Sorge um ein versehentlich getätigtes Plagiat führt dabei dazu, dass die Studierenden Sätze aus dem Quelltext besonders sorgsam überarbeiten und in ihren eigenen Worten wiedergeben. Es gibt in den Korpusdaten dennoch auch Hinweise darauf, dass nicht alle Studierenden bei der Vermeidung unnötiger Textüberschneidungen erfolgreich sind. Die vorliegenden Ergebnisse erweitern somit unser Wissen über das Patchwriting als Lernstrategie um Erkenntnisse über den Einsatz des Kopierens als Scaffolding-Strategie, die vornehmlich dazu dient, Textüberschneidungen zu vermeiden. Es ist wahrscheinlich, dass sich diese Ergebnisse auch auf andere Studierende mit einer ähnlichen Schreibkompetenz erstrecken.

Aus den vorliegenden Erkenntnissen ergeben sich methodische Implikationen für die L2-Schreibforschung; zudem lassen sich wichtige Hinweise für die Vermittlung des quellengestützten Schreibens an Universitäten ableiten. Die Studie zeigt auf, dass es bei der Erforschung intertextueller Strategien von entscheidender Bedeutung ist, verschiedene Formen der Quellenverweise und deren Kombination zu berücksichtigen, um studentische Strategien im quellenbasierten Schreiben vollständig zu erfassen. Die L2-Schreibforschung kann durch die Kombination von Analysen von Schreibprozessen und -produkten aussagekräftige Informationen über das Schreiben in der Fremdsprache liefern, um die Vermittlung wissenschaftlicher Schreibfertigkeiten zu optimieren und die Schreibberatung auf die individuellen Bedürfnisse der Studierenden zuzuschneiden. Dies gibt Forschenden und Schreibcoaches gleichermaßen die Möglichkeit, die Ursachen transgressiver und unkonventioneller Intertextualität zu identifizieren und somit einen wichtigen Schritt hin zur wirksamen Förderung wissenschaftlicher Schreibkompetenz zu gehen (Ranalli, Feng & Chukharev-Hudilainen 2018: 87). Die Ergebnisse dieser Studie geben wichtige neue Impulse für die Vermittlung intertextueller Strategien und können sowohl in wissenschaftlichen Schreibkursen als auch in fachwissenschaftlichen Kursen einfließen. Die Verbindung von Korpus- mit Prozessdaten ermöglicht die Ermittlung geeigneter didaktischer Maßnahmen, die Studierende im Schreibprozess und im Schreibprodukt für die Bearbeitung konkreter Schreibaufgaben in der L2 Englisch benötigen. Dieser methodologische Ansatz ebnet somit den Weg zu fundierter und umfassender Schreibforschung, die die pädagogische Praxis in allen Bereichen des wissenschaftlichen Schreibens unterstützt.

# Danksagung – Acknowledgements

Allen voran gilt mein Dank meiner Familie: meinen Eltern, meinem Bruder und meinen Tanten. Ohne ihre Förderung, Begleitung und Liebe in meinem gesamten persönlichen und akademischen Werdegang wäre dieses Projekt unmöglich gewesen.

Natürlich danke ich ganz besonders meinem Doktorvater Herrn Prof. Dr. Callies für die außerordentlich gute Betreuung und Zusammenarbeit. Durch sein Vertrauen, sein konstruktives Feedback und sein Engagement konnte ich mein Promotionsprojekt mit viel Freude voranbringen. Für die umfassende Unterstützung und die bereichernde und rundum schöne Zeit in seiner Arbeitsgruppe bin ich sehr dankbar.

I am deeply grateful to Prof. Casey Iezzi for being my co-supervisor and for her always valuable feedback that helped me refine my research. Our fruitful discussions during my research stay at Boise State University had a profound impact on my project. I am furthermore very thankful to Prof. Diane Pecorari for her advice in the early stages of my PhD, for kindly agreeing to be the third evaluator of my thesis, and for travelling to Bremen for the defence.

Meinen Dank aussprechen möchte ich außerdem meinen Doktorschwestern, meinen Kolleg\*innen und Kommissionsmitgliedern sowie dem BYRD-Team für ihre Unterstützung bei der Durchführung meines Promotionsprojekts. Weiterhin bedanke ich mich bei den Studienteilnehmer\*innen für ihr Vertrauen und den drei studentischen Mitarbeiterinnen für die sorgfältige Transkription der Interviews.

Eine besondere Stütze waren mir in den letzten Jahren auch meine wunderbaren Freundinnen und Freunde, die mich mit Heiterkeit und einem offenen Ohr begleitet und entscheidend zum Gelingen dieses Projekts beigetragen haben. Nicht zuletzt danke ich meinem lieben Partner, der mir mit seinem Humor und seiner Geduld zu jeder Zeit eine echte Stütze war.



# 1 Introduction

Intertextuality is an integral and defining feature of the academic register. It refers to the overt reference and response to prior writing in the field, to other authors' words, ideas, research, and theories. Voices from cited literature are omnipresent in academic texts across academic disciplines.<sup>2</sup> Writers<sup>3</sup> create new meanings and arguments by combining knowledge from the literature with their own ideas in a variety of ways. The formally correct and diligent acknowledgement of source use is a central component of good scientific practice and a requirement of academic publications. Disobeying academic conventions can lead to accusations of plagiarism and thus may have considerable repercussions for a scholar's academic success.

Source use can be direct or indirect and is predominantly manifested in direct quotes, paraphrases, summaries, and generalisations from multiple sources. Paraphrasing entails rephrasing and restructuring source text material. It is a necessary skill for academic writers, yet difficult to accomplish because the writer has to change the wording and structure without distorting the proposition of the source text and, ideally, without (extensively) copying from it. Summarisation and generalisation additionally require the writer to abstract from the source text so as to provide an overview of its contents. These indirect intertextual strategies often rely on strategic borrowing of textual material from the source text.<sup>4</sup> Direct quotes are passages from an author's text that a writer has integrated into their text in quotation marks. The writer repeats someone else's words verbatim without changing the original wording or proposition. They must carefully select the passage to be quoted and elegantly integrate it into their text. Direct quotation allows the writer to include other authors' voices to support their argument or to distance themselves from the quoted content.

Writers can choose from a variety of means of acknowledging intertextuality in order to meet the standards of the scientific community. By documenting their sources via references, attributing the content to the source in explicit ways, and using reporting structures to signal to the reader the nature of cited passages, writers make explicit the extent to which their text relies on previous literature in the field. The origin of source text material and ideas is conventionally indicated via references both in the text and the bibliography. In addition, source content is attributed and reported through the use of a range of grammatical structures and reporting verbs. By attributing ideas to sources, writers are also able to evaluate and position themselves in reference to them, signal objectivity, and point out research gaps. Source use takes many forms and, despite underlying conventions, is often subject to individual preferences. Source text material is organised, selected, connected and interwoven in such a way that suits the writer's intended message and argument as well as the communicative and discursive context (Spivey 1990: 256).

An intertextual passage, also referred to as a citation, establishes cohesion with the research community's past and present discourse. Academic writers use intertextual reference to persuade their readers of the validity of their arguments and to substantiate their claims by referring to authors who have argued in a similar way. They employ intertextual devices to position themselves in relation to particular theories and assumptions. This is necessary in order to secure the discourse community's acceptance of their claims. Citations are important points of reference as they help the reader identify the disciplinary and narrative context in which the writer's claims

<sup>2</sup> Intertextuality is of course also a feature of many literary works – also in the form of plagiarism, see e.g. Tufescu (2008).

<sup>3</sup> Following the conventional terminology of studies on intertextuality, the word *author* is used here for the person who created the source text, while *writer* refers to the person citing an author in their own text. The word *writer* is understood to include all genders, which is why singular *they* is used as the corresponding epicene pronoun.

<sup>4</sup> Note that the term 'textual borrowing' is sometimes used in a broad sense to refer to any kind of intertextuality or citation in a text: "Textual borrowing refers to collecting, remixing, and transforming previously published texts—print, digital, and aural—to create new texts, which can then be borrowed again to create even more new texts. [...] All texts, both spoken and written, incorporate textual borrowing to varying degrees." (Bloch 2018: 1) As such, 'textual borrowing' can function as a cover term for intertextual strategies such as paraphrasing, summarisation, and direct quotation. In a narrower sense, textual borrowing refers to words in a text that have been copied, i.e. borrowed, from a source text (Weigle & Parker 2012: 121). It is in this narrower sense that textual borrowing is understood and used in this dissertation.

## 2 Introduction

are located (P. Thompson 2001). They thus carry significant rhetorical power and convey intricate social meanings. Reference to prior literature in the field plays a vital role in the reception of academic papers by peers and is accordingly an important factor in reaching publication. The previous literature is instrumentalised to provide “a forceful warrant for current innovation” (K. Hyland 2004: 32).<sup>5</sup> Creating intertextuality can be conceived of as a social process of negotiating the writer’s claim to membership in a discourse community and their contribution to its knowledge (K. Hyland 1999; Jakobs 2003). Since the discourse community’s conventions determine what is appropriate in terms of citations, there are differences across disciplines in terms of how and why academic writers cite.<sup>6</sup>

Intertextual competence, i.e. the ability to reference discipline-specific discourse, is a central aspect of academic writing and seen as a prerequisite of membership in the respective discourse community (Pickard 1995). University students are usually required to write from sources across the curriculum (Leki & Carson 1997; Ruiz-Funes 1999). Intertextual literacy constitutes a branch of academic literacy, and academic success as measured in grades is often closely tied to appropriate and effective source use (Kirkland & Saunders 1991; Shaw & Pecorari 2013). In university contexts, source use often forms the basis for measuring a learner’s competence in academic writing.<sup>7</sup> As a result, mastering intertextual strategies is an important aspect in students’ acquisition of academic literacy, alongside knowledge of relevant academic conventions. In order to create intertextuality successfully, a writer must critically assess their source texts, have a command of appropriate academic language, and document their sources diligently. The rhetorical functions of citations in student writing diverge from those of writers whose aim it is to publish their work and may vary from text type to text type (Petrić 2007; Petrić & Harwood 2013). As their work will usually only be read by their lecturer for the purpose of grading, student writers have to showcase their knowledge of the topic, combine citations to weave a “new intellectual fabric” (Borg 2000: 29), and show that they have understood the cited material by their ability to correctly paraphrase its meanings.

The use of sources in academic writing is complex both linguistically and rhetorically and requires an extensive set of reading and writing skills (Grabe & Zhang 2013; Samraj 2013). It comprises literacy skills that are different from those learnt at secondary school and must be newly acquired by students at university. Academic writers are required to read and understand academic articles of varying linguistic intricacy as well as interpret new and potentially complex information in the light of their topic. Writers often must integrate similar, contrasting or even contradictory information from different sources. In order to write a comprehensive and critical literature review, students have to be able to “evaluate the entire problem-identification-and-resolution enterprise that constitutes research” (Dovey 2010: 51). This is a difficult task for many novice writers because it requires knowledge of previous research, the ability to identify relevant sources, an understanding of the validity of research methodologies, and a discernible authorial identity. Source-based writing is increasingly based on electronic sources, making it all the more challenging for novice writers to identify sources which are reliable and of an acceptable academic standard (Pecorari 2015). Furthermore, though writing from sources is regarded as a core component of students’ writing abilities in various educational contexts, it is not always fostered as such. In Germany, explicit teaching of academic writing is still the exception and students are often expected to develop academic writing literacy on their own account (Breuer 2014: 41; see also Pohl 2007).

Emerging challenges of novice writers in the face of the manifold requirements of source-based writing are sometimes evident because they re-use ideas from only one source or do not connect ideas from various source texts in a meaningful way. A typical feature of inexperienced writers’ academic texts is a tendency to make claims without proper support from the literature. Often, this is paired with patchwriting, a practise that

---

5 Academic research papers contain a considerable number of citations, with an average of seventy citations per paper across disciplines (K. Hyland 2004), which demonstrates writers’ endeavour of lending validity to their claims by providing intertextual warrants.

6 See Erikson & Erlandson (2014) for a study of motives to cite across scientific disciplines.

7 Increasingly, this is also the case at secondary school as source-based tasks are on the rise in university entrance exams such as the German Abitur. In Lower Saxony, so-called material-based writing tasks (“materialgestütztes Schreiben”) have been administered in the Abitur for the subject German since at least 2016 (<https://tinyurl.com/material-ls>). They will become a part of Abitur exams in German in Baden-Württemberg in 2021 (<https://tinyurl.com/material-bw>).

entails copying a string from a source text and then making only minimal syntactic and lexical alterations to it, e.g. deleting or replacing individual words (Howard 1995: 788). This results in passages that are very similar to the source text. Consequently, novice writers' texts may be in breach of academic writing conventions because their ideas and words are not clearly delineated from those of cited authors (Borg 2000; Groom 2000). Developing writers also tend to find it difficult to evaluate and integrate their sources in genre-appropriate ways. They struggle to position themselves as competent and legitimate members of the discourse community because they regard the authors as authoritative experts whom they do not dare to criticise (Abasi & Akbari 2008; Dovey 2010; Plakans & Gebril 2012; Solé et al. 2013). Ineffective or illegitimate intertextuality can thus be revealing indicators of students' insufficient understanding of a source text and lack of familiarity with academic writing conventions (McInnis 2009).

Writing academically is especially challenging in a foreign language, particularly when learners have to refer to other texts in their writing (Abasi & Akbari 2008, C. Thompson, Morton & Storch 2013; Davis 2013). L2 writers<sup>8</sup> are not only expected to be able to identify ideas that need to be cited, but "they must also recognize *how* the idea should be borrowed and the ways in which the language expressing that idea can become their own" (McInnis 2009: 1, original emphasis). Furthermore, not every type of intertextuality is evaluated positively,<sup>9</sup> and L2 writers' inappropriate textual borrowing may result in accusations of plagiarism. In writing research, their issues have often been discussed in the context of plagiarism studies (see overview in Pecorari & Petrić 2014). While both L1 and L2 writers are faced with similar challenges with regard to integrating source text material into their writing and attributing their sources appropriately, L2 writers have been found to rely more heavily on source text material and to be prone to patchwriting (e.g. Campbell 1990; Keck 2006; Shi 2004). Problems of L2 writers' texts reported in previous studies include difficulties in expressing oneself using academic language, a limited repertoire of reporting verbs, lack of stance-taking, unmasked and inappropriate direct criticism, lack of references for paraphrases, copied strings without quotation marks, extensive direct quotation without an apparent function, and insufficient acknowledgement of intertextual material (Borg 2000; Cheng 2006; Dong 1996; Groom 2000; Petrić 2012; Uccelli, Dobbs & Scott 2013; Verheijen 2015).

In recent years, scholars in L2 writing research have become interested in the causes of apparent cases of plagiarism. Initially, the focus was on why L2 writers plagiarised and whether cultural differences and linguistic background could account for their illegitimate source use. However, studies show that the absence of references in source-based student writing is often neither due to cultural background nor to an intention to deceive or laziness. Rather, it is caused by the students' lack of knowledge pertaining to academic writing conventions (e.g. Pecorari 2006). As developing academic writers, they have simply not yet acquired sufficient academic vocabulary and intertextual competence. That issues of source use are connected to literacy development is clear from the fact that insecurities often decrease with advanced writing expertise. However, fear of plagiarism also plays a role in student writers' source use, even if they have internalised academic citing conventions. While L2 writers at earlier stages of development often resort to illegitimate forms of textual borrowing such as patchwriting, students at more advanced stages tend to over-rely on direct quotes because they want to avoid facing accusations of plagiarism (see Petrić 2012). Such findings have compelled L2 writing researchers to reconsider their views of inappropriate cases of intertextuality. This paradigm shift has led researchers to re-conceptualise apparent plagiarism and to distinguish between deceptive and non-deceptive forms of transgressive intertextuality in consideration of the writing context (see e.g. Chandrasoma, Thompson & Pennycook 2004).

In L2 writing research, it is now a commonly held view that inappropriate intertextuality is mostly a developmental issue and in the majority of cases should not be considered an issue of honesty. Accordingly, patchwriting is viewed as a learning strategy used by developing academic writers as they are acquiring academic literacy and not as evidence of plagiarism (Campbell 1990; Currie 1998; Keck 2006; Flowerdew & Li

8 L2 writers in this dissertation are understood as both foreign-language and second-language writers. The present study investigates source-based writing in a foreign-language context.

9 See for example Crocker & Shaw (2002) for a discussion of the evaluation of patchwriting by educators and linguists.

2007). This new perspective on student borrowing practices has led to an interest in intertextuality in L2 academic writing beyond plagiarism, specifically in L2 writers' source use strategies. Recent research has looked at the reasons behind textual borrowing and considered the intertextual strategies of successful students. As a result, L2 writing research has shifted to a more positive perspective on students' intertextual strategies in the broader context of source-based writing (e.g. Ädel & Garretson 2006; Davis 2013; Harwood & Petrić 2012; Hirvela & Du 2013; Li 2013; Petrić 2007, 2012; Petrić & Harwood 2013). It is in this discourse that the present study is situated.

Intertextual practices and textual borrowing by L2 writers have thus far been investigated mostly in small-scale corpus studies and case studies on the basis of text-based interviews. Many of these are located in the discourse on plagiarism and have foregrounded textual borrowing with a view to identifying reasons for transgressive behaviours. Most process studies of source-based writing tasks have been conducted in testing and assessment contexts. There is a persistent lack of research into students' strategies of source use, as Keck (2015: 177) asserts: "Though it is widely recognised that much of what we write in the academy is in some way based upon what has been written before, little is known about when, how often, and in what ways academic writers re-use the language of others". Although they have the potential to offer rich insights into academic literacy development, L2 writers' source use strategies beyond plagiarism, both in the writing process and the product, are yet to receive detailed attention in research. Researchers thus far have spent little time on exploring students' rationale behind the use of specific strategies (Leijten et al. 2019: 577). There is a need for studies that aim at identifying L2 writers' declarative and procedural knowledge about appropriate and inappropriate source use (Pecorari 2015: 95). Such research beyond plagiarism will pave the way for effective academic writing pedagogy that foregrounds writing strategies and fosters understanding of academic conventions and their implementation rather than focus on decontextualised warnings of plagiarism.

This dissertation utilises a mixed-methods design to investigate intertextuality in foreign-language academic writing in English (Figure 1). It combines a corpus and a process study of disciplinary writing assignments based on a single source text, so-called 'reading reports'. The reading reports analysed in the corpus study were composed by L1-German university students. They constituted a real-life assignment in an introductory course in linguistics in the framework of a B.A. programme in English studies. For the process study, students recruited from another first-year cohort wrote reading reports in an experimental setting. The writing process was recorded using screen recording software. Subsequently, the students were asked to comment on their writing processes in a stimulated recall procedure for which the recording of the writing process served as a stimulus. The research methods were carefully selected with a view to conducting qualitative and quantitative analyses of L2 writers' source use, source acknowledgement, and textual borrowing.

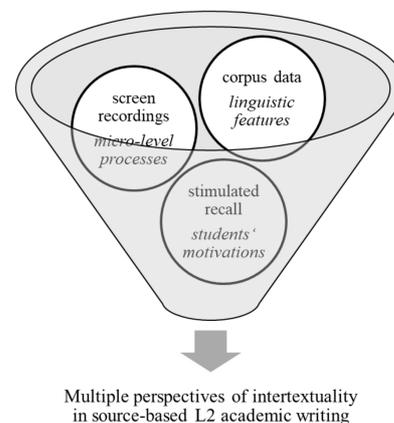


Figure 1: Mixed-methods design.

For the corpus study, reading reports based on a published linguistic research article from the *Corpus of Academic Learner English* (CALE; Callies & Zaytseva 2013) were analysed to identify the types of intertextual links L2 writers use and how they are combined with attribution, documentation, and reporting structures. The reading reports were divided into sentences, which were then categorised according to type of intertextual link. Each intertextual link was scrutinised to identify the amount of textual overlap, the section of origin, the encoding of evaluation, the addition of new information, and students' (mis-)interpretations. They were also annotated for integral and non-integral citation, the kind of attribution, and the type and position of reporting

structures. The individual categories of intertextual links were subjected to further in-depth analysis. For example, the number of sentences on which a paraphrase was based was identified and the length and content of a direct quote were recorded. The findings were quantified and subjected to additional qualitative analyses, for example concerning the types of passages that had been borrowed or quoted from the source text. The goal was to identify individual and common strategies of integrating source text material into a reading report. A further goal concerned the identification of patterns of intertextuality both in individual texts and across the corpus.

In the process study, observation is combined with retrospection to gain a more fully-fledged picture of L2 writers' processes in reading reports. Six first-year students wrote a reading report based on the same source text as in the corpus study. The writing took place in an untimed experimental setting and the process was recorded using screen recording software. The participants were provided with the source text and the instructions for writing a reading report in digital format. They were allowed to use additional digital language resources of their choice. One day after the writing sessions, the students participated in a stimulated recall procedure that used the screen recordings as a prompt. They watched the screen recording together with the researcher and commented on their writing process, especially on aspects of interest identified by the researcher. The stimulated recall was supplemented by a small set of interview questions. The screen recordings were segmented so that each segment contained the writing process of an individual sentence. Micro-level processes were recorded for each segment in chronological order of occurrence. These were then analysed to identify the L2 writers' intertextual strategies, their general approach to the task, and common patterns across the group. The findings were contextualised in reference to the students' comments in order to illuminate their motivations, intentions, and insecurities when writing from sources. The goal of the stimulated recall procedure was to discover students' explanations for their own writing processes and to establish why certain process strategies were used. This procedure would ideally also uncover areas in which they feel they need support in order to specify potential pedagogical interventions.

The triangulation of these three research methods allows the researcher to marry multiple perspectives on the subject matter under scrutiny, in this case intertextuality in L2 source-based writing. The different data offer multiple perspectives on a broad range of intertextual phenomena and effectively compensate individual weaknesses, especially because both qualitative and quantitative analyses are performed. The goal of this triangulated approach is to generate reliable, reproducible, and detailed results regarding L2 student writers' use of the source text and creation of intertextuality. Corpus data allow for a deep analysis of the linguistic features of source-based texts. It is possible to quantify the features of reading reports, compare the textual material to the source text, and draw comparisons across texts. At the same time, the data are static and the researcher can only surmise as to the processes that led to the final product and the reasons behind certain manifestations of intertextuality. It is difficult to evaluate the exact nature of reliance on the source text, especially if a sentence has been radically rephrased and restructured. Data from screen recordings, while being less suited to detailed analyses of linguistic patterns, offer insights into the evolution of the text, and especially into changes that are made in the course of writing. They also provide a clear link to source text passages and allow the researcher to gauge the frequency and timing of interaction with the source text. The stimulated recall elicits additional information about the students' intentions and thought processes as well as underlying motivations.

The goal of the present study is to address a gap in research regarding focused investigations of L2 writers' intertextual strategies from a proficiency-oriented<sup>10</sup> perspective, i.e. a positive perspective focusing on competent source use. By mixing methods, the study sheds light on how intertextuality is realised both in the writing process and the product. It provides in-depth insights into student writers' strategies of source use and acknowledgement, their knowledge and subjective theories. The present study is a comprehensive exploration of complementary and interacting manifestations of intertextuality in L2 academic writing. It aims at contributing to the existing body of research by uncovering the micro-level processes of source-based writing, by exploring

---

10 Note that 'proficiency' here does not refer to proficiency in the L2, but to the proficient use of sources.

the interactions and interconnections between different markers of intertextuality, and by offering insights into students' conceptualisations of writing from sources.

The dissertation is structured as follows. Sections 2 and 3 contain a review of relevant literature from writing research and beyond. First, intertextuality in academic writing is considered in section 2, beginning with an overview of the forms and functions of citations in section 2.1. Section 2.2 focuses on the most common types of intertextual strategies, namely direct quotation, paraphrasing and summarisation, and generalisation. This is followed by an overview of the ways in which intertextuality is signalled and evaluated in section 2.3. Research investigating intertextuality in L2 academic writing is presented in section 3. Section 3.1 provides an overview of the discussion surrounding the issue of legitimate and illegitimate forms of intertextuality and section 3.2 discusses research findings with regard to unintentional textual borrowing and learning strategies. The use of intertextual strategies such as direct quotation, paraphrasing and summarisation and of attribution in L2 student writing are reviewed in section 3.3. Section 3.4 looks at the use of the source text as a language resource. These perspectives are complemented by an overview of models of the writing process and a discussion of research into the processes of L2 source-based writing in section 4. Section 5 contains the rationale for the study and the study design. It introduces the two studies that form the basis of the present dissertation, namely a corpus study investigating intertextuality in the writing products (section 6) and a process study of intertextuality (section 7). In section 8, the findings of both studies are discussed and interpreted together. This is followed by an examination of implications for L2 writing research and pedagogy as well as limitations and directions for future studies. Concluding remarks are presented in section 9.

## 2 Intertextuality in academic writing

It is a conventional expectation among readers of all but the most playfully postmodern of Anglophone academic texts that it will be clear at any given point whose ‘voice’ is ‘speaking’, and thus whose textual subjectivity is being put ‘on the spot’, that is, being made open to potential interlocutory challenge. (Groom 2000: 15)

Intertextuality is the incorporation of linguistic elements, aspects, and concepts from other texts into a text (Fairclough 1992).<sup>11</sup> It is created by overtly referring to and embedding the words and ideas of others into one’s writing. Writers select and reconfigure content from sources and interweave it with their own insights on the topic to create a “textual tapestry” (Spivey 1990: 260). Intertextuality arises in the form of manifest intertextuality, which is created by explicit ways of referring to other texts (e.g. direct quotation and paraphrasing), and constitutive intertextuality, which is implicit and arises from the topics, concepts, and language a text shares with other texts (Fairclough 1992). Manifest and constitutive intertextuality typically co-occur and are found in all types of reproductive writing, i.e. writing that relies on and integrates sources (Jakobs 2003).<sup>12</sup> One type of reproductive writing is academic writing. As stated in the above quote by Groom, this register is characterised by a relatively strong presence of other authors’ voices, making intertextuality a distinctive feature of academic texts. The use of intertextuality goes beyond attributing ideas and methods and clarifying which previous research exists in the field. Academic writers identify and position their sources in a complex interplay between citations and their own comments, evaluations, and propositions (Groom 2000).

According to Crocker and Shaw (2002), there are three dimensions of intertextuality, namely quantity, closeness of wording to the source text, and documentation. First, in terms of quantity, writers can choose between the extensive and the sporadic use of sources, or they can limit themselves to the strategic incorporation of key expressions. Second, there are varying degrees of closeness to the source text. Academic writers may choose between paraphrasing and copying word-for-word, for instance. Copied passages may differ in length as a writer may decide to copy a term, a short fragment or a longer passage. Copying is not necessarily illicit since writers may mark copied passages as direct quotes by inserting quotation marks. However, paraphrases may contain large chunks of the original text whose structure has not been altered, a phenomenon which has been referred to as near-copying (Campbell 1990) or patchwriting (Howard 1995) and is usually deemed problematic in academic texts or even viewed as plagiarism (see section 3). Third, writers may document their sources – for example by referencing – or they may not. These dimensions focus on the actual manifestations of a source text in a writer’s text.

In contrast, Chandrasoma, C. Thompson, and Pennycook (2004) propose a view of intertextuality as “multiple strands of knowledges within texts designed to produce desired meanings”. They argue that intertextuality should not be reduced to the occurrence of surface forms of citation because it is constitutive to academic writing. They distinguish three modes of intertextuality. First, according to the authors, intertextuality is conceptual because concepts from other texts are incorporated, appropriated or advanced. Second, it is complementary because sources are used to support the writer’s claims and arguments. Third, it is metalinguistic in the sense that intertextual manifestations such as stance-taking and the adoption of terminology are achieved via linguistic resources. These three modes are influenced by and related to the respective discourse in which the

<sup>11</sup> The term ‘intertextuality’ was first coined in the 1960s; see Fairclough (1992) for a history of the term.

<sup>12</sup> Jakobs (2003) distinguishes between two types of reproductive writing, namely 1) texts whose sole purpose it is to provide an overview of another text’s content, e.g. abstracts and book reports, and 2) texts in which the writer uses other authors’ ideas to underpin his or her own argument, e.g. research papers and dissertations. Most academic text types belong to the latter category.

writer's text is situated. Intertextuality should thus be seen as acting on several levels of an academic text, both in terms of the manifested textual surface and of the character of academic discourse.

Citations have long been an integral feature of academic research articles and fundamental in the construction of persuasive arguments. Late 19<sup>th</sup> century research articles were characterised by frequent use of references, although they were employed mostly for general remarks rather than to refer to specific findings or topics. This meant that in the discourse of the time there was little sense of a coherent research community (Bazerman 1984, cited in Swales 1986). The early 20<sup>th</sup> century saw a decrease in the number of references cited, but a shift in focus to recent sources which were directly relevant to the respective study. Since then, there has been a steady increase in the number of references reported in research articles, leading to a stronger sense of embedding of research articles in the previous literature by means of citation (Swales 1990).<sup>13</sup>

As scientific research became increasingly contextualised in disciplinary discourse, the Discussion section gradually replaced the Methods section as the persuasive focal point of academic research articles (K. Hyland 1999). This shift in conventions was accompanied by an increasing use of citations throughout the paper, instead of their occurrence being restricted to the Introduction, as was the preference in the late 19<sup>th</sup> century. At the same time, the average length of articles doubled from the 1950s to the 1980s (Swales 1990). Overall, the increased use of citations in research articles has led to a publication culture in which a discourse community discusses and expands on previous studies. Academic text types, for example dissertations, indispensably contain explicit reference to prior literature and draw on sources to support their arguments, making intertextuality one of the most conspicuous features of academic writing (see e.g. P. Thompson 2001). K. Hyland (2002) even goes so far as arguing that academic texts without reference to prior literature in the field would be unlikely to reach publication.

The centrality of intertextuality in academic writing originally sparked scientific interest in the study of citations. Citation analysis as a sub-field of applied linguistics<sup>14</sup> emerged around eighty years ago out of an interest in the development of statistical laws pertaining to bibliographical entries. It was hoped that such laws would allow for a quantitative description of citation behaviour and would help explain phenomena such as the dominance of a small number of authors who published the majority of contributions in a particular field. Some scholars examined citations in order to uncover differences and similarities between different fields of inquiry and diachronic developments of citation practices (see e.g. Bazerman 1988; Salager-Meyer 1999). Other studies compared the use of citations across journals in the same discipline to make deductions regarding the journals' standing in the discourse community (see detailed review in Swales 1986). Citations were seen as indicative of the quality of a writer's work based on the assumption that more citations meant a larger impact (Okamura 2008a). Swales (1986) regarded the potential of citations for grading, evaluating, and ranking expert and student papers as the main motive for citation analysis in the 1980s.<sup>15</sup>

Today, the status of citations as a quality measure is contested. The assumed correlation between citation count and a publication's quality is complicated by certain aspects of citation behaviour, for example by the fact that information which is considered to have become general knowledge in the field is no longer accompanied by a reference (Swales 1990). Citations differ in length and in the weight they are given in publications, which may have an impact on a scholar's perceived worth. Furthermore, the role of self-citation is not accounted for. It has been argued that writers have different motivations to cite and that some citations actually indicate a negative stance, which is contrary to the assumption that a citation is evidence of a publication's quality and impact (see

13 For example, in a study of research articles published in *TESOL Quarterly* in the first twenty years of publication, Swales (1990) found that the average reference count in 1968 was 4, which had increased to 34 in 1986. In the same period, he observed that the authors' reliance on books decreased, a development which was accompanied by a parallel increase in the number of citations from articles and chapters from edited volumes. Many of the latter type came from previous articles published in *TESOL Quarterly*, which created an impression of the publications being part of the same discourse.

14 Citations have also been studied by scholars in information science and sociology of science. However, there is still very little integration of the disciplines, despite the potential mutual benefit (Okamura 2008a; Swales 1986).

15 Swales (1986) assumes a number of different reasons for scholars' interest in citations. One of the earlier motivations for citation studies may have been sparked by historians' "attempts to establish the origin and distribution of particular ideas and discoveries, and to trace major networks of influence, collaboration, and dependence" (1986: 39).

Okamura 2008a for an overview of the discussion). Due to these problems with citation counts, content analysis has emerged as a complementary approach that takes into account the linguistic context of a citation. It has focused, for instance, on the rhetorical functions of citations, their persuasive potential, and the significance of evaluative language (e.g. Harwood 2009; K. Hyland 2008; G. Thompson & Ye 1991).

As suggested by content analyses, citations are not simply “nods all round to previous researchers” (Swales 1990: 6), but fulfil a range of rhetorical functions in academic writing. Academic writers use intertextual reference to attribute information, to name the originator of an idea, finding or concept, or to direct the reader to a source of further information. References may identify sources which have used a similar research method, have a similar focus, contain a related (or contrasting) proposition or argument or have produced a similar (or diverging) result.<sup>16</sup> However, academic writers also rely on intertextuality to create an authorial self and to establish a research niche (Harwood 2009; K. Hyland 1999; Lim 2012; Mansourizadeh & Ahmad 2011; P. Thompson & Tribble 2001). Most importantly, the use of intertextuality is closely tied to the epistemological function of writing in general and academic writing in particular. According to Shi (2008: 21), “[c]iting a source text is more than providing a name and a date; it is a subjective process of deciding how to make meaning out of the available resources.” Intertextual reference allows academic writers to construct and extend previous knowledge, to acknowledge and evaluate the work of other authors, to lend validity to their claims, and to situate their writing in a particular stream of scientific thought.

Academic writers construct an authorial self by responding to other authors, by engaging with, elaborating, and refuting their claims:

By acknowledging a debt of precedent, a writer is also able to display an allegiance to a particular community or orientation, create a rhetorical gap for his or her research, and establish a credible writer ethos. (Hyland 1999: 342)

Writers use intertextual devices to position themselves in relation to the views and assumptions held by other authors in the same discipline and to signal conformity and dissent with particular theories and assumptions. They use citations to mark alignment with or demarcation from a particular school of thought. In this process, the writer is foregrounded, but still relies on others’ contributions to the discussion. Intertextuality is thus a means of establishing oneself as a legitimate member of the relevant discourse community. In many academic disciplines, a scholar’s success crucially depends on how they position themselves towards previous research in the field (K. Hyland 1999; 2004; 2005). Academic merit is usually measured in the number of publications and conference papers, both of which traditionally depend on favourable peer-reviews. Citation has been found to be a predictor of abstract evaluations and to have an influence on students’ grades (Egbert & Plonsky 2015; Plakans & Gebriel 2012; Plakans & Gebriel 2013). This means that in order to be admitted to conferences or have their paper published, academic writers across disciplines are required to convince their peers that their research is in line with and constitutive to the respective research community’s knowledge. The citation of relevant previous literature and the creation of links to one’s own research is consequently a prerequisite to an academic text being judged worthy by other members of the discourse community.

How academic writers integrate and reference their sources in their writing appears to be determined by an intricate set of factors. Differences in citation practices can to a certain extent be attributed to different academic conventions between the so-called soft sciences, that is, the humanities and social sciences, and the hard sciences, that is, the natural sciences and engineering. The citation conventions of a field of study are a reflection of conceptions of research and representations of knowledge, which differ across disciplines (K. Hyland 1999). As members of the discourse community, academics are the ones who construct the discipline’s knowledge. Consequently, writers from different disciplines cite differently because “their discursive decisions are influenced by, and deeply embedded in, the epistemological and social conventions of their disciplines” (K. Hyland 1999: 341).

---

<sup>16</sup> See Erikson & Erlandson (2014) and Harwood (2009) for empirical studies of the rhetorical functions of citations. See also Swales (1986) for a detailed overview of earlier citation taxonomies based on function.

Within their field of expertise, academic writers can assume a certain knowledge of the theoretical foundations, methodologies, and expert language amongst their readership. This is particularly the case in the natural sciences, which is reflected in scientists' reliance on a standardised code to frame their experiments. The persistent view held in the natural sciences is that the purpose of research is to discover the truth, not to create it (K. Hyland 1999). In the arts and humanities, on the other hand, the reiterative nature of academic writing means that writers are likely to refer to a broader range of topics. Hence, writing in these disciplines references a more diffuse body of previous literature from different fields and eras which cannot be assumed to all be known by the reader. As readers do not necessarily share the same assumptions and interpretative basis, there is a higher general importance of exact and detailed citation in the soft sciences to create a framework for the writer's research. The diverging practices between areas of research reflect differences in the ways in which knowledge is typically constructed in scientific disciplines (K. Hyland 2004). Despite general tendencies of hard and soft sciences, K. Hyland's (1999; 2004) studies suggests that the distinctions are more complex than can be captured by such a simple dichotomy. There are also striking differences within the two groups, which are to some extent due to the different citation styles which are conventionally used in the individual scientific domains (K. Hyland 2004). Variation is found even between texts of the same genre from the same discipline (see e.g. Swales 2014).

While some variation in the use of citations may be attributable to discipline-specific preferences, it is clear that the text type also influences the density and form of citations. Comparing a study of research articles (K. Hyland 1999) and one of doctoral theses (P. Thompson 2000), P. Thompson and Tribble (2001) attest that the density of citations is considerably lower in doctoral theses than in comparable research articles. They speculate that this is due to the different length of these text types. Their comparison can only provide a general idea, however, because the texts in the surveyed studies are from similar, but not from the same disciplines. Factors besides discipline and text type, such as individual writer's preferences, audience expectations, or publisher guidelines, also play a role in the use of citation practices in academic writing. Finally, academic disciplines impose stringent regulations regarding the use of sources, some of which are implicit. These shape the ways in which intertextuality is manifested in both published and unpublished texts because disobeying these regulations, for example by committing plagiarism, may be severely punished (Crocker & Shaw 2002; see discussion in section 3). Despite the undeniable influences of the factors discussed above, it is important to bear in mind that many of the factors influencing the use of sources are very individual and are not generalisable across research communities, cultures, or text types (see e.g. Jakobs 2003).

Intertextuality can take on a range of different forms and is signalled in a variety of ways. In section 2.1, citation as a feature of academic writing is illuminated with a focus on the forms of citations and their respective functions in a text. Section 2.2 focuses on the four major types of intertextual strategies, namely direct quotation, paraphrasing, summarisation, and generalisation from multiple sources. Finally, section 2.3 is devoted to the ways in which intertextuality is signalled and discusses features of academic writing such as attribution, documentation, and reporting.

## 2.1 Forms and functions of citations

Intertextual passages, also referred to as citations, serve to create connections to other texts produced in the same discourse community. The term 'citation'<sup>17</sup> is used as an umbrella term for intertextual strategies used by academic writers and generally refers to the act of referencing another author's words, ideas, theories or opinions in one's text. In a narrower definition, a citation is a passage in a writer's text which conveys an idea or

---

17 The term 'citation' is sometimes used as a synonym for 'quotation' or 'referencing' (see Jakob 2003: 897 for a discussion). It is used here as a general term for manifestations of intertextuality. A quotation is a source text passage that is repeated verbatim and inserted into a text embedded in quotation marks (Petrić 2012); a reference is "a formal acknowledgement of the sources of other writers' words and thoughts" (Borg 2000: 27), i.e. a bibliographical reference.

finding from another author's text.<sup>18</sup> It is an instance of intertextuality, for example a direct quotation or a paraphrase. Because of their conventional nature, citations can easily be discerned in academic texts (P. Thompson 2000).

Citations are commonly categorised using Swales' (1990) distinction between integral and non-integral citations, which is based on the way the writer attributes content to a source. Swales' revised version of the *Create a Research Space* (CARS) model posits three Moves for the Introduction sections of research articles: 1) Establishing a territory, 2) Establishing a niche, and 3) Occupying the niche. Based on his corpus analysis, Swales asserts that the reference to relevant previous literature in the field is an obligatory step in Move 1. In this step, the writer of a research article "needs to relate *what has been found* (or claimed) and *who has found it* (or claimed it)" (1990: 148, original emphasis). This means that the writer is required to relate relevant previous research in the field in the given amount of detail and attribute it to the researcher(s) in question. At the same time, writers are expected to indicate their evaluation of the cited literature by signalling their stance towards it. Swales proposes that in order to resolve this tension, the writer can choose between two citation forms: integral and non-integral citation. This distinction is based on the surface form of citations.

An integral citation is a citation in which the name of the cited author or the respective personal pronoun appears in the sentence and carries a syntactic function. In integral citations, the authors' names can occur in three different sentence positions: in subject position (see the constructed example in 2.1), in non-subject position (example 2.2), and in noun phrases within adjuncts (example 2.3), e.g. in combination with the complex preposition *according to* (K. Hyland 1999).

(2.1 ) Linguis (1977) claims that linguistics students are the smartest.<sup>19</sup>

(2.2 ) The theory that linguistics students are the smartest was put forward by Linguis (1977).

(2.3 ) According to Linguis (1977), students of linguistics are the smartest.

Integral citation can be combined with a considerable number of reporting structures – G. Thompson and Ye (1991) identified more than 400 reporting verbs alone. This affords writers with a wide range of choices as regards the embedding of citations into their texts.

In contrast, non-integral citations do not contain the author's name in the sentence. Authors' names are given in parentheses either directly following the cited statement or at the end of the sentence (see examples in 2.4). They may also be referred to elsewhere, for example by means of a foot- or endnote indicated by a number in superscript (see 2.5). Consequently, in non-integral citations the researchers in question do not play an "explicit grammatical role" (P. Thompson 2000: 93) in the respective sentence. In other words, the cited authors "are not allowed by the creators of the texts to take part in the arguments presented" (Tadros 1993: 106). Two variants are possible: a non-integral citation is either embedded in a sentence with a reporting structure or a sentence without such a structure; compare (2.4) and (2.6). In (2.6), the cited individual is only named within parentheses, while the sentence itself does not contain a reporting structure and consequently no acknowledgement of the original author.

18 A special type of citation is self-citation. Self-citations are passages in which the writer cites their own work. Typically, the writer refers to themselves in the third person as with any other cited author. Researchers may cite their past work on the same topic to show how their ideas have developed or to point to drawbacks and inadequacies of their previous approaches (Tadros 1993). Self-citation allows the writer to forego lengthy reiterations of their prior research. It is also conceivable that writers use self-citation to establish themselves as experts on their topic on account of their previous publications, though this assumption has not yet been tested empirically. It is interesting to note that K. Hyland (2004) disregarded self-citation in his analysis of citation practices in academic writing on the basis that it is of negligible importance in comparison to the citation of other authors. He did suspect, however, that self-citation is quite different from other-citation in terms of motivation and frequency of occurrence. In his 2008 study, K. Hyland found that self-citation was less common in the hard than in the soft sciences, which he explained in reference to the hard science's general tendency to downplay the role of the individual and the focus on scientific facts.

19 This and the following constructed examples are based on those presented by Swales (1990).

(2.4 ) Previous studies have alleged that linguistics students are the smartest (Linguis 1977).

It has been proposed that linguistics students are the smartest (Linguis 1977), but other researchers have called this claim into question (Lit 1981).

(2.5 ) According to some researchers, linguistics students are smarter than others.<sup>21</sup>

Linguistics students are assumed to be the smartest.<sup>21</sup>

(2.6 ) Linguistics students are probably the smartest (Linguis 1977).

It should be noted that even though the made-up examples above are all paraphrases, this is not the only intertextual strategy that occurs with either of the two citation forms. Direct quotes may also be combined with both integral and non-integral citation.<sup>20</sup> Integral citation is more frequent with direct quotation (Pickard 1995), and usually quotation fragments are integrated with elements of paraphrase in the same citation.

In rare cases, an integral and a non-integral citation may co-occur in the same sentence, for example when the writer wants to draw the reader's attention to further relevant literature in the field that is not cited in the sentence itself. The combination is sometimes employed in cases of contrasting views; see (2.7).

(2.7 ) Linguis (1977) has claimed that linguistics students are smarter than others (but cf. Lit 1981).

This special type of citation was first described by Jacoby (1987) using the term contrastive reference because the two citations in the sentence offer diverging views on the same issue. Its structure is similar to generalisation, which is discussed in section 2.2.3.

Aside from integral and non-integral citation, there is a third pattern which is a hybrid of the above two. It is predominantly found in science papers that use endnotes or a sequential numbering system in which each reference is given a number in the bibliography, which is then referred to in the text instead of the authors' names (see K. Hyland 1999; 2004). In this pattern, writers attribute knowledge not to a particular author, but to a particular reference from their list of works cited. It is exemplified in the sentences in (2.8) (quoted from K. Hyland 2004: 34). The reference number may also accompany a sentence without occupying a grammatical function therein, similar to a non-integral citation; see (2.9).

(2.8 ) According to ref. [11] the coupling parameters in the free electron ...

(Physics)

...using the highly efficient techniques described in [22], [23], and [25].

(Electrical Engineering)

...properties of a line trajectory in spatial motion are researched by Refs [21-23]

(Mechanical Engineering)

References [4, 5] reveal points with special kinematic meanings in the main body

(Mechanical Engineering)

(2.9 ) Linguistics students have been proposed to be the smartest [21].

Swales (1990) claimed that this sequential numbering system (which he referred to as a numerical/superscript system) does not integrate easily with reporting structures. This clearly does not apply to the science papers analysed by K. Hyland (1999), as several of the examples quoted in (2.8) are in fact

---

<sup>20</sup> See Tadros (1993) for a corpus study of integral and non-integral citations including direct quotes. There are apparently different views as to whether or not integral and non-integral citations may contain direct quotations. Okamura (2008a; 2008b) proposes that the inclusion of direct quotations in an analysis of citations would necessitate a third category. However, no details as to what such a category may entail are provided.

introduced by a reporting phrase, attesting to the considerable freedom afforded by such structures (see also section 2.3.2).

The choice of citation pattern determines the relationship between writer and author that is established in the text. It is generally assumed that writers consciously choose between integral and non-integral citation forms: “The use of one form rather than the other appears to reflect a decision to give greater emphasis to either the reported author or the reported message” (K. Hyland 2004: 23). From a pragmalinguistic perspective, the integral style is chosen when the writer wants to draw attention to the researcher or the particular study, for example to acknowledge the value of their contribution to the discipline. Integral citation puts the cited author in the focus of the sentence and thus gives them prominence. For this reason, integral citation has also been termed author prominent citation (Weissberg & Buker 1990, cited in Groom 2000). By placing the name in the focus, the writer positions the author in a dominant position in relation to themselves. Because integral citation is a salient feature of a sentence, it is likely to be interpreted as crucial to the argument (Okamura 2008a). Thus, by opting for integral citation, writers can exploit the syntactic structure of the citing sentence to place emphasis on the cited author (K. Hyland 2004), for example to shift responsibility for what has been stated, to align oneself with the particular author’s school of thought or simply to show that one is aware of important research in the field. The balance between writer and author responsibility, however, ultimately depends on the context (Ridley 2012).

Conversely, the non-integral form foregrounds the content and is a more implicit form of reference. It has also been referred to as information prominent citation (Weissberg & Buker 1990, cited in Groom 2000). Writers select non-integral citations to avoid breaking the flow of argumentation, because they can be integrated into a text more smoothly. Non-integral citation is the predominant pattern in both the hard and the soft sciences. In the natural sciences, scholarship is based on the assumption that academic writers report on observations of nature, but have no proprietary claims to them. It is therefore circumstantial who exactly it was that published it:

Removing the agent helps remove the implication of human intervention, with all the other influences of personal interest, social allegiance, faulty reasoning and other distorting factors beyond the empirical realm which that might suggest. (K. Hyland 1999: 356)

Authors’ names inadvertently draw attention to the person and potentially away from scientific argument. By using non-integral instead of integral citation, scientists direct the attention of the reader towards the subject of inquiry and emphasise the writer’s own argument or research finding rather than the work of others (K. Hyland 1999; Okamura 2008a; P. Thompson 2000).

The choice between integral and non-integral citation allows the writer to attribute material while creating a hierarchy of intertextual reference and the writer’s own voice (Groom 2000). Nevertheless, non-integral citation is not necessarily a sign that the cited information – or indeed the cited author – is less important. As P. Thompson (2000) points out, functional considerations may be overridden by stylistic ones. The alternation of integral and non-integral citation may create variety much like the use of diverse reporting verbs (see section 2.3.2). It is possible that both structures exist alongside precisely because of their stylistic and rhetorical potential (see also Swales 1990). Apparently, writer preference also plays a role: some prefer integral, some non-integral, and some mix both styles due to what may be “stylistic unease” about always repeating the same pattern (Swales 1986: 47). In addition, as suggested above, there are disciplinary preferences at play that determine a writer’s use of integral and non-integral citation.<sup>21</sup>

As discussed above, both integral and non-integral citations occur with a range of intertextual manifestations. Intertextual reference is usually created either in the form of direct quotations, i.e. directly copied excerpts from the source text which are marked by quotation marks, or paraphrases and summaries, which constitute the writer’s rendering of ideas from a source text in their own words. Writers may also use generalisation to refer to a publication without giving a detailed account of its contents. The following section provides an overview of these four major intertextual strategies in academic writing.

<sup>21</sup> For cross-disciplinary studies of citation patterns see K. Hyland 1999; Pickard 1995; P. Thompson 2000; P. Thompson & Tribble 2001.

## 2.2 Types of intertextual strategies

Academic writers' reference to previous literature in the field entails a complex interplay of intertextual strategies employed in the construction of knowledge. The use of intertextuality is determined by an intricate set of factors which tend to vary across research communities and cultures (Shi 2011). When integrating source material into their writing, writers have to identify ideational overlap between their own work and previous literature in the field. They have to make decisions about their intended line of argumentation and how information from their sources can be used to build it. References to other authors' work create a context for the writer's argument and establish links with the existing literature. They show which theories the writer uses as a framework for their research and whose work they were inspired by. At the same time, intertextual strategies can be employed to provide support for one's argument and to show why one's research is relevant and new, which is usually achieved by pointing out shortcomings of previous studies. Different intertextual strategies afford writers with different ways of highlighting the ideas they are reporting. This means that the choice of intertextual relationship is determined by the way in which a writer intends to frame their argument and consequential in terms of whether or not it convinces the reader (K. Hyland 1999).

Across disciplines, the most important intertextual strategies are direct quotation, paraphrasing, summarisation of a single source text, and generalisation from several sources (Dubois 1988; K. Hyland 2004). In analogy to the terms of direct and indirect speech, these strategies can be divided into strategies of direct source use and indirect source use (Gebriel & Plakans 2009), see Figure 2. They are viewed as core academic writing activities (see e.g. Hirvela & Du 2013) and are discussed in the following sections.

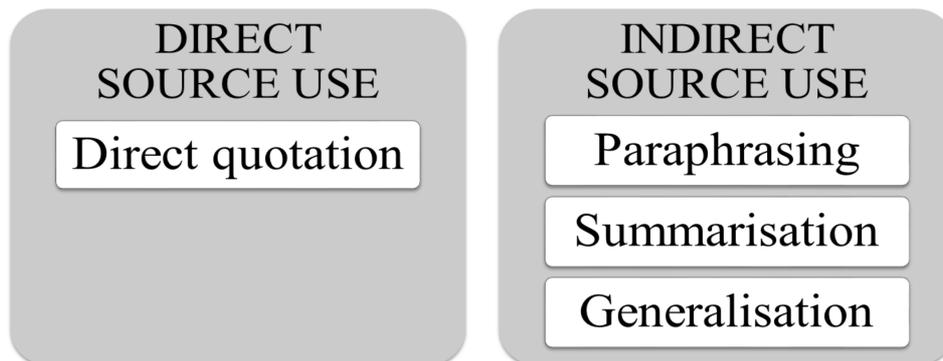


Figure 2: Distinction of intertextual strategies based on direct and indirect source use.

### 2.2.1 Direct quotation

Direct quotation is defined as the “verbatim repetition of textual material taken from a source” (Petrić 2012: 102). It occurs in speech (see Clark & Gerrig 1990) and a variety of written genres. A direct quote is identical with the original statement and thus faithful to both its form and its content. To identify a passage as a direct quote, it is placed within quotation marks or otherwise typographically marked, e.g. by indentation, wider margins, and font size.<sup>22</sup> Like other intertextual strategies, direct quotes are commonly introduced by means of reporting structures (see section 2.3.2) and accompanied by a reference in an integral or a non-integral citation.

<sup>22</sup> In a relatively broad definition, Dubois (1988) also considers borrowed sequences of at least three words identical with the source text to be direct quotes, even if they are not embedded in quotation marks. The more common definition of direct quotes as passages in quotation marks or indented as block quotes (see e.g. Petrić 2012; Verheijen 2015) is adopted here.

They allow writers to “partly or wholly detach themselves from what they depict” (Clark & Gerrig 1990: 792) by using someone else’s words and attributing the respective proposition to that person.

Direct quotation is a characteristic feature of academic writing. It has been explored in expert and novice academic writing alongside other strategies of source use, but it is usually not discussed in depth because direct quotes are relatively marginal in quantity compared to other intertextual manifestations (see e.g. K. Hyland 2000). Studies of direct quotation in academic texts by English native speakers have focused on the principles that govern the choice of direct quotation over other intertextual strategies, on the rhetorical functions of direct quotes, and on disciplinary differences (see e.g. Hu & Wang 2014; K. Hyland 1999; Jakobs 2003). Another aspect of study has been the comparison between L1 and L2 speakers of English, both in novice and expert writing (e.g. Ädel & Garretson 2006; Campbell 1990; Borg 2000; Shi 2004; T. A. Hyland 2009).<sup>23</sup>

In academic texts, as in most written genres, direct quotes are conventionally signalled by quotation marks on either side of the unmodified original utterance. Most style guides recommend double quotation marks (“”) for direct quotes and inverted commas (‘’) for quotes inside quotes (e.g. the APA and MLA style guides, see Purdue Online Writing Lab 2016).<sup>24</sup> Nevertheless, some writers prefer single quotation marks (e.g. Tadros 1993), which is the standard format in British English (Oxford Dictionaries 2016). Quotation marks indicate to the reader that the words between them are someone else’s. However, their function goes beyond being merely identificational, since quotation marks also “promise the reader that the words are repeated accurately, that they have not been taken out of context, that they come from the work of the author who is named” (Pecorari 2006: 6). Direct quotes are distinguished from illegitimately borrowed material by being fully acknowledged according to academic citing conventions.

Direct quotes fulfil a variety of interactional purposes. They are used when it is considered necessary to maintain the original wording, for example if the words are judged as very appropriate or elegant. Writers also choose to quote if changes to a quoted passage would alter its proposition or if it is believed to lend authority to a claim. Direct quotes are relatively infrequent in most disciplines because writers only employ them only when they “consider them to be the most vivid and effective way of presenting their case” (K. Hyland 2004: 26; see also Dubois 1988). A writer may choose direct quotation for stylistic reasons, for example to make a text more lively or dramatic (Petrić 2012). Furthermore, direct quotation may serve to shift responsibility for what is said, to indicate an attitude towards it or to signal objectivity through the attribution of the quoted idea to a reliable source. Direct quotation thus functions as a linguistic strategy to provide an authentic reproduction of an utterance while at the same time distancing oneself from its proposition, if required.

Unlike paraphrases, in which the writer makes grammatical and lexical changes to the cited excerpt (see section 2.2.2), direct quotes introduce another author’s voice into the text in an unmediated way. Thus, quotes have to be carefully selected for content as well as tone to be smoothly integrated into one’s writing. For many academic writers, “it is the vivid and succinct expression of an important idea, where both the content and its linguistic expression are equally effective, that renders the source suitable for a direct quotation” (Petrić 2012: 111). The choice of which passage to quote and how to embed it is usually guided by functional considerations. Varied and purposefully embedded direct quotes are seen as a sign of sophistication and advanced academic literacy (Davis 2013; Verheijen 2015), because the seamless co-textual and contextual integration of direct quotes into a new environment is relatively complex.

In order to effectively integrate a direct quote into the text, it has to be adapted not only to the communicative context, but also to the linguistic co-text (Jakobs 2003). Academic writers must have a high degree of linguistic competence and academic literacy in order to make the necessary linguistic adjustments to the selected source text passage (Petrić 2012: 103). The integration of quotes also entails making grammatical,

<sup>23</sup> See section 3.3.1 for a survey of research into direct quotes in L2 writing.

<sup>24</sup> Despite these widely accepted conventions, short strings copied from a source text are not always indicated by quotation marks, at least in some disciplines (see Dubois 1988).

lexical, and/or syntactic adjustments to the co-text in accordance with the linguistic make-up of the quote.<sup>25</sup> Such changes are necessary because insufficient adaptation leads to morphological and/or syntactic discontinuities which disrupt the flow of a text. In addition to formal adaptations, writers have to achieve a meaningful incorporation of a direct quote into the new context in the sense that the quoted excerpt has to be tailored to its intended rhetorical function. In order to achieve this, writers have to ensure that their intention in quoting is clear, that the quote is not taken out of context, and that its purpose is interpretable by the reader (Jakobs 2003). This presupposes that the writer frames the direct quote in a way that allows the reader to decode its purpose and the writer's stance towards it, e.g. by suitable reporting structures or comments on the quote. Finally, it may be desirable to link a quote to related disciplinary discourse, for example by combining quotes from different sources in the same sentence (see Verheijen 2015).

An important choice a writer has to make when selecting a direct quote regards the length of the quoted passage. The writer has to take care not to change the meaning of the direct quote by quoting too little. Conversely, too lengthy a quote could negatively affect readability and might lead to redundancy or the impression that the writer relies too heavily on the ideas of others. Based on their length, direct quotes can be sub-categorised into quotation fragments (shorter than a t-unit), brief quotations (one t-unit or longer, but less than forty words), and extended quotations (longer than forty words and usually inserted as block quotations) (Borg 2000; Petrić 2012). Quotation fragments have to be incorporated syntactically and semantically in order to create coherence, whereas the other types may be used without co-textual modification. The integration of quotation fragments thus requires more intervention, but also affords the writer more syntactic freedom (Petrić 2012). Of course, which length is appropriate for an effective direct quote depends on the writer's goal and the context as well as the linguistic make-up of the quoted passage.

There are three main ways of incorporating direct quotes into one's writing. They may be integrated into the sentence structure as an embedded quote, set off from the body of the text and indented as a block quote or placed before the text as an independent quote. Quotations of all of the above types can be nested, as a quoted passage may already contain a direct quote. Embedded quotations are passages from the source text which have a syntactic function in the writer's text. The quote can constitute an entire clause or a phrase or a component of a phrase (Verheijen 2015). Style guides typically recommend that embedded quotations should be shorter than forty words or no more than three lines of running text.<sup>26</sup> Writers may also choose to combine quotes from several sources in the same sentence for a more balanced argument. This is viewed as a sophisticated use of direct quotes and is common in expert writing (Petrić 2012; Verheijen 2015). Embedded direct quotes often take on the syntactic function of objects in sentences in which they are introduced by reporting verbs (Clark & Gerrig 1990). Of course, sentences may also be quoted in full. This type of embedded direct quote is not uncommon in L2 writing, yet is avoided by expert writers (Verheijen 2015).

The second type are block quotes, which are also referred to as long quotes, independent quotes, extended quotes, extensive quotes, non-integrated quotations, and longer quotations (K. Hyland 1999; Petrić 2012; Pickard 1995; Swales 1990; Verheijen 2015). Block quotes form part of the text, but are not always syntactically embedded into a sentence. They are typically preceded by a complete sentence or a reporting structure which introduces them. Usually, block quotes are indented on the left-hand side, set in smaller font, and used without quotation marks (Verheijen 2015). Most style guides recommend block quote formatting for direct quotes that are comprised of four or more lines of running text or forty or more words (Purdue Online Writing Lab 2016). Independent quotes, in contrast, occur extra-textually at the beginning or end of a paragraph. They stand alone and are neither introduced nor incorporated into the text, though they are sometimes referred to (Verheijen 2015).

25 Orthographic or lexical changes may have to be made to the direct quote itself, for example when an excerpt beginning in a capital letter is quoted in the middle of a sentence. In such a case, the upper-case letter is replaced by its lower-case equivalent in brackets. Words that have been added to the quote by the writer, for example for clarification of deictic references or for reasons of syntactic adaptation, are added in brackets as well. Conversely, a quote may have to be shortened so as to create a smooth transition to the writer's own words. Undesired or irrelevant material may have to be elided, but in such a way that the original proposition is not altered. Ellipsis is conventionally signalled by three full stops in brackets: [...] (see Purdue Online Writing Lab 2016).

26 See Purdue Online Writing Lab 2016 for a summary of APA and MLA recommendations.

Other terms for independent quotes without textual integration are ‘hortatory quotes’ (Borg 2000) and ‘stand-alone quotations’ (Petrić 2012). They, too, are usually indented. Independent quotes are placed at the beginning or end of a text to “exhort or vividly express the article’s point of view by the selection and presentation of another’s words” (Borg 2000: 33). In expert writing, they may also appear in footnotes to give further support to a statement (Verheijen 2015).

Further issues to be considered by academic writers when using direct quotes relate to the frequency of direct quotations in the text. Though it is often regarded as a central feature of academic texts, direct quotation is peripheral in comparison to other intertextual strategies (Petrić 2012). It is virtually non-existent in natural sciences and engineering, used sparingly in economics and psychology, and constitutes a minority of citations in the arts and humanities (Dubois 1988; K. Hyland 2004; Pecorari 2006; Swales 2004; P. Thompson & Tribble 2001). There are also disciplinary differences in the preferred length of direct quotes as well as in the understanding of how long a borrowed passage has to be to require quotation marks. Studies focusing on the frequency of direct quotation in L1 student writing have shown that disciplinary differences also exist at the undergraduate and postgraduate level (Ädel & Garretson 2006; P. Thompson 2001).<sup>27</sup> Academic writers thus not only have to decide whether or not to quote on the basis of the salience of the excerpt, but also on account of whether direct quotation is a desired intertextual practice in the respective academic community.

In most disciplines, especially the natural sciences, it is considered more appropriate to rephrase previous literature using one’s own words (see e.g. Dubois 1988; P. Thompson & Tribble 2001). Writers tend to prefer restating another author’s words because this affords them with more flexibility in framing their claims (K. Hyland 2004). This entails the use of paraphrases and summaries, which are presented in the following section.

## 2.2.2 Paraphrasing and summarisation

Paraphrasing and summarisation are two further core intertextual strategies in academic writing. In a study of research articles, summaries were found to be the most frequent intertextual manifestation across eight academic disciplines (K. Hyland 1999).<sup>28</sup> A study of biomedical journal articles (Dubois 1988) has also suggested that summarisation is the most frequent intertextual strategy, whereas paraphrasing is not used by all writers. Nevertheless, paraphrasing is also considered to be a “core academic reading/writing technique” (Hirvela & Du 2013: 87). Both strategies usually co-occur in summary and other academic writing. Paraphrasing and summarising material from a source requires the writer to reconcile a range of complex and recursive (meta-)cognitive processes, for example planning and reflecting on their text and evaluating and revising it (Kirkland & Saunders 1991). Summarisation and paraphrasing have been referred to as indirect strategies to contrast them with direct quotation, a direct source use strategy (see Gebril & Plakans 2009). As noted above, direct quotes may occur within paraphrases or summaries (G. Thompson 1990).

Summaries and paraphrases are distinguished either in reference to their specificity or their length. In Hirvela and Du’s (2013) definition, paraphrasing is the act of re-writing an original sentence using a combination of source text material and the writer’s words. As a result, a paraphrase retains the level of specificity provided in the source text. Summarisation, on the other hand, condenses the source text by focusing only on the main points without recasting detailed information. Summaries contain information from several source text sentences or even paragraphs, but typically do not re-use the original wording. While Hirvela and Du’s distinction is based predominantly on the specificity of the reported content, Dubois’ (1988) definitions take into focus the length of

<sup>27</sup> Quotation is generally more frequent in student than in expert writing. In linguistic articles, 10% of citations are direct quotes in published expert writing (Hyland 1999) and exceeded 20% in L1 student writing (Ädel & Garretson 2006), which points to differences between writer groups within disciplines. The students’ proportionally heavier reliance on verbatim reproduction distinguishes them from experts in this aspect of academic writing.

<sup>28</sup> Note that K. Hyland (1999) does not distinguish between summaries and paraphrases and it is hence likely that his summary category includes citations that may also be considered paraphrases.

the sentence. A paraphrase, in her taxonomy, contains words that are different from those in the source text, but the length of the sentence is the same. A summary, in contrast, is described as an “abbreviated statement” (Dubois 1988: 183) of an idea found in a source text. In the following, the two strategies are presented in turn with the definitions presented above in mind.

The ability to paraphrase, as Hirvela and Du (2013: 88) assert, is “an important marker of a scholar’s (or a student’s) understanding of a source text”. Paraphrasing is used to acknowledge previous research, to provide a context for the writer’s argument, and to situate it in the disciplinary discourse (see K. Hyland 1999). Combining the author’s vocabulary and grammatical structures with the writer’s own language allows the writer to identify a proposition as someone else’s without being any less specific than the source text (Hirvela & Du 2013). Paraphrases are also a means of conveying an evaluation of another author’s ideas and thus positioning one’s study in relation to the cited research (see Groom 2000).

In a paraphrase, the writer recasts a passage from an author’s text using their own words and grammatical structures; see example (2.10). The writer typically works with an individual sentence or clause, which is rephrased and restructured in several iterations with the goal of retaining or at least approximating the original proposition (Flower et al. 1986). Keck (2006) defines paraphrases as passages reproducing the source text that contain at least one lexical change, e.g. the substitution with a synonym. Her definition thus regards a certain lexical and grammatical overlap as a given.

(2.10 ) *Source text:*

Listening to this group [...] reveals how Māori English helps to create and define Māori students’ identity within the confines of physical and social spaces both on and off campus. The study shows that for these Māori students, Māori English functions as an important emblematic marker of their group identity. (King 1999: 36)

*Paraphrase:*

In conclusion, the most important finding of King’s research is that ME is a mechanism used to create group identity in physical and social spheres, especially where Māori constitute a minority. (CALE; RR1.G.HB.104)

Paraphrases are similar to indirect speech in the same way that direct quotes share characteristics of direct speech (see extensive discussion in G. Thompson 1990). For example, statements are usually paraphrased in *that*-clauses and questions in *wh*-clauses. Writers may also introduce modals or *to*-infinitives to encode future reference or requests made in the cited original. According to G. Thompson (1990), paraphrases are distinct from other forms of intertextuality in that the form and style of the message are adapted to the needs of the writer. Features of the original context are not necessarily carried over into the paraphrase, as questions may be paraphrased as statements, for example. Reporting structures can be additionally introduced to attribute the content and signal the writer’s stance towards the reported information (Charles 2006b; see also section 2.3.2). However, since reporting is not obligatory with paraphrases, they may be difficult to distinguish from the writer’s own ideas in an academic text (G. Thompson 1990; see also Wiemeyer 2017b).

There are diverging views in the literature on what constitutes a paraphrase and how much source text material it may contain. This is possibly caused by the fact that most paraphrasing studies focus on student writing. To some researchers, paraphrasing does not entail textual overlap with the source. Campbell (1990) categorised sentences whose wording was close to the source text as near copies. She treats these separately from paraphrases, which, in her definition, entail syntactic changes. In reference to her categorisation, Crocker and Shaw (2002) emphasise that if the writer retains the structure of the original sentence and the sentence contains chunks from the source text, the writer is near-copying, not paraphrasing or summarising. They position this practice between quotation and paraphrasing on their second dimension of intertextuality, closeness of wording to the source text. Howard, Serviss, and Rodrigue (2010: 181) define paraphrases as consisting of “fresh language”, but concede that key words may be carried over from the source text.

Shi (2004: 178) shares the view that paraphrases may, but must not necessarily entail textual borrowing from the source text passage. She regards sentences that do not contain copied strings of three words or more as total paraphrases and does not include them in her analysis of textual borrowing. She includes the term ‘paraphrase’ in only one of her subcategories of textual borrowing, namely “closely paraphrased by reformulating syntax or changing the wording of the original text”. This category suggests an understanding of paraphrases as being either reformulated or restructured, not both. There is no clear-cut distinction between this and another subcategory which is applied to sentences to which slight modifications were made, for example via synonym substitution. Such sentences are not considered to be paraphrases in Shi’s study, suggesting that a certain amount of modification is viewed to be a prerequisite for a sentence to be considered a paraphrase.

In contrast, Keck (2006) argues that the full scope of paraphrasing can only be explored if paraphrases that contain substantial textual overlap with the source as well as those that do not contain traces of the original wording are considered. In her taxonomy, paraphrases are categorised based on the percentage of textual overlap. She used software to identify and count so-called ‘unique links’, i.e. single words or strings which are unique to one source text sentence. Paraphrases consisting of unique links to more than 50% were categorised as Near Copies, those consisting of 20 to 49% unique links are Minimal Revisions, 1 to 19% unique links define a Moderate Revision, and no unique links are found in Substantial Revisions. Unlike Keck, Cumming et al. (2005) do not assume that the amount of textual borrowing is a defining feature of paraphrases. Instead, they treat textual borrowing as a separate variable that may be found in sentences functioning as declarations, quotations, paraphrases, and summaries.

These diverging definitions show that although it is generally assumed that paraphrases are created by rephrasing and restructuring source text material, there is no consensus on whether and how much textual borrowing may occur and whether it is defining of paraphrases. This is in line with studies showing that there are different opinions on how much textual overlap is appropriate in source-based writing both among lecturers and among students (see survey of plagiarism studies in Pecorari & Petrić 2014).

The definitions above place their focus on lexical and syntactic features of paraphrases. However, it has been argued that paraphrasing is more than changing wordings and syntactic structures. According to Yamada (2003), skilled paraphrasing necessarily entails inferential thinking, as expert writers form links and analogies between the information presented in a source text and their world knowledge. They rely on the results of their inferential thought processes to create paraphrases in which the original content is re-framed and rephrased in resourceful ways. In addition to using one’s own words, academic writers thus also need to introduce their own interpretation in paraphrasing (Shi 2012).<sup>29</sup> Nevertheless, it remains difficult to define what makes a paraphrase successful and hence what is required of students in paraphrasing (Hirvela & Du 2013). This is evident from the varying assumptions in paraphrasing studies with regard to how textual borrowing defines a paraphrase, as discussed above.

Summaries are sentences which provide a general summary of a particular text or of a part thereof. They contain the key points of a publication; see example (2.11).

- (2.11 ) In her research paper “Talking bro: Maori English in the University Setting” from 1999, King describes and discusses the findings of her study on how Maori students perceive the ways in which Maori English (ME) is used in different social contexts. (CALE; RR1.G.HB.106)

Like paraphrases, summaries express the meaning of the source text in a different form. They may be viewed as distinct from paraphrases because the writer summarises the content of a publication or its parts rather than providing a rendered version of individual passages from it. Summarisation requires the writer to identify important ideas in a source text and to condense sentences and paragraphs from a source text into a brief statement (Kim 2001; Shi 2012). The purpose is to “get to the gist” (Shi 2012: 136) while maintaining the tone

<sup>29</sup> Shi (2012) argues that paraphrasing is similar to translation as in both situations the writer has to interpret the original text before they can write their own version.

intended by the author. This is achieved by operations such as the selection of important points, the omission of negligible or redundant information, and the collapsing of several ideas into a unit. The outcome is a sophisticated, recursive cognitive process of concisely rendering the original information (Kim 2001).

The resulting summary has to be meaningfully embedded into one's own text. Depending on its purpose and location in the text, it may contain minimal information or constitute a lengthy summary of the original text. Structurally, summaries may range from a noun phrase or prepositional phrase to a complete sentence (G. Thompson 1990). G. Thompson (1990) maintains that the reporting verbs used with summaries do not necessarily occur with paraphrases or direct quotes and may carry part of the meaning of the summary, for example the verb *criticise*. He also contends that summaries are functionally different from paraphrases: in paraphrases, the focus is on what was said, while a summary creates a broader focus, for example to include an explanation of what was said. Though summaries are investigated as one type of intertextual link in a number of studies of academic writing, they tend to be used as a category without definition (see e.g. K. Hyland 1999, 2004; Wette 2010).

Paraphrasing and summarisation are more versatile than direct quotation and tend to be preferred when writers use source text material to provide support for their own arguments. While direct quotation may be appropriate when the writer aims to argue the same point as the original author, "paraphrasing or summarizing enables him/her to absorb a cited proposition into his/her argument with rhetorical flexibility" (Hu & Wang 2014: 22). This is assumed to be the reason why these two strategies are preferred in academic writing (K. Hyland 1999). In paraphrasing and summarisation, academic writers incorporate aspects from the literature into their writing in a way that is best suited to providing evidence for their own claims. Writers can make their interpretation of the cited excerpt more explicit than when quoting while at the same time emphasising only those aspects which they consider most relevant (K. Hyland 2004).

While illicit copying and patchwriting have been studied quite extensively, paraphrasing and summarisation have thus far received relatively little attention in the linguistic literature (see section 3.3.2). One reason for this may be that their distinction is not as clear-cut as suggested by the definitions presented above. Since paraphrasing often entails omission of source text material, definitions based on the comparative length of summaries and paraphrases compared to the original source text passage (e.g. Sherrard 1986) are difficult to apply. Furthermore, because the level of specificity is a gradual phenomenon, the distinction of the two strategies based on this criterion (e.g. by Hirvela & Du 2013) is also problematic.

It thus comes as no surprise that the terms paraphrasing and summarisation are sometimes collapsed into one category or used interchangeably (see e.g. K. Hyland 1999; Petrić 2007), especially since paraphrasing is a major strategy employed in summary writing. Scholars have decided in the past to discuss them together because raters disagreed on their categorisation (Gebril & Plakans 2009) or because the lines were blurred and the frequency of potential summaries was comparatively low (Wette 2017). It is clear that though definitions for the two strategies as distinct categories exist, paraphrases and summaries are similar in many respects. Because they share many features, paraphrases and summaries are presented in the same section here. It remains arguable whether they really constitute two separate categories, as discussed further in the study presented below.

### 2.2.3 Generalisation

A generalisation is a general statement about a publication made by the writer without providing a summary of the cited publication and without paraphrasing from it. It is based on a large part or the entire text rather than individual passages (Dubois 1988). Often, this type of intertextuality draws on and is accompanied by references to two or more sources, typically in a non-integral citation. It thus ascribes the cited information to more than one author (Charles 2006a; K. Hyland 2004; Petrić 2007; see example 2.12). Generalisation in such

cases serves to provide information about the current state of research in the discipline instead of attributing ideas to an individual researcher (Jacoby 1987). It emphasises the writer's awareness of what the respective discourse community considers to be common knowledge (Petrić 2007).

- (2.12 ) It is generally acknowledged that the relationship between the writer of an academic text and the sources cited in that text is not entirely straightforward (Howard, 2000; Hyland, 2000; Pecorari, 2006; Pennycook, 1996; Shi, 2010). (Wette 2010: 159)

Generalisations are not necessarily cases of attribution in the narrow sense as they may fulfil the purpose of mentioning a source rather than attributing content to it. They are nevertheless important cohesive and referential devices in academic writing (Ädel & Garretson 2006). A generalisation can also be a "statement of similarity" (Dubois 1988: 183) in that it either compares the writer's own work to that of others or compares several cited sources. P. Thompson classified this type of citation as a 'Refer' citation, which he considered a kind of "shorthand device" (2000: 94). The writer employs this to refer the reader to another publication which may not be discussed in the writer's own text and which is usually, although not necessarily, accompanied by the word *see*. The excerpt in (2.13) is an example of how generalisation can be used to provide the reader with an overview of similar studies in the field. In the excerpt, the writer alludes to the content of the referenced publications without providing detailed information about it. There is no summary of either the scope of the studies or the findings. The reader is thus informed about related and relevant research, but it is left to them to consult the reference should they be interested in details. Generalisation in such contexts may be used to establish a research niche and to document the extensive literature review on the part of the author.

- (2.13 ) The nature and functions of direct quotation in both speech and writing have been examined in pragmatics, linguistic theory, and philosophy of language (e.g., Clark & Gerrig, 1990; Thompson, 1996). (Petrić 2012: 103)

- (2.14 ) Many scholars have argued that...  
It is commonly believed that...

Generalisations are often introduced by specific reporting structures to highlight that the writer is attributing knowledge to a certain school of academic thought or a group of researchers, as in the examples in (2.14). Dubois (1988: 185) asserts that the high degree of abstraction in generalisations and summaries may sometimes make them "difficult for a lay reader to verify, with some cases requiring several detailed rereadings". They are, however, the second most frequent type of intertextual link in research papers across disciplines. This has been explained in reference to the flexibility generalisation offers to the writer in terms of emphasising and interpreting cited material in a way that most suitably supports their argument (K. Hyland 2004). Some disciplines may lean more towards generalisation than others; for instance, generalisation was the most frequent strategy in biomedical journal articles in Dubois' (1988) study. Natural science papers generally contained a larger number of generalisations than the humanities and social sciences in K. Hyland's (2004) study. This preference for generalisations can be explained in reference to the natural sciences' tendency towards general statements about the state of the art instead of the more detailed summaries of other studies that are preferred in the humanities and social sciences (see K. Hyland 2004). Generalisations allow the writer to contextualise their research without giving greater prominence to other authors.

An overview of the features of the four core intertextual strategies direct quotation, paraphrasing, summarisation, and generalisation, appears in Table 1.

Table 1: Overview of major intertextual strategies and their features.

	<b>Direct Quotation</b>	<b>Paraphrasing</b>	<b>Summarisation</b>	<b>Generalisation</b>
<b>Definition</b>	Direct quotation is the verbatim repetition of a string of words from a source text, conventionally embedded in quotation marks highlighting the words as someone else's.	Paraphrasing is the act of rephrasing individual sentences or short source text passages using one's own words and alternative syntactic structures	Summarisation is the act of summarising the content of entire sections, articles or even books using one's own words, thereby providing an overview of extended sections of the source text.	Generalisation is the act of making a general statement about one or more publications without providing details of the content and without paraphrasing from them.
<b>General Features</b>	<ul style="list-style-type: none"> <li>explicitly marked as quoted through the use of quotation marks (“”)</li> <li>modifications such as omissions and insertions marked using square brackets, e.g. [...]</li> <li>usually embedded in a paraphrase</li> </ul>	<ul style="list-style-type: none"> <li>restatement of the original idea in a new context, often characterised by a restructuring of information</li> <li>usually based on individual sentences and short passages of the source text</li> <li>combines ideas (and words) from the source text with those of the author: identical proposition as the original but different linguistic form</li> <li>same level of specificity as the source text</li> </ul>	<ul style="list-style-type: none"> <li>contains the gist of the summarised section without detailed recasting of information</li> <li>typically does not entail paraphrasing of individual passages</li> <li>less specific than the source text: focus only on main points without details</li> </ul>	<ul style="list-style-type: none"> <li>a publication is mentioned e.g. as a possible reference to consult</li> <li>content is neither paraphrased nor summarised, no details provided</li> <li>less specific than the source text: focus only on general content without details</li> <li>often found in introductory sections</li> </ul>
<b>Textual Borrowing</b>	<ul style="list-style-type: none"> <li>quoted passage is identical with source text passage in both form and content</li> </ul>	<ul style="list-style-type: none"> <li>sometimes contains instances of linguistic overlap with the original, especially in terms of terminology</li> <li>textual overlap of 3+ words is usually avoided</li> </ul>	<ul style="list-style-type: none"> <li>occasionally contains instances of linguistic overlap with the original</li> <li>textual overlap of 3+ words is usually avoided</li> </ul>	<ul style="list-style-type: none"> <li>rarely contains instances of linguistic overlap with the original</li> <li>textual overlap of 3+ words is usually avoided</li> </ul>
<b>Reporting and Documentation</b>	<ul style="list-style-type: none"> <li>reference usually provided</li> <li>typically framed by reporting structure</li> </ul>	<ul style="list-style-type: none"> <li>reference may be provided in the same sentence or in the co-text</li> <li>optionally framed by reporting structure</li> </ul>	<ul style="list-style-type: none"> <li>reference may be provided in the same sentence or in the co-text</li> <li>optionally framed by reporting structure</li> </ul>	<ul style="list-style-type: none"> <li>reference(s) provided</li> <li>optionally framed by reporting structure</li> </ul>
<b>Comment</b>		<ul style="list-style-type: none"> <li>sometimes used synonymously with summarisation</li> </ul>	<ul style="list-style-type: none"> <li>sometimes used synonymously with paraphrasing</li> </ul>	

## 2.3 Signalling intertextuality: attribution, documentation, reporting structures, and evaluation

According to academic writing conventions and expectations, intertextual passages must be clearly marked as such to allow the reader to identify their nature. Instances of intertextuality are indicated using a range of means such as reporting structures and overt attribution of textual material. The presence of other authors in academic writers' texts is largely dependent on their use of these intertextual signals. Academic writers interweave their own ideas and arguments with those of other authors in the field. A given source-based academic text contains two types of text passages: those which consist of the writer's own words and purvey their own ideas, and those which cite the ideas and opinions of others. The first kind is averred, the second is attributed. It is paramount for academic writers to clearly mark these passages so that the distinction is clear to the reader. In attributed passages, academic writers thus have to acknowledge the source in order to show where the knowledge in their writing stems from.

There are three main ways of achieving this. The perhaps most visible is the documentation of sources, also referred to as referencing. This entails providing a bibliographical reference to the cited source in the text (example 2.15) as well as in the bibliography. Another way to acknowledge the origin of an idea is the use of attribution, i.e. making explicit in the text who the originator of the idea is. Attribution can occur in a variety of ways, for example via overt reference to the author, their name or their text; see examples in (2.16). Writers provide references to both attribute content and document their sources; that is, providing a reference is one way of attributing content to a source (see section 2.3.1). Attribution is thus possible in integral and non-integral citations.

A third important means of highlighting the source of cited content is via reporting structures, i.e. phrases which make explicit that something is reported from a source. Reporting structures allow academic writers to frame cited material in a way that supports their argument. Typically, these phrases contain a reporting verb (example 2.17), but other structures, for example containing complex prepositions (example 2.18), are also attested (see section 2.3.2). Because reporting structures often contain author names, they may also be used as a means of attribution. The choice of reporting verb plays a decisive role in encoding the writer's stance towards another author's idea. Writers have two basic choices when embedding citations in reporting structures: voicing approval, that is, endorsing the author and committing to the content of the citation, or voicing scepticism and distancing themselves from it. The choice signals whether or not the writer regards a cited claim as justified, as discussed in section 2.3.3.

(2.15 ) King (1999: 18) claims that  
... are often the cause of perceived plagiarism (Müller 2003).

(2.16 ) The authors claim that ...  
According to Meyer, ...  
The study shows that ...  
The chapter highlights...

(2.17 ) Thompson and Ye (1991) acknowledge that counter-factive use of reporting verbs is rare.

(2.18 ) According to Thompson and Ye (1991: 372), academic writers can use reporting verbs to encode their stance towards the cited information.

As Figure 3 shows, there is a certain amount of overlap between these ways of signalling intertextuality. Both reporting structures and documentation may entail attribution. A reporting structure may contain a reference

documenting the writer's source. Finally, evaluation is often achieved through the choice of a pertinent verb in a reporting structure.

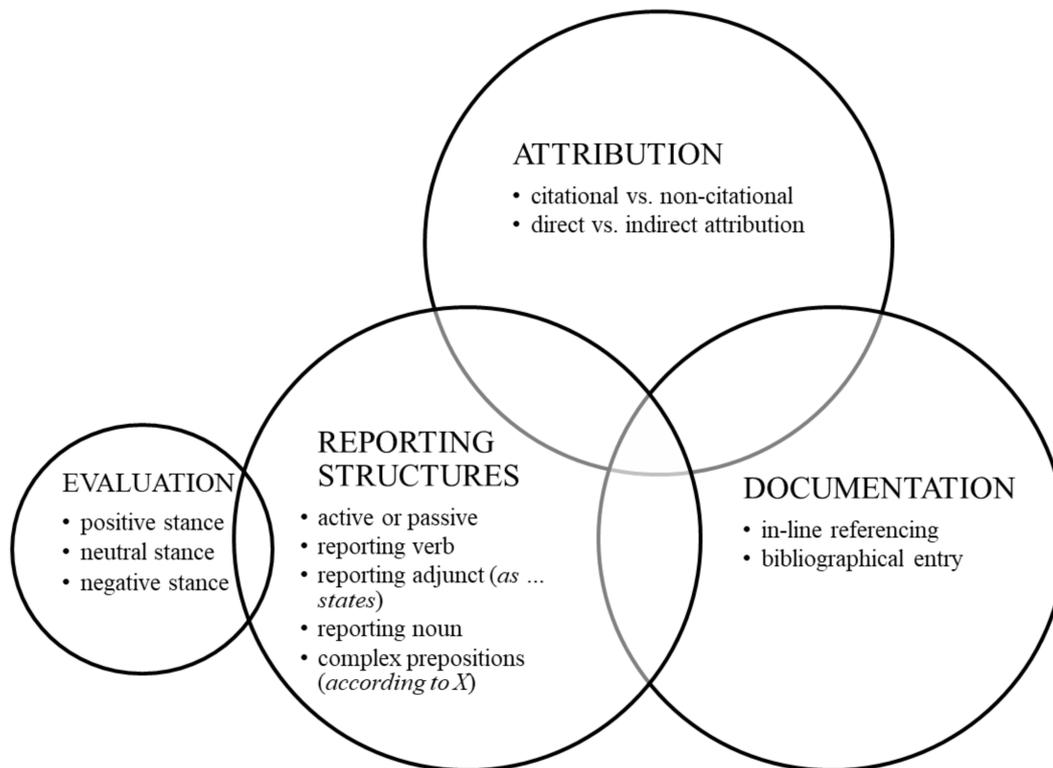


Figure 3: Strategies of signalling intertextuality in academic writing.

### 2.3.1 Attribution, averral, and documentation: Distinguishing voices in the text

Readers of academic texts are aware that these are almost indispensably based on sources, and that source-based passages are blended into the writer's own argumentation. Passages containing the writer's own thoughts, ideas, and concepts are instances of averral, while passages containing material from clearly signalled sources are instances of attribution. The distinction between averral and attribution goes back to Sinclair (1988). When reading an academic text, the reader assumes that the writer "avers every statement in his or her text so long as he/she does not attribute these statements to another source – whether that source is other or self" (Tadros 1993: 100). Thus, unmarked passages in the text are assumed by the reader to embody the voice of the writer, who assumes full responsibility.<sup>30</sup> According to Tadros (1993), averral can be identified by the reader via the use of first person pronouns, the writer's comments or the statement of an attitude (positive averral). Alternatively, any passage which does not contain attribution is by default averred (negative averral).

Aside from averred statements, any academic text will contain passages which are overtly and transparently attributed to other authors by naming, alluding to and/or referencing them. These passages are

<sup>30</sup> The ultimate responsibility for a published academic text lies with its writer, with some minor responsibility taken by the editor and publisher (Sinclair 1988).

intertextual in nature. Attribution is defined as “the use of a manifest intertextual marker to acknowledge the presence of an antecedent authorial voice” (Groom 2000: 15). Through attribution, the writer signals that a proposition, a term or a passage of the text is taken from or based on another source and acknowledges someone else’s authorship. Attribution plays a crucial role in academic writing because it is a central means for the writer to interact with their reader (K. Hyland 2004). Its general purpose is to help the reader identify the originator of a particular proposition, but attribution fulfils several functions: it may be used for support of an argument, to give credentials, or to report diverging points of view (Tadros 1993). The purpose of a particular attribution depends on the writer’s goal for their text in general (see section 2.3.3).<sup>31</sup> Its success depends on the reader’s correct interpretation of this purpose, which sometimes requires reading between the lines (Hunston 1995).

Averral and attribution are interdependent (Groom 2000). As a consequence, they may co-occur in the same sentence without rendering it ambiguous (Tadros 1993). Averral signals to the reader that the writer does not feel the need to draw support for their view from other authors and that they are confident that their work is eligible and credible in and of itself. In contrast, attribution defers responsibility for a proposition to someone else, at least to a certain extent (Sinclair 1988; Groom 2000). Tadros argues that writers consciously choose to aver statements rather than to attribute them to others because “[c]itation will shake their position as possessors of the key to knowledge” (1993: 108). Groom (2000) expands on this argument by conceptualising averral and attribution as two poles of a continuum of what he refers to as propositional responsibility. The different points on the continuum are distinguished by different degrees of presence of the writer, the writer’s positioning towards the author, and the writer’s evaluation of the proposition (or lack thereof). According to Groom, successful argumentative writing positions the writer as dominant in the text, while writers of less effective texts tend to defer responsibility to the author instead and assume a subordinate position.

Attribution is often realised in the form of a reference. Tadros (1993) distinguishes two broad categories of attribution based on whether a reference is given or not: citational and non-citational attribution. In non-citational attribution, the writer attributes content not to an actual source but to a group of people such as a school of thought or an entire discipline without explicit mention of individuals. Accordingly, no reference is given and the writer neither paraphrases nor quotes others. Non-citational attribution is frequently introduced by reporting phrases containing cognitive verbs such as *believe* or *assume*. It is commonly found in textbooks where it is desirable to provide a broad overview of a topic rather than a detailed account of individual studies. In this specific genre, which is characterised by a relatively large difference in knowledge between writer and reader, citations might be assumed to “weaken the authoritative voice of the textbook writer” (Tadros 1993: 113).

Citational attribution, on the other hand, is achieved through the use of a reference. In Tadros’ definition, both integral and non-integral citation are forms of citational attribution and both may occur with direct quotation. In both styles, the names of one or more authors are specified, i.e. the content is attributed to concrete individuals. In integral citations, citational attribution to the author may be direct, for example via the overt mentioning of the author’s name, or indirect, for example in a passive construction (Wiemeyer 2017b). In a broader definition of citational attribution, a source text proposition may be attributed not just to the author(s), but to the type of publication, for example *the article* or *the book*, or the type of research, for example *the study*. It is also possible to attribute content to the type of text (e.g. *the paper*), to the informants of the cited study, or to a particular section of a source text (e.g. *in the introduction*) (Wiemeyer 2019).

Attribution is distinguished here from the documentation of sources. While attribution serves to acknowledge the originator of a claim or an opinion, documentation refers to the act of providing a reference to a cited work, either within the text or in the bibliography, or both. There are a number of conventional styles of documentation which are regulated in style guides. In the arts and humanities, references are usually formatted following either the citation and formatting style of the Modern Language Association (MLA) or the American Psychological Association (APA). Academic journals often base their own style sheets on one of these style

---

31 Attribution may also be viewed as the rhetorical purpose of a paraphrase or a direct quotation (see Petrić 2007).

guides. University guidelines on academic writing usually prescribe the use of references, especially with direct quotes (see Crocker & Shaw 2002 for an example).

The documentation of intertextuality renders source use legitimate (Crocker & Shaw 2002), though not providing documentation for an intertextual passage may be viewed as acceptable in some cases. According to Crocker and Shaw (2002), documentation is one of three dimensions of variation of intertextuality. An attribution may or may not be accompanied by a reference documenting the source. Expert writers tend to provide references for all their direct quotes and the majority of paraphrases in their writing (Campbell 1990). However, while documentation entails attribution (since a reference is an intertextual marker and thus a form of attribution), the opposite is not necessarily the case. As discussed above, academic texts have been found to contain non-citational attribution, i.e. undocumented attribution which is not accompanied by a reference (Tadros 1993). Attribution is sometimes used instead of referencing to signal the intertextual nature of undocumented passages, e.g. via the use of author names and reporting structures (see e.g. Shi 2004). When citing a well-known Greek philosopher whose words are understood to have been passed on orally, the need for documentation is relinquished as long as the words are quoted correctly and are attributed. The use of such non-citational attributions presupposes that the reader recognises the content or at least the author's name. Certain forms of intertextuality may be considered legitimate without referencing and without attribution, but only if they refer to common knowledge (Crocker & Shaw 2002; Tadros 1993). In other cases, however, the absence of reference to one's sources may raise suspicion and the reader may be inclined to question apparently unsubstantiated claims. At worst, the writer may even be accused of plagiarism (see section 3).

References help the reader identify the scientific community with which the writer associates their research and thus relevant theoretical frameworks and schools of knowledge: "The embedding of arguments in a network of references not only suggests an appropriate disciplinary orientation, but reminds us that statements are invariably a response to previous statements and are themselves available for further statements by others" (K. Hyland 2004: 21). By explicitly referring to a specific context of knowledge, writers signal alignment to a particular discourse and acknowledge prior research. Attribution is often achieved via reporting structures, which also afford the writer with the opportunity to evaluate the content of their sources.

### 2.3.2 Reporting structures and reporting verbs

The act of attributing propositional content to someone else is referred to as reporting (K. Hyland 1999). The act of reporting can be made explicit in academic writing by the use of reporting structures.<sup>32</sup> Reporting verbs are the most obvious choice when creating a reporting structure (G. Thompson 1990), and early research on reporting structures focused predominantly on the types of verbs found in published articles and the enactment of academic criticism (see e.g. Thomas & Hawes 1994; Thompson & Ye 1991).

A prototypical reporting sentence consists of a reporting clause, a reporting verb, and a reported clause (Thomas & Hawes 1994), as in example (2.19). Reporting verbs are verbs that describe the (supposed) action of the cited author.<sup>33</sup> They serve to introduce the reported clause as reported; that is, the passage is highlighted as being based on or taken from another author's work. The writer's choices regarding reporting verbs also provide "interpretive cues" as to their evaluative stance (Calsamiglia & López Ferrero 2003: 21). The reporting verb may be used in an active structure (see underlined structure in examples in 2.20) or a passive structure (2.21). The reporting structure may also be a subordinate clause introduced by *as*, also referred to as a reporting adjunct, which can occur with active and passive constructions; see examples in (2.22). Other reporting structures may

<sup>32</sup> Note that a reporting structure does not necessarily entail attribution as the origin or originator of the reported information does not have to be named. Likewise, attribution does not necessarily entail the presence of a reporting structure since it can occur outside of reporting structures, e.g. in adverbial clauses.

<sup>33</sup> In a narrower definition proposed by P. Thompson (2000: 92), reporting verbs are only those that are "directly controlled by the citation subject, in the particular sentence".

contain a reporting noun (Verheijen 2015), see example (2.23), or a complex preposition such as *according to*; see example (2.24).

- (2.19 ) [ Baba [ argued ]<sub>reporting verb</sub> that ]<sub>reporting clause</sub>  
[ this variation in lexical diversity is due to ... ]<sub>reported clause</sub>  
(Gebril & Plakans 2016: 80).
- (2.20 ) For example, Source Text 1 states that ‘the destruction of a language is the destruction of a rooted identity’. (Chan 2017: 20)
- Skehan further claimed that these three aspects of performance are somewhat independent of one another. (Mackey & Gass 2013: 240)
- Research has proposed that integrated tasks improve the authenticity of academic writing assessment since they simulate, in part, actual practices in academic contexts (Gebril, 2009; Knoch & Sitajalabhorn, 2013). (Gebril & Plakans 2016: 78)
- (2.21 ) It has been suggested that when checking assignments, markers focus more on the copying of words, rather than ideas (Angélil-Carter, 2000). (Davis 2013: 126)
- (2.22 ) As Swales and Feak (1994) show, each of the range of practices discussed in the introduction has its own place on a scale of legitimacy (Figure 1). (Crocker & Shaw 2002: 51)
- As suggested by several researchers (Fox, 2003; Upshur & Turner, 1999), the source of this difference is attributed to rater variability. (Yang & Plakans 2012: 95)
- (2.23 ) The discussion supports the claim that it might be misleading to treat all citations as equal in quantitative citation analysis. (Erikson & Erlandson 2014: 625)
- (2.24 ) According to the supervisors, the fundamental nature of a thesis in Agricultural Botany is that of a *report*. (P. Thompson 2001: 79)

In his cross-disciplinary study, K. Hyland (1999; 2004) identified 400 different verbs that were used in reporting structures, of which almost 50% were found only once in the corpus. The most frequent verbs were *suggest, argue, find, show, describe, propose, and report*, which together made up over 25% of all instances. Such verbs play a crucial role in reporting because “they allow the writer to clearly convey the kind of activity reported and to precisely distinguish an attitude to that information, signalling whether the claims are to be taken as accepted or not” (K. Hyland 1999: 344). The examples in (2.20) illustrate that research verbs can be marked for various tenses. Research has shown that the choice of tense in reporting verbs does not necessarily follow ‘textbook rules’ and can have important implications for their interpretation (see Swales 1986; Swales 1990: 151f. for a discussion). Swales (1986) suggested that the integral or non-integral form of the citation governs the choice of tense in the reporting verb. Individual disciplines may have specific preferences with regard to citation (see Bazerman 1984, cited in Swales 1986; see also Hawes & Thomas 1997 for a study of tense choices in medical journal articles).

Numerous attempts at describing and classifying reporting verbs have been made. Swales (1990) subdivided reporting verbs, which he proposed to be a set of around fifty verbs of different frequency, into two main categories. The first category contained verbs which indicate the writer’s agreement with the cited proposition, e.g. *demonstrate* and *establish*. This distinguishes them from verbs in the second set, such as *suggest* and *propose*, which do not convey accord. Swales suggested that the stylistic variation afforded by reporting verbs is the reason why writers alternate between integral and non-integral citation. An oft-cited

taxonomy proposed by G. Thompson and Ye (1991) sub-categorises reporting verbs based on three processes: textual, mental, and research. Textual verbs refer to obligatory verbal expression, e.g. *state, write, point out, deny*. Mental verbs describe mental processes (e.g. *believe, think, consider, prefer*). Finally, research verbs such as *measure, calculate, quantify, obtain, and find* give expression to processes related to research. One issue with describing and classifying reporting verbs is that they are not necessarily easy to identify. P. Thompson (2000) explains his decision to forego a study of reporting structures in PhD theses with reference to the difficulty in distinguishing reporting verbs used for extra-textual reference from those used to make intra-textual reference and the lack of a clear dividing line between verbs of reporting and verbs of narration.

In the natural and technical sciences, reporting structures tend to be less frequent than in the humanities and social sciences (K. Hyland 2002). On average, each paper in K. Hyland's (2004) corpus contained 28.6 reporting structures, with 42.6% of citations being introduced by them, though the average in physics papers was 6.6 (27% of citations). At the other end of the spectrum, philosophy papers contained an average of 57.1 reporting structures introducing 67% of citations. These differences arise because other authors are seldom mentioned and referred to in the hard sciences. The authors' importance is understated to "maintain the legitimacy of scientific knowledge as built on non-contingent pillars such as strict procedures, replication, falsification, and rigorous peer review" (K. Hyland 2004: 34). Across disciplines, authors used reporting and non-reporting verbs in relatively equal numbers, although physicists leaned more towards non-reporting and philosophers more towards reporting. The kind of reporting verbs that was preferred and the density of verbs differed substantially, however.

K. Hyland (1999; 2004) observed substantial variation across disciplines with regard to those reporting verbs which occurred most frequently, with very little overlap between them. The reporting verbs found in research articles from the humanities and social sciences varied much more and were much more argumentative and discursive, thereby providing an evaluation of the claims and establishing common ground with their readers. Writers in the natural sciences and engineering preferred verbs related to research practices and relied on a smaller set of verbs overall. K. Hyland explained this difference in reference to the expectation in the soft sciences that the context of research and the theory that is drawn on is elaborated through citation, whereas in the hard sciences the writer can assume knowledge of the state of the art on the reader's side. As the previous findings on which research is based are usually well-known to readers within the community, there is more shared knowledge between scholars in the hard sciences. Hence, writers are not expected to lay out how reliable a claim is and can thus reduce their reporting of previous studies to a minimum.<sup>34</sup> In sum, K. Hyland's (2004) study corroborates the general tendency that the humanities tend to rely more on citation of other authors to support their arguments than the natural and technical sciences.

Because of the large number of reporting verbs, structural variety, and tense options available, reporting structures offer a range of stylistic options for academic writers. The use of citations with reporting structures allows academic writers to employ intertextual reference for different rhetorical purposes (see e.g. Harwood 2009). For instance, reporting structures have a certain evaluative potential that is exploited by academic writers to position themselves towards a citation.

### 2.3.3 Evaluation and writer's stance

Evaluation in academic writing is defined as "the conveying of the writer's view of the status of the information in her text" (G. Thompson & Ye 1991: 367). By citing another author, a writer implicitly evaluates what they cite, simply by the fact that they have deemed it worthwhile to report a particular statement. In this

<sup>34</sup> For studies of discipline-specific use of reporting verbs, see Pickard 1995 (applied linguistics); Thomas & Hawes 1994 (medicine). For cross-disciplinary studies of reporting, see K. Hyland 2002; P. Thompson 2001. For a study of cross-disciplinary citation practices in student writing, see Ädel & Garretson 2006.

sense, the attribution of an idea per se constitutes a general form of evaluation (Hunston 1995). On the other hand, a writer can also explicitly evaluate a particular cited proposition. The writer's commitment towards this proposition can vary as they may personally take position towards it or attribute a certain position to the author without revealing their own (K. Hyland 1999). When taking a personal stance, the writer has to select one of two options in relation to the author's assumption: accepting it or challenging it. A challenge is usually realised in the form of another assumption, which may, in a continuation of the cycle, again be accepted or challenged (Sinclair 1988). Evaluation may be woven through the text in a cumulative way. It creates a further layer of meaning to a text that signals its purpose to the reader (G. Thompson & Ye 1991).

Because it is employed as a central discursive strategy for creating a research niche (Swales 1990), writers employ citation to construe what Groom (2000: 19) has referred to as "manifest relations of power between textual and intertextual voices". Consequently, explicit evaluation is a general feature of academic research articles. The writer's evaluation may be destructive or mitigated (Tadros 1993). Destructive evaluation serves to discredit another author or even a school of thought in order to boost one's own viewpoint and accentuate the deficiencies of previous approaches. In contrast, an evaluation is mitigated when it is preferable to provide a comprehensive and balanced account without undermining others and to go about the literature review in a more cautious and polite way. A writer may use mitigated evaluations even when pointing to the inadequacies of others' work. In this case, the writer emphasises that a previous approach is dismissed, not because it is generally lacking, but because it is not suited to the writer's current purposes (Tadros 1993).

In introductions, evaluations often occur after the writer has established the necessity of the research and provided a summary of previous literature – usually in a detached manner (Swales 1986). They fulfil the function of emphasising gaps in the existing research that the writer intends to fill. Usually, this is achieved by pointing out what the authors omitted rather than explicitly criticising their work's shortcomings. However, both types of evaluation exist and are selected for different purposes. Accordingly, "space-creating, omissional, pseudo-negational citation" is distinct from openly critical citation, but the difference can only be established by thorough analysis of the context (Swales 1986: 46). Though evaluation occurs across disciplines, stance was more frequently attributed in the soft than in the hard sciences in K. Hyland's (1999) study. In the soft sciences, evaluation tends to occur in the vicinity of noun phrases containing the author's name in integral citation, a strategy which is much less frequently observed in the hard sciences (K. Hyland 2004).

There are different ways in which evaluation can be realised linguistically to encode stance, for instance through the choice of reporting verb, through negation, through quantifiers preceding the subject, such as *little* or *very few*, through lexical words with an evaluative connotation, e.g. *fail*, *neglect* and *lack*, or by asking questions or formulating diverging hypotheses (Charles 2006b; Swales 1986).<sup>35</sup> Depending on the linguistic expression, the reported message is consequently marked as true or untrue, although in some cases no clear evaluation by the writer is signalled. The latter case allows the writer to shift responsibility to the author, whose attitude towards the message is purported instead of the writer's. The most obvious way to assign a particular stance to the cited author is through evaluative reporting verbs which convey a positive, tentative, critical or neutral author stance (K. Hyland 1999).

By taking a positive stance, the writer assumes a subordinate position in relation to the author. When giving a negative evaluation, the writer instead takes the dominant position they usually occupy in an argumentative text (Groom 2000). Reporting verbs lend themselves to this purpose as their semantics often imply a positive or negative assessment. Their choice reflects the kind of evaluation intended by the author. Hunston (1995) proposes two classes of evaluative reporting verbs: argumentative verbs whose meaning is clearly evaluative, either in a positive or in a negative way (e.g. *acknowledge* and *insist*), and those that do not carry an evaluative connotation themselves but open up a space for a later positive or negative evaluation (e.g. *claim* and *argue*). Other authors have also identified neutral reporting verbs. Groom (2000) claimed that neutral verbs fully transfer the responsibility for the content citation onto the author. At the same time, they impel the

35 See also Cheng (2006) for an analysis of other evaluative devices academic writers have at their disposal.

writer to provide an evaluation of its proposition at a later stage because their neutrality creates a temporary absence of writer responsibility. The verb *say*, which is infrequent in expert writing (G. Thompson & Ye 1991), is considered to be one such neutral verb.

The introduction of a citation by a reporting structure creates an “evaluative space” for the writer to emphasise a contrast between their own view and that of the reader (G. Thompson & Ye 1991: 369). Reporting structures that include evaluative elements afford writers not only with a certain stylistic freedom, but also with an opportunity to use them as semiotic devices to signal their stance towards a particular author and, as a consequence, their alignment with a particular school of thought. They thus create a contract between writer and reader (Groom 2000), on whose interpretation the effectiveness of the evaluation ultimately depends (see also Hunston 1995).

The review of the literature on intertextuality in academic writing in the present chapter has provided an overview of the manifold intertextual markers present in an academic text and the strategies academic writers can employ to create a space for their research via reference to sources. It has also outlined the considerations academic writers have to make to contextualise their research and situate themselves as contributing to a particular disciplinary discourse. Reconciling these different demands of source-based academic writing is a challenge to most academic writers, especially to those who are managing the task in an L2. Much of the research into source-based writing has focused on the ways in which novice writers cope with the peculiarities of this register. Studies on intertextuality in academic texts written by learners of English are reviewed in the next chapter.

### 3 Intertextuality in L2 academic writing

At universities around the world, learning to write academic texts that are typical of the respective discipline is an integral part of degree programmes. Many of the academic texts that student writers have to produce rely on sources. Intertextual competence, i.e. the ability to felicitously appropriate and reference discipline-specific discourse, is thus a central aspect of academic attainment. Competent source use can be defined in terms of “varied use of citation, substantial paraphrasing, a range of reporting verbs and appropriate attribution to avoid plagiarism, with very rare errors” (Davis 2013: 128). At the same time, sources are integrated into academic texts not just conventionally, but because they are integral to the writer’s argument and fulfil a range of functions (Bloch & Chi 1995). In order to write effective academic texts, novice writers need to develop a thorough understanding of the purpose of intertextuality and how it is made explicit in academic texts. They must strike a balance between staying true to the meanings intended in their sources and presenting them in their own words. They are expected to create an authorial self and interpret their sources so as to create a narrative in which their sources support their intended argument. This requires them to employ strategies of creating intertextuality and of acknowledging their sources in a competent way.

Learning to write from sources is important for student writers because it prepares them to write expert genres, but also because their intertextual competence may actually determine their academic performance. For both L1 and L2 student writers, effective source use is a central academic outcome (Cumming, Lai & Cho 2016). Students’ competence in academic writing is often measured in terms of their critical use of their sources and the appropriate integration of source text material with their own ideas (Shaw & Pecorari 2013). It has been proposed that source use influences grades in source-based writing tasks (Plakans & Gebril 2012; see also Petrić 2012). In a study of writing tasks using multiple source texts (Plakans & Gebril 2013), paraphrasing was a predictor of high scores, as was the incorporation of source text ideas which had received a high importance rating. Verbatim source use was only mildly predictive of score, but, importantly, exact copying negatively correlated with test results. Students’ performance in a timed and graded integrated writing task was found to correlate not only with the number of source ideas incorporated, but also with content accuracy (Uludag et al. 2019).<sup>36</sup>

There appears to be a complex relationship between source interaction and text quality, and the correspondence between grades and source use is not always straightforward. Copying from the source text – a potentially problematic strategy – positively impacts the lexical diversity of L2 writers’ texts, so that teachers may perceive students’ writing to be of a higher quality even if it contains potentially inappropriate textual borrowing (Gebril & Plakans 2016). In addition, language proficiency appears to have an impact on source use and thus on writing quality. While source interaction in the writing process – as measured via keystroke logging – was shown to be a significant determiner of text quality in Dutch L1 writing in Leijten et al.’s (2019) study of a timed writing task, this relationship was not found when the same students wrote texts in their L2 English. In their native language, the writers’ use of a large variety of sources had a profound impact on the quality of their texts. In their L2, the use of language-related resources apparently took away time from content-related source use. The students may have simply had insufficient time to understand and interpret the L2 source texts in such a way as to produce a text of a higher quality. These findings suggest that the comprehension, interpretation, selection, and accurate incorporation of source text materials play a central role in distinguishing achievement, especially at higher levels of attainment. It seems to be the case that the quality of L2 writers’ source use affects or even determines their outcomes, which emphasises the importance of source-based writing skills for academic success.

---

36 Interestingly, linguistic modifications did not positively predict performance in Uludag et al.’s (2019) study.

Source-based writing tasks are popular with both students and teachers in L2 writing contexts because the source texts serve as models for language use and offer support for students' arguments (Chan 2011; Gebril & Plakans 2016). However, such tasks also create challenges to L2 writers, first and foremost in terms of inadequate source use and apparent plagiarism. This is reflected in research into source use in L2 student writing in English, which can be broadly divided into two areas. Most studies in this area, especially earlier ones, have investigated issues of plagiarism and inappropriate copying. Research into transgressive forms of intertextuality and explanations for such student behaviours are surveyed in section 3.1. Other studies have focused on effective strategies of source use beyond plagiarism or the use of sources by successful students, as discussed in sections 3.2 to 3.4.

### **3.1 The complex issue of apparent plagiarism: Explaining transgressive intertextuality**

The appropriation and integration of sources is a demanding task, especially in the L2 (Abasi & Akbari 2008; Davis 2013; Hirvela & Du 2013; C. Thompson, Morton & Storch 2013). Much of the research into source use in L2 academic student writing has focused on problematic aspects, particularly on the acknowledgement of sources – or lack thereof. Many studies have pointed out issues of transgressive copying and patchwriting in L2 writers' attempts at paraphrasing and summarisation (e.g. Campbell 1990; Keck 2006; Pecorari 2003; Shi 2004). At the undergraduate level, L2 writers' paraphrasing strategy mostly consists of rephrasing sentences by replacing individual synonyms and deleting strings and without changing their structure (Keck 2006, 2010; Pecorari 2003; Wette 2010). They are often insecure as to what is acceptable in terms of closeness of wording to the source text and even prefer direct quotes as a safer option for accurately reproducing the original author's ideas (Hirvela & Du 2013).

Some students struggle to avoid copying from the source text and the syntactic structure of their paraphrases is often identical to that of the original sentence (e.g. Campbell 1990; Crocker & Shaw 2002; Davis 2013; Shi 2012; see extended discussion in section 3.3.2). Others even regard patchwriting as a legitimate strategy that can be applied generously (Currie 1998; Pennycook 1996). Students' texts are sometimes characterised by features such as extended yet unmarked copied passages, especially from internet sources, and long stretches of paraphrased text without references. This lack of references intensifies the problem as the origin of ideas in students' text is often not transparent (Shi 2004). This creates issues for L2 student writers insofar as poor paraphrasing, especially if insufficiently referenced, may be viewed as plagiarism in the same way as intentional academic theft and sanctioned accordingly (see Pecorari & Petrić 2014).

Other issues are not as clearly and unambiguously definable as plagiarism but – if identified – may be viewed as inappropriate because they are not in line with disciplinary expectations. An example is unacknowledged secondary citation (Pecorari 2003), the practise of citing a source text without having consulted the original. Some problematic uses of sources are apparently caused by a fear of committing plagiarism, for example over-citation, i.e. providing references too frequently and redundantly (Davis 2013; Wiemeyer 2017b). L2 student writers have been shown to have difficulties in expressing their own authorial voice and often do not encode their stance towards their sources (Borg 2000; Groom 2000).<sup>37</sup> They tend to view their sources as representing the truth (McCulloch 2012; Abasi, Akbari & Graves 2006). As a result, they sometimes report

---

37 It must be noted that a lack of authoritative stance in student writing may be caused by the task description. Dovey (2010) observed that her students, whom she had asked to write a literature review that was not tied to a research project, relied heavily on the information found in the sources, not on the results of the research presented. She concluded that this was caused by their literature review "not being driven by the need to generate and provide the rationale for a genuine set of research questions" (2010: 53). Students' awareness that they were not writing for an authentic audience but for their lecturer made it difficult for them to negotiate an authentic position towards the cited sources.

statements from their sources as facts rather than as points of view. This perception apparently leads L2 writers to cite extensively from sources whose content they assume to be reliable and truthful, and to omit references (T. A. Hyland 2009; McCulloch 2012).

At the graduate level, students more aptly integrate source text material with their own in order to re-contextualise its content, but illicit copying is still an issue (Shi, Fazel & Kowkabi 2018). Even postgraduate students have attested to lacking explicit instruction in intertextual strategies. Davis (2013) found that postgraduate L2 students also attributed their sources insufficiently, especially those they had accessed online. They tried to follow their teachers' advice of using quotations sparingly, but often did not manage to paraphrase successfully due to their limited knowledge of alternative syntactic structures, related expressions, and synonyms. These postgraduates resorted to patchwriting because of a lack of vocabulary, but also because of insufficient awareness of academic conventions and lack of teacher guidance.

Based on such empirical findings, it has been proposed that L2 student writers tend to employ paraphrasing for knowledge construction, but do not critically dissect their sources (Borg 2000; Hirvela & Du 2013). Their ideas may be so closely intertwined with those lifted from source texts that it is difficult for the reader to differentiate between them, or they are clumsily linked without a clearly discernible narrative (Li & Casanave 2012; McCulloch 2012; Wette 2010). Students' ability to paraphrase effectively is further complicated by challenges in comprehending complex source texts, in identifying important source text ideas, and in selecting relevant source text passages for paraphrase (Borg 2000; Karbalei & Amoli 2011; Li & Casanave 2012; Sherrard 1986; Wette 2010).

On the other hand, it has been emphasised that there are individual differences between students, and that many show awareness of citation conventions. There is often a small number of students in a group that copy extensively, while others' paraphrases are inconspicuous in terms of textual overlap (Keck 2014; Weigle & Parker 2012). In addition, there is evidence that students who commit apparent plagiarism often readily reveal their sources, which has led L2 writing researchers to stress the importance of distinguishing between intentional and unintentional forms of apparent plagiarism (e.g. Chandrasoma, Thompson & Pennycook 2004; Pecorari & Shaw 2012; see also Pecorari & Petrić 2014 and further discussion in section 3.2). Furthermore, students improve over time and through teaching (e.g. Wette 2010). While the participants of Davis' (2013) longitudinal study showed no improvement in terms of rephrasing and restructuring passages from the source text, they did increase their proficiency in the use of reporting verbs. They also improved with regard to attributing their sources appropriately over the course of their studies as they became more familiar with academic writing conventions.

Taking a perspective on certain incidences of apparent plagiarism as being non-deceptive, researchers have zoomed in on the reasons behind problematic source use. These have been predominantly discussed in reference to students' native language and proficiency in English as well as to their academic writing expertise, as specified in sections 3.1.1 and 3.1.2.

### **3.1.1 Intertextuality, cultural background, and language proficiency**

Apparent differences in source use between L1 and L2 student writers have prompted L2 writing researchers to consider the influence of the students' background on their strategies of constructing intertextuality. L2 writers' source use and the effectiveness of their citations have frequently been discussed in reference to their language proficiency in English, their non-nativeness, and their culturally-shaped conceptualisations of knowledge.

According to some scholars, L2 writers' comparatively lower language proficiency may be the explanation for at least some of their issues and the differences to L1 writers' texts, for example excessive textual

borrowing. It has frequently been observed in plagiarism studies that some L2 student writers – even at postgraduate level – copy extended strings from the source text, even entire sentences, into their text. They acknowledge the source, but do not mark these excerpts as direct quotes (Chandrasoma, C. Thompson & Pennycook 2004; Davis 2013; Pecorari 2003, 2006). Lower-scoring L2 writers generally tend to make less use of strategies of rephrasing, to quote more frequently, and to prefer longer direct quotes than their higher-scoring peers (Campbell 1990; Cumming et al. 2005; Gebril & Plakans 2009; Petrić 2012). These issues have been explained by a lack of vocabulary on the part of student writers and an unconventional understanding of what has to be placed in quotation marks (Davis 2013).

It has been proposed that patchwriting is related to low language proficiency and that illicit copying is a strategy employed by students with limited linguistic resources who are still acquiring the language of composition (e.g. Campbell 1990; Cumming et al. 2005; Currie 1998; Johns & Mayes 1990; Pennycook 1996; see Gebril & Plakans 2009: 49 for further discussion). Such interpretations are supported to some extent by the students' own comments. Some L2 student writers feel that direct quotation is a safer choice that is easier to accomplish than paraphrasing, especially if their linguistic resources are limited, because they feel insecurity regarding how to reword source text excerpts (e.g. Hirvela & Du 2013). Other students have even reported feeling at a disadvantage in comparison to L1 students because of their lower language proficiency (Borg 2000).

Other researchers have emphasised that linguistic background may not sufficiently explain source use issues. A student's proficiency level in their L2 English may be an ambiguous indicator for proficient source use. A study of integrated writing tasks in the context of language assessment (Plakans & Gebril 2012) found no significant differences across L2 proficiency levels in terms of the use of direct quotation and copied strings of three words or longer. However, the students' proficiency in English significantly impacted the extent of their source use, with more proficient students being less reliant on the source text in terms of lexis and syntax. Clearly, language proficiency is only indicative of source-based writing skills to a certain extent and cannot explain all issues. Multiple influences may cause differences between students.

The differences between L1 and L2 students have also been suggested to be the result of unconventional perceptions of inappropriate intertextuality on the part of L2 students. These perceptions sometimes diverge from those of L1 students, and it has been argued that this may lead to questionable uses of source texts such as illicit copying (e.g. Chandrasegaran 2000; Marshall & Garry 2006; Pecorari 2003). For instance, the Chinese L2 students in Shi's (2004) study generally approached intertextuality in significantly different ways than their peers with L1 English. The L2 writers tended to patch together sentences from copied excerpts with little modification and referencing, while the L1 writers attempted to create their own sentences, modified more extensively, and borrowed considerably fewer words.

Such observed contrasts between groups have sometimes been said to be culturally influenced since cultural attitudes shape the ways in which sources are integrated into English texts. Not all cultures share the citation conventions of Anglophone academic writing and the corresponding views on what should or should not be cited (Bloch & Chi 1995; Chandrasoma, C. Thompson & Pennycook 2004; Pennycook 1996). As a result, cultural norms have been widely discussed as reasons for apparent plagiarism, and issues of textual borrowing and problematic referencing have been attributed to different concepts of knowledge and ownership in different cultures (e.g. Abasi, Akbari & Graves 2006; see overview in Pecorari & Petrić 2014). The unfamiliarity with English writing norms and the subtleties of English expressions and grammatical structures may put L2 writers at a disadvantage (Bosher 1998; Cumming 2002; Rowley-Jolivet & Carter-Thomas 2014). It has also been claimed that L2 writers develop their authorial voice in English-language writing more slowly than L1 writers because of their different educational backgrounds and because they may regard sources as containing incontestable truths (Abasi, Akbari & Graves 2006; Currie 1998).

Differences between L1 and L2 speakers of English in the use of integral and non-integral citations (Hryniuk 2016) as well as of attribution and stance (Rowley-Jolivet & Carter-Thomas 2014) are identifiable even in published expert writing, which suggests that native language and culture do shape academic writing

strategies to a certain extent. The general consensus in the literature is that cultural understandings and the students' ability to internalise Western norms regarding source use play a role in non-deceptive forms of plagiarism, though it remains controversial to which extent this is the case. It is clear that the relationship between intertextuality, plagiarism, and culture is more complex than sometimes assumed.<sup>38</sup>

The distinction between L1 and L2 writers when explaining problematic source use has been called into question because many of the issues experienced by L2 writers are shared by L1 writers. Though L2 writers' texts generally contain more borrowed material than those of their L1 peers, both L1 and L2 student writers have been shown to copy when summarising, and to patchwrite (Campbell 1990; Howard, Serviss & Rodrigue 2010; Shi 2004). Students from both groups make formal errors, for example leaving out quotation marks, omitting references from the bibliography, and using inconsistent referencing styles (Borg 2000).

Irrespective of their L1, student writers struggle with the rhetorical demands of source use such as separating their own voice from that of their cited authors and reconciling originality with synthesis. Across these groups, reading skills and knowledge of topic seem to play a decisive role. L2 writers can apparently compensate for some potential issues in their source use by applying systematic reading strategies (T. A. Hyland 2009). As early as 1988, Dubois argued that apparent plagiarism is the result of weakness rather than maliciousness – in both groups:

Both L1 and L2 students of my experience can find themselves so far out of their intellectual or linguistic depth – in the former case, unable to synthesize or reason; in the latter case, unable to paraphrase, summarize, or generalize – that sheer desperation occasions, although it does not excuse, their misconduct. (Dubois 1988: 189)

Many aspects of intertextual writing are the same across L1 and L2 writers' texts, for example text structure as well as the length and functions of citations. Similarities in source use by L1 and L2 writers may simply be overlooked as research into these two groups is not as intertwined as it could be (see also Pecorari 2016: 341). Furthermore, there are conflicting findings regarding source use by L1 and L2 student writers. Some studies have observed more consistent referencing in L1 student writing, even with short citations (Shi 2004), while others have identified L2 student writers as the more diligent (Campbell 1990). As Cumming, Lai and Cho (2016) argue in their state-of-the-art article, absolute distinctions between L1 and L2 writing from sources are difficult to draw on account of the small cohorts and limited number of writing tasks examined in existing studies.

On account of the similarities between the two groups and individual differences between students that have emerged from research, T. A. Hyland (2009) has proposed to desist from making assumptions about differences between L1 and L2 student writers in future research since, as she has argued, all writers appear to go through various stages of development towards becoming proficient academic writers. A variety of aspects appear to influence the amount of textual overlap, including text type and task type (see synthesis of research in Cumming et al. 2016: 52). Textual borrowing differs across task types and depends on the role the source text plays in these tasks (Weigle & Parker 2012). For example, the extent of students' source text dependence was found to be different in opinion pieces and in summary tasks for both L1 and L2 student writers (Shi 2004). L2 writers have been observed to re-use much more source text material in reading-into-writing tasks than in listening-into-writing tasks, probably because the visual presence of the source text leads them to reiterate the exact wording (Cumming et al. 2005). Such observations compel the conclusion that source-based writing skills are also tied to individual understandings of appropriate textual borrowing, context, and academic writing experience.

---

38 See e.g. Currie (1998) and Pennycook (1996) for discussions of cultural factors that may induce inappropriate textual borrowing, and Pecorari (2016: 339ff.) for a review of the debate.

### 3.1.2 Intertextuality and academic writing expertise

Writing research has shown that certain differences between L1 and L2 student writers are related to the extent of experience in academic writing rather than language proficiency. K. Hyland (2002: 116) has suggested that illicit copying is not necessarily a problem of limited vocabulary but rather “symptomatic of a larger issue of how to appropriately acknowledge sources in academic writing”. A growing body of L2 writing research has thus considered writing ability rather than linguistic background as a factor explaining questionable features of students’ texts. The insights from studies into the issues underlying inappropriate textual borrowing have led to the re-conceptualisation of apparent plagiarism as a developmental issue.

Academic writing expertise strongly influences the writing process and the use of sources, especially in the construction of stance (Abasi, Akbari & Graves 2006). More experienced L2 writers use a wider range of intertextual strategies more proficiently.<sup>39</sup> Research suggests that inexperienced L2 writers are often overwhelmed by the diverse demands of writing from sources in an L2. It has been argued that their problematic source use may be caused by underlying issues such as unfamiliarity with the topic and general lack of understanding of academic discourse. Students who are unsure of what kind of argument they could construct may rely on points made in their sources without connecting them in a meaningful way. They frequently struggle to encode an evaluative stance, which in turn leads to issues such as over-quoting. Their sources are placed at the centre of their papers at the expense of their own authorial voice (McCulloch 2012; see also Abasi & Akbari 2008).

Ineffective paraphrasing in L2 source-based writing has in the past been attributed to the writers’ language proficiency (Petrić 2012; Shi 2012), yet research shows that this is clearly an issue that is remedied by increased writing experience, since graduate L2 writers often paraphrase smoothly and employ knowledge-transforming processes (Shi, Fazel & Kowkabi 2018). Pecorari and Petrić (2014: 278) have emphasised that

the anti-plagiarism dictum ‘say it in your own words’ oversimplifies the task for students encountering a new discourse, who do not yet possess the appropriate linguistic repertoire necessary for writing about academic or discipline-specific topics and who thus need to rely on source texts in order to acquire it.

Issues such as textual borrowing are thus related to insufficient knowledge of academic language, not to L1 background. It has also been proposed that students resort to copying when they do not understand the content of the source text (Howard 1995; Howard, Serviss & Rodrigue 2010). L2 students have been shown to have varied concepts of what constitutes ‘common knowledge’ and to struggle with distinguishing relevant from unimportant source text information. These students use patchwriting as a survival strategy as their writing skills are not sufficiently developed to negotiate the demands of their academic tasks (Chandrasoma, C. Thompson & Pennycook 2004; see also Currie 1998). Students’ limited experience with citation and their positive attitude towards unattributed copying are further potential sources of patchwriting (Shi 2004). That apparent plagiarism is often caused by a lack of academic writing experience is confirmed by several studies showing that both L1 and L2 students have problems identifying plagiarised excerpts in textual examples, especially in paraphrases (Roig 1997; Chandrasegaran 2000). Sometimes, L2 student writers are unable to cope with source-based writing tasks simply because they have not received explicit instruction on the skills they are expected to demonstrate (Currie 1998; Shi 2004). Their knowledge about plagiarism may be purely declarative without access to successful source use strategies (Pecorari 2003; Li & Casanave 2012).

These findings attest to the various challenges faced by novice writers when writing from sources, which is a far more complex task than plagiarism warnings may insinuate. It is evident that L2 writers’ issues, especially patchwriting, are caused by a range of factors, the vast majority of which are non-deceptive in nature. Several studies have emphasised the role of individual preferences and educational background in source-based writing (e.g. Choi 2016; Keck 2014; McInnis 2009; Shi 2004). Students’ varying abilities in writing from sources

---

<sup>39</sup> See Gebрил and Plakans (2009) for an overview of studies investigating the relationship between expertise and writing process.

have been linked to factors such as field of study, prior work and writing experience, and knowledge of topic (see review of research in Cumming et al. 2016). The approaches to source-based writing employed by individual students may vary depending on the task type and the context of writing (Cumming et al. 2005; Davis 2013; Li & Casanave 2012; McCulloch 2013; Petrić & Harwood 2013; Shi 2004). The quality of a student's intertextual writing task is a reflection of their reading comprehension, their self-conception as a writer, and their writing ability (Hirvela & Du 2013; T. A. Hyland 2009). Not all L2 writers use sources in the same ways, and not all are prone to inappropriate forms of intertextuality, as is further discussed below.

### **3.2 Beyond plagiarism: Unintentional textual borrowing and learning processes**

There is an ongoing debate as to whether instances of intertextuality that violate academic conventions should be treated as purposeful deception, i.e. plagiarism. Studies into the reasons for apparent plagiarism have often revealed a lack of intentionality on the part of student writers. For example, Pecorari (2003) found that even though the students in her study were aware of the conventions of academic writing in English, they did not always attribute their sources correctly and consistently. Most of her students' texts contained patchwritten passages with more than 50% copied but unattributed source text material. This apparently resulted from the fact that the students were inexperienced novice writers and were often overwhelmed by the manifold demands of composing academic texts. Pecorari's interviews with the student writers revealed that their inappropriate citing behaviour was unintentional, leading to the conclusion that L2 writers tend to use sources inappropriately while still learning how to use them in conventional ways. Such findings have paved the way towards a re-conceptualisation of certain transgressive forms of intertextuality as features of developing writers' texts.

Indeed, patchwriting has been identified as a learning strategy that novice writers apply, especially when they are confronted with unfamiliar texts and have to make meaning out of the source text content (Howard 1995; Pecorari 2003). The requirement to use complex language in paraphrasing may explain L2 writers' inclination towards patchwriting (Shi 2004). Furthermore, Shi's (2004, 2012) and Pecorari's (2003) studies of textual borrowing practices of L2 undergraduates and postgraduates showed that students whose texts contained unattributed copies from original articles did not hesitate to reveal and talk about their sources, which suggests that they had no intention to deceive. In a similar vein, the presence of references in sections containing inappropriate intertextuality may be a sign that the student is still developing their paraphrasing skills, but also implies that the transgression is not the result of dishonest intentions (see Crocker & Shaw 2002). These findings support the assertion that patchwriting is not a prototypical form of plagiarism.

Even professors have been found to patchwrite in a summary task when confronted with difficult source texts from another discipline (Roig 2001). Considerable textual borrowing from source texts is also traceable in some published research articles, irrespective of discipline and native language (Sun 2013). This establishes that source-based writing – especially in unknown contexts – is challenging even to expert writers and that patchwriting is not necessarily a sign of low competence. L2 student writers have also been found to strategically use the source text to vary their wordings in source-based writing (Gebril & Plakans 2016; see extended discussion in section 3.4).

Researchers now generally agree that many problems identified in L2 writing do not constitute plagiarism in the narrow sense (see discussion in Pecorari & Petrić 2014). Student writers are often acutely aware of the need to avoid plagiarism, but lack the means of circumnavigating it successfully or have unconventional understandings of appropriate source use. Some students even cite fear of inadvertent plagiarism as a motivating factor for some of their source use strategies (Li & Casanave 2012). These findings substantiate the assumption

that “for both L1 and L2 academic writers, copying from source texts is a necessary phase through which developing writers must pass before they acquire more sophisticated ways of integrating sources into their writing” (Keck 2006: 262).

It has been proposed that because problematic, yet non-deceptive textual borrowing is a developmental stage on the way towards becoming a competent academic writer, it should be regarded as non-prototypical plagiarism, precluding sanctions (Pecorari 2003; see also Chandrasoma, C. Thompson & Pennycook 2004; Shi 2004). This view is shared at least by some university teachers who have stressed that deliberate plagiarism should be distinguished from other forms (Pecorari & Shaw 2012; Sutherland-Smith 2005). Chandrasoma, C. Thompson, and Pennycook (2004) have proposed altogether abandoning the notion of plagiarism, which in their view does little to clarify issues of inadequate source use, neither for students nor their teachers. They have suggested to consider the context of such issues instead and treat them in terms of the distinction between what they have termed transgressive and non-transgressive intertextuality. In their view, this re-conceptualisation allows researchers and teachers to focus on the issues that underlie questionable intertextuality, for example language, academic literacy, authorial identity, and power. They argue that transgressive intertextuality is very much context-dependent and can only be judged taking the specific situation into consideration.<sup>40</sup> According to the authors, drawing a line between transgressive and non-transgressive behaviours also entails uncovering different understandings on the part of teachers and students of what is transgressive and what is not.

In order to avoid illicit source use, novice writers have to learn the conventions associated with academic discourse and academic writing, specifically when using sources. They also have to acquire the necessary academic language and authorial identity (Pecorari 2015). What is sometimes viewed as plagiarism in L2 writing can thus be seen as the surface form of students’ learning processes. Because of the manifold influences at play, Pecorari and Petrić (2014) have emphasised the importance of complementing research of student plagiarism with studies that take a positive perspective on students’ source use and the factors shaping their intertextual practises. It is thus important to consider not only problematic manifestations of intertextuality, but also the effective ways in which students integrate and acknowledge source text ideas in their texts.

### **3.3 Manifestations of intertextuality in L2 academic writing**

When writing from sources, L2 writers employ a range of strategies to create intertextuality. Beyond the discussion of plagiarism in L2 writing, there is a considerable body of research into L2 writers’ approaches to source use in academic writing. The following sections contain a review of L2 writing research on the (successful) use of the core intertextual strategies – direct quotation, paraphrasing, summarisation – as well as on attribution, reporting structures, and documentation.

#### **3.3.1 Direct quotation in L2 student writing**

While various studies have addressed the use of direct quotes in published and unpublished academic writing by L1 writers (see e.g. Dubois 1988; K. Hyland 2000), there are few studies concentrating on direct quotation in L2 student writing. Often, direct quotation is investigated as one manifestation of intertextuality in studies of source use by student writers in general. The existing studies of direct quotation have mostly relied on corpus analyses (Borg 2000; Davis 2013; Verheijen 2015; Wiemeyer 2019), some of which have been

---

<sup>40</sup> In their study, Chandrasoma et al. cite ten concerns that foster an understanding of transgressive intertextuality: intentionality, development, identity, resistance, student epistemologies, common knowledge, mediated discourse, interdisciplinarity, variability, and task type (2004: 189).

triangulated by interviews (Petrić 2012; Shi 2008). Direct quotation has also been discussed in text-based interview studies (e.g. McCulloch 2012). A small number of studies have included introspective methods, for example by combining think-aloud protocols with text-based interviews (Hirvela & Du 2013) or by analysing reading and writing logs and stimulated recall interviews as well as student texts (Ruiz-Funes 1999).

Direct quotation is relatively common in L2 writing, though it is less frequent than paraphrasing. Studies have established that direct quotes constitute an average of around 35% of all citations in L2 writers' texts (Petrić 2012; Wiemeyer 2019), though there are vast differences between individual writers. Direct quotes are typically embedded into sentences and showcase the students' ability to syntactically integrate these intertextual links in a successful way. Rarely, students also use free-standing quotes, for example at the beginning of the text. L2 writers' direct quotes vary in length and range from individual words, for example terminology, to several sentences. Combined quotes, i.e. several quotes from one source in the same sentence, occur in both student and expert writing. Though they feature significantly more often in experts' texts than in those by L2 students, students are apparently often aware of this relatively sophisticated way of integrating source text information (Verheijen 2015; Wiemeyer 2019).

Most commonly, L2 writers produce direct quotes of short source-text excerpts in non-integral citations, which are often accompanied by reporting structures or another form of attribution. This kind of quote is often very similar to those of experts (Wiemeyer 2019). Furthermore, it is clear from the existing studies that unlike other forms of intertextuality, direct quotation is almost always acknowledged and referenced across learner populations (Petrić 2012; Verheijen 2015; Wiemeyer 2019). Transgressive intertextuality appears to be a minor issue in the context of direct quotation, perhaps with the exception of secondary citation (Wiemeyer 2019; see also Pecorari 2003).

Research suggests that there is a correlation between the use of direct quotation and students' grades. In Petrić's (2012) study of high- and low-rated Master's theses of L2 learners of English, there is a clear connection between writing skills and use of direct quotes. The more successful writers were found to quote significantly more frequently than those with low-rated theses, a surprising finding in the light of previous research (e.g. Campbell 1990; P. Thompson 2001). In their texts, direct quotation made up for almost half of all instances of citation, as compared to 20% in the low-rated theses. The writers of high-rated theses were more likely to construct their own discourse. They incorporated mostly short and succinct fragments from the source texts into their own writing, while the less successful students tended to quote complete finite clauses and longer, unmodified stretches of text. The average length of direct quotes was consequently lower in high-rated theses (see also Borg 2000).

There are also striking qualitative differences in Petrić's (2012) study. The writers of the high-rated theses quoted much more terminology than their peers. Some overgeneralised citation rules and used quotation marks every time they mentioned a term. Through this somewhat overcautious practice, they simultaneously avoided accusations of plagiarism and acknowledged the originators of terminology, which created stronger links to the cited publications and may have been viewed favourably by the examiners. While some students in the study struggled with direct quotation, others used direct quotes proficiently and purposefully, for example to instil stylistic variation into their writing, to distance themselves from the content or to lend support to their own claims. In the interviews, some students reported to quote directly if the excerpt was considered especially vivid or succinct and contained an important idea, a motivation that has previously been reported for expert writing (see K. Hyland 2004).

A varied picture of L2 writers' motivations to quote has emerged from interview studies. While some of the L2 student writers interviewed by Shi (2008) used direct quotation because they did not know how to rephrase, others did show awareness for the general preference of paraphrasing over direct quotation in academic writing and adjusted their citation practices accordingly. Yet, there were also students in the same study who chose paraphrasing because they were unsure of whether direct quotation was possible for a particular excerpt. According to one student in Hirvela and Du's study, direct quotation guaranteed a more accurate reproduction of

the source text because it “ensured the specificity, technicality and authority of the information” (2013: 94). The author’s exact words were assumed to make for a more powerful argument than a paraphrase. Similar findings were reported by Ruiz-Funes (1999), whose subject explained using direct quotes to make her text more memorable. The student aimed to lend credibility to her argument by quoting others who shared her perspective, but also to make it explicit if she was the only one who held a specific view. In contrast, a tutor’s preference led one postgraduate student in Davis’ (2013) study to avoid direct quotation. These studies have highlighted the fact that many L2 writers generally have a good grasp of disciplinary expectations as to direct quotation, but sometimes hold insecurities concerning the more fine-grained functional aspects of this intertextual strategy.

According to Petrić (2012), who conducted the first study focusing exclusively on direct quotation in L2 student writing, L2 writers’ use of direct quotation differs remarkably from that of the expert writers examined in the studies by K. Hyland (2000) and Pickard (1995) as well as from that of L1 novice writers (Ädel & Garretson 2006). Ädel and Garretson (2006) have tentatively named the relatively higher frequency of direct quotation as the most marked feature distinguishing novice from expert writing. The differences between novice L2 writers and L1 expert writers were at the focus of a study by Verheijen (2015) investigating the language of quoting of L2 writers and experts using two written corpora. Verheijen observed differences in the ways in which the groups embedded quotes into their writing. Reporting constructions were significantly more frequent in the L2 writers’ texts, and they preferred direct quotes in subordinate clauses, whereas the experts preferred to use them in main clauses. The L2 writers’ texts also contained more non-embedded direct quotes.

It must be noted that the text types were not directly comparable and the experts’ texts were peer-reviewed and proof-read, so that at least some of the differences are likely to be a result of the different text types and writing contexts. It emerged from Verheijen’s corpus analyses that the L2 writers’ errors were mostly due to their unfamiliarity with academic conventions of quoting and their limited linguistic repertoire, as observed in other studies (e.g. Borg 2000; Petrić 2012). In a study of linguistic research papers written by German EFL<sup>41</sup> writers, Wiemeyer (2019) found that direct quotes were more frequent than in the expert writers’ texts analysed by K. Hyland (2004). However, the students’ direct quotes were qualitatively quite similar to experts’ quotes, suggesting that differences only apply to certain features of quotation such as frequency.

Studies comparing intertextual strategies in L1 and L2 student writing have established certain similarities. Both groups use direct quotes to construct an argument, but sometimes confuse paraphrasing and direct quotation or resort to patchwriting (T. A. Hyland 2009). Disciplinary differences in the frequency of direct quotation are consistent across L1 and L2 cohorts. Student writers’ direct quotation practices mirror typical differences between hard and soft sciences, with quotation being more frequent in the arts and humanities and typically being avoided in natural sciences and engineering (Ädel & Garretson 2006; Pecorari 2006). Generally, both groups use direct quotation more frequently across disciplines than expert writers of the same fields in comparable corpora (Ädel & Garretson 2006; see also K. Hyland 1999, 2004). The similarities between L1 and L2 writers stress the influence of L2 writers’ novice status on their use of direct quotation.

There are nevertheless notable differences in direct quotation between these two groups. Texts written by L1 students contain a higher number of quotes than those written by L2 students, even at high levels of L2 proficiency (Borg 1999; Campbell 1990; Shi 2004). It is possible that the strengths and weaknesses of individual students account for the perceived differences in quoting practices (T. A. Hyland 2009; Keck 2014). The use of copied material without quotation marks may also explain the relatively low frequency of correctly signalled direct quotes in L2 writing. In Shi’s (2004) study, the L1 student writers used more direct quotes, but the L2 writers copied more strings word-for-word without attribution, pointing towards a lack of guidance as to academic conventions. Aside from frequency, there are also differences in terms of preferences for certain types of direct quotes. The L2 writers in Borg’s (2000) study used fewer citations than their L1 peers overall, but their texts contained significantly more extended and independent quotations.

---

41 EFL = English as a foreign language

The existing research has revealed that L2 student writers tend to quote in ways that are different from those of L1 student writers and experts, especially in terms of frequency, length, and embedding. L2 writers sometimes have less training and practice in academic writing than their L1 peers, at least in English. Nevertheless, frequent direct quotation is not necessarily a sign of poor intertextual writing skills, just as minimal use of direct quotation does not necessarily entail good paraphrasing. Integrating several direct quotes can actually be necessary for building a strong argument by contrasting different authors' positions and when discussing terminology and definitions (McCulloch 2012). The frequency of direct quotation also appears to depend on text type (compare studies by McCulloch 2012; Petrić 2012; Wiemeyer 2019). It is thus important to consider both the frequency and the effectiveness of direct quotes in L2 writing. The studies surveyed above show that a variety of factors influence the use of direct quotation in L2 student writing.

Since it does not require the writer to make any changes to the original wording, direct quotation is sometimes regarded as comparatively easier to accomplish than paraphrasing, summarisation or generalisation (Petrić 2012). However, empirical evidence suggests that this intertextual feature, despite its apparent simplicity, poses problems for some students when composing academic text types in an L2. It has been claimed that L2 student writers may not be aware of the stylistic and linguistic differences between their own and the author's text (Verheijen 2015). It has also been suggested that L2 student writers struggle to embed direct quotes into their texts effectively and sometimes use lengthy direct quotes where paraphrasing would have been more appropriate (Borg 2000; Hirvela & Du 2013). Apparently, L2 writers sometimes quote previously unencountered combinations of common words because they interpret them as unique to a given source text (Petrić 2012). A student in McCulloch's (2012: 60) study used direct quotes without a discernible purpose that did not contribute to her argument and appeared to substitute her own findings. L2 student writers apparently rely on direct quotes when they are unsure of how to paraphrase a source text or afraid of committing plagiarism (Borg 2000; Hirvela & Du 2013). Even postgraduates have been found to over-rely on direct quotes, especially for key terms (Davis 2013).

In Petrić's (2012) study, the L2 writers showed a certain awareness of their own tendency to rely perhaps too heavily on direct quotation, but were nevertheless inclined to quote excessively because of external pressures such as a lack of time or fatigue.<sup>42</sup> Some felt that they lacked the linguistic means to express the proposition in a sophisticated way and were afraid they might accidentally change it. They also worried that they might inadvertently reproduce chunks that they had memorised from their sources. Petrić's (2012) study shows that neither high-rated nor low-rated L2 writers always use direct quotes effectively, even when using them in legitimate ways. Students from both groups sometimes include unnecessary quotes which could have been paraphrased. Several students also admitted having used direct quotes to meet supervisor requirements despite insufficient understanding of a passage or inability to paraphrase. This corroborates the previous finding that faculty advice influences students in their decisions about direct quotation (see Davis 2013).

Some L2 writers apparently hold insecurities with respect to the functions of direct quotation. Some students' sentences consist almost exclusively of direct quotes, which Petrić (2012: 111) viewed as a "more sophisticated version of patchwriting" that is acknowledged in conventional ways. In Wiemeyer's (2019) study, students sometimes used the same quote twice in the same text and clumsily integrated quotation fragments into the co-text. L2 writers tend to use direct quotes even in the discussion sections, where they are not typically found in published papers (McCulloch 2012; Hyland 1999). Like other problematic aspects of L2 student writers' source use, ineffective and excessive direct quotation has been tied to insufficient knowledge of academic conventions and writing strategies.

However, Wiemeyer (2019) found that – despite certain individual challenges – undergraduate L2 writers are often acutely aware of the academic conventions regarding direct quotation and use them in formally correct ways. Their quotes were mostly embedded into the text in syntactically and semantically sound ways. The direct quotes varied in length, but the majority were short phrases and thus mirrored experts' writing practices. There

42 See also Abasi & Akbari (2008) for a discussion of stress as a cause of the related issue of patchwriting.

were very few instances of excessively long quotes and quotes which were not syntactically integrated. These mostly occurred when students quoted definitions or excerpts from methods chapters, making it likely that the L2 writers relied on these strategies when they struggled to understand the source text.

Overall, L2 writers' attempts to use sources in transparent ways indicate a desire to acquire a new discourse while adhering to academic standards. On the other hand, students often lack academic writing expertise, which is reflected in their problems in deciding what should or should not be quoted. Similarly to patchwriting, the over-reliance on direct quotes is likely a symptom of the "tension between students' aspirations to write like experts in the field and their endeavor to distance themselves from the language of the sources" (Petrić 2012: 114). It is thus a developmental stage of novice writers in their socialisation into the disciplinary discourse community. That this over-reliance can be remedied has been exemplified in a study of texts written by novice L2 writers before and after a teaching unit on intertextual strategies (Wette 2010), in which the students used fewer direct quotes after instruction.

Despite proficiency-related issues, L2 writers often opt for direct quotation in careful consideration of academic standards and execute them successfully. Problems with direct quotation are commonly caused by external pressures, understanding of academic writing conventions or strict abidance by teachers' advice. While the existing studies offer insights into direct quotation as employed by L2 writers, more in-depth research is needed, especially with regard to the factors governing the choice of direct quotation beyond inability to paraphrase and the ways in which quoted material is selected and incorporated into texts.

### **3.3.2 Paraphrasing and summarisation in L2 student writing**

There is a relatively large body of research on paraphrasing and summarisation in L2 student writing due to the preponderance of summary writing in academic pedagogy. Summaries are a typical text type assigned to university students because they provide insights into a student's ability to understand and condense source text material. Paraphrasing activities provide students with an opportunity to practise close reading and rephrasing of source texts and allow teachers to judge their academic reading and writing skills, especially in an L2 (Hirvela & Du 2013). It has even been proposed that paraphrasing "is the main means through which writers express the ideas of other authors, and therefore is one of the most important language issues in academic writing for international students" (Davis 2013: 126). Both paraphrasing and summarisation have also been observed to be used by students to help them better understand their sources (Plakans, Liao & Wang 2019). This emphasises their significance as central strategies in the source-based writing process.

Paraphrasing and summarisation strategies in L2 academic writing have been explored in several corpus-based studies (Keck 2006, 2010, 2014; Shi 2012) as well as studies combining textual analysis with student interviews (e.g. Davis 2013; Hirvela & Du 2013). Such investigations have taken into focus both undergraduate (see e.g. Keck 2006, 2014; Wette 2017) and postgraduate writing (Davis 2013). Other studies have considered paraphrasing and summarisation among L2 student writers' general source use strategies (e.g. Borg 2000; Campbell 1990; Wette 2017). The existing studies have focused on paraphrases and summaries in a range of text types, such as reports (Davis 2013), research papers (Hirvela & Du 2013; Shi 2008), summaries of textbook chapters (Campbell 1990), and summaries of non-fiction texts (Keck 2006, 2007, 2010, 2014). Because many studies of source use in L2 writing have been based on short summaries, their focus is usually on paraphrasing. Many do not take summarisation into account or do not consider it a distinct strategy (but see Campbell 1990; Wette 2017). For this reason, the main focus of this section is on paraphrasing in L2 student writing.

Despite the common understanding that certain forms of transgressive intertextuality are caused by a lack of knowledge of citation conventions and not by deceptive intentions, such transgressions are still widely regarded as severe academic offences (Pecorari & Petrić 2014). Much research on paraphrasing and

summarisation in L2 writing has been from the perspective of plagiarism (e.g. Li & Casanave 2013; Pecorari 2003; C. Thompson 2009). Plagiarism research has identified patchwriting and flawed referencing as features of L2 students' paraphrases (Abasi & Akbari 2008; Campbell 1990; Currie 1998; Li & Casanave 2012; C. Thompson 2009), even at the doctoral level (Pecorari 2003). More recently, however, L2 writing research has taken a less evaluative perspective on paraphrasing and summarisation in the wider context of source-based student writing (see Davis 2013; Harwood & Petrić 2013; Hirvela & Du 2013; Li 2013).

In university settings, students are required to paraphrase and summarise when composing a variety of text types, ranging from in-class exams to summaries, reviews, and research papers (Keck 2006). Their choice between a paraphrase and a summary in such writing tasks is apparently determined by a number of variables, including proficiency, text type, and use of sources.<sup>43</sup> In Campbell's (1990) study of student compositions, summaries were the most frequent intertextual manifestation in the more proficient L2 writers' texts. Like paraphrases, which were also relatively frequent, they occurred across the text. In contrast, summaries and paraphrases were relatively rare in the less proficient students' texts and occurred mostly in the body of the text. The less proficient students preferred direct quotation and used a larger amount of inappropriate copying. Cumming et al. (2005) also observed that more proficient writers' texts were characterised by a higher percentage of summaries, whereas less proficient writers tended to paraphrase source texts sentence-by-sentence.

Paraphrasing may be easier to accomplish than summarisation because it does not require the writer to comprehend the entire paragraph or text, but only those sentences to be paraphrased (Howard, Serviss & Rodrigue 2010). Undergraduate L2 writers have generally been found to rely mostly on paraphrases of individual sentences from a single source (Hirvela & Du 2013; Howard, Serviss & Rodrigue 2010; Shi 2004, 2008). Expert writers, in contrast, rely mostly on summaries because these offer the most flexibility of using one's own words to make a convincing claim and situating the cited ideas in a suitable context (K. Hyland 1999). Issues with reading comprehension may explain the minimal use of summaries in non-proficient L2 student writing. In discussing the absence of summarisation in the undergraduate research papers analysed in their study, Howard, Serviss, and Rodrigue (2010) argued that this is problematic because the reader has no way of telling whether the writer had understood or even read the source text. If their paraphrases follow the source text very closely, students may expose themselves to the danger of inadvertent plagiarism.

At secondary school level, students are expected to paraphrase in summary writing, but are rarely taught ways of accomplishing it (Feilke 2012). This creates challenges for them at university level, where there is an even stronger focus on source-based writing. University students may be expected to use referencing when they have not yet received any instruction (Crocker & Shaw 2002). They are typically advised to paraphrase in order to avoid plagiarism, but the findings of text-based interview studies indicate that many L2 writers have vocabulary issues, difficulties in reading comprehension, and limited knowledge of paraphrasing strategies (Davis 2013; Hirvela & Du 2013; Li & Casanave 2012; Shi 2008).

Many of the issues identified in L2 student writers' paraphrases and summaries have been explained in reference to their limited lexical repertoire that prevents them from substantially rephrasing source text material (Abasi, Akbari, & Graves 2006; Chandrasoma, C. Thompson & Pennycook 2004; Keck 2006, 2014; McInnis 2009). Nevertheless, L2 writers are often aware of the need to avoid textual overlap and employ strategies to avoid it. In a study of the grammatical strategies used by students in their paraphrases, Keck (2010) found that substitution, deletion, and addition of material were the most common, especially in Near Copies. In Substantial Revisions, students often formed nominalisations and added new clause elements, but relied less on substitution,

---

43 Cumming et al. (2005) found the choice between paraphrase and summary to be determined by text type and source use. It is important to note that the study only compared two text types that were distinguished by their source use, namely independent writing tasks and source-based tasks. As a result, their observation is not necessarily accurate when comparing different source-based text types.

deletion, and addition. Keck concluded that both L1 and L2 writers used paraphrases as a means of avoiding textual overlap while attempting to maintain the original meaning of source text excerpts.<sup>44</sup>

Keck's (2006) study showed that paraphrasing is a major strategy used by L1 and L2 undergraduate students in summary writing. Her study was the first which compared paraphrasing by L1 and L2 writers based on a corpus of student writing. Using 165 summaries written by L1 and L2 undergraduate writers at a U.S.-American university, Keck investigated the frequency of paraphrases, their use of paraphrases versus exact copies, and differences in the use of paraphrase types. She developed a taxonomy of paraphrases consisting of four mutually exclusive types, namely Near Copies, Minimal Revisions, Moderate Revisions, and Substantial Revisions. Each category is defined by the amount of unique links contained, which she defined as words or strings of words copied from the original excerpt which do not occur elsewhere in the source text.

While the mean length and number of paraphrases was comparable, the use of the different paraphrase types was significantly different across the two groups. The L2 writers relied more heavily on Near Copies and used a lot more exact copies than their native-speaker peers. Only a minority of them used any Substantial Revisions at all. The L1 writers, on the other hand, used significantly more Moderate and Substantial Revisions and hardly any exact copies. Keck attributed the L2 writers' reliance on Near Copies to their limited linguistic resources which, she argued, prevented them from making elaborate lexical and syntactical changes. In a later publication, however, Keck (2014) amended the taxonomy to include linguistic features in the definitions of the four paraphrase types so as to make the classification more linguistically founded. Using this refined definition and taking the students' meta-data into account, Keck was able to show that copying predominantly occurred in individual texts written by L2 student writers in their first year in the U.S., while those who had been in the country for longer also used the more advanced paraphrase types.

The differences between L1 and L2 students identified in Keck's (2006) study were thus to a certain degree caused by a small number of L2 writers who copied extensively in their paraphrases. The 2014 study also shed light on many similarities between the groups, e.g. concerning the selected source text excerpts and the retention of the chronology of the source text. A study of paraphrasing by high-proficiency L2 students (Shi, Fazel & Kowkabi 2018) showed that even a very good grasp of English does not stop students from copying from the source text when paraphrasing. It has also been suggested that under timed conditions, student writers may simply not have the time to rephrase thoroughly and may instead rely on simple source text ideas (Uludag et al. 2019: 5). L1 student writers tend to be more successful in differentiating their own ideas from those of other authors (Shi 2004). Nevertheless, like their L2 peers, they struggle to condense source text information and closely rely on the structure of the source text (Sherrard 1986). It is clear that linguistic competence is one, but not the exclusive reason for L2 writers' paraphrasing choices.

Studies have revealed that L2 writers often do not have a clear understanding of the reasons for paraphrasing. This issue is caused by superficial – or even non-existent – teaching of the purposes of intertextual reference in general and of paraphrasing and summarisation in particular. Instead of actively engaging with their sources, L2 student writers tend to view paraphrasing as a way of demonstrating knowledge and recounting ideas from source texts to complete an assigned task. Paraphrases tend to be employed for the sole purpose of attributing an idea because L2 student writers are often unaware of the functions of paraphrases beyond avoiding copying. They do not necessarily integrate and connect source text ideas to develop a coherent argument and create a common ground with the reader (Du 2013; Wette 2010). Both L1 and L2 student writers have conceded that they used source texts as a model for their own text because they were unfamiliar with the topic (T. A. Hyland 2009). Some students employ paraphrasing because they feel it is required of them (Shi 2008). As a consequence, their own voices are overpowered by those of the cited authors (Abasi & Akari 2008; Hirvela & Du 2013; Wette 2010). Without clear guidelines from their teachers, L2 student writers appear to find themselves

---

<sup>44</sup> Another observed strategy of avoiding extensive textual overlap is to replace a patchwritten paraphrase with a paraphrase that contains a direct quote. As the example of the subject in Chan's (2017) study shows, this practice is not always successful, however, and may impede the quality of the text.

in a challenging process of deciding between citing and not citing and between paraphrasing, summarisation, and quotation (Shi 2008).

One of the issues identified in the teaching of source use strategies is that paraphrasing is commonly taught exclusively as a means of avoiding plagiarism (Abasi & Graves 2008; Yamada 2003). Often, universities' plagiarism policies and brochures simply provide abstract rules and short lists of examples. These have been deemed unsuited to helping students acquire effective paraphrasing and summarisation strategies because they are more likely to instil them with fear of accidental literary theft (Abasi & Graves 2008; see also the survey of relevant studies in Pecorari & Petrić 2014). L2 students themselves have reported not to have found formal instruction on plagiarism useful (Shi, Fazel & Kowkabi 2018). Even academic writing textbooks often focus on decontextualised examples of paraphrasing and referencing that fail to capture the reality of source-based writing in the disciplines (Chandrasoma, C. Thompson & Pennycook 2004). 'Good' and 'bad' paraphrases are sometimes juxtaposed in order to instruct students how to restate source text material without patchwriting (Keck 2006; Yamada 2003).

Writing resources usually do not show paraphrases in the context of extended, authentic academic texts (Keck 2010). They thus fail to provide students with a clear idea of how paraphrases may be composed and for what purpose – other than avoiding plagiarism. This may be seen as a reflection of the understanding of intertextuality prevalent in many academic contexts (Abasi & Graves 2008). Furthermore, there is no consensus among lecturers on how much source text material a paraphrase may contain (Shi 2012), which makes it difficult for students to assess the acceptability of their textual borrowing. In some settings, students are forced to develop an understanding of paraphrasing on their own due to a lack of adequate instruction. It comes as no surprise that these efforts are not always fruitful and may lead to inappropriate conceptualisations of intertextuality (Shi, Fazel & Kowkabi 2018).

As reported by Wette (2010), teaching interventions to improve L2 student writers' source use can be successful in increasing their declarative knowledge and decreasing illicit textual borrowing, though difficulties remain. These manifest themselves especially in the selection and rephrasing of relevant source text passages and the clear distinction of source text ideas from the students' own. In a related study (Choy & Lee 2012), a ten-week course on paraphrasing and summarisation skills did lead to limited improvement of skills for some, but not all students. The students' confidence in their own ability to learn intertextual strategies and succeed in the summary task was not necessarily reflected in their actual paraphrasing skills. These studies suggest that the complexity of paraphrasing and summarisation is difficult to master within short time frames. Successful source-based writing depends on extensive academic writing experience, and L2 student writers sometimes compensate by patchwriting and copying on their way to becoming experts (see Bloch & Chi 1995; Campbell 1990; Keck 2006; Li & Casanave 2012). However, many also carefully attribute and reference their intertextual links, as outlined in the following section.

### 3.3.3 Attribution, reporting structures, and documentation in L2 student writing

The use of attribution, reporting structures, and documentation by L2 writers has not garnered much attention in linguistic research, even though this has long ago been claimed to be a feature of academic writing that novice writers "find difficult to handle effectively, both in reading and writing" (G. Thompson & Ye 1991: 366). Like direct quotation, these practices of making intertextuality explicit are usually treated as one aspect in studies investigating source use more generally. This is especially true of documentation, which is usually discussed only in the context of paraphrasing and direct quotation. Most studies of attribution have exclusively considered reporting structures, which is also reflected in the focus of this chapter. Reporting structures and reporting verbs in both L1 and L2 writing are typically investigated in corpus studies (Callies 2016; Hawes & Thomas 1997; Hyland 1999; Ishikawa 2016; Jabulani 2014; G. Thompson & Ye 1991; Thomas & Hawes 1994)

and oftentimes analysed in isolation removed from their context. A very small number of studies have used corpora to investigate reporting structures alongside other means of attribution and documentation, for example in the context of direct quotation (Wiemeyer 2017b, 2019).

The acknowledgement of sources via attribution, reporting structures, and references is an integral aspect of creating intertextuality in line with academic conventions, yet L2 writers' understanding of these practices often diverges from conventional understandings of transgressive intertextuality. Students tend to differ from experts in their views of what constitutes a good paraphrase and when it should be attributed (Pecorari 2003; Roig 1997). In Pecorari's (2003) study, for example, one student assumed that background readings did not necessarily have to be attributed. There is a tendency for L2 student writers to provide references for direct quotes, but not for paraphrases and summaries, especially if the source text is known to the reader (Campbell 1990). Shi (2008) found that her students were worried about being accused of plagiarism when their paraphrases resembled the source text too closely. At the same time, they did not believe that referencing was necessary in cases where they had made slight changes to the copied source text excerpts (see T. A. Hyland 2009 for similar findings). The students' undeveloped authorial voice may cause further issues in referencing (McCulloch 2012).

Much of the research into reporting structures in L2 writing has focused on variation in the frequency of reporting verbs and reporting structures, especially as compared to expert writing. The different patterns of use in L2 student writing in comparison to published texts are often discussed as problematic. An early study by Thompson and Ye (1991) into reporting verbs in expert writing was inspired by the assumption that L2 writers had a limited set of reporting verbs at their disposal and had issues implementing their evaluative functions. Pickard (1995) reported an over-representation of the verb *say* in L2 writing. She viewed this over-representation as indicative of students' lack of knowledge of alternative reporting verbs and limited awareness of the possibilities for stance-taking afforded by them in academic writing. Several studies have proposed that L2 student writers tend to rely on a small set of neutral reporting verbs which are also frequent in spoken English, irrespective of their L1 (Biber & Reppen 1998; Callies 2016; Ishikawa 2016; Jabulani 2014).

That the repertoire of reporting structures of some L2 writers may be limited is put forward in several studies (e.g. Davis 2013; P. Thompson & Tribble 2001; Verheijen 2015). A small number of student assignments analysed by P. Thompson and Tribble (2001) contained little variety in reporting structures and repeating the same structures, for example *according to*. Verheijen (2015: 112) reported generally lower lexical variation rates and significantly fewer reporting nouns and stance adverbs in students' writing, which she viewed as indicative of their smaller lexical and phraseological repertoire. Repetitive use of reporting structures was also observed by Davis (2013) in her study of postgraduate writers. It has been claimed that some writers do not seem to be aware of the need to use reporting structures at all (McInnis 2009), and that others sometimes use reporting verbs incorrectly, either semantically or grammatically, or combine them in unusual ways (Jabulani 2014; McCulloch 2012; Thompson & Tribble 2001; Verheijen 2015). As many of these issues appear to be the result of insufficient teaching, it is unsurprising that the use and functions of reporting verbs and structures have repeatedly been proposed to be included in writing syllabi (e.g. Davis 2013; Mei 2007; Pickard 1995).

In the light of these findings, it is important to emphasise that L2 writers employ a range of means of acknowledging intertextuality besides reporting structures. Student writers often pay close attention to referencing and attributive markers. Wiemeyer (2019: 144ff.) found that direct quotation in L2 writers' research papers is almost always accompanied by a reference, which is either provided in the same or in a neighbouring sentence. In addition, the students employed a range of means of attribution, both in and outside of reporting structures. Almost half of all direct quotes were attributed, usually to the author (35% of all quotes in the corpus), but also to the type of publication (8%) or the type of research (2%). These results emphasise that L2 writers are often aware of attribution as a conventional feature of source-based writing and may use a variety of means to indicate intertextual relationships in their texts to their reader besides reporting structures.

It is furthermore significant to stress that limited variation in reporting structures does not necessarily entail inappropriate intertextuality. Most student writers across proficiency levels regularly use reporting structures and referencing in their writing. The range of reporting verbs in EFL learners' writing appears to increase during the course of their secondary education, and these verbs are used in very similar ways by L1 and L2 students in university writing (Callies 2016). A range of reporting verbs encoding research acts, cognition acts, and discourse acts have been found in Master's theses (Manan & Noor 2014). Keck (2006) found that a majority of paraphrases in L2 writers' summaries were introduced by reporting structures. 36% of direct quotes in Wiemeyer's (2019) study were embedded in a reporting structure. These usually preceded the quote and included a reporting verb, but other structures such as *according to X* and *as X states* were also attested. Similarly, Jabulani (2014) reported that 70% of citations in students' texts were accompanied by reporting verbs or nouns. The degree of variety of reporting verbs before *that* in speech and in writing was very similar in a comparison of L1 speakers and Asian learners of English (Ishikawa 2016). There is clear evidence that L2 writers are aware of the potential of reporting verbs in disclosing intertextual relationships and utilise them to attribute content even when patchwriting (Chandrasoma, Thompson & Pennycook 2004; Davis 2013).

Generally, L2 writers at university level appear to be aware of reporting structures, attribution, and documentation and strive to apply them in line with academic conventions. As with other aspects of intertextuality, the use of reporting structures and attribution is tied to writing expertise, not language proficiency. Research has found there to be more similarities than differences between L1 and L2 student writers' citation practices, especially as far as referencing and connecting of ideas are concerned (T. A. Hyland 2009). Ädel and Garretson (2006: 279) found little overlap of reporting verbs between L1 student and professional writing, even within the same discipline. Similarly, there was no statistically significant difference in the use of reporting verbs with direct quotes between L2 students and experts in Verheijen's (2015) study. In both studies, each group used different verbs.

These findings accentuate that there is variation in novice writers' texts irrespective of native language. L2 writers' texts contain fewer references than those of L1 students, but neither group always references their sources consistently. Both L1 and L2 student writers have previously conceded having omitted references because of the task type or because they knew that their reader had access to the source texts (Shi 2004; T. A. Hyland 2009). Moreover, students from both groups who accept the content of their sources as facts tend to leave out references for their paraphrases (T. A. Hyland 2009). The studies by Keck (2014) and T. A. Hyland (2009) both indicate that differences are more pronounced between individual members of each group than between the groups. Apparently, the differences lie mostly in the cause of the difficulties: L1 student writers have named fear of plagiarism as a major reason for questionable intertextuality in their writing, while L2 writers have mostly attributed their mistakes to content- and vocabulary-related issues (T. A. Hyland 2009). As Pecorari and Petrić (2014: 285) assert, "while understanding our students' backgrounds may inform us about their plagiarism, a more fruitful enterprise may be studying their plagiarism for what it can teach us about their experiences as writers".

It is also possible that L2 student writers are sometimes judged more harshly in textual analyses than L1 students and experts when it comes to referencing. In McCulloch's (2012) study, a paragraph containing only one reference is described as possibly inciting accusations of plagiarism because the reader has to infer that the content of the following sentences is from the same source. This is despite the fact that the paragraph is coherently written and can be easily connected to the reference. Even expert writers do not provide a reference for every paraphrase (Campbell 1990) so that their intertextual nature regularly has to be inferred. Fear of plagiarism may lead some students to over-cite in the process of learning expert-like reporting strategies (Davis 2013). These aspects not always taken into consideration when L2 writers' intertextual practices come under scrutiny.

Many apparent differences between L2 student writing and expert writing are likely to result from disciplinary and individual differences. On this basis, the suggestion that a small set of reporting verbs or

reporting structures in L2 writing is an issue has been contested. In discussing the variation reporting verb frequency between the students in her study, Davis (2013: 133) raised the question of “whether this level of citation use is acceptable or whether it is essential for students to use of wider range of citation forms and functions”. She argued that variation in reporting verbs depends on discipline and that a wide repertoire may not be necessary for all students, especially in the natural sciences and technology. In L1 writing, there is indeed variation in reporting practices across disciplines. According to a study of academic texts by proficient L1 student writers (Ädel & Garretson 2006), attribution is typically realised via reporting structures containing a reporting verb across disciplines. Nominal forms are generally much less frequent than verbs and typically found in the soft sciences. There does not appear to be a set of common reporting verbs, and many verbs are specific to particular disciplines. It is likely that these disciplinary differences are also reflected in L2 student writing, yet this has not yet been investigated.

Aside from the students’ novice status, language- and culture-related aspects appear to be influential factors. In a study of writer stance by Neff et al. (2003), the students with a Romance language as their L1, especially Spanish, tended to use reporting verbs in very different ways than the American experts from the control group, while there was no significant difference between the Dutch and German students and the experts. The Chinese L2 writers in Shi’s (2004) study preferred implicit attributions such as *it has been stated* over explicitly mentioning the author in the text, which was the strategy preferred by L1 writers from the United States. In a study of attribution and stance in French expert writers’ published research articles and pre-publication manuscripts written in English, Rowley-Jolivet and Carter-Thomas (2014) found that features specific to French academic discourse had been imported into the English articles, which resulted in a weaker argumentative structure. The fact that the French expert writers were apparently often unaware of slight semantic differences between stance markers in the two languages made it difficult for the reader to identify their stance towards the cited literature. Based on the observation that the influence of French citation conventions were clearly discernible in the French writers’ English texts, the study authors concluded that “citing in English is far from straightforward for writers of other languages, and citation practices are neither language- nor culture-free” (Rowley-Jolivet & Carter-Thomas 2014: 18).

Much like the employment of reporting structures and attribution, evaluative awareness is an aspect of academic writing that novice writers gradually acquire in their academic literacy development. In a case study of academic criticism, Cheng (2006) observed that her L2 student used indirect and direct criticism in his writing to signal reservation. He engaged with this academic practice in the writing process. This led to a higher awareness of disciplinary conventions, which he tried to emulate into his own writing. Though he encoded criticism in unusually overt ways in his text, he explained that this helped him more clearly understand the current discourse on his topic and the various positions of scholars in the field (Cheng 2006: 300). Expressing criticism in a variety of ways was thus a learning strategy for Cheng’s subject, and may well be for other L2 writers as well.

Student writers are often hesitant to voice their own opinions or unaware of the significance of writer’s stance, for example because they regard the cited authors’ statements as facts or do not consider themselves to be righteous members of the academic community (see e.g. Abasi, Akbari & Graves 2006; Bloch 2010). This assumption is reinforced by the obvious lack of evaluative stance that characterises many L2 writers’ texts (Abasi & Akbari 2008; Borg 2000; Shi 2004; Wiemeyer 2019), which sometimes leads to the impression that the student is supporting arguments made by the author and not constructing their own (McCulloch 2012). There is often a strong reliance on neutral reporting verbs such as *say*, *state* or *explain* (e.g. Neff et al. 2003; Verheijen 2015; Wiemeyer 2019).<sup>45</sup> It has therefore been proposed that it may be difficult for students to understand the role of reporting verbs in subtly expressing stance in academic writing. Some researchers argue that L2 writers base their choice of reporting verb on different criteria than expert writers. Instead of focusing on their argument and on expressing a certain stance, their selection of reporting verb is possibly driven rather by stylistic concerns

45 Note that there is variation across learner populations as *state* was underrepresented in student writing in Neff et al.’s (2003) study, but very frequent in Verheijen’s (2015) and Wiemeyer’s (2019) studies. This may be connected to different writing expertise of the learner groups.

(Bloch 2010: 221). Students' choices may not always be conscious because they are not necessarily aware of the rhetorical functions of reporting structures (Pecorari 2008). Further research into reporting phrase use is needed to trace L2 writers' decisions when using these structures and their awareness of their evaluative and rhetorical potential.

Even in L2 writing research beyond plagiarism, a focus in the study of source use has been on comparing students' sentences to the source text and identifying textual borrowing as well as lexical and structural changes. What some of these studies have disregarded, however, is that some textual overlap in source-based writing is unavoidable and even necessary, for example when it comes to expert terminology, compounds, and fixed combinations of grammatical words (see Cumming et al. 2005; Shi 2004). Another aspect that is often overlooked is that textual borrowing is not necessarily a symptom of insufficient rephrasing. It can in fact be a useful strategy for improving one's vocabulary and appropriating an academic style of writing. The following section contains a survey of studies into strategic and legitimate forms of borrowing from source texts and into students' recourse to the source text as a language repertoire.

### **3.4 Towards proficiency-oriented approaches: The source text as a language resource in L2 student writing**

The emphasis on developmental aspects in explaining apparent plagiarism has spurred interest in students' strategies of writing from sources beyond plagiarism. A growing body of research has looked at students' strategies of coping with the demands of academic tasks, thus adopting a proficiency-oriented approach to source-based writing in general and textual borrowing in particular. This strand of research thus focuses on proficient source use and students' approaches to writing from sources rather than their deficits. The strategies of student writers who receive high grades in their assignments have been at the focus of a small number of studies (Ädel & Garretson 2006; Harwood & Petrić 2012; Petrić 2007, 2012; Petrić & Harwood 2013). They have shown that many L2 writers have a firm understanding of the need to avoid plagiarism, are familiar with academic writing conventions both declaratively and procedurally, and use their sources in conscious and acceptable ways, for example as a language resource.

When writing from sources, the writer connects semantic content from source texts with their knowledge from other texts and their world experience and makes inferences and elaborations (Spivey 1990). In this process, the source text functions as a repository of information as well as of ideas for the generation of new arguments. Beyond the content, the writer may also be inspired by source text expressions when connecting the language of the source text with their own: "Through mimesis the novice writer can try on the cloak of scholarly language and learn to feel comfortable in it" (Pecorari 2015: 97). Indeed, some textual overlap is unavoidable or even necessary, for example in the case of expert terminology, compounds, and fixed phrases (see Shi 2004; Cumming et al. 2005: 36). When asked about the acceptability of re-using common expressions in academic texts, academics across disciplines agree that the borrowing of general academic phrases of up to six words is acceptable and not plagiarism (Davis & Morley 2015).

The re-use of lexis from source texts appears to be a general feature of academic texts, especially those written in a foreign language. Research has shown that there is always lexical overlap between the source text and the text that is based on it, not only in L2 student writing (Currie 1998; Flowerdew & Li 2007; Gebril & Plakans 2016; Keck 2006, 2007, 2014; Pennycook 1996; Shi 2004; Vieyra, Strickland & Timmerman 2013; Weigle & Parker 2012), but also in L1 expert writing (Sun 2013). The source text regularly serves as a language

repository for academic writers, especially for those with limited vocabulary resources.<sup>46</sup> Apparent plagiarism has been found to be mostly unintentional, but students in text-based interviews have conceded that they sometimes use the source text as a lexical and phraseological resource (see Plakans & Gebрил 2012). Especially at undergraduate level and in content courses, L2 writers use the source text as scaffolding which provides them with vocabulary, grammatical structures, and academic expressions. It thus reduces the cognitive load with respect to finding the right words (Leki & Carson 1997: 56).

In testing contexts, language support has been found to be one of the central reasons behind students' source text use (Plakans & Gebрил 2012). Gebрил and Plakans (2016) have argued that this function is one of the reasons why source-based writing is so popular among instructors and students. Although some authors have portrayed textual borrowing in L2 writing as unconditionally unacceptable (e.g. Flowerdew & Li 2007), copying from the source text has regularly been recognised as a writing and learning strategy (see e.g. Currie 1998; Keck 2006, 2014; Pecorari 2003; Pecorari 2015). Nevertheless, L2 writers' strategic recourse to source text lexis has received considerably less attention in L2 writing research than apparent plagiarism.

Interviews with students have shown that some lexical similarities between the students' text and the source text are neither coincidental nor the result of an attempt to survive academically (Currie 1998). They are instead the result of students' strategic approach to writing. Participants in several case studies have attested to their use of the source text as a language resource, which is also observable in their writing processes. The student in Ruiz-Funes' (1999) case study understood the source text as a model which allowed her to imitate and copy an academic style of writing, for example with regard to syntax. During reading, she underlined expressions and grammatical structures to use in her own writing. This was evident in the final product, which shared resemblance with both the source text and the written instructions.

In Stapleton's (2010) study of the process of writing a 4000-word essay using sources, the participant documented in her log that she copied and pasted sentences from her sources. She then modified these in several ways, for example by replacing words and changing the syntax, to make the sentences adhere to her own style. The student explained that she also copied adjectives and verbs from sentences which she was not citing. She imitated syntactic structures found in her sources in order to elevate the style of her own writing to a more academic level. This strategy was recursive, as the student frequently made changes on the lexical and syntactic level throughout her composing process. The effectiveness of her strategy was confirmed by both her grade – she received a B+ from two independent graders – and an inconspicuous plagiarism check via *Turnitin*. The result of the latter indicated a low level of textual overlap with texts in the database, suggesting that unlike the participant of Ruiz-Funes' (1999) study, this student's manipulation of borrowed material was so thorough that hardly any resemblance to the source text remained. This use of source text material was thus much more successful than the transgressive intertextuality observed in many plagiarism studies.

Source texts provide L2 student writers with expressions for their writing and help them expand their vocabulary. Plakans and Gebрил (2012) reported L2 test takers' high agreement with a questionnaire item addressing the re-use of words from the source text. The study participants also agreed strongly that the readings helped them write better. The texts provided them with English expressions for concepts they were familiar with in their mother tongues, especially technical terms. Text-based interview studies (Plakans & Gebрил 2012; Wette 2017) furthermore show that students consciously search the source text for useful expressions and even use it as a model to correct spelling errors. The undergraduate students interviewed for Wette's (2017) study explained that they read adjacent sentences to those they were citing to find synonyms and terminology to use in their own writing. They also reported searching the source text PDF file for key words using the search function (CTRL+F) as a strategy to identify sections worth citing.

---

46 Whether or not the resulting intertextuality is appropriate is highly dependent on the context of writing and the prevalent interpretation of what constitutes acceptable borrowing (Polio & Shi 2012). Weigle and Parker (2012) have argued that the differences are nuanced. Some instances of textual borrowing are clearly acceptable in one context, but not in another. For example, textual borrowing may be seen as problematic if the low frequency of a phrase across academic texts marks the particular passage as an author's unique choice of words.

The use of the source text as a language resource has notable effects on students' writing. Cumming et al. (2005) found that integrated tasks resulted in texts that differed significantly from independent tasks in terms of argument structure as well as lexical and syntactic complexity. There was a higher lexical diversity in L2 writers' independent writing tasks than in source-based ones. This finding was explained by Gebril and Plakans (2006), who found that their text takers intentionally utilised the source texts of their integrated writing tasks in order to achieve more lexical diversity. In the most in-depth study of the re-use of source text vocabulary by L2 student writers to date, Gebril and Plakans (2016) looked at the effect of source text borrowing on scores in L2 academic writing assessment. Their study found a positive effect of vocabulary re-use on lexical diversity in reading-into-writing tasks, but only at lower levels of attainment.<sup>47</sup> There were significantly lower means for lexical diversity once the borrowed words had been removed. These findings empirically confirmed the self-reports by L2 writers from Plakans and Gebril's earlier (2012) study.

Gebril and Plakans' (2016) systematic identification of borrowed source text vocabulary revealed that L2 writers are most likely to borrow technical terms and proper nouns, high-level general academic vocabulary, and phrases. There was also overlap of general topic-related words. Students at higher proficiency levels were able to make their texts more lexically diverse than beginners, who compensated through borrowing. This, in turn, may have led to a higher perceived quality of their texts during grading. Interestingly, intermediate-level learners had the most benefit from their borrowing, with the resulting lexical diversity comparable to that of high-level learners. Gebril and Plakans have called for qualitative and mixed-methods studies to complement their results in order to shed light on the range and frequency of borrowed words and on the reasons why students select them.

Student writers' use of source text vocabulary – both in legitimate and illegitimate ways – in source-based writing tasks is widely attested and its effect on scores is documented in the literature. Yet, most of what is known about source text vocabulary use comes from analyses of the products of writing. The processes of strategic re-use of source text material in L2 student writing have not yet received due attention in research. It remains largely unknown why L2 student writers re-use particular lexical items, especially technical terms and general academic vocabulary, and why some are used more than others. This highlights the need for research into L2 writers' use of the source text as a language resource.

The studies reviewed above have helped elucidate L2 student writers' practices of writing from sources. Their findings are highly relevant to academic writing pedagogy because understanding these practices is a prerequisite for designing realistic and goal-oriented source-based writing support, despite certain shortcomings of individual studies (see also section 5.1). Much of the existing research has focused on the products of writing, i.e. the texts produced by L2 writers in various settings and genres. The processes of creating intertextuality and interacting with the source text in such writing tasks have not received an equal amount of attention. The attention to products instead of processes in the literature is mirrored in English for Academic Purposes (EAP) courses, where the focus is usually on the structures and linguistic characteristics of written genres, probably because procedural knowledge is much more difficult to convey (see also Dovey 2010 for a discussion of a lack of process approaches in genre pedagogy). However, exclusively teaching learners of academic English about what the products should look like does not equip them with the complete skill set necessary for successfully managing academic writing tasks (Dovey 2010). It is therefore equally important to consider the processes of source-based academic writing. The next section summarises previous research into these processes.

---

<sup>47</sup> In other studies, the concern has been raised that textual borrowing may lead to improper assessment of L2 writers' ability in testing contexts if testers do not notice the copied passages or do not regard copying as an issue (see Cumming et al. 2005; Weigle & Parker 2012).

## 4 Writing processes of source-based academic writing

Writing is a multilayered cognitive process that combines a variety of different actions, such as planning, writing, and editing, and results in the written product. Some phenomena and issues that occur during the writing process may not be observable in the product, which is why it is important to consider the process in its own right. This section complements the previous remarks on the products of source-based academic writing with a perspective of the writing process. Writing processes are typically studied using observational, introspective, and retrospective approaches (Polio 2012; see section 4.3). Based on empirical findings, typically from think-aloud protocols, several models have been developed to analyse and describe the writing process, both in the L1 and the L2.

### 4.1 Models of L1 and L2 writing processes

Traditionally, models of the writing process have taken into focus the cognitive processes of writing (see overview in Galbraith & Vedder 2019). The most widely acknowledged model of the L1 writing process was developed by Flower and Hayes (Hayes & Flower 1980; Flower & Hayes 1981). They argued that the traditional assumption of a linear succession of pre-writing, writing, and re-writing does not fully capture the recursiveness of the writing process in which writers are constantly planning and revising as they write. The authors conceive of the writing process as a “set of distinctive thinking processes which writers orchestrate or organize during the act of composing” (Flower & Hayes 1981). Based on the analysis of writers’ think-aloud protocols (Hayes & Flower 1980), their model identifies three major subprocesses of writing on the macro-level, namely *PLANNING*, *TRANSLATING*, and *REVIEWING* (see Figure 4).

These subprocesses are hierarchically organised and highly embedded, which means that one subprocess may occur within another. *PLANNING* refers to combining information from the task and the writer’s world knowledge to set goals and determine steps for completing the writing task. It is comprised of the subordinate processes of generating, organising, and goal-setting. *TRANSLATING* refers to the subprocess in which information is verbalised into written text. *REVIEWING* serves to improve the textual quality and entails subprocesses of reading and editing. All these subprocesses of writing are to be seen in the context of the task environment, which comprises the topic, audience, and motivational factors of the writing assignment, but also the text that has already been produced. The subprocesses of writing are additionally shaped by the writer’s long-term memory, i.e. their knowledge about topics and audiences. According to the model, writing is goal-driven and the writing process is constantly monitored to ensure the fulfilment of these goals. Because writing goals are individually different, so are writing styles and consequently writing processes. The model attempts to account for these differences as it does not specify the order and length of the subprocesses.

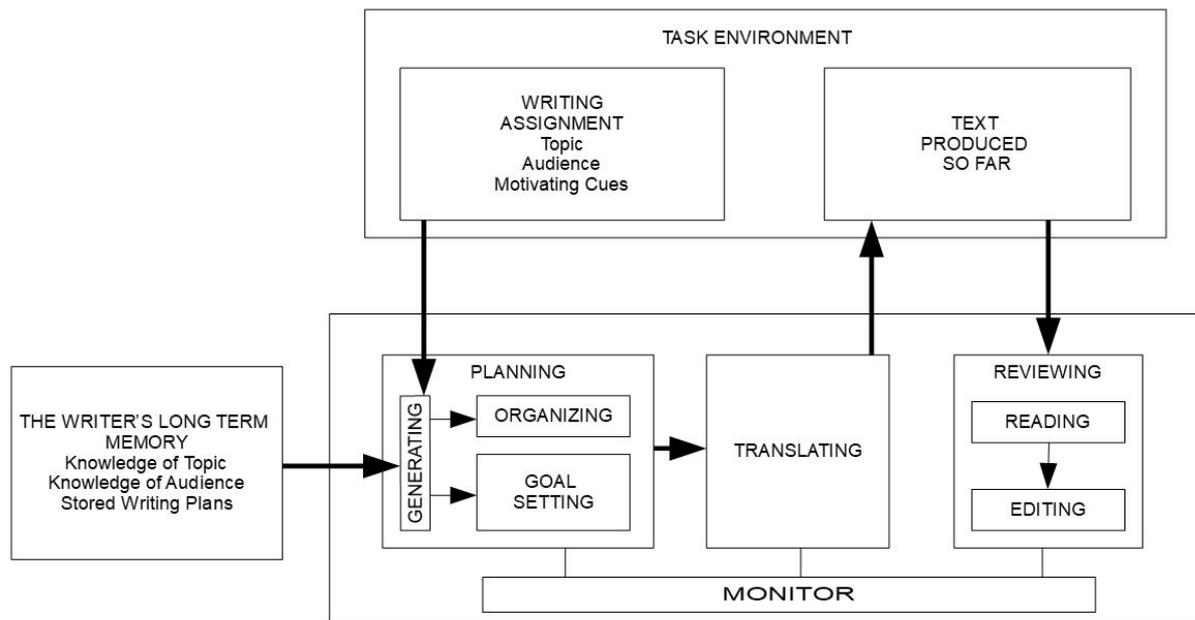


Figure 4: Hayes and Flower's model of the writing process (author's own visualisation based on Hayes & Flower 1980: 11).

According to Flower and Hayes (1981), the writing process is structured into three phases which are distinguished by different subprocesses. In the first phase, generating is the dominant process, resulting in single words and incomplete sentences. In the second phase, organising processes lead to a more systematically structured text. Finally, in the third phase, translating processes are employed to create complete and correct sentences. As a result of different processes being foregrounded, the form of the text changes from phase to phase. Hayes and Flower (1980) emphasise that the actions that constitute the process are repeated and interrupt each other in the course of writing until the writer is content with the final product. Some subprocesses, for example generating and editing, appear to be given a higher priority than the processes which they interrupt. This led Flower and Hayes (1981) to propose that writing is a recursive and non-linear process. The phases in their model may be overlapping and embedded and, as the authors emphasise, are not completed one after the other. Despite the pervasiveness of the model, it must be noted that its conception was based on a very small set of empirical data of very advanced L1 writers and that its focus is on planning, while the other phases are theorised in much less detail.

The influential model by Flower and Hayes prompted research into its applicability to writing in an L2 and was adapted by Börner (1987), who integrated aspects such as teaching agendas, the writers' writing experience in their L1 and L2, and their L2 proficiency. An important addition to the model as pertaining to L2 writers was the role of intertexts both in the L1 and the L2, which allows for a consideration of linguistic issues such as the identification of appropriate expressions and grammatical structures as well as the correction of language errors.<sup>48</sup> The subprocesses of identification, evaluation, and solution of L2 problems were also posited as a central difference between L1 and L2 writing processes in Krings' (1989) empirical study of L2 writing processes as revealed by think-aloud protocols. Krings accordingly placed planning and problem-solving subprocesses at the centre of his model. Zimmermann (2000), on the other hand, focused on the subprocesses of formulation in the conception of his model of the (L2) writing process. The model acknowledges the external factors proposed by Hayes and Flower (1980) as given and situates formulating (the equivalent of TRANSLATING

<sup>48</sup> See Zimmermann (2000) for a translation of Börner's (1987) model into English.

in Hayes and Flower's model) between planning and reviewing. According to Zimmermann's model, all three subprocesses are subject to L2 problems and their solutions.<sup>49</sup>

The model assumes global and local planning to precede writing, though these processes may be recursive. Formulating is conceived of as the prerequisite to writing and to consist of typical and peripheral subprocesses. Typical subprocesses include the formulation of a tentative form, its evaluation, and subsequent acceptance, which are followed by writing it down (Zimmermann 2000: 85ff.). In Zimmermann's model, only very few processes, namely the formulation of tentative passages in the L1, the simplification of formulations, and L2 problem solving, are regarded as L2-specific. L2 problems, e.g. word searches, and their solutions are viewed as being most dominant in formulating and revision sequences, but their position is posited to be unpredictable (see also Krings 1989: 397ff. for a description of L2 problems in the writing process). It is noteworthy that writing in the L2 prompts more revision processes than writing in the L1, most of which are grammatical and orthographic changes rather than reformulations (Zimmermann 2000: 89). Zimmermann stresses that individual differences are clearly more influential than L2-induced differences, confirming the assumptions as to the individuality of writing processes made by Hayes and Flower (1980).

Despite the recursiveness of the writing process recognised in process models and empirical studies, there does appear to be a certain linearity to it in terms of planning, writing, and editing, which has prompted various L2 writing scholars to subdivide the writing process into pre-writing, writing, and post-writing stages (e.g. Chan 2011; Choi 2016; Plakans 2008). Focusing on the subprocesses of formulating at the micro-level, which were not considered in previous models, Zimmermann (2000) argues that the writing process can be linear and non-linear at the same time. Certain micro-level processes of formulating, especially processes of editing, may occur at any time during the process and are somewhat unpredictable. Other micro-level processes do not occur as flexibly and are restricted to characteristic sequences, often at specific points in the writing process (Zimmermann 2000: 84). Based on these observations, it can be assumed that writers follow certain patterns in their writing processes and that its structure is not arbitrary.

## 4.2 The influence of source use on the writing processes

Recently, there has been a shift in research away from cognitive processes to the processes of text production. This shift has been accompanied by a diversification of research methods employed in process studies (see overview in Galbraith & Vedder 2019). By employing methods such as keystroke logging and eye-tracking, often in combination with interviews or stimulated recall, writing scholars have investigated aspects of the writing process such as pausing behaviour, error revision, reading and writing strategies, writer identity, L1 vs. L2 fluency, and L1 use in L2 writing, among others (e.g. Breuer 2014; Chan 2017; Leijten & Van Waes 2013; Sasaki 2000; Spivey 1990; Wang & Wen 2002; Zhao 2011). One area of interest has been writing from sources in L2 English, especially the processes of integrated writing tasks in testing contexts (see detailed overview in section 4.3).

Source-based writing processes are a special type of writing process. They encompass an interplay of reading, thinking, planning, and writing, i.e. of receptive and productive processes (Dovey 2010; Jakobs 2003; Solé et al. 2013). The writer reads the source text and makes sense of it, then aligns it or contrasts it with their own arguments to create new meanings. Information is organised, selected, and integrated so that new connections are created (Dovey 2010; Spivey 1990). These actions result in a complex, multifaceted writing process that is characterised by the recursiveness and embeddedness of subprocesses and in which productive and reproductive aspects of creating intertextuality are interconnected. In both L1 and L2 writing, the interaction with source texts and integration of information results in an increased cognitive load for the writer. This

<sup>49</sup> See Zimmermann (2000: 85) for a visualisation of his model of the (L2) writing process.

necessitates the use of adequate strategies in the writing process to balance the demands of source-based writing (Kirkland & Saunders 1991; Leijten et al. 2019). Transforming ideas from the source in a meaningful way requires the writer to employ a combination of conceptual, rhetorical, and linguistic operations. Writers have to monitor both what they want to write and how they want to write it, and they have to comprehend, integrate, and elaborate source text ideas accordingly in their writing process (Solé et al. 2013). This interplay of content and rhetoric presupposes an understanding of the source text as well as the ability to integrate different ideas to create a consistent line of argumentation in one's own text.

Source-based texts are created in complex writing processes which entail a variety of actions and decisions on the macro- and the micro-level. On the macro-level, the writer engages in processes such as understanding the task, planning and structuring their text, and reading the source text (see Choi 2016; Plakans 2010). In addition, the writer has to perform linguistic operations on the micro-level, e.g. paraphrasing, integration of direct quotes, and formulation of reporting structures. Research into L2 writing processes has confirmed that such processes, too, are recursive (e.g. Ruiz-Funes 1999). The writer moves back and forth between the different subprocesses and between their own and the source text during the composition of a text (Chan 2017; Choi 2016; Plakans 2008). Mastering the writing process on both the macro- and the micro-level allows the writer to move beyond knowledge display towards knowledge transformation (Bereiter & Scardamalia 1987: 10), a commonly assumed goal of writing education (Segev-Miller 2004; Solé et al. 2013). This transition may, however, be hindered by cognitive overload, which is a constricting factor in the writing process. Because of limited cognitive capacity, writers may not always complete all subprocesses effectively (e.g. Flower et al. 1986: 44). This is especially true of novice writers, who often focus their energy on individual aspects of writing, neglecting others (Galbraith & Vedder 2019: 634). With experience, novice writers tend to develop strategies of reducing cognitive overload such as note-taking to more successfully cope with the demands of the writing process (see Heine et al. 2014: 131)

It has been argued based on empirical studies that the process of source-based writing in the L2 can be broadly divided into two major stages: the pre-writing stage, in which the writer finds and reads source texts and plans their text, and the writing stage, which consists of processes such as planning, formulation, re-reading, writing from the source text, and in-text revision (e.g. Chan 2011; Plakans 2008). A third stage dedicated to post-writing in which the writer edits and revises their text has been proposed in other studies (Choi 2016; Leijten et al. 2019), though revisions are viewed by some scholars as a part of the writing stage (e.g. Plakans 2008). Because of the recursiveness of the writing process, it is assumed that there are neither clear boundaries between the stages of source-based writing nor that there is a clearly linear succession of the strategies employed. Instead, the subprocesses frequently overlap and interrupt each other, especially in more experienced writers' processes, as proposed in Hayes and Flower's (1980) model. The recourse to other authors' texts evokes a writing process that tends to be even more recursive and less continuous than processes of writing which are not source-based (Chan 2011; Plakans 2008). Source-related processes such as finding sources, reading, highlighting, taking notes, and composing from sources are interwoven with other composition processes and reiterated. Their order and succession is goal-oriented, determined by the writer's choices, and culminates in the completion of the writing task (Krings 1992; Li 2013).

It is evident from the literature reviewed in section 3 that the majority of studies has focused on the products of source-based L2 writing. The processes of writing from sources in an L2 have been at the focus of a smaller number of studies. These studies have considered a variety of tasks and settings and used a range of methodological approaches. They are presented in the next section.

### 4.3 Process-oriented perspectives on source-based L2 student writing

The processes of source-based writing in the L2 have been at the focus of a small number of studies, especially in the context of language assessment. Macro- and micro-level processes of L2 writers' source use have been investigated via one-time integrated writing tasks both in the context of English language assessment (Chan 2011, 2017; Gebril & Plakans 2009; Plakans & Gebril 2012) and in composition classes (McInnis 2009; Ruiz-Funes 1999). Other research has compared the processes of writing from sources to the processes of writing-only tasks (Plakans 2008) and examined the actions that L2 writers take to resolve contradictions in source-based writing from the perspective of activity theory (Li 2013). A handful of studies has looked at the real-life processes of writing academic texts in either L1 student writing (Solé et al. 2013) or L2 student writing (Choi 2016; Li 2013; McCulloch 2013; Stapleton 2010) or both (McInnis 2009). Text types under investigation have included research articles (Choi 2016), Master's theses (McCulloch 2013), and argumentative essays based on either non-fictional (Bosher 1998; Chan 2017; Gebril & Plakans 2009; McInnis 2009; Plakans 2008; Plakans & Gebril 2012) or fictional source texts (Ruiz-Funes 1999). Participants completed their tasks either using pen and paper (e.g. Bosher 1998) or on screen with the help of word processing software (e.g. Stapleton 2010).

L2 writing scholars have employed a variety of methods to gain insights into the ways in which sources are used when undergraduate and postgraduate L2 students write in a variety of contexts. Observational approaches to L2 writers' intertextual processes have included screen recording (Chan 2011), video recording (Bosher 1998), and keystroke logging (Chan 2017; Leijten & Van Waes 2013). Think-aloud protocols are typically the method of choice for introspection and have been used in a number of process studies (e.g. McCulloch 2013; Plakans 2008). Others have employed reading and writing logs that L2 writers filled in during a task (e.g. Choi 2016; Ruiz-Funes 1999). Studies relying on retrospection have used instruments such as stimulated recall (Ruiz-Funes 1999; Sasaki 2000), interviews (e.g. Choi 2016; Li 2013), retrospective questionnaires (Gebril & Plakans 2009), and process logs filled in by the students after writing (Li 2013). Some scholars have combined data from observation, introspection, and retrospection for triangulation, e.g. video recordings and stimulated recall (Bosher 1998; Sasaki 2000), process logs and stimulated recall (Ruiz-Funes 1999) or video recordings, keystroke logs, process logs, and retrospective interviews (Choi 2016). Combining process methods allows researchers to capture various aspects of the writing process and to compensate for the drawbacks of individual methods (Bailey & Withers 2018).

Most of what we know about L2 writers' processes of writing from sources has come from process research into reading-into-writing tasks, also referred to as reading-to-write, reading-writing or integrated writing tasks (see e.g. Asención Delaney 2008; Kim 2001; Plakans 2009a; Weigle 2004).<sup>50</sup> While reading and writing generally elicit different processes, it has been shown that many processes are shared between reading tasks and writing tasks, such as meta-cognitive monitoring, using background knowledge, re-reading, and summarising (Plakans, Liao & Wang 2019). Moreover, writing tasks which include the use of sources – and thus inevitably require reading – have been shown to give rise to a very characteristic set of processes (McCulloch 2013). Reading-into-writing tasks can be viewed as a combination of a receptive and a transformational process. The receptive process presupposes and also promotes basic academic literacy, while the transformational process presupposes and also facilitates the acquisition of critical literacy (Flower et al. 1990: 4f.; see also Hirvela 2016: 128). Reading-into-writing processes have been approached from an input-based and from an output-based perspective in the literature (Hirvela 2004). While the input-based view focuses on how writers learn about writing by reading, the output-based view, which was adopted in the studies reviewed here, focuses on “the act of writing and the text-production processes that enable the writer to appropriate source text material in accepted ways” (Hirvela 2016: 128).

---

50 Spivey observed that “using texts as sources is a very common way of going about writing and a very common reason for reading” (1990: 258). Despite the fact that source-based writing is situated at the nexus between reading and writing, these two lines of research have often been followed separately.

Unlike impromptu writing-only tasks, which are also referred to as independent writing tasks, reading-into-writing tasks require learners to read and summarise or synthesise one or more source texts (Weigle 2004). They are often assigned so that students can gain a more profound understanding of a topic (Solé et al. 2013). The source text, which can be either written or spoken or both, is typically included in the writing prompt (Plakans 2008). Reading-into-writing is of interest in language testing and assessment because this type of task is often viewed as more accurately reflecting the learners' language proficiency than writing-only test items. Multiple skills in the areas of reading comprehension and writing are necessary to complete these tasks and influence the writer's success (see e.g. Plakans 2009a, b). More importantly, such tasks elicit specifically academic writing processes and thus require a command of strategies of appropriately introducing ideas from one or more sources into one's text (Plakans & Gebril 2013). Because reading-into-writing tasks in language tests require students to perform processes that are – at least to a degree – reflective of actual academic contexts, they are perceived as more authentically representing the real-life writing conditions of novice academic writers (Chan 2017; Gebril & Plakans 2016; Plakans 2009b). Analyses of the processes of integrated writing tasks have been useful in explaining differences and similarities across student texts and how they relate to proficiency and the validity of assessment measures (Gebril & Plakans 2009).

The use of source texts makes reading-into-writing tasks challenging, especially for L2 writers who have not been formally instructed on this type of assignment. Some students participating in process studies of reading-into-writing tasks have noted finding these tasks difficult due to the demands of selecting, synthesising, and interpreting source text information (see Chan 2011; Ruiz-Funes 1999). Interestingly, Chan's (2011) test taker subjects reported that despite significantly lower scores on the summary than on the essay task, they preferred writing summaries for teaching and learning purposes. This corroborates findings by Plakans (2008), whose students found the source texts helpful for finding information and ideas for their text, and by Leki and Carson (1997) and Ruiz-Funes (1999), whose study subjects reported using source texts as models for academic expressions and syntactic organisation. Apparently, the language learning opportunities and argument support afforded by the source text have a motivational effect on L2 writers that overrides the challenges.

Source-based writing tasks are also popular among teachers because of their epistemic potential: unlike writing-only tasks, they presuppose attentive reading, the critical analysis, selection, and elaboration of information, and the use of intertextual strategies. Reading-into-writing tasks require the writer to strike a balance between content and rhetoric and are assumed to encourage knowledge transforming (Solé et al. 2013; Plakans & Gebril 2013). A growing understanding of the topic at hand often develops in parallel to the writing process (Li 2013). In addition, these tasks are assumed to be less creatively demanding than independent tasks since the source texts provide students with ideas for their writing and support in terms of content and language (Gebril & Plakans 2016; Plakans 2008; Weigle 2004).

For these reasons, integrated writing tasks have been gaining popularity in language testing, though they may create new challenges such as selecting source texts, avoiding patchwriting, and citing correctly (Gebril & Plakans 2009; Plakans 2008). Outside of language testing contexts, reading-into-writing tasks are frequently assigned in composition and EAP classes to create alignment with the writing tasks students have to complete in disciplinary courses, because they are also a typical tasks in disciplinary courses themselves (Plakans & Gebril 2012). In both language testing and EAP contexts, reading-into-writing tasks are increasingly assigned in combination with writing-only tasks and have sometimes replaced them (Gebril & Plakans 2009). This development has led to a growing number of studies into the differences in processes and products between reading-into-writing and writing-only tasks to address issues of development, rating, and construct validity.

Process studies of reading-into-writing tasks have allowed researchers to create representations of the source-based writing process and explore the functions of its subprocesses. According to Ruiz-Funes (1999), processes of writing from sources in a foreign language are comprised of composing processes present in any writing process (planning, writing, revising, and editing), but also additional processes which are determined by source use. Source-based tasks trigger processes of elaborating (by making inferences and integrating previous

knowledge), structuring (at the linguistic and content level), monitoring, and synthesising. In Ruiz-Funes' study, students employed the two central processes, elaborating and structuring, to balance source text information and their world knowledge by evaluating and judging the source text and carefully selecting and generating ideas for writing. Monitoring was used to ensure that the information and ideas taken from the source text were complete and accurate, while synthesising served mainly to generate an understanding of the source text, abstract from it, and create a plan for writing.

That source text interactions during reading predetermine the composition process was shown by McCulloch (2013), who examined the reading component of real-life reading-into-writing processes using concurrent think-aloud protocols. The two M.A. students used a range of reading strategies and engaged actively with their sources. The students showed very different profiles of strategy use, each of which was reflective of the student's individual writing goals (see also Choi 2016). Their pronounced intertextual awareness was apparent from their making explicit connections across the source texts and paying attention to author's names. They appeared to judge source text excerpts in terms of their usefulness for their theses and took notes on aspects of interest, implying a focused and goal-oriented reading process on the part of these advanced students. McCulloch's study is suggestive of complex source text interactions during reading which shape the writing process.

Solé et al.'s (2013) process study, though it examined young L1 writers' composition processes, offers further interesting insights into the organisation of source-based writing processes. The participants, teenagers aged 15 to 16, relied closely on the source text in their writing process, which meant that the written products closely resembled them in terms of both content and structure. The students either rapidly switched between theirs and the source text to select ideas to immediately copy into their text or they re-read more thoroughly and then elaborated the information. Three different patterns of integration emerged: Students employing a linear/reproductive pattern followed the source text sequence, re-read sparingly, and tended to copy directly and paraphrase closely with little revision. A linear/elaborative pattern entailed more elaborate use of paraphrases, deeper re-reading, and a more structured approach in which ideas were selected and then paraphrased and elaborated, with more extensive checking and revision. Finally, a linear/elaborative pattern with elements of recursiveness involves a more complex interplay of reading, planning, re-reading, and elaboration, and frequent recourse to the source texts, which continued through the final revision. The resulting texts also differed in terms of how thoroughly the information from the source text was combined and integrated with the students' own ideas.

The study found that more complex processes lead to more expertly synthesised texts, suggesting a correlation between successful re-reading and integration processes and the quality of the product. Importantly, borrowed excerpts were manipulated repeatedly during writing to decrease their similarity to the source text. However, the participants' integration strategies, e.g. paraphrasing, synthesising, and elaborating, were not always successful and sometimes lead to the misrepresentation and omission of source ideas, especially among students who copied and paraphrased without restructuring or mediating information. Solé et al.'s study bears evidence to the fact that the product does not necessarily reflect the process. In fact, many of the ideas and connections the students generated during reading did not make their way into the final product. Research from L2 writing (e.g. Chan 2011) implies that a similar correlation holds for non-native writers, though this has not been explored in detail.

Comparisons of reading-into-writing and writing-only tasks have established that though the composition processes are largely the same on the macro-level, there are some differences in their organisation and realisation. Plakans (2008) identified subtle differences between the processes of the two task types in both the pre-writing and the writing stage. In the pre-writing stage, students spend more time studying the prompt after reading the source text to remind themselves of the task. In the writing stage, reading-into-writing processes are less continuous and include more online planning than the processes of writing-only, but the students do not seem to have to reorient themselves as frequently. Studying the processes triggered by two test items – a

summary and an essay task – used in a high-stakes English language test, Chan (2011) observed that the source-based writing task elicited different cognitive processes: macro-planning and discourse synthesis were common in the source-based task, while micro-planning, monitoring, and revising dominated the independent task. The differences between writing-only and reading-into-writing tasks have furthermore been found to extend to the written product. For example, the discourse features of the products of writing-only differ significantly from those of integrated writing tasks, e.g. in terms of syntactic complexity, lexical variation, and argument structure (Cumming et al. 2005).

Interestingly, phenomena which are typically discussed in the context of plagiarism, e.g. patchwriting and textual borrowing, have emerged as strategies in reading-into-writing tasks (e.g. Esmacili 2002) and were considered to be so-called test-wiseness strategies in a study of integrated tasks used in the TOEFL iBT (Yang & Plakans 2012). Such source-related subprocesses are the most important difference between reading-into-writing processes and writing-only processes (see Plakans & Gebril 2013 for a discussion of how this difference relates to score). Despite observable differences in the processes and products, the task representations of some L2 writers for these two types of tasks have been found to be quite similar, even identical. Other L2 writers approach them in very different ways, specifically by awarding a central role to the information from the source text. This points towards a certain degree of individuality in the respective task representations that is reflective of academic writing experience and only mildly related to source use (Plakans 2010).

Many of the insights into the strategies of source text use that occur in reading-into-writing tasks has come from a series of studies by Plakans and Gebril, who used think-aloud protocols, retrospective process questionnaires, and interviews with test takers to investigate the processes and products of source-based writing in language testing contexts (Gebril & Plakans 2009, 2016; Plakans 2008; Plakans & Gebril 2012, 2013). Their research was based on reading-into-writing tasks from university-administered English placement tests and the Internet-based Test of English as a Foreign Language (TOEFL iBT). In the first study of the series (Gebril & Plakans 2009), source use in the writing process was quite uniform across proficiency levels as measured via scores. Irrespective of language proficiency, writers used the source texts as an inspiration to form opinions and find ideas for their texts. They felt comfortable with integrating others' ideas into their text. In the process questionnaires, some students reported using the source texts to structure their essays and to return to the source text for ideas occasionally during writing. They indicated moderate amounts of paraphrasing and re-use of some source text vocabulary.

Interestingly, this study also revealed that students across levels strongly rejected having copied directly from the source text though they had in fact done so. This contradiction suggests that the students were either unaware of their own patchwriting or were unwilling to disclose it, rendering the process questionnaires used in the study unreliable for the investigation of this sensitive issue. The researchers identified only minor differences in source text use between writers of different proficiency levels. Writers on the highest proficiency level were concerned about not knowing the scientific terminology required for the writing task, while lower-achieving students paid more attention to spelling and understanding source text terminology. The use of authors' names was most pronounced among the most proficient group, who also most frequently reported knowledge of the issue of plagiarism.

A follow-up study (Plakans & Gebril 2012) used a mixed-methods approach to the functions of source use in integrated writing tasks. The study confirmed that source texts helped students shape their opinions for the argumentative essay. The texts provided ideas and support for the students' arguments, which were important in meeting the content- and discourse-related task requirements. The study brought to light that the students used the source text as a constant companion, frequently checking for ideas and correctness during writing, which was also observed by Solé et al. (2013). Writers at all three proficiency levels also reported language-related learning from the source texts, as the source texts provided them with general and technical expressions and were helpful for spelling.

Other research on integrated writing tasks in language assessment has corroborated the finding that students use source texts to generate ideas and as a linguistic resource. In Chan's (2011) study, much of the students' interaction with the source text, such as selecting and organising ideas, took place in the pre-writing stage and was subject to individual differences. The test takers frequently paused writing to re-read both the source text and their own and to check their text against the source. Though all test takers employed discourse synthesis strategies, some students' low scores suggest that not all used them effectively.

A related case study (Chan 2017) produced similar results in that both participants' discourse synthesis processes were closely intertwined with revision processes. However, individual differences emerged in terms of how they selected and organised ideas from the source texts. The higher-rated student organised and connected ideas autonomously before writing, whereas the other lifted ideas from the text, retained their order, and connected them using short formulaic expressions without any attempt at interpretation or elaboration. After a planning phase, the higher-scoring student produced the text in a fluent process without recourse to the source text, adding his own interpretations and stance. He stated that he had selected and organised ideas in his mind while reading and used the readings as reminders of ideas and spellings, but recounted their content in his own words.

The lower-scoring student, in contrast, frequently consulted the source texts and the task prompt while writing. She relied on them for content and copied extensively from them, making only minor changes to lexis and grammar. In the interview, the student explained that she intentionally based her text closely on the sources to show that she had read them thoroughly. She also misinterpreted a warning of plagiarism in the prompt as a suggestion to generously quote from the text. The quotes' minimal embedding made them difficult to interpret. Furthermore, she employed mostly deletions to reduce similarity to the source. Despite general similarities between students, then, their overall success seems to be impacted by the quality of their selection and paraphrasing processes.

The cited studies show that L2 student writers employ a range of processes in their source-based writing and are aware of their own strategies, especially at higher proficiency levels. More experienced writers transform knowledge and reorganise and interpret source text information in the writing process. An important implication of previous research is that proficiency in the L2 is not necessarily a predictor of intertextual strategies. Issues in reading comprehension and lack of paraphrasing skills may lead to limited and inappropriate source use at lower proficiency levels, but not necessarily for more advanced learners. Writing experience and knowledge of citing conventions appear to positively impact the way the students approach integrated writing tasks, and misconceptions may lead to inappropriate source use. The source texts' role in the writing process as a model in terms of ideas and evidence as well as vocabulary, spelling, and textual organisation has become evident, pointing towards two important functions of source text use: one of supporting content generation and argument construction and one of supporting language use (see also Gebril & Plakans 2016).

A growing body of research has investigated the processes of source-based writing in experimental settings beyond the language assessment context, though such studies of L2 writers are still relatively rare. Boshier's (1998) study was the first experimental study of L2 writers' intertextual strategies outside of testing and EAP contexts. Three South-Asian L2 writers' strategies of coping with problems in the writing process were investigated using videotaping and retrospective interviews with stimulated recall. The students' processes differed in the number and quality of problem-solving strategies as well as in the extent of planning and textual organisation. One student focused mostly on generating text and attended to surface features, while the other two were most concerned with structure and content. There were quantitative and qualitative differences in the extent to which source text ideas were integrated into the essays, which according to Boshier reflected the students' individual strengths and weaknesses. Despite lower fluency in English, one student synthesised and challenged source text information. Another copied some passages verbatim and integrated them with her own opinions. The third wrote a summary that mirrored the same stance and did not integrate source text ideas with their own. Like

the product-based studies reviewed above, this study suggests that similar language proficiency is not necessarily indicative of comparable intertextual strategies and that writing experience has a much more profound influence.

McInnis' (2009) study also found that writing processes in source-based writing are individually different, but she argued that native language plays a role in shaping these processes. Using a mix of corpus data, stimulated recall, and think-aloud protocols, she compared the processes of paraphrasing of three L1 and three L2 community college students in an experimental setting. Her students were asked to paraphrase individual excerpts of a source text they had read at home. Both groups showed awareness of illegitimate textual borrowing and saw paraphrasing as a way to avoid plagiarism. All participants were eager to use their own words, find synonyms, and abstain from copying from the source, as evident from their verbal protocols, though they sometimes expressed uncertainty as to how much copying would be acceptable. The L1 writers made a conscious effort to identify the main idea of the excerpt, to include all the key points in their paraphrases, and to restructure the original sentences. They also declared a preference for converting each source text sentence into a paraphrase. Notably, the L1 writers experienced little difficulty with finding alternative expressions and grammatical structures.

The L2 writers, in contrast, did not appear to place emphasis on grammatical restructuring, but made frequent use of dictionaries and thesauruses. Cultural aspects also seemed to play a role in the students' processes of paraphrasing, as the L1 writers were familiar with the cultural concepts discussed in the article, while the L2 writers were not.<sup>51</sup> All three voiced concern over re-using source text lexis, but they took different approaches to this challenge in the writing process: One L2 student asserted that every word from the source text had to be changed in a paraphrase and took much effort to do so. Another L2 student left a copied sequence unchanged despite being aware that her teachers may disapprove, explaining that if she felt unsure, she did not make changes. The third student opted for a source text expression even though she had identified a suitable synonym in a dictionary. Synonym substitution was a much more challenging task for the L2 writers, who spent more time and cognitive energy on rephrasing. Though they showed awareness of subtle differences in meaning, their synonyms did not always accurately convey the original proposition. The study showed that limited vocabulary sometimes prevented the L2 writers from fully understanding the source. McInnis' (2009) findings are line with previous research showing that cultural background paired with limited vocabulary has an effect on paraphrasing strategies and may lead to patchwriting and incorrect presentation of source text information.

That source use strategies may be related to native language was also proposed in a study using keystroke logging to record L1 and L2 writers' processes. Studying Dutch Master's students writing in Dutch and English, Leijten et al. (2019) found that the students spent more time reading texts in their L2 than in their L1, but that the ratio between reading and writing was roughly the same in both languages. Apparently, source use takes more time in L2 writing. The average quality of texts was very similar in both language and did not change throughout the academic year, which shows that students do not necessarily improve their writing skills in this time frame. More interestingly, the quality of the texts was only positively impacted by source use in the students' L1, possibly because they used more sources. This suggests that the major cause of differences between L1 and L2 source-based writing is time: Writers need more time to process their sources in their L2, so reading as many sources as in their L1 may not have been possible in this timed writing task.

Recently, the processes of intertextual writing in authentic settings have garnered attention, as evident from a small set of studies of real-life source-based writing tasks. The motivation behind such studies of naturalistic writing tasks, as McCulloch (2013) argues, is that the one-time tasks assigned in the majority of process studies insufficiently reflect the complexity of real-time source-based writing. The processes and strategies typical of long-term writing processes, e.g. of writing research papers, cannot be discerned from isolated tasks that are completed in one session (Choi 2016). Due to the breadth of naturalistic writing sessions in terms of time, materials, and perhaps locations, they are typically investigated in case studies in which the students are instructed on the use of process data collection tools and allowed to choose the settings and time

51 See also Plakans (2008) for a discussion of the impact of culture and educational background on writing processes.

frame in which to complete the task. The assignment is usually from one of the participants' real-life university classes.

Stapleton's (2010) study of a student writing a source-based essay complements McInnis' (2009) findings with insights into the intertextual processes of real-life tasks. Stapleton used a combination of process logs, a questionnaire, and interviews to investigate the time allocated by the student to different micro-level processes in the entire course of a real-life writing task. Copying from sources, either by pasting or by typing, emerged as a major activity that the student performed during her research. Though the study did not investigate how these copied excerpts were used in her writing process, the interview data suggests that she tended to copy-paste excerpts from sources and then combined, restructured, and reformulated these, for example by substituting words with synonyms. More changes were implemented in revision phases. The student used the source text for language support and borrowed individual words and expressions. She was inspired by its sentence structures and used these at other points in her essay to attain a sophisticated style. A *Turnitin* check of the essay showed acceptable source use with little textual overlap. The student pondered issues of source material integration at several points during writing, for example whether summarisation would be better than direct quotation and how a quotation could be embedded in a meaningful, reader-friendly way. Stapleton's study suggests that the processing of sources is a central component of the writing process in a naturalistic setting and integral to the generation of ideas and consecutive composition of the text.

L2 writers' preoccupation with issues of textual borrowing and their efforts to resolve these tensions were at the focus of Li's (2013) study combining process logs and interviews to investigate students writing a real-life term paper. The study showed that the three L2 students approached potential source texts with the goal of writing a paper already firmly in mind. Li interpreted their interactions with sources as attempts at addressing certain tensions that arose during writing, e.g. between their status as a student and the requirement of assuming an expert's stance. The students resolved the issue of selecting relevant and reliable sources by conducting purposeful keyword searches, labelling source texts with adjectives such as 'important', and relying on books rather than online resources because of their perceived authoritativeness. The challenge of writing from sources in their L2 English was met by recourse to online dictionaries and websites in the participants' native language Chinese. One student highlighted source text vocabulary she wanted to learn for later use.

Though this was not foregrounded by Li, her study is indicative of student behaviours in the writing process that potentially lead to unacceptable forms of textual borrowing. Two out of three students copy-pasted from source texts when taking notes, like the student in Stapleton's (2010) study. One student explained that she copied passages from the source into her document to gain a deeper understanding thereof. She regarded copying as a thinking process. Incidentally, she also assumed that as long as she provided a reference, copying extensive passages from sources was acceptable. Overall, she prioritised her goal of completing the paper over her goal of adhering to citation conventions. Another student explained that copy-pasting was her preferred strategy and she only sometimes summarised from sources. Her 'notes' consisted of copied passages that she collected in case she might need them. She added page numbers to sentences copied from books, but page ranges to sentences copied from journal articles, an approach that was also taken by the third student.

Li's study is revealing in terms of students' misconceptions about citing conventions and acceptable language re-use, which are reflected in the writing process. It thus bears further evidence that even if students aspire to follow the rules, questionable micro-level processes can be the cause of transgressive intertextuality. Having to balance the manifold demands of completing an assignment sometimes means that certain tasks are completed with less diligence than others, and students sometimes resort to copying when stressed (see also Abasi & Akbari 2008; Currie 1998; Petrić 2012).

An interest in the use of writing strategies with multiple source texts was at the core of a comprehensive study by Choi (2016) that integrated process questionnaire and process log data with keystroke logging, video recording, and retrospective interviews to examine graduate students' processes of writing a research paper over the course of a semester. Highlighting/note-taking, saving/organising sources, and printing/organising sources

were found to be major source interaction strategies. They were considered to be prerequisites to writing a research paper, but were not concurrent with composition processes. Two groups of students emerged: The first group organised, read, and highlighted sources in printed format, while the second group processed sources digitally. Choi identified paraphrasing as the main composition strategy. Avoidance of plagiarism was one motivating factor for paraphrasing in the writing process. Especially the more experienced students devoted a considerable amount of time to referencing, which was a task requirement, and attempted to follow the recommended citation style as accurately as possible. Synonym searches in dictionaries typically occurred during paraphrasing and very rarely during post-writing. The students used different combinations and varying numbers of strategies in each writing session, with some occurring only during certain stages of the process and others occurring throughout. Though micro-level strategies such as borrowing and rephrasing were not considered, Choi's study shows that long-term writing processes are multifaceted in nature and combine a variety of resources and strategies.

These studies reviewed here have shown that L2 writers engage with their sources consciously and continuously in the writing process. They have identified the creation of intertextuality as a central concern of students to which they devote much energy. Across studies, the preoccupation with acceptable source use is evident from students' processes and self-reports which have revealed their strategies of rephrasing source text ideas and associated insecurities. Studies of authentic writing tasks have been useful for identifying individual differences in the writing process. They have helped in creating a representation of source-based writing processes in authentic writing scenarios unaffected by generic prompts, unfamiliar environments, and time limits.

Process-oriented research sheds light onto the factors and assumptions guiding students' decisions when writing from sources. For example, the student in Stapleton's study deliberated what to do when she felt that she had not explained a direct quote sufficiently well, which she regarded as "irresponsible" (2010: 302). Her thoughts centred on the intended interpretation of the quote by a reader. Apparently aware of the rhetorical functions of direct quotes and the necessity to embed them in a meaningful way, she was dissatisfied with her own effort and eager to make changes to her text to rectify the issue. Such insights are confined to process-oriented approaches such as think-aloud protocols and process logs. Limiting the analysis to the product in this case would only have disclosed whether the student succeeded in meaningful integration of the direct quote or not. An observation of the process can bring to light the cognitive processes and decisions made by the students while writing which lead to those products. In this case, the process log revealed that the student's awareness of citing conventions influenced her evaluation of her own text and prompted her to make changes.

Researchers have emphasised the need for more studies of the processes of source-based L2 writing (Cumming et al. 2005; McCulloch 2013; Pecorari & Petrić 2014). The review of the literature shows that the majority of process studies has focused on EAP and language assessment contexts, with very few investigating the writing processes of disciplinary assignments. Tasks designed specifically for experiments or language tests are often quite different from those assigned in disciplinary courses. The processes of disciplinary writing thus merit further study. In addition, most studies – with the exception of McInnis (2009) – have considered intertextual strategies and micro-level processes of writing from sources only in passing. Some insights on how L2 writers incorporate sources into their writing in the process have emerged, but these are limited in scope due to the small-scale nature of case studies. It has been stressed that more research is needed to provide insights into how L2 writers incorporate source text information via paraphrasing, direct quotations, and summarisation as well as their copying practises (see Ruiz-Funes 1999: 56). In order to improve our understanding of source text use and integration in academic assignments, it is important to study intertextual strategies across larger learner groups, both in the products and processes of source-based writing. This is the goal of the present study. The following section explains the rationale for the study, which is followed by the presentation the research questions and an overview of the study design.

## 5 A mixed-methods study of intertextual strategies in L2 academic writing

The present study of intertextual strategies in L2 academic writing in English is situated in the discourse focusing on intertextuality from a positive, proficiency-oriented perspective. It is based on the assumption that many instances of apparent plagiarism are unintentional and non-deceptive in nature. The study adopts an output-based view of source-based writing (see Hirvela 2016: 128) by focusing on the writing processes of L2 student writers who are integrating source text material into disciplinary assignments in accordance with academic citation conventions.

The study is comprised of two complementary empirical investigations, namely a corpus and a process study, and uses three different types of data. The first is a corpus study of intertextual features of reading reports written by German learners of English from the *Corpus of Academic Learner English* (CALE; Callies & Zaytseva 2013). Major strategies, such as direct quotation, paraphrasing, and summarisation, are compared with regard to attribution and documentation as well as the use of reporting structures and evaluation. The study also investigates textual overlap between students' reading reports and the source text. The process study triangulates the corpus study by providing a complementary perspective on these phenomena of source use in disciplinary assignments. It looks at the processes of writing a reading report, especially the micro-level processes of interacting with the source text and creating intertextuality. This second study uses screen recordings created during the writing of reading reports in order to investigate the composition process in terms of how L2 writers integrate others' words and ideas. Stimulated recall and retrospective interviews conducted with the students provide additional insights into their choices when selecting, rewording, restructuring, and integrating source text material.

The rationale for this study is presented in section 5.1. On this basis, the research questions are derived in section 5.2. Reading reports, the text type chosen for this study, are characterised in section 5.3. Finally, section 5.4 is devoted to the conception of the study design.

### 5.1 Rationale for this study

Academic writing is a central part of academic literacy, and the teaching of source-based writing can only be improved if we know more about how L2 writers use their sources. Despite general recognition that intertextuality in academic texts is a highly complex and multifaceted phenomenon, much of the existing research into L2 writing has failed to view intertextuality in a multidimensional way, especially in corpus studies. Despite the re-conceptualisation of certain types of transgressive intertextuality as a developmental issue, research into intertextuality, especially into its illegitimate forms, is still rarely oriented towards solutions and practical implications, but rather focuses on the problems experienced by L2 writers (Wette 2010: 159). This choice of focus veils that “what is considered plagiarism is not necessarily an ethical issue but rather a multidimensional phenomenon, involving aspects of reading and writing processes, learning to write in an academic environment, disciplinary learning, and language learning” (Petrić 2015: 108).

As the literature review shows, many studies thus far have placed their focus on source use in general. While there are studies that have looked at individual strategies, e.g. paraphrasing or direct quotation, these have typically been investigated in isolation and not with regard to how they are combined with other strategies.

Instances of intertextuality have typically been identified via markers such as quotation marks and referencing, making it likely that other, less conspicuous intertextual links were missed. Citation has often been explored as a general phenomenon without consideration of finer-grained aspects of intertextuality such as lexical and structural similarity, attribution, and reporting verbs. None of the existing corpus-based studies have investigated the various intertextual phenomena in interaction in order to uncover L2 writers' broader approaches to creating intertextuality. McCulloch's (2012) study for instance, while being the only one to provide an in-depth textual analysis of direct quotation, paraphrasing, and summarisation in students' texts, did not take into consideration whether and how these were framed via references and reporting structures.

What research into plagiarism has also frequently overlooked is that the appropriation of source text language is a frequent and beneficial strategy in L2 writing. It is generally acknowledged that learning is promoted by imitation and repetition, but research has nevertheless focused on copying from the source text as a deficit, not as a learning strategy. Students generally do not intend to plagiarise, but they may still intentionally re-use passages from a source text. L2 writers have attested to using the source text as a lexical and phraseological resource in text-based interview studies (see e.g. Flowerdew & Li 2007; Stapleton 2010), and similar observations have been made in textual analyses (Keck 2006, 2014), which is why this facet of textual borrowing should be taken into account. Scholars have additionally called for research into borrowing from the source text in the writing process, especially of copying behaviours and of instances that are no longer traceable in the final product (Leijten et al. 2019: 577). In order to capture L2 student writers' source use adequately, research must also consider individual and contextual variables that have a bearing on their conceptions of intertextuality (Hirvela 2016: 130). That individual variables influence source use is an important finding of previous research (e.g. Keck 2014; Weigle & Parker 2012).

Text-based studies of paraphrasing in L2 writing have mostly been based on summaries of non-academic non-fiction texts written in second language contexts, e.g. in the U.S. Often, texts written in preparatory composition courses and EAP writing courses have been used as data in such studies (see e.g. Keck 2006; 2014). Much of the previous research has neglected to consider the writing contexts of non-Anglophone countries in which writing courses are not traditionally part of the curriculum and such assignments are not the norm. Students writing in English in non-English medium universities do not necessarily have previous academic writing experience, neither in their first language nor in English (Pecorari 2015). In addition, students studying in non-English-speaking countries such as Germany often have to write disciplinary assignments which require them to situate their texts in the broader context of the literature they have read for that particular class, e.g. term papers. Competent source text use is usually a requirement of such disciplinary writing tasks and commented on by lecturers in the feedback given for marked assignments (see Wette 2017), but is not always taught in these courses.<sup>52</sup> Pedagogy for source use in disciplinary writing must be based on research findings based on such text types, which is why it is important to study intertextuality in disciplinary contexts.

Aside from a lack of corpus-based studies of the interplay of intertextual phenomena, especially in disciplinary assignments, there is also a lack of process-oriented research. The existing process studies of integrated disciplinary assignments are relatively limited in scope, for example because of a focus on reading (e.g. McCulloch 2013). Most of the research has been conducted on summary and essay tasks produced in testing contexts. Thus far, L2 writing research has only provided few insights into the micro-level processes that lead to intertextual relationships in L2 writers' texts (but see Leijten et al. 2019 for an exception) and have mostly focused on the organisation of the intertextual writing process on the macro-level. Studies of L2 writing processes have not yet addressed what happens to source text material in the course of integrating and editing and which role source texts play during these operations. Furthermore, though it is widely attested that L2 writers refer to the source text for lexical and grammatical support, it remains unclear exactly how exactly they use it as a language resource during the writing process (see Plakans & Gebriel 2012).

52 Adequate source use is, as Wette (2017) notes, not the only requirement, however. Apart from source text use, student assessment in the assignments analysed for her study was also based on the way they addressed the issue, on their understanding of the topic at hand, on the accuracy of descriptions of their own research projects, and on the conciseness and logic of their argument.

Scholars in L2 writing research have called for studies of intertextual writing processes in order to shed light on students' strategies when writing from sources (e.g. Pecorari & Petrić 2014; Cumming et al. 2005), and especially for larger-scale studies of disciplinary writing tasks (Plakans 2008). Such studies are needed to gain insights into students' conceptualisations of academic writing in general and intertextuality in particular. Larger scale studies can lay the groundwork for inferences about L2 writers' academic writing abilities and can also provide validity evidence in assessment contexts (Gebriel & Plakans 2009; Plakans 2008). By combining perspectives of the processes and products of intertextual disciplinary writing in English as a foreign language, we will gain a more accurate impression of the relationship between writing process and product. This will allow us to determine the ways in which the product is a reflection of the L2 writers' writing process and how issues in the product can be explained in relation to the process. More in-depth research of integration processes will improve our understanding of the micro-level processes of source-based writing and how they lead to the intertextuality manifested in the final product.

A set of desiderata emerge from the gaps in research highlighted above. There is a need for focused research into intertextual strategies in source-based disciplinary writing and into legitimate and non-transgressive strategies of textual borrowing. Such research should mix methods in order to explore not only the products of writing from sources, but also the processes and the students' own perspectives on their source use. Corpus analyses of larger cohorts are useful to investigate trends identified in case studies in L2 writing research. They should be triangulated to provide additional perspectives of students' writing processes in the same task. In order to fully capture L2 student writers' use of sources, we need to focus on their intertextual strategies on the micro-level of the writing process – on strategies such as paraphrasing, summarisation, direct quotation and others – and on the interaction and complementarity between different intertextual phenomena in their texts.

This is necessary because intertextuality as a characteristic of a text is created via the interplay of manifold phenomena. These manifestations are neither created in isolation nor intended to function on their own. They rely on different degrees of source text interaction and integration. They thus should be investigated together in the context of the entire text, both in the process and the product. Capturing intertextuality comprehensively also presupposes close attention to individual differences in students' source use and in conceptions thereof. This entails taking the context of writing into account. Such an approach will allow educators to take an objective perspective on such differences and draw meaningful conclusions of practical relevance for their teaching practise. Finally, if we want to capture students' approaches to writing from sources in a neutral manner, we have to address intertextuality based on the assumption that transgressive forms are not necessarily plagiarism and that unconventional source use is a sign that further teaching is in order – and not penalties.

## 5.2 Research questions

In order to address the desiderata outlined above, the present study investigates a type of source-based assignment typical of disciplinary courses in linguistics. The multiple dimensions and interactions of intertextual strategies are at the core of the present study. In order to capture intertextuality in student writing comprehensively and gauge the ways in which L2 writers employ source text material, this study explores both the product and the process. It aims to explore the micro-level processes of source-based disciplinary assignments and the students' perspectives thereon in order to capture how intertextuality is created and conceptualised. It takes a perspective of apparent plagiarism as being mostly unintentional and adopts a focus on strategic, legitimate re-use of source text words and the students' reasons for textual borrowing.

In this dissertation, a corpus study of reading reports written in linguistics courses is combined with a process study of L2 writers composing reading reports based on the same source text in an experimental setting.

The study combines quantitative and qualitative analyses in a mixed-methods design (see Figure 5). In the corpus study, intertextual strategies and markers are identified and analysed so as to provide insights into the various dimensions and interrelations of intertextuality in the product. The corpus study informs the process study, in which writing processes are recorded on-screen so that the creation of intertextual phenomena and source text interactions can be observed. This process data is complemented by retrospective interviews with stimulated recall using the screen recordings as a stimulus. It is expected that this triangulated approach will offer complementary and converging insights into students' intertextual strategies and their motivations in source-based writing.

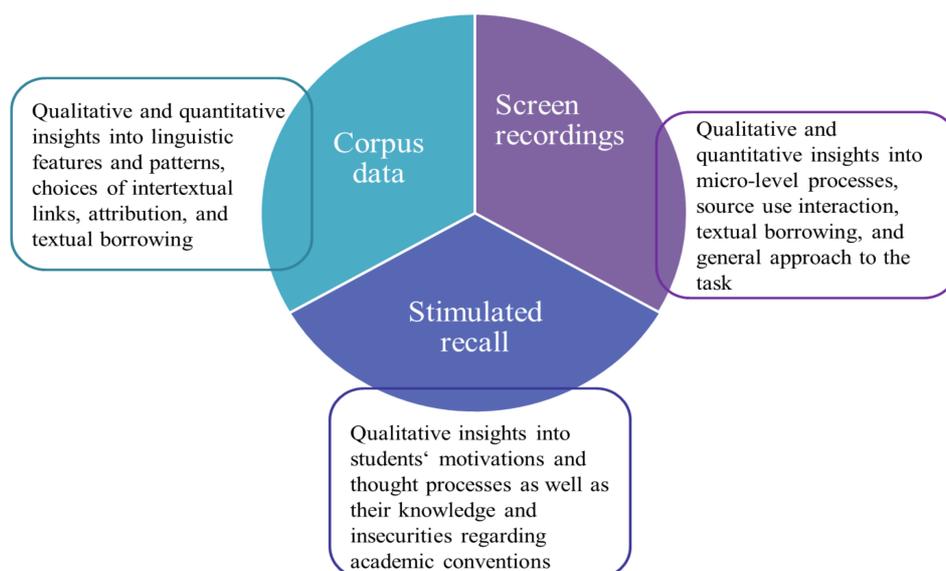


Figure 5: Triangulation of research methods.

To my knowledge, no previous study has carried out a corpus study of source-based academic texts in order to inform a consecutive process study of the same task and text type. This mixed-methods study addresses the calls for research taking into consideration both the process and the product of source-based writing tasks in order to provide informed recommendations for teachers. The study accounts for the multifaceted nature of intertextuality by investigating a range of phenomena such as paraphrasing, direct quotation, summarisation, referencing, attribution, textual borrowing, and strategic re-use of source text vocabulary. Manual identification and quantitative and qualitative analyses allow for well-founded and reliable interpretations that accurately represent L2 writers' actual writing practice. Finally, the combination of product and process data with stimulated recall reflects the need to take into account the declarative, procedural, and conditional knowledge of L2 writers.

The following research questions are at the core of the present study:

1. Which intertextual strategies do L2 student writers employ and how are they combined, both in the writing process and in the finished reading report?
2. In what ways do L2 student writers document, attribute, and report source text information and material in the processes and products of writing reading reports?
3. In what ways do L2 student writers copy and re-use source text material, both in the writing process and in the finished reading report? Is textual borrowing employed strategically, and if so, how and why?

The goal of the study is, first, to obtain a detailed picture of the dimensions of source use in L2 academic writing, and second, to uncover students' common and individual approaches to creating intertextuality. The

study helps answer the questions of how students interact with the source text in the process of writing and how information is integrated and acknowledged on the micro-level. It sheds light on the fine-grained processes of re-using source text material and ideas and on the complex interactions of intertextual markers in L2 student writing. The study also adds to our understanding of students' individual conceptualisations – and perhaps misconceptions – of source use in academic writing and how these influence the writing process and final product. It thus offers new perspectives on cases of apparent plagiarism. The overarching practical purpose of this study is to gain detailed insights into students' approaches in process and product that provide educators with a solid basis for goal-oriented academic writing pedagogy. Reading reports, a kind of summary of a published research article, were deemed to be a useful text type for the envisaged study for the reasons presented below.

### 5.3 The text type 'reading report'

A reading report is a source-based text pedagogical type that is typical of undergraduate courses. It is a part of the curriculum in various disciplines at universities in countries where English is the language of instruction and may be based on one or more texts. It is regularly assigned in English linguistics courses at the University of Bremen, where this study was conducted. In the task description used in these courses, reading reports are defined as "an academic text type that contains the main points, hypotheses, methodology and general line of argumentation of a research paper" (reading report instructions; see Appendix 11). The students are asked not to copy from the source text and to use their own words instead. In essence, a reading report is a summary of an individual academic research article, though with a focus only on the study presented. It is intended to roughly follow an IMRD structure (Introduction, Method, Results, Discussion; see Swales 1990). Reading reports are typically assigned throughout the semester as part of the course work, especially in first-year courses. They are based on research articles from the course reading list. The reading reports analysed for the corpus study were submitted as ungraded disciplinary assignments in linguistics courses by first-year students enrolled in a B.A. programme in either English-Speaking Cultures or Language Sciences in the summer semesters of 2016 and 2017.

A reading report is a type of integrated writing task. Integrated writing assignments are increasingly common in academic writing assessment. They are considered more authentic of academic contexts than writing tasks which are not source-based (Gebriel & Plakans 2009; Plakans 2008; Weigle 2004) because they trigger processes such as the extraction of key points from one or more sources and the integration of textual material (Plakans & Gebriel 2013: 226). Most academic writing is source-based, and source texts provide a common knowledge base for the students to help them generate ideas (Weigle 2004). Plakans and Gebriel (2012: 18-19) have argued that integrated writing tasks more authentically represent language use by L2 writers than independent tasks.

Reading reports replicate aspects of writing literature reviews and share some of their intertextual features. For example, they allow students to practise various kinds of intertextual relationships, e.g. paraphrases and direct quotes, and they may contain both integral and non-integral citations. Reading reports thus constitute a pedagogical genre (Johns & Swales 2002: 21). The purpose of assigning this pedagogical genre is to prepare students for writing a literature review in a research paper or a B.A. thesis<sup>53</sup>. Pedagogical task types of a low complexity allow students to practise synthesis in a clearly defined context. In reading reports, the focus is on practising intertextual strategies and extracting information so as to provide a summary of an individual empirical study.

---

53 Such an approach of using a succession of different genres of increasing complexity in writing education is widely practised, though Dovey (2010) has suggested that it does not foster the transfer of knowledge from one genre to the next.

Reading reports were selected as the basis for the present study for several reasons. First, reading reports are a text type which does not permit plagiarism in the sense of intentional deception. The source text is provided by the lecturer, which means that the students know that their source use can easily be traced (see also Shi 2004). Writers of reading reports thus cannot illicitly and deceptively copy from the text, though patchwriting is of course a possibility (see Howard 1995 for similar observations on summaries). Reading reports are thus suited to an investigation of intertextuality beyond apparent plagiarism. Second, reading reports allow learners to focus on intertextuality in their writing. A major issue in intertextual writing is that novice writers are faced with a variety of different challenges, only one of which is appropriate intertextuality:

It is easy to imagine inexperienced writers taking their eyes off the source-use ball when there is so much else to learn about the writing process, as well as the content of the courses in which writing is done. (Pecorari 2015: 95)

While using formal and correct language is, amongst others, a requirement of reading reports, building an argument and constructing knowledge from sources are not. This means that the writers can expend their energy on appropriating the source text without changing its meaning. This makes reading reports attractive for a study of intertextual strategies. Third, reading reports are a useful diagnostic tool for problematic ways of using sources because the students' texts are based on an individual source. While inappropriate copying might be overlooked in a text based on several sources or an unknown source text (Pecorari 2015: 95), it can be unambiguously identified in reading reports. In teaching contexts, such issues can then be addressed through formative feedback. For the purpose of the present study, using a genre that is based on a single source text known to and selected by the researcher has the major advantage of allowing for a direct comparison of different students' strategies of summarising the same text. The source text can be exploited as the basis for the identification and categorisation of direct quotes and paraphrases. It is also easy to identify instances of textual and structural overlap. While previous studies have often used short non-fictional source texts (e.g. Keck 2006, 2014), reading reports require students to read and summarise full-length research articles and thus provide insights into writing from academic source texts in a disciplinary context.

The majority of studies of intertextuality in L2 writing has focused on summaries, making them the only source-based writing task on which there is extensive research (see Spivey 1990). While this has often been cited as a disadvantage because too little is known about other text types, it also means that because of the similarity between reading reports and summaries, the results of this study can be compared to previous research. It is important to note, however, that unlike the argumentative essays used in some previous studies (e.g. Boshier 1998; Chan 2017; Gebril & Plakans 2009; McInnis 2009; Plakans & Gebril 2012; Ruiz-Funes 1999), reading reports do not have an evaluative element. The students are not required to argue a position or to voice their own opinion. This is important to bear in mind when comparing the findings of the present study to those of previous research using argumentative text types.

It has previously been emphasised that term papers based on students' own research projects represent a more common pedagogical text type. A composition process over several weeks, in different locations, and with various resources is arguably more representative of expert academic writing than a prompt-based one-time writing task (Stapleton 2010). However, it is unfortunately not always possible to access all of the sources used by the students for their term papers. References may accidentally or intentionally be missing from the text and/or bibliography. Furthermore, paraphrase identification is more problematic in term papers because it is uncommon to provide references for every paraphrase (see T. A. Hyland 2009). Finally, while they certainly resemble published research articles to some extent, term papers are written for a very different purpose. Students are expected to showcase their knowledge of a particular field, their understanding of research methodologies, and their ability to conduct a research project and critically interpret its results. Such research papers are not intended to be published, although this is of course a possible later step. This means that students are not necessarily required to identify a gap in research, which is a common move in published research articles (Lim 2012). Their literature reviews are likely to be structured in a different way to published research papers,

which also means that their intertextual strategies are likely to be different. As a result, comparisons of students' and experts' papers must be conducted with caution. Since students' intertextual strategies are not compared to those of experts in this study, reading reports were deemed a more suitable type of data for the reasons outlined above.

A clear advantage of reading reports, though they are a one-time pedagogical writing task and do not resemble expert genres, is that they require students to transform source text material and can provide insights into how students accomplish this task. For this reason, they are a useful genre for examining writing processes (see also Hirvela 2016: 130). They lend themselves to a study of the micro-level processes of source-based writing, which are more difficult to identify in other text types. The fact that they are relatively short, based on only one specific source text, and can be completed within less than three hours makes them especially suitable for process studies of intertextuality, as Stapleton (2010: 297) has suggested:

[T]he type of tasks which lend themselves well to employing think-aloud and stimulated recall protocols as well as keystroke logs are those which set a prompt with learners completing the task within a limited period of time while recording takes place.

The conception of the study design is presented in the following section.

## 5.4 Conception of the study design

The methodological design of the present study was conceived of in several phases. Initially, the goal of the project was to investigate intertextuality using corpora of academic writing. Several preliminary studies were conducted using reading reports (Wiemeyer 2015a, 2015b, 2017a, 2017b) and research papers (Wiemeyer 2016a, 2016b, 2019). Because previous studies had not considered the interplay between manifestations of intertextuality, e.g. of paraphrases, direct quotes, and summaries with attribution and referencing, these preliminary investigations produced valuable new insights into L2 writers' source use in academic assignments. However, they proved to be limited in scope because some of the observed intertextual practices could not be explained without insights into the writing process. For this reason, it was decided to triangulate a corpus study with a process study that included a stimulated recall procedure. The observations made in a corpus study of reading reports based on a different source text (Wiemeyer 2017a) informed the design of this process study.<sup>54</sup> Three source texts were piloted for the process study (see section 7.2). A research article on the word-forming element *eco-* (Kettemann, König & Marko 2003) was consequently selected as the source text for both studies. Data compilation for the corpus and the process study took place in 2017. The data were analysed sequentially, with the process data analysis following the corpus analysis.

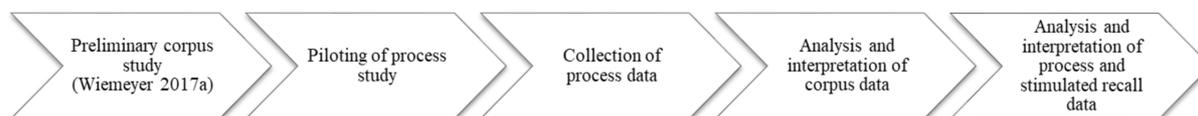


Figure 6: Study design process.

Figure 6 shows the study design process, which can be summarised as follows: Drawing on an earlier corpus study (Wiemeyer 2017a), the process study was piloted in May and conducted in June and July 2017. First-year students wrote reading reports based on the chosen source text under experimental conditions and

<sup>54</sup> A preliminary corpus study of reading reports from the CALE was conducted in early 2017 (Wiemeyer 2017a). Its results were to inform the design of the process study, which was to be conducted in a linguistics class offered to second-year B.A. students. However, because of low enrolment, the process study instead had to be conducted with first-year students enrolled in an introductory linguistics module. Thus, for reasons of comparability, a different sub-corpus of reading reports was used. Ultimately, the results of the first corpus study (Wiemeyer 2017a) were not used in this dissertation, but they informed the collection of process data.

were subsequently interviewed using a stimulated recall procedure. Corpus data consisting of reading reports based on the same source text that had been written in non-experimental conditions were analysed in the second half of 2017.<sup>55</sup> This was followed by the annotation and analysis of the process data in the first half of 2018. The present study combines a corpus and a process perspective of intertextuality in L2 writers' reading reports based on the same source text. The corpus study is presented in section 6, followed by the process study in section 7.

---

<sup>55</sup> Selected results of the corpus study were presented at the Learner Corpus Research Conference 2017 in Bolzano, Italy, in October 2017 (Wiemeyer 2017c) and at ICAME 39 in Tampere, Finland, in May 2018 (Wiemeyer 2018).

## 6 Corpus study

### 6.1 The corpus

The corpus study is based on 25 reading reports of a research paper by Kettemann et al. (2003) that were taken from the *Corpus of Academic Learner English* (CALE). The reading reports were written by German first-year students as an assignment in an introductory course on research methods in linguistics. The course was the second part of a compulsory two-course module for first-year students and complemented an obligatory Introduction to Linguistics the students had taken in the preceding semester. The students were given one week to complete the assignment. All students enrolled in the course had to write a reading report, but submission to the corpus was voluntary. The reading reports were collected in two separate years; 19 reading reports were written by the 2016 cohort, and the remaining six were from the 2017 cohort. The majority of students professed to having consulted dictionaries and other language resources, but none of the texts had been proofread by a native speaker.

The student writers were native speakers of German. At the time of data collection, they were between 19 and 37 years old, with an average age of 23.6 years. English was their first L2, which they had been learning for nine to sixteen years, most of which had taken place at secondary school. In their self-assessment, the majority of students judged their proficiency of English to be good or very good. The students did not undergo standardised proficiency assessments. In order to be admitted to the B.A. programme English-Speaking Cultures, they had to have a certain grade in the subject English in their university entrance qualification, in their case the German Abitur.<sup>56</sup> Alternatively, they could provide the result of a standardised international test, e.g. the Test of English as a Foreign Language (TOEFL). It is usually assumed that students have reached at least B2 level according to the Common European Framework of Reference for Languages (CEFR) when they complete their Abitur.<sup>57</sup> B2 can thus be assumed to be the level minimally reached by the study participants, though their actual proficiency is likely to be more advanced. Most students stated that they used English almost every day. All spoke at least one and up to three other foreign languages, usually French and/or Spanish. They all indicated English to be their strongest L2, which was also the preferred foreign language of three quarters of the students. Less than half had spent time abroad in an English-speaking country.

---

56 “11 points in the subject English as advanced foreign language option (applicable only for higher education entrance qualifications gained in Germany)” (Website B.A. English-Speaking Cultures, University of Bremen; <http://www.fb10.uni-bremen.de/anglistik/ba2/zulassungsvoraussetzungen.aspx>, last accessed in March 2020)

57 See for example Informationsschreiben Nr. 135/2014 for regulations regarding CEFR levels in English by the Free Hanseatic City of Bremen (<https://tinyurl.com/info135-2014-bremen>). It states that reaching B2 level in English is the curricular goal. The CEFR level is listed in the Bremen Abitur certificate.

Table 2: Corpus overview.

Text code	Year	Word tokens	Word types	Type/token ratio (per 100 words)	Sentences	Words per sentence
RR1.G.HB.009	2016	508	241	47.44	24	21.17
RR1.G.HB.010	2016	605	236	39.01	26	23.27
RR1.G.HB.011	2016	777	261	33.59	38	20.45
RR1.G.HB.012	2016	532	255	47.93	18	29.56
RR1.G.HB.013	2016	593	228	38.45	32	18.53
RR1.G.HB.014	2016	443	193	43.57	13	34.08
RR1.G.HB.015	2016	564	216	38.30	19	29.68
RR1.G.HB.016	2016	560	218	38.93	20	28.00
RR1.G.HB.017	2016	620	251	40.48	26	23.85
RR1.G.HB.018	2016	580	245	42.24	24	24.17
RR1.G.HB.019	2016	515	195	37.86	22	23.41
RR1.G.HB.020	2016	599	247	41.24	31	19.32
RR1.G.HB.021	2016	470	196	41.70	24	19.58
RR1.G.HB.022	2016	442	186	42.08	15	29.47
RR1.G.HB.023	2016	392	193	49.23	11	35.64
RR1.G.HB.024	2016	581	233	40.10	21	27.67
RR1.G.HB.025	2016	600	245	40.83	23	26.09
RR1.G.HB.026	2016	745	296	39.73	32	23.28
RR1.G.HB.027	2016	590	240	40.68	22	26.82
RR1.G.HB.082	2017	595	225	37.82	26	22.88
RR1.G.HB.085	2017	491	198	40.33	21	23.38
RR1.G.HB.090	2017	494	196	39.68	24	20.58
RR1.G.HB.092	2017	513	193	37.62	23	22.30
RR1.G.HB.189	2017	347	171	49.28	12	28.92
RR1.G.HB.190	2017	599	251	41.90	29	20.66
<b>TOTAL</b>		<b>13,755</b>	<b>5,609</b>	<b>41.20</b>	<b>576</b>	<b>24.91</b>

In total, the corpus contains 13,755 words in 576 sentences. The average word count per text is 550 (SD = 93.68) with a range from 347 to 777. The average number of sentences per text is 23 (SD = 6.35) with a range from 11 to 38; the average length of a sentence is 25 words (SD = 4.45). For each reading report, the word token and type frequency were determined using the *Compleat Lexical Tutor's VocabProfile* (<http://www.lextutor.ca/vp/eng/>). Titles and references were not included in the word count. An overview of the corpus is provided in Table 2.

## 6.2 Data annotation and analysis

Each reading report was divided into sentences and the number of sentences and words per sentence were counted. The body of each text was then manually annotated for intertextual links using the sentence as the basic unit.<sup>58</sup> Each sentence was scrutinised in order to determine in which way it was based on source text material and was manually coded using one of ten categories: direct quote (Q), paraphrase (P), paraphrase with embedded direct quote (PQ), summary (S), summary paraphrase (SP), meta-level observation (M), exact copy (E), exact copy embedded in a paraphrase (PE), and new content (NC). The codes PQ and PE were introduced to adequately categorise sentences containing multiple strategies so that such co-occurrences would be apparent after coding.<sup>59</sup> The code new content (NC) was used for sentences which do not convey information from the source text, i.e. which contain content added by the student. The data were coded manually because unlike automated coding this allows the researcher to record the degree of textual modification. Manual coding promotes the recognition of all manifestations of intertextuality, from copied passages to excerpts which have been thoroughly rephrased on the morphological, lexical, and structural level (Weigle & Parker 2012: 120). Each category is described below; an overview of codes appears in Table 3.

Direct quotes (Q) are defined as verbatim reproductions from the source text which are marked by the writer via conventional signals such as quotation marks, author names, third-person pronouns, and page numbers (see K. Hyland 2004). A passage was annotated as a direct quote if the student had used quotation marks to identify it as a direct quote (following Borg 2000) or, in the case of independent quotes and block quotes, if it was indented without quotation marks. Following Petrić (2012), certain passages were not considered direct quotes even if they were placed within quotation marks, namely text in scare quotes (non-intertextual uses of inverted commas), publication titles, and examples, e.g. from corpus data. Only those sentences which consisted of only a direct quote or a direct quote and a reporting structure were coded as Q. Sentences in which a direct quote was embedded in a paraphrase were categorised as PQs. Each direct quote was then categorised using the scheme for direct quotes proposed in Wiemeyer (2019). This hierarchical scheme makes a broad distinction between free-standing quotes and embedded quotes. Free-standing quotes are set off from the text and are subdivided into block quotes, which are indented, and independent quotes, which stand alone before or after a chapter and are not integrated into the text. Embedded quotes are integrated into the text and may be either syntactically integrated fragments, not syntactically integrated fragments (e.g. following a colon), or complete sentences.

Sentences that had been copied word-for-word from the source text but not marked as a quote were coded as exact copies (E) using Keck's (2006, 2014) terminology. This was applied even if the sentence contained morphological changes to the source text material, e.g. to tense markers on the verb or plural markers on nouns, and if individual grammatical words were replaced or added. This decision was made once it became evident that exact copies of entire sentences without alterations were exceedingly rare, but that students did in fact copy sentences from the source text to which they had made minimal grammatical changes. By amending Keck's (2006) definition, it was possible to identify both exact copies in the narrow sense, which are identical to the source text (see example 6.1)<sup>60</sup>, and exact copies in a slightly broader sense to which the students had only made morphological or grammatical changes (example 6.2, change underlined).

58 Earlier studies have often disregarded instances of intertextuality which were not followed by a reference (e.g. Okamura 2008a; 2008b). As this approach is unlikely to capture all instances of intertextuality, every sentence was considered to be a possible intertextual link here.

59 In Keck's (2010) study, a similar approach was taken in that paraphrases were annotated for the occurrence of multiple grammatical strategies.

60 All examples from the corpus are reproduced as in the original, i.e. typos and other errors were not corrected. Where changes were made, e.g. underlining to highlight certain passages, this is indicated in the text.

## (6.1 ) Exact copy of entire clause or sentence without changes (E):

*Source text excerpt:*

Originally, *eco* occurred in neo-classical compounding, where new lexemes are composed of morphemes that have a hybrid status between affixes and free morphemes. (Kettemann, König & Marko 2003: 138)

*Student paraphrase:*

Originally, *eco* occurred in neo-classical compounding. (RR1.G.HB.010)

## (6.2 ) Exact copy of entire clause or sentence with morphological or grammatical changes (E):

*Source text excerpt:*

The selection is based on the criterion of institutionalization. (Kettemann, König & Marko 2003: 137)

*Student paraphrase:*

This selection is based on the criterion of institutionalization. (RR1.G.HB.024)

There were also instances of copied clauses and sentences from the source text that were embedded in paraphrases. In order to account for these constructions, a further category of paraphrases with embedded exact copies (PE) was introduced. This category was only used for paraphrases containing copied excerpts that constitute and are identical with entire sentences or clauses in the source text; see example (6.3).

## (6.3 ) Exact copy of entire clause or sentence embedded in a paraphrase (PE):

*Source text excerpt:*

The OED is the largest dictionary of the English language. Publication started in 1884 and took until 1928. (Kettemann, König & Marko 2003: 138)

*Student paraphrase:*

The OED is the largest dictionary of the English language and was first published in 1884. (RR1.G.HB.011)

Paraphrases (P) are defined as passages that are based upon and linguistically linked to one or more excerpts of the source text and convey the same proposition, but have been altered lexically and/or syntactically by the writer (see examples in Table 3). Following Keck (2006; 2014), any passage reproducing an idea from the source text that contained at least one lexical alteration, e.g. a synonym<sup>61</sup>, was coded as a paraphrase. Depending on the origin of the proposition(s) in a paraphrase, it was noted whether it was based on one sentence, two to three adjacent sentences, or constituted a gist statement based on several sentences or paragraphs.

An earlier study of reading reports (Wiemeyer 2017b) had shown that Shi's (2004) and Keck's (2006; 2014) categories for paraphrase types, which rely on textual overlap in the form of identical strings of words as one criterion for paraphrase identification, were too broad for this particular text type. The writers in Wiemeyer's (2017b) study selected content for paraphrase from across the source text and did not necessarily re-use its vocabulary. It was often impossible to identify one particular sentence in the source text on which a paraphrase was based. Some sentences contained linguistic links and propositions from two or more source text sentences.

61 The terms 'synonym' and 'synonym substitution' are used in this dissertation following Keck (2006, 2010, 2014) because the words used to replace source text words are commonly referred to as synonyms in L2 writing research, even if they are not synonyms in the narrow sense (e.g. in Chan 2017; Keck 2006; Howard 1995; Leijten et al. 2019; McInnis 2009). The term 'synonym substitution' is also found elsewhere in the literature (e.g. Bailey & Withers 2018; Pecorari 2003). It must be noted that from a semantic point of view, many of the terms regarded here as 'synonyms' would, strictly speaking, be more aptly referred to as either near-synonyms or semantically related terms.

In other words, these students used different, perhaps more elaborate paraphrasing strategies than the L2 writers in Shi's and Keck's studies. For this reason, these existing paraphrase taxonomies were modified and expanded in order to accommodate the strategies observed in this corpus.

Unlike in previous studies, textual overlap with the source text was not considered to be a defining feature of a paraphrase. Textual borrowing, i.e. the re-use of words from the source text, was consequently treated as a phenomenon separate from paraphrasing. In the narrow definition adopted for the purpose of this study, only the direct transfer of textual strings of three or more words is considered textual borrowing. It was annotated independently on a different layer so as to also capture such re-use of vocabulary if it occurred in other intertextual strategies besides paraphrases (see description below). As a result, the taxonomy used in this study allows for sentences to be classified as paraphrases that do not contain any textual material from the source text, but do convey the same meaning.

Summaries were initially defined as sentences which summarise the source text without paraphrasing individual paragraphs. Signals such as the authors' names, the title of the source text, and research nouns (e.g. *study, paper, research*) were relied on for the identification of summaries. During the initial coding process, some sentences were found to be problematic because the students paraphrased the authors' summarising sentences in order to summarise the source text in their reading reports. The sentence in (6.4), which considerably overlaps with the sentence it paraphrases (see underlined passages), serves as an example. While the student's sentence has a summarising function, it is clearly based on an individual source text sentence and thus akin to a paraphrase. To avoid overlap between the categories summary and paraphrase, a third category was introduced for summary paraphrases. A summary (S) was defined as a sentence in the student's text that provides a general summary of the text (not of individual paragraphs) without going into detail with regard to method, participants, and findings. If a sentence had a summarising function but contained elements of paraphrase of an individual source text sentence, it was coded as a summary paraphrase (SP) instead. All other rephrased sentences, including gist statements which summarised longer stretches of the text, but not the entire text, were treated as paraphrases. Generalisations were not included in the annotation scheme because reading reports do not require reference to other research literature in the field and were unlikely to contain generalisations from multiple sources.

(6.4) *Source text excerpt:*

In this paper we aim to show the relevance for lexical studies of comparing dictionary and corpus data [...]. (Kettemann et al. 2003: 135)

*Student summary paraphrase:*

The present paper aims at demonstrating the relevance of comparing dictionary and corpus data when carrying out lexical studies. (RR1.G.HB.023)

During coding, it became apparent that not all sentences in the reading reports were based on source text sentences, but were still intertextual in nature. In such sentences, the students made the act of writing a reading report and summarising a source text explicit by commenting on the authors' approach or on the structure of the source text on the meta-level, see the examples in (6.5). Such instances were coded as 'meta-level observation' (M).

(6.5) The research is divided into six sections, giving both introduction and conclusion, two kinds of analysis, the difference between the OED (Oxford English Dictionary) and the BNC (British National Corpus) and for the most part the semantics of "eco". (RR1.G.HB.009)

The author introduces the two media which are going to be investigated. (RR1.G.HB.011)

Finally, there were also sentences in the corpus containing information that was not in the source text. For example, students sometimes gave their opinions on the source text or volunteered explanations based on their previous knowledge of the subject (see examples in Table 3). These were coded as new content (NC) (see Keck 2014 for a similar category of ‘invented’ information). An overview of coding categories with definitions and examples is provided in Table 3.

Table 3: Coding categories.

Category	Code	Definition	Examples from the corpus
Paraphrase	P	A sentence that is based upon parts of the source text (one or more clearly identifiable sentences or paragraphs) and conveys the same proposition, but has been altered lexically and/or syntactically by the writer	Paraphrase without copied strings of 3+ words: The authors <u>suggest that this could</u> be because of the <u>short live</u> of products or words used in advertisement. (RR1.G.HB.090)  Paraphrase with copied strings of 3+ words: In contrast to the OED, <u>the BNC proved</u> to be a much higher <u>source of eco-words</u> whose background and development can be studied. (RR1.G.HB.009)
Direct quote	Q	A sentence constituting a verbatim reproduction from the source text which stands alone or is embedded in a reporting structure and is either <ul style="list-style-type: none"> <li>marked with quotation marks (single or double) by the writer, or</li> <li>a block quote (longer, indented quote in the text) or an independent quote (set off from text), even if not marked with quotation marks</li> </ul> <p>Exceptions: The following instances are not coded as Q even if marked with quotation marks:</p> <ul style="list-style-type: none"> <li>direct quotes embedded in paraphrase (→ coded as PQ)</li> <li>text in scare quotes (non-intertextual use of inverted commas)</li> <li>publication titles</li> <li>quotations from questionnaires or interviews, e.g. test items</li> <li>examples, e.g. from corpus data</li> </ul>	From their analysis, the Kettemann et al. conclude “that the corpus and the dictionary do not represent the same version of reality” (Kettemann et al. 2003: 146). (RR1.G.HB.010)
Direct quote embedded in paraphrase	PQ	A sentence in which a direct quote is embedded in a paraphrase	The BNC as „the largest commercially-available collection of empirical data in British English“ (136) provides all words starting with a specific prefix and also allows to quantify. (RR1.G.HB.090)
Summary	S	A sentence that provides a general summary of the entire source text without going into detail with regard to method, participants, findings, etc. and is <u>not</u> based on an individual source text sentence.	The text is a scientific examination and comparison of the differences between a corpus and a dictionary, which is based on the analysis of one specific morpheme, using the British National Corpus (BNC) and the Oxford English Dictionary (OED) as data sources. (RR1.G.HB.019)
Summary paraphrase	SP	A sentence that provides a general summary of the entire source text without going into detail with regard to method, participants, findings, etc. and contains elements of paraphrase of an individual source text sentence	In the article „The BNC and the OED. Examining the usefulness of two different types of data in an analysis of the morpheme eco“, by Bernhard Kettemann, Martina König and Georg Marko, the authors demonstrate that the comparison of dictionary and corpus data is a relevant aspect in lexical studies. (RR1.G.HB.024)

Category	Code	Definition	Examples from the corpus
New content	NC	A sentence which exclusively conveys information that is not in the source text, including student opinions Such sentences may make reference to the source text, but do not convey a proposition found in the source text.	In contrast, the formation of a corpus is based on explicit design criteria. (RR1.G.HB.019) Again I would assume that the higher amount of “eco-words” in the BNC results from the green-wave that hits the globe since the late eighties and that green-activism and green-awareness get more important day after day which is supported by the media. (RR1.G.HB.026)
Meta-level observation	M	A sentence in which the student makes the act of writing a reading report and summarising a source text explicit or describes the structure of the source text Note: Paraphrases of meta-level observations from the source text are not coded as M.	The author introduces the two media which are going to be investigated. (RR1.G.HB.011) The research is divided into six sections, giving both introduction and conclusion, two kinds of analysis, the difference between the OED (Oxford English Dictionary) and the BNC (British National Corpus) and for the most part the semantics of “eco”. (RR1.G.HB.009)
Exact Copy	E	A complete clause or sentence that is copied word-for-word from the source text, but is not marked as a quote This category includes copied clauses and sentences to which only morphological/grammatical changes have been made, e.g. to tense markers on the verb or plural markers on nouns.	<u>Originally, eco occurred in neo-classical compounding.</u> (RR1.G.HB.010)
Exact Copy embedded in Paraphrase	PE	A sentence that contains a complete clause or sentence copied word-for-word from the source text that is not marked as a quote and that is combined with paraphrased passages This category includes copied clauses and sentences embedded in paraphrases to which only morphological changes have been made, e.g. to tense markers on the verb or plural markers on nouns.	Overall, <u>the corpus is likely to contain a high proportion of words that are fashionable for a time and then fade out of use</u> whereas the dictionary shows more a selection on long-term used words. (RR1.G.HB.190)

Independent of their type, the intertextual links were then coded for additional characteristics (see overview in Table 4). First, they were annotated for the SECTION from which the information stemmed. For all intertextual links except NC, DOCUMENTATION, ATTRIBUTION, USE OF REPORTING STRUCTURES, and EVALUATION were recorded. DOCUMENTATION referred to whether or not the writer provided a reference for their source. ‘Documented’ intertextual links were subdivided into integral and non-integral citations based on Swales’ (1990) well-established distinction. A third category was used for sentences in which students combined integral and non-integral citation of the same author(s), which had previously been identified as an issue in L2 writing (Wiemeyer 2019).

Because previous studies have shown that referencing is not the only way in which L2 writers acknowledge their sources (Wiemeyer 2016a, 2017b, 2019), each intertextual manifestation was also coded for ATTRIBUTION. Attribution refers to the acknowledgement of the origin or originator of cited material. For instance, a source text proposition may be attributed to an author, to the type of publication, to the informants of the cited study, or to a particular section of a source text. Attribution may also be achieved indirectly via a passive structure. The target of attribution was recorded using the categories AUTHOR(S) (DIRECT/INDIRECT), TYPE OF PUBLICATION, CITED AUTHOR / SECONDARY CITATION, TYPE OF RESEARCH, SECTION OF TEXT, and OTHER.

Furthermore, each intertextual link was coded for USE OF REPORTING STRUCTURE, TYPE OF REPORTING STRUCTURE, and whether the student had encoded EVALUATION. A reporting structure is understood to be a phrase that introduces a piece of reported information from the source text (see Thomas & Hawes 1994 for a similar definition). Following the procedure for the identification of reporting structures used by G. Thompson and Ye (1991), canonical reporting structures containing the authors’ names or an equivalent expression (e.g. third-person pronouns and semantically related nouns such as *researchers*, *text*, and *study*) as the subject; passive structures containing reporting verbs (e.g. *are examined in detail*); and impersonal constructions (e.g. *it has been argued that*) were considered to be reporting structures. Because students are likely to use non-canonical and creative types of reporting structures, the presence of a reporting verb was not relied on as a defining criterion. Non-canonical structures were coded as reporting structures if the writer had in some way marked information as reported, e.g. via a complex preposition (*according to*), phrases containing reporting nouns (e.g. *another finding was*) or subordinate clauses (e.g. *as X claims*) (see also Verheijen 2015). Creative structures marking information as reported were also recorded in order to fully capture the students’ practices.

A further coding category of INTERPRETATION was used if the writer stated something that was not explicit in the text. This category was used if the interpretation was considered legitimate with respect to the source text. MISINTERPRETATION OR MISREPRESENTATION was noted if the intertextual link reproduced the source text incorrectly or contained an inappropriate interpretation, as accurate representation is a key component of effective source use (Plakans & Gebriel 2013: 228). Because of the subjective nature of this category, it was not quantified, however.

Finally, textual overlap was identified in each sentence using *Wcopyfind* 4.1.5 (Bloomfield 2016). The goal was to identify the number and length of copied strings per reading report in order to gain an insight into the extent to which the L2 writers lexically relied on the source text and into possible patchwriting. For this purpose, textual overlap was defined as strings of three or more words of length that were copied from the source text and were identical or identical except for morphological changes. Three or more words were deemed as a useful cut-off point based on Cumming et al.’s (2005) findings that this length neither over- nor under-identified strings copied from the source text<sup>62</sup>. In addition, this definition made it possible to distinguish between compounds and other strings which could have been rephrased. As noted above, it was assumed that textual overlap could be found in all intertextual links, not just in paraphrases. Consequently, each sentence was analysed in terms of whether there was textual overlap in the shape of strings of 3+ words. The copied string(s) and the number of

62 Some categorisation schemes for paraphrases used a narrower definition of two or more words (e.g. Shi 2004). This meant that student paraphrases containing compounds from the source text had to be coded as near paraphrases even if they were otherwise thoroughly rephrased. As this results in a rather inaccurate and fuzzy categorisation (see discussion in Wiemeyer 2017b), this cut-off point was considered too low for the purpose of the present study.

words copied from the source text were then recorded for each sentence to allow for a qualitative and quantitative analysis of textual borrowing practices.

Table 4: Additional coding categories applied for each type of intertextual link.

Annotation category	Type of intertextual link								
	P	Q	PQ	S	SP	M	NC	E	PE
Section of origin	✓	✓	✓	✓	✓	✓		✓	✓
Documentation (integral, non-integral, combined integral+non-integral)	✓	✓	✓	✓	✓	✓		✓	✓
Attribution, type of attribution (direct/indirect), target of attribution	✓	✓	✓	✓	✓	✓		✓	✓
Reporting structure (yes/no), type of reporting structure	✓	✓	✓	✓	✓	✓		✓	✓
Evaluation, type of evaluation	✓	✓	✓	✓	✓	✓		✓	✓
Type of source text excerpt (one sentence, adjacent sentences, gist statement / paraphrase based on several sentences from across the text)	✓	✓	✓					✓	✓
Copied strings of 3+ words (yes/no), total no. of words, copied string(s)	✓	✓	✓	✓	✓	✓		✓	✓
Interpretation by the student (yes/no) (if interpretation is supported by and implicit in the source text)	✓	✓	✓	✓	✓	✓		✓	✓
Misinterpretation / misrepresentation (yes/no) (if the source text is incorrectly or inappropriately interpreted)	✓	✓	✓	✓	✓	✓		✓	✓
Added information (yes/no)	✓	✓	✓	✓	✓	✓	✓	✓	✓
No. of direct quotes, quoted string(s), category of quote, type of embedded quote		✓	✓						

To ensure the reliability of the annotation scheme, six reading reports were selected at random from the corpus, separated into sentences, and independently annotated for intertextual links, documentation, attribution, and reporting structures by the researcher and by a student assistant. Each annotator coded the first text individually. Following a discussion of issues and disagreements in the annotation, the definitions of paraphrases with copied strings and exact copies were amended so that a clear distinction was possible. Finally, the definitions for S and SP were specified so that these codes were only applied to sentences that provided a summary of the entire source text, but not of individual sections.

Each annotator then coded the remaining five reading reports. An inter-annotator agreement of 82% was reached for intertextual links. Disagreements mostly regarded meta-level observations with elements of paraphrase. As a result, the definition for meta-level observations was refined to cases where it was discernible that it was the student who commented on the study or the structure of the text, even if there was some lexical overlap with the source text. If a student paraphrased a sentence that contained a meta-level observation by the authors, however, this was coded as a paraphrase. In addition, an early category called ‘added information’ caused disagreement. There were indeed sentences in the corpus which consisted exclusively of new content added by the student. However, there were also intertextual links such as paraphrases which contained information that was not found in the source text. The category of ‘added information’ was thus replaced by

‘new content’ – for sentences that contained exclusively new information and were not based on source-text information. An additional layer of coding for ‘added information’ was introduced for all intertextual links.

The coders reached 100% agreement for documentation and 83% for attribution. Differences in the annotation of attribution occurred because the second coder did not take into account attribution to the text and indirect attribution, i.e. via passive structures. An inter-annotator agreement of 82% was obtained for reporting structures, for which disagreements were often the result of different understandings of attribution, e.g. regarding reporting structures in the passive. The categories in which disagreements arose were discussed and their definitions subsequently refined. The remaining reading reports were annotated independently by the researcher.

### 6.3 Results and discussion

Among the eight coding categories for intertextual links, paraphrasing and direct quotation emerged as major strategies in the reading reports (see overview of intertextual links in Table 5). Paraphrasing is by far the most frequent intertextual strategy in the corpus, as 79% of all sentences (N=454) were coded as paraphrases. 8% of sentences (45) contain direct quotes – either embedded in reporting structures or in paraphrases. Meta-level observations are the third most common type of intertextual link in the data (37; 6%), pointing to the students’ desire to make the act of summarising explicit and guiding the reader through the text. Summaries and summary paraphrases are even less frequent and together constitute less than 5% of all intertextual links. Exact copies (E and PE) and sentences with new content (NC) are very rare, which attests to the students’ general understanding of the foundations of source-based writing.

Table 5: Overview of quantitative findings.

Manifestation type	Code	N	%
Paraphrase	P	454	78.7
Direct quote embedded in paraphrase	PQ	40	6.9
Meta-level observation	M	37	6.4
Summary paraphrase	SP	15	2.6
Summary	S	9	1.6
New content	NC	8	1.4
Direct quote	Q	6	1.0
Exact copy embedded in paraphrase	PE	6	1.0
Exact copy	E	2	0.4
<b>TOTAL</b>		<b>577</b>	<b>100.0</b>

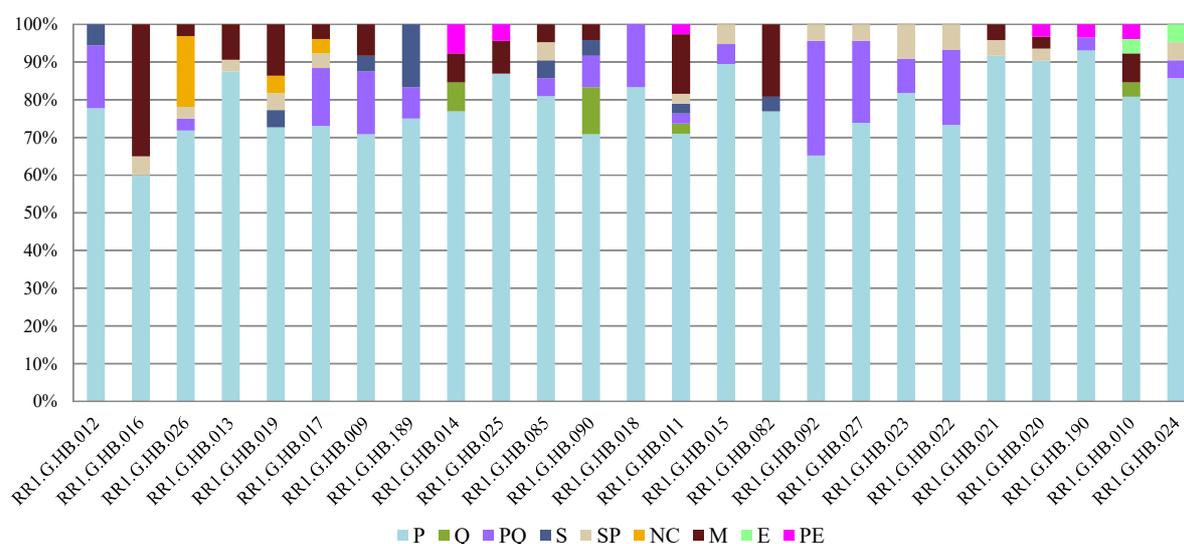


Figure 7: Inter-learner variability in the use of intertextual links.

Though paraphrasing was the most common type of intertextual link across texts, individual preferences for specific combinations of intertextual links emerged from the data. Figure 7 shows that while some reading reports consist almost exclusively of paraphrases, other texts contain direct quotes, while yet others comprise a mix of different strategies. Meta-level observations and exact copies, for example, are absent from the majority of texts. That one combination of intertextual strategies is not necessarily more effective than another shows the potential for individual approaches in intertextual writing, which was exploited by these students. In the following, the intertextual strategies of these L2 writers will be presented and discussed in detail in the light of the research questions. Quantitative findings will be integrated with insights from the qualitative analysis.

### 6.3.1 Strategies of source text material selection and structuring

The students selected the information for their reading reports from across the source text. As visualised in Figure 8, the majority of intertextual links were traced back to the third and fifth sections, i.e. the data and the results and discussion section. While the percentages generally correspond roughly to the length of the respective section, the conclusion stands out as the third most commonly cited section. Apparently, the condensed and summarising nature of the text in this section was considered especially useful as the basis for intertextual links by these L2 writers. Of those intertextual links which combined information from several sections, the majority was based on either the abstract or on the conclusion and another section. Certain intertextual links were preferred in connection with certain sections. For example, summaries were usually based on the abstract or the introduction of the source text.

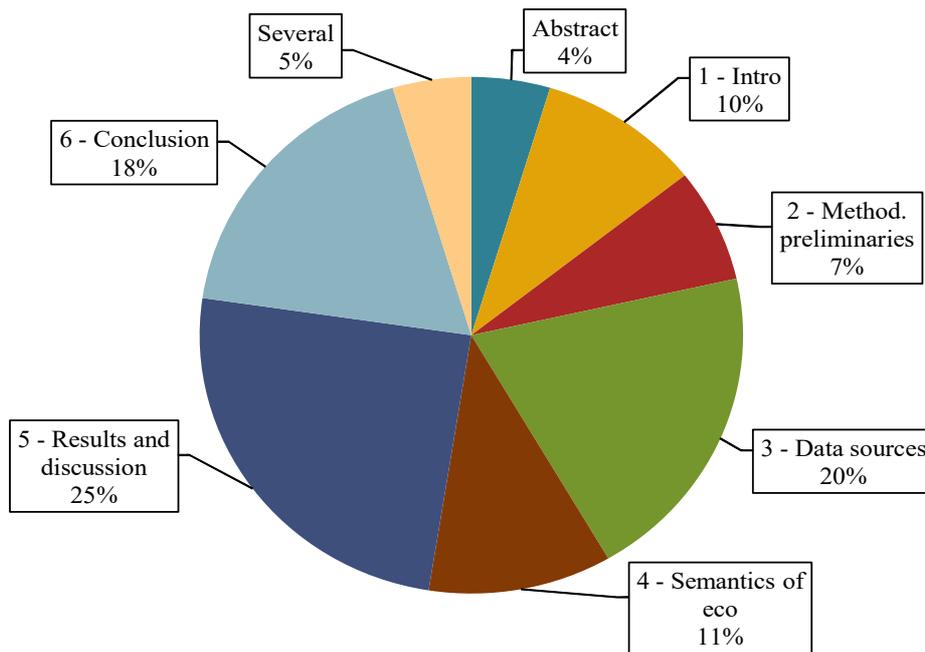


Figure 8: Use of material from different sections of the source text.

Typically, the students selected information from the source text chronologically and derived the structure of their reading reports from that of the source text. The reading report RR1.G.HB.190 is a case in point. It consists of 29 sentences, the first of which is a paraphrase that includes a direct quote from the abstract (PQ). All remaining sentences are paraphrases. The first paraphrase is based on the introduction of the source text. It is followed by five which are based on section 3, then three which are based on information from section 4, then ten which paraphrase material from section 5. The chronological order is only broken in the last part of the reading report, in which the student combines paraphrases from sections 5 and 6 for their conclusion. Apart from this, the student presented the information in the order in which it was presented in the source text, using information from all sections but the second.

A chronological order is to be expected and to some extent determined by the task description for the reading report, which asked students to summarise the study and explicitly lists data, methodology, and findings as points to be included (see reading report instructions in Appendix 11). Hence, it is only natural for students to follow the source text's IMRD structure when writing a reading report about a research article (see also Sherrard's 1986 study of L1 summary writing). Some students, however, chose a non-chronological approach in which they selected excerpts from across the source text, but did not use them in the original order. This is illustrated by the excerpt from reading report RR1.G.HB.022 in Table 6.

Table 6: Example of non-chronological structuring of information in the reading report.

Student text (RR1.G.HB.022)	Section of origin
Second, the authors list the differences between corpus and dictionary data.	Section 2, p. 137
The authors figured out that the BNC proved a richer source than the dictionary data which I think is interesting to know, because it shows that often it is more useful to not always look for words in the most common source – the dictionary.	Section 6, p. 145
After a short glimpse of the semantics of the morpheme <i>eco</i> , the morpheme is analysed by regarding three different meanings, e.g. the “environmentally friendly” meaning or <i>eco</i> in the sense of “pertaining to the environment”.	Section 4/5, pp. 138, 141ff.
The “environmentally friendly” meaning rarely appears in the OED whereas <i>eco</i> -words in the sense of “pertaining to the environment” and “pertaining to the ecological movement” were found to be more often.	Section 5, pp. 141ff.
The researchers explained that if once a word appears in a dictionary, it hardly is going to be removed in a short time.	Section 6, p. 145
The authors touch on the open question why some words are contained in dictionaries and some are not.	Section 6, p. 145

Such diversions from the original text’s chronology may be indicative of issues of structuring or an incomplete understanding of the characteristics of the text type, as suggested by the writer giving their opinion in the second sentence. However, since the reading reports were not graded and the quality thus not assessed, such an interpretation can only be tentative.

An interesting finding is that a small number of students used information from the abstract to summarise the conclusion; see example (6.6).

(6.6) *Source text excerpt from abstract:*

[...] on the assumption that corpora contain lexemes as raw data whereas dictionaries contain a specific selection of words chosen by lexicographers on the basis of their perceived institutionalisation in the language. (Kettemann et al. 2003: 135)

*Student paraphrase from summary:*

The initial hypotheses is proved right, as we could clearly see that a corpus contains more the raw material and the dictionary has been proved to be a selection of a certain word, not the whole word collection, which are at the same time the most important findings of the research. (RR1.G.HB.014)

As both sections summarise the content of the publication in some way, this appears to be caused by the desire to condense the information from the source text in an efficient way. Basing paraphrases on sentences of a summarising nature was a strategy that was observed across reading reports. Because it relies on the summarisation skills of the authors, it is a much more economical strategy than re-reading the source text sections and paraphrasing from sections that are less condensed. The students are seemingly aware that the authors have already summarised their findings and do not feel the need to repeat this work, possibly because it could result in misrepresentation.

Some students appear to look for source text material that helps make their text accessible to the reader. The paper by Kettemann et al. (2003) begins with a general statement about linguists’ over-reliance on corpus data. Many students decided to begin their texts with a paraphrase of said statement (see example 6.7), perhaps because they were orienting towards the source text structure.

(6.7) *Source text excerpt:*

Corpus linguists have a tendency to regard the corpus, if not as a panacea to all linguistic problems, then at least as the most valuable kind of data in the study of language, sometimes forgetting that other, different forms of data are equally valuable. (Kettemann et al. 2003: 135)

*Student paraphrases:*

According to these three, linguists rely heavily on corpora even though a broader set of tools might improve the quality and depth of their work. (RR1.G.HB.012)

According to Kettemann, König & Marko, corpus linguists are often forgetting that there are other valuable kinds of data than a corpus. (RR1.G.HB.015)

It is possible that this tendency results from the students' more intimate familiarity with other genres, such as newspaper articles, in which such introductory sentences are commonly used. By paraphrasing this type of clause, they are accommodating features of two distinct genres without crossing the line.

There are some texts in the corpus which show quite different intertextual strategies than the majority. For example, in RR1.G.HB.018 it is not apparent that the text is based on an empirical research article. The writer uses elaborate paraphrasing, but the intertextual nature of the report is not made explicit. The use of the passive combined with an absence of referencing and relative paucity of attribution and reporting structures contributes to this impression (see example 6.8). As a result, it may be difficult for the reader to deduce who conducted the cited study, which is certainly not the intention of writing a reading report.

- (6.8) By combining these research methods in a contrastive way, phenomena can be observed that would have remained unrecognized if only one of the tools had been accessed. This is illustrated by the example of the morpheme *eco*. Both recourses offer a variety of lexemes formed with *eco* that can be categorized into six different meanings of the lexeme *eco*. (RR1.G.HB.018)

A second reading report that exhibits unusual intertextuality given the task description is RR1.G.HB.016. The text is a kind of meta-level summary that does not go into detail regarding the content of the research article. The authors are regularly attributed, but often this occurs in meta-level observations: The student describes the surface structure of the research article instead of summarising the results and arguments presented by the authors; see example (6.9).

- (6.9) They end their descriptions of the word *eco* with six different meanings of the term *eco*, to set up a basis from where their analysis starts. They list up all findings regarding three out of his six different meanings of the word, and compare and explain what is shown on the tables. They show similarities between the OED and the BNC, they name the differences, in meaning and number, and point out special or unexpected findings. (RR1.G.HB.016)

The writers of these two texts seem to have had a specific interpretation of the task description that led to atypical ways of creating intertextuality. It is possible that these approaches, which are not intended by the task design, would have an impact on these students' grades. Another approach that diverges from the task description is found in RR1.G.HB.026. It consists of the student presenting their own interpretations and conclusions that diverge from the authors' (see 6.10).

- (6.10) In conclusion we might say that *eco* often appears in a context that is based biological and environmental but in todays times is ofen used in a context that inculdes the so often and so fashionable green-awareness thus rather in spoken language. (RR1.G.HB.026)

Finally, the kind of essay style addressing reader directly found in RR1.G.HB.026 appears to be an unusual choice for the genre of reading report that impacts on the manifestation of intertextuality; see example (6.11).

- (6.11 ) Using not only one of the files named above carries a lot of advantages as well see later. [...] You'll recognize that both, if both the information is added properly will offer complementary results.  
(RR1.G.HB.026)

Such examples point to the difficulties experienced by L2 writers when completing source-based writing tasks. It is clear that the students' understanding of the genre of reading report has an impact on the way in which they create and frame intertextuality. The students may have misunderstood the task or simply did not receive sufficient instructions on creating intertextuality in a reading report. The intertextual practices described in the preceding paragraph may also be the result of students' problems in selecting and presenting source text information due to difficulties in understanding the source text. It is quite likely that such issues can be prevented by providing students with detailed explanations and illustrative examples as to the genre expectations and giving them the opportunity to ask questions both on the task and on the content of the source text.

The findings presented in this section lead to the conclusion that the structure of the source text served as orientation for the structure of the students' reading reports. Though some students combined information from different sections in the same sentence, the most common approach was to follow the order in which information was presented in the source text. All sections of the source text were consulted, though preferences emerged with regard to summarising sentences and sections such as the conclusion. Finally, atypical approaches include superficial summaries describing the structure of the source text, addressing the reader directly, and presenting one's own interpretation, all of which are objectionable in a reading report.

The next section is devoted to the students' use of the three major intertextual strategies, i.e. paraphrasing, direct quotation, and summarisation, as well as the ways in which the students marked information as taken from the source text.

## **6.3.2 Intertextual strategies**

Each of the intertextual strategies identified in the data will be discussed in turn in this section, beginning with paraphrasing.

### **6.3.2.1 Paraphrasing**

Paraphrases constitute the vast majority of intertextual links (454 of 577; 78.7%). They are mostly undocumented. That paraphrases are so frequent was to be expected given the task description (see Appendix 11). As the students were asked not copy from the text and use their own words, paraphrasing was an implicit requirement.

Previous studies of paraphrasing have mostly focused on summaries of short non-fiction source texts (e.g. Keck 2006, 2014). As a result, the strategies for selection of source text material for paraphrasing by L2 writers working with longer source texts have not yet been explored. Certain similarities between the strategies used in summary writing of short texts and those used here emerge from the analysis. The analysis shows that the majority of paraphrases are based on individual sentences (255; 56%). This preference is noteworthy because it shows that L2 writers rely on individual sentences for paraphrasing even when summarising research articles of more than ten pages of length. The students selected sentences from across the source text, but also used

headings as the basis for paraphrases. There are several individual source text sentences that were paraphrased by several learners.

That different students used the same sentences for paraphrase is certainly no coincidence. They seemed to strategically select certain sentences for paraphrase. Amongst these are topic sentences and sentences with a summarising function whose condensed nature apparently allows students to capture the gist of longer source text passages. A similar finding was made by Sherrard (1986) in a study of undergraduate L1 writers. The students in her study tended to base their paraphrases mostly on single sentences from the source text. Some of these source text sentences were paraphrased by several of the writers, though they were not typically taken from the beginning or the end of a paragraph and thus not necessarily topic sentences. In a consecutive judging task, the expert judges confirmed that this particular set of sentences was coherent and meaningful. This supported Sherrard's interpretation that "in longer texts, certain sentences are universally selected for inclusion in summary because they contribute crucially to the meaning of the text" (1986: 338). The similarity between the L2 writers in this study and the L1 writers in Sherrard's suggests that this strategy of selecting individual, meaningful sentences is common to novice writers in general.

In the present study, there are also instances of paraphrasing of adjacent sentences (68; 15%). In this type, information from two, sometimes three adjacent sentences is integrated into one paraphrase. Evidently, the motivation for writing paraphrases based on adjacent sentences is to condense information, a necessary strategy for summary writing. In Sherrard's (1986) study, as in this study, the paraphrasing of adjacent sentences was the second most frequent strategy after the paraphrasing of individual sentences. The fact that the L1 writers infrequently selected sentences from across paragraphs to write more complex paraphrases was interpreted by Sherrard in reference to the expository text type. She surmised that because the meaning of expository texts may depend on their sequential ordering and that the students selected adjacent sentences for paraphrasing for this reason. It is likely that the L2 writers in this study also selected this strategy because of the technical nature of the source text and the sequential presentation of information.

The third type of paraphrase is a kind of gist statement of a longer part of the source text. 29% (131) of paraphrases fall into this category. There are two types of gist statements in the data. The first is a paraphrase that synthesises information from individual sentences from across a section. In example (6.12), the information stems from two sentences from Kettemann et al.'s (2003) article, one from the abstract and one from the third section on the data sources. Textual overlap (underlined in the example) clearly links the student's paraphrase to these sentences.

(6.12 ) Words are included on the basis of the perceived institutionalization of lexicographers, thus whether or not they conclude that a certain lexeme or its specific meaning has a commonly accepted inter-subjective status. (RR1.G.HB.027)

For the other type of gist statement, it was difficult to identify individual sentences from which the information was taken. This type of paraphrase is based on information from across a section without a clear link to individual sentences and little to no lexical overlap, see (6.13).

(6.13 ) The authors not only analysed which words are shared, but they also examined what kind of words overlap in both resources. (RR1.G.HB.092)

Words with this meaning are often used in advertisement to imply that a product is environment friendly and „clean“. (RR1.G.HB.090)

Gist statements usually presuppose inferences and thus abstraction from the source text. In example (6.14), the student summarises a section on the origin and range of meanings of the morpheme *eco-* and makes an inference about an increase in semantic nuances that is not explicit in the source text. Such inferential thinking is a sign of an expert-like writing strategy (Yamada 2003).

- (6.14 ) A notable observation is the slowly growing range of acceptable meanings and uses of the word.  
(RR1.G.HB.012)

However, gist statements are difficult to achieve and require careful selection and rephrasing strategies. A ‘meta-approach’ taken by some students in gist statements highlights the importance of paying attention to both content and linguistic operations. The following paraphrases, for example, do not report details of the content of the source text despite being based on its wording and are vague as a consequence, see (6.15).

- (6.15 ) Next, two main types of analysis were shown and then explained them as well as transferring them on the subject matter explained why the contrastive analysis was used, not the additive analysis.

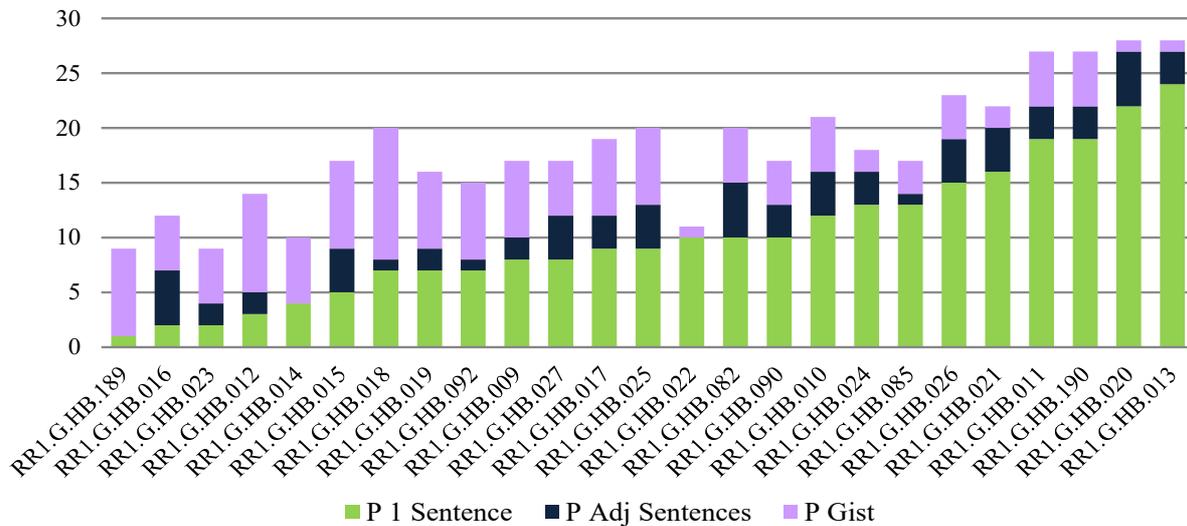
Next, the semantic of the word *eco* is been discussed and a brief history of the word is given as well as the range of the word is shown and said that the research was concentrated on words with the prefixation *eco* and not every word with *eco* in it. (RR1.G.HB.014)

Gist statements, while much less likely to occur in summaries of short texts, are a useful strategy in summarising long texts. They allow the writer to synthesise information from sections of the source text without going into detail. Because they require a thorough understanding and overview of the source text coupled with more complex rephrasing skills than necessary for paraphrases of individual sentences, gist statements are telling indicators of academic writing ability. Nevertheless, it is important to bear in mind that gist statements fulfil a different purpose than paraphrases of individual sentences, so that the presence of gist statements does not necessarily correspond to an advanced writing proficiency (compare discussion in Sherrard 1986).

These selection strategies for paraphrases are combined in different ways in the reading reports; see Figure 9. One approach is to chronologically select individual sentences from across the source text and construct the reading report from this type of paraphrase (e.g. RR1.G.HB.072). Another is to rely predominantly on gist statements for the majority of the reading report (e.g. RR1.G.HB.012), although gist statements always occur alongside paraphrases of the other two types. There are also reading reports that combine all three types in roughly equal proportions (e.g. RR1.G.HB.009, RR1.G.HB.010).

These findings suggest that paraphrasing strategies are to a certain extent influenced by the length of the source text and the expected length of the summary. Gist statements are a useful means of summarising longer texts, but impractical for a summary of a short source text. Paraphrases of individual sentences, however, seem to be popular among L2 writers in summaries of both long and short texts, irrespective of whether they are non-fictional or academic in nature.

a) Absolute numbers of paraphrases by selection strategy.



b) Relative frequencies of paraphrases by selection strategy.

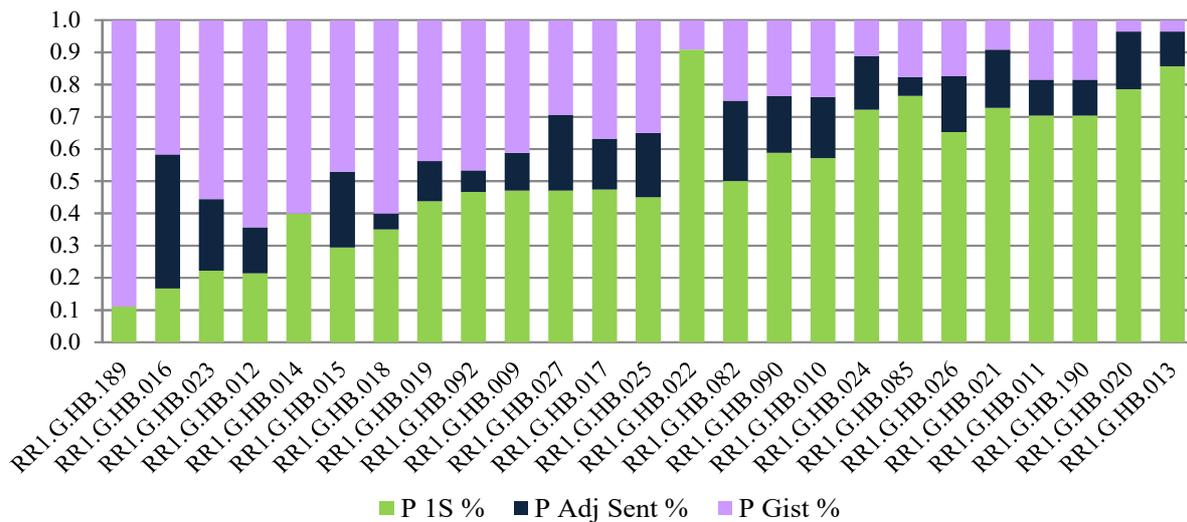


Figure 9: Inter-learner variability in selection strategies for paraphrases.

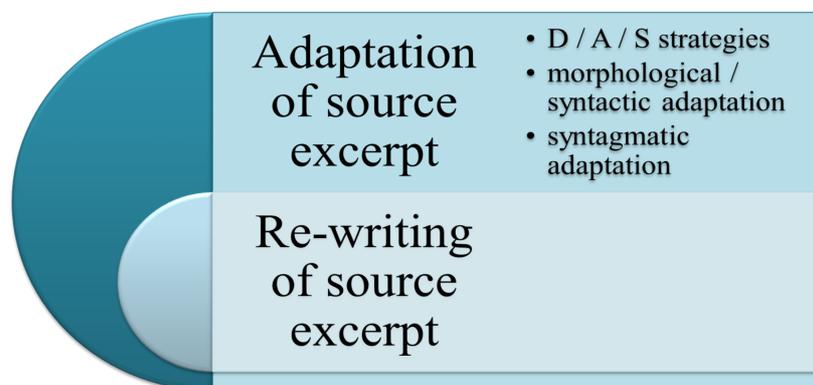


Figure 10: Approaches to paraphrasing.

The data allow for a broad distinction of two approaches to paraphrasing taken by the students (Figure 10). The first approach, ADAPTATION OF SOURCE EXCERPTS, is to use source text excerpts as templates to which the students apply changes in order to minimise textual overlap. This approach comprises three different strategies, as discussed below. The second approach, RE-WRITING OF SOURCE EXCERPTS, is to re-write source text sentences so that little to no lexical and grammatical similarity remained.

**ADAPTATION OF SOURCE EXCERPTS.** Three different strategies of source excerpt adaptation emerge. In many paraphrases, it is possible to identify strings that were borrowed from the source text and consequently underwent deletions (D) as in (6.16), additions of words (A) as in (6.17), or substitution (S), e.g. with an alternative expression or synonym, as in (6.18).

(6.16 ) *Source text excerpt:*

The BNC constitutes the largest commercially-available collection of empirical data on British English as it is actually used by different speakers, in different situations, for different purposes. (Kettemann et al. 2003: 136)

*Student paraphrase with deletions:*

First, they take a look at the BNC which contains the largest collection of data on British English. (RR1.G.HB.020)

(6.17 ) *Source text excerpt:*

[...] on the assumption that corpora contain lexemes as raw data whereas dictionaries contain a specific selection of words [...] (Kettemann et al. 2003: 135)

*Student paraphrase with addition:*

The idea of complementarity is coupled with two assumptions which say that corpora comprise raw linguistic data whereas dictionaries incorporate precise choice of words due to their institutionalization (= widely accepted lexemes). (RR1.G.HB.009)

(6.18 ) *Source text excerpt:*

The first is that [...] the BNC proved a much richer source of eco-words than the OED [...]. (Kettemann et al. 2003: 145)

*Student paraphrase with substitution:*

In contrast to the OED, the BNC proved to be a much higher source of eco-words whose background and development can be studied. (RR1.G.HB.009)

These D/A/S strategies were first described in Keck's (2010) study of summaries of 1,000-word non-fiction texts. In the present study, they mostly occur in paraphrases of individual and of adjacent sentences, rarely in gist statements. The changes differ in their extent: some students only change individual words, resulting in considerable lexical overlap; in other students' paraphrases there are only traces of the original lexis and grammatical structure. Due to the nature of the source text, paraphrases that consist entirely of copied material with changes only to individual words are very rare. The students much more commonly combined passages from one or several sentences and deleted, replaced or added words within those strings. They linked them using passages that conveyed propositions of the source text but had been more thoroughly re-written. In their attempts to rephrase the source text and avoid copying, some students overshot the target. For example, common linguistic terminology was sometimes replaced (see the replacement of *meaning* with *understanding* in example 6.19).

(6.19 ) *Source text excerpt:*

These reflections lead us to posit six meanings of *eco*: [...]. (Kettemann et al 2003: 139)

*Student paraphrase:*

They record six different understandings for the word and categorize these into three groups, “BNC only”, “OED only” and “BNC & OED”. (RR1.G.HB.012)

Clearly, insecurities regarding what is acceptable to copy and what is not prevail at this level of academic writing proficiency. L2 writers need to be made aware of the distinction between technical terminology that is acceptable to re-use and general academic expressions that can and should be rephrased.

The second strategy of source excerpt adaptation is to adapt excerpts morphologically and/or syntactically. Most of these changes are implemented on the paradigmatic level. Usually, they are confined to the verb phrase and manifest as morphological adaptations (6.20), removal or insertion of hedges (6.20 and 6.21), and changes in voice (6.22). Changing the word class of source text words was another common strategy (6.23 and 6.24)

(6.20 ) *Source text excerpt:*

Additive analysis involves simply combining the data from the two sources so that what is left is a superordinate pool of data to be analysed. (Kettemann et al. 2003: 136)

*Student paraphrase:*

Additive analysis combines the data of two sources to end up with a superordinate pool of data which can be analyzed. (RR1.G.HB.020)

(6.21 ) *Source text excerpt:*

The corpus is likely to contain a high proportion of words that are fashionable for a time and then fade out of use. (Kettemann et al. 2003: 146)

*Student paraphrase:*

In contrast the corpus contains words which are more fashionable and fade away after they are out of trend. (RR1.G.HB.021)

(6.22 ) *Source text excerpt:*

For the purpose of illustration, we analyse the semantics of the morpheme *eco* [...] (Kettemann et al. 2003: 135)

*Student paraphrase:*

For illustration, the semantics of the morpheme *eco* was analyzed. (RR1.G.HB.014)

(6.23 ) *Source text excerpt:*

The careful annotation and concordancing possibilities enable us [...] (Kettemann et al. 2003: 136)

*Student paraphrase:*

[...] the British National Corpus (BNC), which constitutes the largest collection of empirical data which is carefully annotated [...] (RR1.G.HB.022)

(6.24 ) *Source text excerpt:*

In this paper we aim to show the relevance for lexical studies of comparing dictionary and corpus data [...]. (Kettemann et al. 2003: 135)

*Student paraphrase:*

In the article [...] the authors demonstrate that the comparison of dictionary and corpus data is a relevant aspect in lexical studies. (RR1.G.HB.024)

The third adaptation strategy consists of changes on the syntagmatic level. Such syntactic strategies of avoiding copying include presenting unaltered copied strings in an order that is different from that of the source text, see (6.25).

(6.25 ) *Source text excerpt:*

Originally, eco occurred in neo-classical compounding, where new lexemes are composed of morphemes that have a hybrid status between affixes and free morphemes. These morphemes are called combining forms and are usually of Greek or Latin origin (cf. Bauer 1983:213 ff.). (Kettemann et al. 2003: 138)

*Student paraphrase:*

Here, new lexemes are composed of morphemes of usually Greek or Latin origin which have a hybrid status between affixes and free morphemes. (RR1.G.HB.010)

Some of the students made syntactic changes to individual copied phrases to minimise the overlap between theirs and the source text, for example within coordinated noun phrases. The student in example (6.26) was intent on avoiding copying, but this did not go so far as to result in the replacement of the coordinated noun phrase with a semantically equivalent string. Instead, they changed the order of the constituents within the noun phrase. A similar case is presented in example (6.27). Lexically, the first clause is largely identical with the source text. However, by changing the order of constituents, the student avoided a copied string of the excessive length of nine words by dividing this string into several shorter ones. There are also instances in the data in which the student accidentally left out or misspelt words from the source text when copying, which leads to a lower degree of lexical overlap but faulty syntax (see e.g. RR1.G.HB.024 for several such errors).

(6.26 ) *Source text excerpt:*

However, in a situation where the two data sources are different, as in a dictionary and a corpus, the picture may be distorted by an additive approach. (Kettemann et al. 2003: 136)

*Student paraphrase:*

However when it comes to a corpus and a dictionary, the results could be falsified out of different features from the data sources. (RR1.G.HB.013)

(6.27 ) *Source text excerpt:*

The BNC makes no distinction between institutionalized and less widely accepted lexemes. (Kettemann et al. 2003: 137)

*Student paraphrase:*

In corpora like the BNC there is no distinction between less widely accepted and institutionalized lexemes and only a careful sampling can lead to conclusions about the status a word. (RR1.G.HB.024)

The minimal alterations discussed above certainly impede the automated identification of copied strings, but they do not result in effective paraphrases. The examples suggest that the students' attempts at rephrasing

were grounded in a desire to minimise the overlap between theirs and the source text to avoid suspicions of plagiarism, not in a desire to use their own words to create an elegant paraphrase. This points to an insufficient understanding of the principles of paraphrasing, coupled perhaps with a lack of alternative formulations. Issues of textual borrowing are discussed further in section 6.3.4.

As the examples above illustrate, none of these strategies occurs in isolation. Paraphrases based on individual sentences often contain strings from the source text which have undergone deletions, synonym substitution, and additions. Furthermore, the students make grammatical adjustments and add their own material around those modified strings. Different students have different preferences, but since many paraphrases are largely written in the students' own words and a considerable amount of lexical overlap is confined to compounds and fixed expressions (see section 6.3.4), it would be difficult to assign one strategy per paraphrase. Unlike in the short summaries investigated by Keck (2010), a clear-cut distinction of paraphrases based on grammatical strategies does not seem feasible for this proficiency level and task.

**RE-WRITING OF SOURCE EXCERPTS.** The second approach to paraphrasing is to radically re-write source text sentences, which is achieved in elaborate ways, for example by inverting the focus of the source text sentence; see (6.28).

(6.28 ) *Source text excerpt:*

Obviously, a dictionary does not contain token frequencies, there is a restricted number of co-texts (in examples), and some information about meaning is explicitly given rather than having to be inferred, as in the corpus. (Kettemann et al. 2003: 137)

*Student paraphrase:*

Moreover, corpora give frequencies of a given word in a variety of text types, its meaning and status has to be inferred from that data. (RR1.G.HB.027)

While the focus on the original sentence is on the dictionary, the student chose to foreground corpora instead, resulting in a paraphrase that is clearly based on that particular sentence but that the student has made their own.

Re-writing source text excerpts completely is certainly a strategy requiring advanced academic writing proficiency, both in terms of vocabulary and grammar (see also Keck 2014). The L2 writers frequently made inferences about the source text which are manifest in their paraphrases. This strategy of 'reading between the lines' apparently helped them formulate paraphrases that expressed the content appropriately without re-using the same expressions; see (6.29).

(6.29 ) *Source text excerpt:*

When comparing data on *eco* in a dictionary (the OED) and a corpus (the BNC), it seems to make more sense to embark on a contrastive analysis. (Kettemann et al. 2003: 136)

*Student paraphrase:*

In order not to distort the results of the study it is reasonable to apply the contrastive analysis. (RR1.G.HB.009)

While the desire not to distort the results is not expressed in the source text, this interpretation is certainly warranted. The student thus adequately captures the motivation for the authors' choice of a contrastive analysis in this cleverly formulated paraphrase.

Considerable rewriting is not a warrant for a successful paraphrase, however. Some paraphrases in the corpus have been so vigorously rephrased that they feel somewhat removed from the source text and are on the verge of misrepresenting its content. For example, one student speaks of an "experiment" and of "tools and sources" instead of naming corpora and dictionaries (as in example 6.30). The resulting paraphrase is rather

vague and presents a desideratum proposed by the authors as more generalised than it is intended in the source text.

- (6.30 ) In conclusion, Kettemann, König and Marko express their lasting support for the positive impact of a larger range of tools and sources on the results of linguistic research as their experiment confirmed their hypothesis about the comparably more extensive results through contrastive analysis, yet admitting visible problems that resulted from the different tools they used. (RR1.G.HB.012)

The correct representation of source text content is a crucial goal of any academic summary, yet certain ideas and concepts from the source text were seemingly misinterpreted by the students and thus misrepresented in their texts. Smaller infidelities include the reference to *eco-* as a word rather than a bound element. Such errors are likely due to the students' imperfect knowledge of linguistic terminology and not related to intertextual issues. There are, however, also sentences in the data in which the students incorrectly convey the content of the source text. The misrepresentation in example (6.31) is the result of an inappropriate synonym substitution of *sometimes* with *often*. There seems to be a tendency to overstate ideas from the source text. This is noticeable in example (6.32), in which the student exaggerates the severity of an unsolved issue discussed by the authors.

- (6.31 ) *Source text excerpt:*

[...] sometimes forgetting that other, different forms of data are equally valuable. (Kettemann et al. 2003: 135)

*Student paraphrase:*

According to Kettemann, König & Marko, corpus linguists are often forgetting that there are other valuable kinds of data than a corpus. (RR1.G.HB.015)

- (6.32 ) *Source text excerpt:*

Why the 'environmentally friendly' meaning of *eco-* should not appear and the 'pertaining to the ecological movement' meaning be included is not entirely clear and does not seem to relate to frequency of occurrence. We have suggested some possible reasons, but it is beyond the scope and indeed objectives of this study, to conclude why dictionaries contain some words and omit others. (Kettemann et al. 2003: 146)

*Student paraphrase:*

The only problem the authors claim is that there is no evidence why dictionaries contain some words and why some words are being left out. (RR1.G.HB.092)

Such inaccuracies may be the result of the students' lack of care in their endeavour to avoid textual overlap and to make their texts sound academic. It is unlikely that there was an actual misunderstanding of the source text. In other cases, however, misrepresentation of source text ideas is quite certainly caused by incorrect comprehension of the source text itself. This results either from misunderstandings of the meaning of individual words in context or from misunderstandings of the foundations of the study more generally. The paraphrase in (6.33) reveals the student's lack of understanding of the original sentence. This student falsely interpreted the word *economy*, which the authors provided as an example of a neo-classical compound, as denoting a semantic field, and distorted the reference to neo-classical compounds. Example (6.34) shows another student's confusion over the source text's second section, in which the authors compare two types of analysis to justify their use of a contrastive analysis. The paraphrase in the example and several that follow it are based on the false assumption that the authors combined both approaches in their study. Due to this misrepresentation, it is obvious from the reading report that the student did not read or understand the source text thoroughly enough.

(6.33 ) *Source text excerpt:*

Note that *eco* also features in neo-classical compounds not related to biology or the environment, e.g. in *economy*. (Kettemann et al. 2003: 138)

*Student paraphrase:*

Eco is not only related to biology or environment, it also relates to economy, called neo-classical. (RR1.G.HB.190)

(6.34 ) *Source text excerpt:*

Basically, there are two routes possible if one has two sources of data: additive analysis or contrastive analysis. [...] When comparing data on *eco* in a dictionary (the OED) and a corpus (the BNC), it seems to make more sense to embark on a contrastive analysis. (Kettemann et al. 2003: 136)

*Student paraphrase:*

To be able to analyze the prefix, they use two different types of analysis: additive and contrastive. (RR1.G.HB.020)

Misrepresentations resulting from misunderstandings are often paired with patchwriting. The student who produced example (6.35) apparently struggled to understand the technical nature of explanations, but felt they were important to include. They decided to copy a longer passage outlining methodological decisions, which results in a nonsensical paraphrase. It is likely that this and other students resorted to patchwriting in an attempt to cover up issues of comprehension, an issue that has been noted in previous research (see further discussion in section 6.3.4). Other paraphrases from the data suggest that students' attempts to avoid copying may result in misrepresentations, for example when they inadvertently change the meaning by omitting integral parts of copied strings; see (6.36).

(6.35 ) *Source text excerpt:*

It is worth noting that we have adopted a broad notion of lemmatization, including different word categories under the same lemma [...]. [...] It is striking that the majority of the terms appear in both sources, which indicates that this meaning with this particular formal pattern (Greek base) is likely to be easily taken on board by lexicographers. (Kettemann et al. 2003: 142)

*Student paraphrase:*

On the one hand, they have a Greek base and on the other hand because they include different word categories under the same lemma e.g. ecosystem, which indicates that lexemes with a Greek basis are easily listed by lexicographers. (RR1.G.HB.190)

(6.36 ) *Source text excerpt:*

Note that *eco* also features in neo-classical compounds not related to biology or the environment, e.g. in economy. (Kettemann et al. 2003: 138)

*Student paraphrase:*

Its origin is from the neo-classical, but it also features compounds which are not related to biology or the environment.

While the number of borrowed words was minimised, the omission of the preposition in the student's version results in a meaning that is not intended in the source text, of which the student was apparently oblivious. Though it is of course possible that this error was an oversight, the pressures of avoiding plagiarism coupled with insufficient knowledge of avoidance strategies may have played a role here.

Some misrepresentations result from students reporting something that is not expressed in the source text. By changing *language* to *languages* and claiming that the study aimed to establish which lexemes are not used, the student who produced the paraphrase in example (6.37) changed two important premises of the original study.

(6.37 ) *Source text excerpt:*

A contrastive analysis of dictionary and corpus data should therefore reveal which lexemes actually occur in authentic language, which are (considered) institutionalised and the level of overlap between these. (Kettemann et al. 2003: 135)

*Student paraphrase:*

An analysis was made to show which lexemes actually occur in authentic languages, which don't and which are overlapping. (RR1.G.HB.014)

This illustrates how small mistakes can lead to misrepresentation of source text ideas that decisively affect the quality of a paraphrase. In this and other cases, they may have been the result of the students' overzealous attempt to reduce similarity between theirs and the source text. Balancing language- and content-related aspects in a paraphrase requires an awareness of the complex relationships between source text and summary that some of these students apparently have not yet acquired.

The paraphrasing strategies used in these reading reports show similarities to those observed in shorter summaries. The lexical and grammatical strategies employed in paraphrasing are a sign of the students' awareness of academic conventions regarding intertextuality and plagiarism. Yet, the present study confirms findings from previous studies showing that L2 writers often rely closely on the lexis and grammatical units of the source text instead of venturing away from the text (e.g. Keck 2006, 2014; Shi 2004). It is obvious that these strategies tend to be employed to reduce textual overlap, most likely out of fear of accusations of plagiarism (see also Li & Casanave 2012), while the formulation of sophisticated paraphrases seems to play less of a role. This is especially noteworthy because this type of assignment – a critical summary of one academic research article – generally precludes deceptive forms of plagiarism and is aimed at training paraphrasing skills. Such issues emphasise the importance of raising students' awareness of the pitfalls of paraphrasing and routes to circumnavigate them. As other researchers have asserted (e.g. Abasi & Graves 2008; Wette 2010; see also discussion in Pecorari & Petrić 2014 and in section 8.1.2), it is important for teachers to teach effective paraphrasing strategies instead of merely warning against plagiarism.

In several reading reports, students used summaries in addition to paraphrases. They are discussed in the following section.

### 6.3.2.2 Summarisation

Summaries and summary paraphrases constitute only a small minority (24; 4.2%) of intertextual links. That summarisation is relatively infrequent in the data was to be expected due to the narrow definition of the two categories S and SP. As discussed above, it was discovered in the early stages of the development of the annotation scheme that many apparent summaries are in fact paraphrases of individual summarising sentences in the source text. This means that they are akin to paraphrases of individual sentences, though with a summarising function. They were assigned to a category separate from paraphrases because they show characteristics not shared by other paraphrases. These SPs constitute the majority of summaries (15; 2.6%). Without exception, they are based on sentences from the abstract or the introduction.

The most typical characteristic of summaries is that they contain the title of the source text. The title is sometimes replaced with a phrase such as *the present text* or *the text/paper/research by* followed by the authors'

names (see example 6.38). This phrase is usually placed at the beginning of the summary. The authors' names are a frequent feature of summaries and summary paraphrases and are sometimes accompanied by their affiliation and the year of publication (see example 6.39). Surprisingly, some reading reports contain more than one summary. Perhaps even more surprisingly, not all reading reports begin with a summary; they are only found in three quarters of reading reports. In several summaries, the students distributed the information about methods, data, and goals of the study across up to three sentences. Those texts that do not contain a summary either use a quote as an introduction (RR1.G.HB.014), use a direct quote with a summarising function (RR1.G.HB.190), begin with a meta-level observation that foregrounds the act of writing a reading report (RR1.G.HB.025) or jump *in medias res* (e.g. RR1.G.HB.010).

- (6.38 ) The research paper “The BNC and the OED. Examining the usefulness of two different types of data in an analysis of the morpheme *eco*” by Kettemann, König and Marko from 2003 deals with showing the importance of comparing dictionary and corpus data to see if there are differences between commonly used words in language and these seen as institutionalized terms. (RR1.G.HB.189)
- (6.39 ) In their study, Kettemann et al. (2003) apply contrastive analysis of corpus and dictionary data to reveal putative differences between language in use and institutionalized collections of words. (RR1.G.HB.027)

Interestingly, all summaries and summary paraphrases are attributed, typically to the authors or the paper. There is a clear preference for attribution to the type of publication for summary paraphrases. Only one summary and one summary paraphrase are documented. Eight out of nine summaries and all 15 summary paraphrases are accompanied by a reporting structure, which usually contains a reporting verb that is followed by a noun phrase. All of these reporting structures precede the cited content. It appears that summaries are interpreted by these students as almost obligatorily being accompanied by attribution and a reporting structure. The summarisation function of these intertextual links apparently prompted them to acknowledge the citation, which is likely an application of the writing conventions these students were taught as part of their secondary school education.

The data suggest that the construct ‘summary’, which is widely used in the literature on intertextual writing (see e.g. Wette 2010), is somewhat problematic. Many summaries in the data paraphrase source text sentences, as evident from the high proportion of SPs. Even for many of those sentences coded as S in this study, it was possible to identify informational or lexical overlap with two or more source text sentences. Most of the information used in summaries in this corpus was taken from the abstract or the introduction and not from across the text. As a result, summarisation may be conceived of as a function of paraphrases rather than a distinct type of intertextual link. It is nevertheless possible that the processes of writing summaries and paraphrases are different, which is why summary was retained as a distinct category in this study.

### 6.3.2.3 Direct quotation

Direct quotation is the second most frequent intertextual strategy in the data. 68 direct quotes were identified in 46 sentences from 17 texts, i.e. 8% of sentences in the data contain at least one direct quote. This percentage is considerably smaller than those found in student research papers and pre-Master’s dissertations in previous studies (McCulloch 2012; Petrić 2012; Wiemeyer 2019), suggesting that the use of direct quotation is dependent on text type as well as discipline. Typically, the students integrated one quote into a sentence (28 sentences; 61%). Combined quotes also occur, with two (14; 30%) or even three quotes (4; 9%) per sentence. The students prefer double quotation marks. Interestingly, definitions are sometimes placed in single quotation marks, probably following the source text’s example.

References are only given in seven sentences containing nine direct quotes, of which all but one are non-integral citations. This is interesting insofar as integral citation has previously been identified as being slightly

more common with direct quotations in applied linguistics articles (see Pickard 1995). All of these documented direct quotes stem from two students' texts. Though these two students' referencing strategies differ from experts' texts, it is not feasible to draw conclusions based on this small number of findings. Attribution is more frequent and occurs in 18 sentences containing 28 direct quotes. In total, only half of all direct quotes are documented and/or attributed to the source text, which is probably the result of text-type influence. Because their reader – the lecturer – knew the source text, these L2 writers may have felt it unnecessary to provide information about whom they were citing.

All direct quotes in the data are embedded quotes, all of which are syntactically integrated fragments. This finding is somewhat unexpected given previous studies that identified syntactic integration as an issue in L2 writers' use of direct quotation, at least in essays and research papers (Verheijen 2015; Wiemeyer 2019). It is possible that the reduced complexity of reading reports that are based on a single source text made it easier for students to focus on such formal aspects. The majority of direct quotes (59; 87% of all direct quotes) are embedded in paraphrases, the remainder is embedded in a reporting structure. There are two types of reporting structures that occur with direct quotes. The first is a canonical reporting structure, usually containing a reporting verb, e.g. *Kettemann et al. conclude that* (RR1.G.HB.010). In the second type of reporting structure, it is made explicit that the direct quote contains a definition, but it is not attributed, e.g. *The first definition of the morpheme is* (RR1.G.HB.090).

A closer analysis revealed that there are three different types of syntactically integrated fragments. Direct quotes of definitions and glosses provided in the source text occur most commonly and constitute an impressive 46 out of 68 direct quotes (68%). For example, these direct quotes contain the authors' glosses of Greek words and definitions for the two data sources, the *British National Corpus* (BNC) and the *Oxford English Dictionary* (OED). Apparently, the L2 writers used quotes for complex or technical excerpts, perhaps because they found them difficult to rephrase (see also Wiemeyer 2019). More notably, many students quoted one or more of the authors' six definitions of the meaning of the morpheme *eco-* as in example (6.40). In the original article, these are set off from the main body of the text and numbered. Three meanings are then discussed in dedicated sections in whose headings the definitions also appear. Sentences which quote two or three of these definitions account for the majority of combined quotes in the data. Interestingly, while most students quoted only the definitions, some students quoted a sentence fragment that contained the definition, as in example (6.41) in which the student quoted from a section heading.

(6.40 ) From corpora and dictionary, they extract these meanings, “pertaining to the (balanced) interaction between organisms / environment”, “environmentally friendly” and “pertaining to the ecological movement” being the most popular uses according to corpora data and dictionary entries. (RR1.G.HB.012)

(6.41 ) “Eco in the meaning ‘pertaining to the ecological movement’” includes words like eco-terrorists which allude to political groups who engage in some way, shape or form in the dispute on ecological correctness. (RR1.G.HB.017)

Something about the definitions drew the majority of students towards them. They may have regarded the six meanings of the prefix *eco-* postulated by Kettemann et al. (2003) as especially worthy of being quoted. It is also possible that they felt they were too risky to paraphrase. In fact, paraphrases of these definitions are the exception in the corpus and sometimes result in patchwriting, as (6.42) illustrates (copied passages underlined).

(6.42 ) In their analysis of the meaning of eco, they find *6 meanings*: concerning the study of interactive relations between organism, concerning any integrative study, concerning the relationship between organisms and the environment, a general interaction between entities in any field, concerning the ecological idea and, lastly, the meaning of environmentally friendly.

Previous research has suggested that students prefer direct quotation if they have not fully understood the respective part of the source text (Hirvela & Du 2013). A study of research papers by Wiemeyer (2019) showed that L2 writers not only frequently quoted definitions and technical explanations, but also struggled to integrate them appropriately into their texts. Petrić (2012: 114) has suggested that L2 writers quote technical passages from their source texts which they cannot process linguistically and use them as “stepping stones to writing like an expert”. The quotation of definitions and technical descriptions appears to be a feature of L2 writing more generally, which is sometimes caused by comprehension issues and lack of paraphrasing skills. In this particular study, the prominence and centrality of the definitions in the source text may have played a role. The observations made here are indicative of complex underlying motivations for using direct quotes, though corpus data alone cannot fully capture these.

Thirteen direct quotes (19%) refer to terminology consisting of one or two words; most commonly the term *eco-*. In the case of *eco-*, the students apparently intended to highlight the fact that the morpheme is discussed by the study authors as an example, but were apparently unsure of the appropriate linguistic notation. Other terminology, e.g. *intersubjective* and *institutionalisation*, was placed in quotation marks to signal to the reader that these are terms used in the source text. Note that only four students quoted terminology. In Petrić’s (2012) study, writers of high-rated theses frequently quoted terminology, while low-rated writers rarely did so. She explained this by the better-performing students’ broader knowledge of the subject matter, which they displayed through the use of key terms.

Petrić suspected that the higher-rated students were aware of the need to define technical terminology. Their use of quoted terminology across their theses showed that they were also able to apply technical terms in their writing, though the repeated use of quotation marks for the same terms was associated with insecurity about the formally correct way to mark terminology. Based on these findings, it would be reasonable to assume that the students in this study who marked terminology were eager to display their understanding of key terms and to avoid plagiarism. However, in Shi’s (2008) study, one student explained that they used quotation marks around individual words because they would not have used those terms themselves. Thus, the use of quotation marks with terminology should be interpreted as a sign of students’ awareness of the need to indicate intertextual relations in academic writing, but also discloses certain insecurities.

The remaining nine direct quotes (13%) are sentence fragments of varying length. The shortest is comprised of five words, the longest 42 words. Though they are all syntactically integrated, many of them are of such a length that would be unlikely to be found in published expert writing, see e.g. (6.48) (see also Verheijen 2015). While the use of direct quotation was formally correct in these cases, it is possible that they would be marked by teachers as lacking in qualitative terms. Clause-based sentence fragments are a feature of inexperienced L2 writers’ texts because their syntactic integration is less demanding (Petrić 2012). While reporting structures usually suffice for embedding clause-based fragments, shorter fragments require more effort. Thus, long sentence fragments point towards the students’ inexperience in quoting and rephrasing academic source texts.

None of the direct quotes in the data are full sentences. This may have been induced by the task description, which explicitly asked students to use their own words, but also by the students’ academic writing expertise. Quoting fragments rather than full sentences is a feature of advanced and expert academic writers’ texts (Petrić 2012; Verheijen 2015), which points to the already relatively advanced intertextual writing expertise of these L2 writers at the end of their first year of study. Furthermore, none of the reading reports contain free-standing quotes, which means that Borg’s (2000) assumption that L2 writers’ texts are characterised by a preponderance of free-standing quotes cannot be confirmed for this learner group (see Petrić 2012; Wiemeyer 2019 for similar findings). The use of free-standing quotes appears to be a feature of certain text types, such as research papers or essays, but not of reading reports.

Errors in direct quotation mostly concern errors in quotation marks and incorrectly quoted passages. As in Verheijen’s (2015) study, the students made errors in the use of quotation marks, though they are not quantified

here. A typical mistake is the use of German quotation marks „“ as in example (6.43). This attests to the lack of awareness of different quotation mark conventions and can easily be remedied. There are a few cases of quotation marks used for terminology in which it is arguable whether they were necessary. This includes the use of quotation marks around the prefix *eco-* and its source word *ecology* discussed above, see (6.43).

(6.43 ) With regard to the morpheme „eco“, the authors consider the term „ecology“ as the most frequent one in the two sources of data. (RR1.G.HB.092)

In one text (RR1.G.HB.090) the student retained the inverted commas from the source text in three direct quotes of definitions, resulting in an awkward combination of double and single quotation marks. Errors in quotation marks are probably an issue of novice writers in general, but can also happen to advanced academic writers. It is important to raise students' awareness of this type of error because it negatively affects the text's sophistication and credibility (Verheijen 2015).

Several errors in spelling and missing words occur in the data that are judged to be a result of sloppy copying. A more severe issue are passages marked as direct quotes for which there is no identical match in the source text, for instance (6.44) and (6.45). Example (6.45) is compelling in this respect because this student, like others in the sample, decided to quote the definitions provided by the authors, but reformulated the definitions despite the use of quotation marks. This inappropriate practice may be the result of an unsuccessful attempt at avoiding plagiarism. It is revealing of the student's insufficient understanding of direct quotation and of what qualifies as inappropriate textual borrowing.

(6.44 ) *Source text excerpt:*

Of the six possible meanings of *eco-* identified at the end of section 4, three proved to be highly productive.

*Student paraphrase with incorrect direct quote:*

Instead of an entry for a common and socially accepted expression, the dictionary had nothing to offer other than the two other “highly productive meanings” which were also part of the corpus data. (RR1.G.HB.012)

(6.45 ) *Source text excerpt:*

These reflections lead us to posit six meanings of *eco*:

1. pertaining to the study of the interactive relations between organisms [...]
  3. pertaining to the (balanced) interaction between organisms / environment [...]
  5. pertaining to the ecological movement
- (Kettemann et al. 2003: 139)

*Student paraphrase with incorrect direct quote:*

Firstly, *eco* can be used in the sense of “balanced interactive relations between organisms” (e.g. as in ecosystem) or secondly, referring to “relations between organisms and their environment”. [...] The fourth possible meaning of *eco*, “associated with the ecological movement”, is mainly connected with politics (e.g. eco-terrorism) and often negatively connotated. (RR1.G.HB.027)

It is widely acknowledged in L2 writing research that L2 writers tend to be unsure of what needs to be placed in quotation marks (Pecorari 2003; Pecorari & Petrić 2014). This assumption is supported by the present study, which reveals some students' issues with this aspect of intertextuality. (6.46) is a particularly interesting example because the student quoted a passage of eight words in their sentence and marked it using single quotation marks, but did not mark the two other copied passages in the sentence as direct quotes despite the fact that they were longer.

There are only a few instances of direct quotation in the data where the string could have been paraphrased, see e.g. the sequence of common words quoted in (6.47). Quotes containing short strings of common words have previously been identified as an issue in L2 writers' quotation practices (Petrić 2012). Such direct quotes are viewed as a reflection of the fact that the students are unsure of what constitutes plagiarism and whether it would be legitimate to borrow such a string without the use of quotation marks (Pecorari 2003). There are also some cases in the data in which the students apparently chose a direct quote where they experienced difficulty in paraphrasing (see also Hirvela & Du 2013). In example (6.48), the student quoted a lengthy sentence fragment from the abstract at the beginning of their reading report in lieu of a self-written summary. As these issues were a marginal occurrence in this study, teachers and researchers should take care to assess whether students really do struggle in this respect, also because the interpretation of whether something should or should not be quoted is very subjective. Instructions and hands-on activities can provide students with reliable guidance on how to quote and paraphrase effectively in a given context.

- (6.46 ) In the end, the authors found out that a corpus and a dictionary 'do not represent the same version of reality', as a corpus is likely to contain a high portion of words that are fashionable for a time and then fade out of use, whereas a dictionary contains words which are more likely to be given a permanent place in the language.
- (6.47 ) At first, the authors mention two types of analyses which are used when you have two sources of data, called "additive and contrastive analyses". (RR1.G.HB.022)
- (6.48 ) The given research paper deals with "the relevance for lexical studies of comparing dictionary and corpus data, on the assumption that corpora contain lexemes as raw data whereas dictionaries contain a specific selection of words chosen by lexicographers on the basis of their perceived institutionalization in the language." (RR1.G.HB.014)

Overall, the use of direct quotes by these L2 writers shows that they have largely mastered the formal aspects of direct quotation, such as quotation marks and syntactic embedding. Relatively few direct quotes are undocumented and unattributed, but this should be viewed as a characteristic of the text type, not as a deficiency of the learners. However, insecurities remain with regard to what, how much, and why to quote at this stage of academic writing proficiency.

#### 6.3.2.4 Meta-level observations

A noteworthy phenomenon that has not been previously discussed in the literature are observations on the meta-level. These were used by the students to make explicit the act of summarising and the intertextual nature of the reading report. Unlike paraphrases, direct quotes, and summary paraphrases, such sentences are usually not based on passages from the source text. Their propositions are not explicit in the source text. Yet, these sentences are clearly intertextual because they describe and comment on its structure and content; see examples (6.49) and (6.50).

- (6.49 ) After that the authors compare the two data in tables. (RR1.G.HB.013)
- (6.50 ) Kettemann, König and Marko close their argumentation with a summary of their text. (RR1.G.HB.090)

By conveying an impression of the external structure of the source text, meta-level observations aid in structuring the reading report. The students draw on the structural features of the source text, e.g. sections and

headings, to provide the reader with a general idea of the structure of the research article. The majority of meta-level observations are not documented, but attribution is quite frequent, occurring in 32 out of 37 meta-level observations. Interestingly, most meta-level observations are framed by a reporting structure, usually a reporting verb followed by a noun phrase, even though the content of the sentences itself is not reported (see 6.51). Structuring elements from the source text such as table captions apparently sometimes form the basis for meta-level observations by the students, as in (6.52).

(6.51 ) The author explains the reader the differences between the two sources. (RR1.G.HB.011)

They point out some advantages and disadvantages of both sources and explain some key vocabulary like institutionalization. (RR1.G.HB.016)

(6.52 ) Source text table captions:

Table 1. *Eco*-words found in both the BNC and the OED

Table 2. *Eco*-words found either exclusively in the BNC or exclusively in the OED

Student meta-level observation:

Now, Kettemann, König and Marko divide all *Eco*-words into three categories, words found in both data sources, words only found in the BNC and words only found in the OED. (RR1.G.HB.015)

While the explanation in (6.52) contains some lexical overlap, a feature of about one third of meta-level observations, it makes a statement on the meta-level and describes the structure of the source text. The proposition of the meta-level observation is not found in the source text, which distinguishes it from the category of paraphrase. That the distinction must be made by careful comparison to the source text is evident from paraphrases of source text sentences in which the authors themselves make their research process explicit, see (6.53). Unusual reporting structures may create the impression that the student is making a meta-level comment, but a comparison to the source text clearly shows that they are paraphrasing, as in example (6.54).

(6.53 ) *Source text excerpt:*

A contrastive analysis of dictionary and corpus data should therefore reveal which lexemes actually occur in authentic language, which are (considered) institutionalised and the level of overlap between these. (Kettemann et al. 2003: 135)

*Student paraphrase:*

An analysis was made to show which lexemes actually occur in authentic languages, which don't and which are overlapping. (RR1.G.HB.014)

(6.54 ) *Source text excerpt:*

Corpus linguists have a tendency to regard the corpus, if not as a panacea to all linguistic problems, then at least as the most valuable kind of data in the study of language, sometimes forgetting that other, different forms of data are equally valuable. (Kettemann et al. 2003: 135)

*Student paraphrase:*

The reader is being introduced to the topic by mentioning that corpus data is the most valuable kind of data in the study of language. (RR1.G.HB.011)

A rather unusual strategy observed in a minority of reading reports is a kind of meta-level approach to the source text's structure. Students who employ this strategy describe the steps taken by the authors and the structure of the text, but do not summarise the arguments, findings, and interpretations presented in the research article. RR1.G.HB.014 is a case in point; see example (6.55).

(6.55 ) The introduction explains what is done, a research comparing Oxford English Dictionary (the OED) and the British National Corpus (the BNC) and used the example of the prefix *eco*, as it is said to be one of the fashionable words today.

Next, two main types of analysis were shown and then explaining them as well as transferring them on the subject matter explained why the contrastive analysis was used, not the additive analysis.

Next, the semantic of the word *eco* is been discussed and a brief history of the word is given as well as the range of the word is shown and said that the research was concentrated on words with the prefixation *eco* and not every word with *eco* in it. (RR1.G.HB.014)

Each of these sentences comments on a structural element of the source text, but important information from the source text is missing, for example why the authors selected a contrastive over an additive approach to analysing their data. This results in a superficial description of the source text's structure rather than its contents, which is of limited informative value to the reader. In this specific case, this strategy seems to be an attempt to compensate for insufficient understanding of the subject matter, which is nevertheless obvious from some of the phrasings, especially in the third sentence of example (6.55).

Meta-level observations are a special kind of intertextual link because they do not convey source text content but rather comment on its structure. They are used by these L2 writers to guide the reader through the text and provide them with information on the structure of the source text. But, as the example above shows, while meta-level observations can be a means of structuring the reading report, they are not an appropriate substitute for other intertextual links in terms of conveying the content of the source text.

### 6.3.2.5 New content introduced by the students

The purpose of a reading report is to summarise a study presented in a research article. Hence, as in other summary tasks, the students are not expected to add ideas of their own. That this does not necessarily stop L2 writers from adding material of their own has been shown in previous studies (e.g. Keck 2010, 2014). In the reading reports analysed here, there are only a few instances in which the students included information in their reading reports that was not given in the source text. Sentences with this kind of 'knowledge display', see for example (6.56), are intertextual in a sense, though they are not based on the source text. It has to be accentuated that six out of the seven sentences coded as 'new content' (NC) occur in a single reading report, namely RR1.G.HB.026. This student regularly provides information that is not in the source text. They also offer their own description and interpretation of the data rather than paraphrasing the authors', see (6.57).

(6.56 ) In contrast, the formation of a corpus is based on explicit design criteria. (RR1.G.HB.019)

(6.57 ) The OED is one of the most respected dictionaries in the English language. [...]

The only common word used in both files is *ecofreak*. Again I would assume that the higher amount of "eco-words" in the BNC results from the green-wave that hits the globe since the late eighties and that green-activism and green-awareness get more important day after day which is supported by the media. (RR1.G.HB.026)

The frequency and nature of sentences coded as NC in this reading report suggest that the student misunderstood the task and attempted to deduce their own interpretation of the study results. The expectation that the reading report should not contain new content is not made explicit in the task description, so it is possible that this is the cause of this student's task representation.

The annotation of 'added information' revealed that knowledge display sometimes occurs within intertextual links, but is also infrequent. There are eight instances of students having added a non-source-based

aspect to a sentence; six are paraphrases and two are meta-level observations. They stem from five reading reports. Example (6.58) shows a paraphrase that is supplemented with expert terminology; in the paraphrase in (6.59) the student displays their knowledge by volunteering a definition for the concept of free morpheme that is not provided in the source text and superfluous in the given context. (6.60) is an example of a meta-level observation that has been extended by the student's interpretation of the authors' reasons for discussing institutionalised words.

(6.58 ) The OED provides ca. 500,000 words that are actually used in the English language and are therefore not exposed to the sudden extinction of words that are only used during a short period of time (vogue terms). (RR1.G.HB.018)

(6.59 ) New lexemes consist of morphemes which are made of affixes (bound) and free morphemes which can stand alone and have their own meaning. (RR1.G.HB.025)

(6.60 ) They highlight the significance of differences between institutionalized words and not institutionalized word, because it is important to know for their later argumentation. (RR1.G.HB.016)

Integrating information obtained from other contexts in this way may be these students' way of demonstrating that they are knowledgeable in this area of linguistics. It is conceivable that the students wanted to display their knowledge to their lecturer in this obligatory assignment by referring to concepts from course readings. It has to be stressed that sentences containing new content were very rare and that the students were seemingly aware that the intertextual nature of reading reports generally precludes the provision of additional information.

### 6.3.3 Documentation, attribution, and reporting structures

The learners used manifold means to indicate to the reader that they were writing from a source text. All texts contain some kind of explicit indication of their intertextuality. Overall, 43% of all sentences in the corpus include documentation, attribution, reporting structures or a combination of those. Attribution of source text content in the text is the students' preferred means and occurs in 41% of sentences (N=239). Reporting structures are used in 223 intertextual links (39%), four of which do not contain attribution to the source text.

Documentation is marginal in the corpus data in the sense that only very few students used it. Only 3.3% of intertextual links (N=19) are documented, i.e. accompanied by a reference. References are provided in only five texts. 12 out of 19 instances of documentation stem from the same text (RR1.G.HB.090), which is the only text in the sample in which documentation is used throughout. The other four students typically used referencing in a sentence at the beginning of their reading report in which they cited the title and publication date of the source text and one or two additional sentences. Documentation predominantly occurs with paraphrases (9; 2% of all P) and direct quotes (7; 15% of all Q and PQ). Most documented intertextual links are integral citations (N=11; 58%). Another seven are non-integral citations (37%).

Unlike in the study by Wiemeyer (2019), the redundant use of an integral and a non-integral citation in the same sentence is not an issue among these learners and only occurs in one instance. References are usually given in parentheses and contain the year, the year and page number or only the page number. The latter is most frequent due to the preference of the student who used documentation throughout their text. In two documented summaries, the year of publication is given in the sentence. As the majority of instances of documentation come from one text, however, it would be misguided to draw conclusions regarding the entire group of learners.

What is evident from these findings is that the majority of the students in the sample did not use documentation. They either did not find it necessary to provide references or were not aware of this convention. Since these students had taken academic writing classes, the most likely explanation is that the lack of referencing is a feature of the text type of reading report. The fact that the students were summarising a single source text and knew that their lecturer was familiar with this text may have made referencing appear superfluous (see also Campbell 1990; Wiemeyer 2019). That the students were writing for their lecturer as the reader was evident, for example, from the informal abbreviations for authors' names, see (6.61).

(6.61 ) According to KK+M we have six differing meanings of *eco* all connected to ecology in a more or less biological context. (RR1.G.HB.026)

Despite the infrequency of references, the students were obviously aware of conventional means of elucidating intertextual relationships in their writing. A sizeable proportion of sentences contain attribution, and a range of targets of attribution were identified in the reading reports (Figure 11).

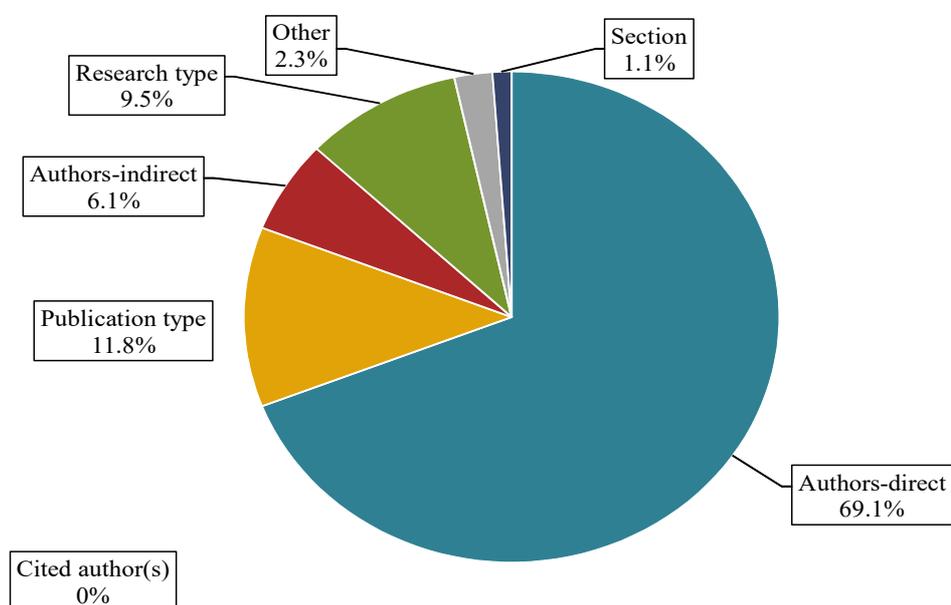


Figure 11: Target of author attribution.

The students most frequently attributed source text content to the authors, either via the noun phrase *the author(s)* (N=71; 30% of attributed sentences), via the pronouns *they* and *he* (N=57; 24%), via the authors' names (N=39; 16%) or via equivalent noun phrases denoting the authors such as *the researchers*, *the linguists*, and *the research team* (N=14; 6%). These choices of noun phrases also add stylistic variety to the students' texts. The pronoun *they* typically occurs following sentences with attribution to the authors. In two texts, there seems to be some confusion over the number of authors and how they should be cited. These students refer to the three authors of the source text as *the author* or *he* in their reading reports, see (6.62).

(6.62 ) To find out about the relevance of using different types of sources, the authors took the morpheme *eco* to analyze its semantic meaning [...]. With the Method of contrastive analysis he will verify his thesis that [...]. (RR1.G.HB.016)

The text by Kettemann, König & Marko is about [...]. The author starts with an introduction about what is going to be examined in the text. (RR1.G.HB.011)

The reason for the misrepresentation of the co-authorship is unclear, but it could be the result of insufficient reading or lack of concentration. Since all three authors are acknowledged in a preceding sentence in both examples, another explanation might be that the phrase *the author* was interpreted as a kind of fixed formula. It is possible that the students learnt this phrase in academic writing classes and employed it as a phraseological means of making the act of citation explicit and for this reason did not make the necessary morphological change.

Some students selected an indirect form of attribution via passive structures (N=16; 7%). In such sentences, the student makes the intertextual nature of the sentence explicit to a certain extent, though the agent is veiled as in (6.63).

(6.63 ) In addition, the general differences of the BNC and the OED are presented.

(RR1.G.HB.014)

Such indirect attribution usually follows sentences with direct author attribution and can be decoded as such by the reader. It is easily overlooked in automated corpus queries based on search terms such as author names (see e.g. the approach taken by K. Hyland 2002, 2004). This is unfortunate because such passive constructions are an effective way of making intertextuality transparent and creating stylistic variation. Studies of attribution in academic writing should take these into account.

Sentences in which the target of attribution is the type of publication constitute the second largest group. The students mostly attribute content to *the text* (N=20; 8%) and to *the (research/academic) paper* (N=11; 5%), both of which occur in modified noun phrases such as *the given text*. Another type of attribution is to the type of research. The students name the *study* (N=15; 6%), the *research* (N=5; 2%), or the *analysis* (N=3; 1%). Research-related adjectives such as *empirical* sometimes modify these nouns. Secondary citation, i.e. attribution to one or several author(s) cited by the authors of the source text (Pecorari 2003), does not occur, though there was one such case in the pilot study. This may be due to the text type, which does not require students to write a literature review based on multiple sources, but it is also possible that these students took care to avoid secondary citation.

92% of attribution (N=219) occur in sentences with reporting structures. In the remaining 20 sentences, attribution occurs in other sentence structures, e.g. in subordinate clauses and modifiers of noun phrases (see e.g. 6.64).

(6.64 ) The data sources of the study are the British National Corpus (BNC) and the Oxford English Dictionary (OED; 2nd edition and online additions up to July 2001). (RR1.G.HB.014)

Contrastive analysis is the chosen method by Ketterman, König and Marko.(RR1.G.HB.026)

10% of sentences contain two types of attribution, e.g. to both the authors and the study. In these cases, attribution is typically encoded in a reporting structure and additionally in an adverbial of place such as *in the text* or *in the study*. This shows that while attribution outside of reporting structures is possible, it is relatively uncommon and usually only occurs in combination with attribution in reporting structures. Generally, the data attest to varied and purposeful forms of attribution in the reading reports. That all students used it clearly shows that they are aware of the need to foreground the origin of information in academic texts, despite the lack of documentation. In addition, they exploited the stylistic potential of reporting structures by alternating different targets of attribution.

Reporting structures serve to accentuate the intertextual nature of the information that was communicated in the reading reports. They occur with all types of intertextual links except new content (NC) and are characterised by remarkable stylistic variation. On average, each text contains seven reporting structures, but individual preferences with regard to their use are apparent. There is at least one reporting structure per text, the

largest number of reporting structures in one text being 17. A variety of different grammatical structures are attested (see Figure 12).

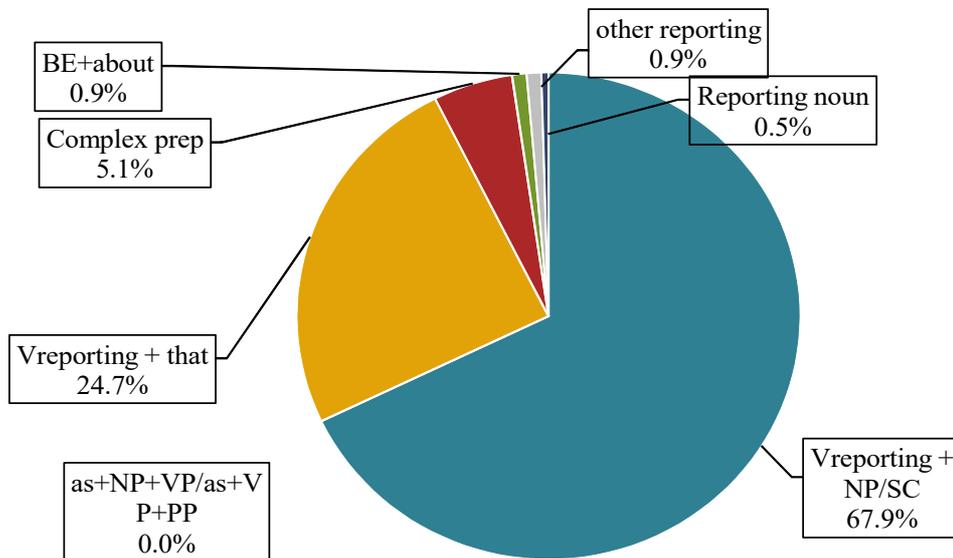


Figure 12: Types of reporting structures.

The vast majority of reporting structures contain reporting verbs, either in the structure  $V_{\text{reporting}} + \text{NP}/C_{\text{subordinate}}$  ( $N=146$ ) or  $V_{\text{reporting}} + \text{that}$  ( $N=53$ ). A total of 98 different simple and complex verbs were identified in 210 instances<sup>63</sup> of reporting verb use. Table 7 lists the ten most frequent reporting verbs overall and those attested with each of the two structures; a complete list of all reporting verbs in the corpus appears in Appendix 1.

Table 7: Most frequent reporting verbs.

Rank	All reporting verbs		$V_{\text{reporting}} + \text{NP}/C_{\text{subordinate}}$		$V_{\text{reporting}} + \text{that}$	
	Reporting verb	Frequency	Reporting verb	Frequency	Reporting verb	Frequency
1.	<i>explain</i>	12	<i>examine</i>	10	<i>state</i>	6
2.	<i>examine</i>	10	<i>explain</i>	10	<i>conclude</i>	4
3.	<i>point out</i>	9	<i>analyse</i>	7	<i>point out</i>	4
4.	<i>state</i>	8	<i>deal with</i>	7	<i>argue</i>	3
5.	<i>analyse</i>	7	<i>compare</i>	6	<i>claim</i>	3
6.	<i>compare</i>	7	<i>point out</i>	5	<i>suggest</i>	3
7.	<i>deal with</i>	7	<i>use</i>	5	<i>explain</i>	2
8.	<i>show</i>	6	<i>describe</i>	4	<i>find out</i>	2
9.	<i>conclude</i>	5	<i>discuss</i>	4	<i>show</i>	2
10.	<i>use</i>	5	<i>show</i>	4	<i>verify</i>	2

It is clear from these lists that certain verbs are preferred with a certain structure. It is also clear that many reporting verbs do not occur in the  $V_{\text{reporting}} + \text{that}$  pattern, calling into question the generalisability of corpus studies which identified reporting verbs via this structure (e.g. Charles 2006a, 2006b).

63 Note that some reporting structures contained more than one reporting verb.

In contrast to previous studies (e.g. Manan & Noor 2014; Verheijen 2015; Wiemeyer 2019), there is no over-representation of specific reporting verbs in this study. The verb *state*, which has been claimed to be ‘overused’ by different groups of learners in Master’s theses (Manan & Noor 2014) and essays (Verheijen 2015), is among the most frequent verbs in this study, but like the other frequent verbs it is distributed across different reading reports. It appears that earlier corpus studies considering reporting verbs without distinguishing between individual texts have misrepresented the actual usage of these verbs by individual L2 writers. In the present study, it was relatively uncommon for individual students to use the same reporting verb more than once. Notably, these L2 writers did not limit themselves to ‘traditional’ verbs of reporting. Among the reporting verbs that occur only once in the corpus are complex ones such as *come to the conclusion* and *test their theory by*. Some of these verbs, such as *proceed by* and *focus on*, may be specific to a summary and simultaneously serve to structure the reading report and attribute cited content.

All other types of reporting structures are comparatively infrequent. The complex preposition *according to*, which was over-represented in the texts by Verheijen’s (2015) learners, occurs only nine times. There is only one sentence that contains a reporting noun followed by *that*, but more reporting nouns are found in phrases of the structure NP + BE (+*that*) (N=8), e.g. *the conclusion is that* and *another important finding is that*. Reporting verbs and reporting nouns are sometimes combined in complex reporting structures, e.g. *close with an explanation for* and *advance the hypothesis that*. Another infrequent type of reporting structure is the pattern *NP is about* found in summaries and summary paraphrases at the beginning of the reading reports. A noteworthy observation is that some of the reporting structures identified are quite long. The students used them to convey additional information about the study or the text’s structure, e.g. *in want to show the relevance of* and *start their analysis with the examination of*. Such structures are atypical of expert writing (see e.g. Charles 2006a; Thomas & Hawes 1994; G. Thompson & Ye 1991) and symptomatic of the students’ developing academic literacy paired with their conceptualisation of the genre of reading report.

There is very little variation in terms of the position of reporting structures relative to the reported clause. The vast majority of reporting structures (203; 91%) precede the citation. Reporting structures in this position often follow extraposed elements, e.g. non-finite clauses such as *regarding the OED* and *to prove their point*. Reporting structures that follow the information (9; 4%) or are placed within (10; 5%) usually occur with passive constructions. Sentences in which a canonical reporting structure or a complex preposition is placed in the middle of or after the reported clause are exceptionally rare. Reporting structures in these positions are also relatively uncommon in native-speaker expert writing (Verheijen 2015).

Evaluation is very infrequent in the data and is only found in twenty sentences (3.5%) from fourteen texts. It is predominantly expressed via adjectives (N=9; 45% of evaluation), which occur and as a modifier in noun phrases and in phrases such as *it is ADJ*. The most common adjective is *interesting*, which occurs five times. This is noteworthy because evaluation tends to be expressed via reporting verbs in expert writing (K. Hyland 2002; G. Thompson & Ye 1991). The students in this study infrequently used evaluative reporting verbs (N=7; 35), *claim* and *suggest* being the only ones identified in the data. It is possible that these verbs were selected for stylistic reasons, not because the L2 writers consciously decided to take a stance. The sole evaluative adverb is *clearly*, used by only two students. The manual annotation of the data allowed for the identification of evaluative expressions that would have otherwise escaped notice. Two such expressions were found, namely the somewhat colloquial phrases *have nothing to offer* and *for sure*.

The students’ approaches to encoding evaluation are in stark contrast to expert writing, in which reporting verbs are used in deliberate ways to denote stance and establish the author’s position in relation to other members of the discourse community (e.g. K. Hyland 2002; Thompson & Ye 1991). At first glance, the marginality of evaluative expressions in this corpus is in line with previous research proposing that L2 writers do not sufficiently evaluate their sources and do not exploit the potential of reporting verbs in academic writing. L2 writers are assumed to avoid stance-taking because they feel that they are not authoritative enough to criticise the published authors that they cite (Abasi & Akbari 2008). However, while this may have influenced the behaviour

of these L2 writers, a more likely explanation for the lack of evaluation in this corpus is the text type. In the reading report instructions, the writers were not explicitly asked to take a stance towards the source text. The paucity of evaluative expressions in the reading reports should be attributed to the students' understanding of this genre as non-evaluative and not viewed as an issue of pedagogic relevance. This has implications for future studies of stance and evaluation, as the text type must be considered as a variable.

In sum, the varied uses of attribution and reporting structures are indicative of the students' procedural knowledge of highlighting source text information. Despite the scarcity of documentation, most sentences in the reading reports are clearly marked as intertextual. As a result, the reading reports are generally easily identifiable as being source-based. The lack of evaluation is a feature of the genre rather than a problem. It is obvious from the data that the students share a general awareness of academic writing conventions with respect to intertextuality and have more than a basic knowledge of how to make intertextuality explicit.

### **6.3.4 Textual borrowing, patchwriting, and strategic re-use of source text vocabulary**

The students made noticeable use of source text vocabulary in their reading reports. On the one hand, they sometimes copied longer strings from the source text and embedded them in their sentences. Some of this copying is illicit and results in patchwriting. Other copied strings represent source text phraseology that the students had to re-use to summarise the source text, e.g. fixed expressions and nominal compounds. On the other hand, the students use the source text as a language repository from which they strategically extract individual words to re-use in their own writing. Each type of textual borrowing is discussed in turn below.

#### **6.3.4.1 Copying from the source text and patchwriting**

Textual borrowing, as outlined above, is defined as the presence of strings of three or more words from the source text. The reading reports contain an average of 20% of words in such copied strings. 65% of sentences (N=376) contain at least one copied string, with a total of 638 strings in the corpus. There is considerable variation in the extent of copying among this group of learners, however. Figure 13 shows that some reading reports consist of copied strings to more than 30%, others to less than 10%. These results suggest that for this particular text type, a median overlap of around 17% is to be expected, because writers are required to re-use certain phrases and compounds in order to accurately report on the topic of the source text. Re-use of less than 10% and more than 20%, however, can be indicative of issues with respect to textual borrowing, as discussed below.



Textual borrowing occurs across texts, making it a characteristic feature of reading reports. A qualitative analysis of the types of strings copied reveals that this is not necessarily the case because students are patchwriting. Many strings that have been copied from the source text are in fact proper nouns and other phrasemes which have to be re-used for accurate reporting of source text content. These strings are captured under the definition of textual borrowing here, but do not constitute transgressive intertextuality. For illustration, the five most frequent copied strings are *the morpheme eco-* (N=23), *in the OED* (22), *in the BNC* (17), *the Oxford English Dictionary* (15), and *the British National Corpus* (15) (see Appendix 2 for a complete list of copied strings). Together, these five strings alone make up 14% of all copied strings. As the word cloud in Figure 14 shows, there are many copied strings in the data that do not classify as illegitimate copying but are in fact desirable.

This highlights the need to distinguish between legitimate, even necessary forms of textual borrowing and illegitimate forms. Most of these L2 writers understood which words and phrases could be legitimately re-used. On account of the data, it is questionable whether Cumming et al.'s (2005) cut-off point of three words really is appropriate for measuring textual borrowing as many of the 3-grams in this corpus are compounds and noun phrases, most of which are technical terms and proper nouns. A qualitative analysis of copied strings seems warranted in any case to distinguish legitimate from illegitimate copying.

Selecting words from the source text and re-using them in acceptable ways can pose a challenge for any writer learning the discourse of a new discipline. Despite the preponderance of legitimately copied phrases, the analysis of textual overlap disclosed that patchwriting and exact copying are an issue among some of these L2 writers at the end of their first year of university study. The most severe type of illegitimate textual borrowing is the act of copying clauses and sentences word-for-word from the source text, which was annotated here as an exact copy. This practice is relatively rare in the reading reports, which is a sign that most students took care to avoid it. Exact copies (E and PE) occur in seven reading reports in a total of eight sentences. A common issue is the copying of the definitions of *eco-* listed by Kettemann et al. (2003: 139) without quotation marks. They were copied in four out of the 25 reading reports and constitute four out of six of the exact copies embedded in a paraphrase (PE). One student copied not the definitions themselves but the headings of the subsections in which Kettemann et al. (2003: 141–144) discuss the individual meanings of *eco-*. The student even maintained the capitalisation; see example (6.65).

(6.65 ) Next up is the word *eco*, used in different ways explained, which are: Eco- meaning ‘pertaining to the (balanced) interaction between organisms / environment’, Eco- in the meaning ‘environmentally friendly’, Eco- in the meaning ‘pertaining to the ecological movement’. (RR1.G.HB.014)

That the definitions were not marked as direct quotes in their reading reports suggests that these four students did not understand them as something that needed to be marked as a direct quote. It is possible that they regarded them as part of general knowledge. Two exact copies of definitions occur in texts which contain more than 30% of text in copied strings (RR1.G.HB.010, RR1.G.HB.020), which means that these students generally relied relatively heavily on the source text. Two students whose texts contain exact copies (RR1.G.HB.20 and RR1.G.HB.25) did not use any direct quotes in their texts, which could be a sign that they are unfamiliar with direct quotation conventions.

Overall, however, the scarcity of exact copies found in this corpus, especially of entire copied sentences, speaks to the writing proficiency of these L2 writers, especially as compared to other studies of writers in their first year (see e.g. Keck 2006). The fact that many exact copies contained the definitions from the source text could be a sign that students copy illegitimately only from certain parts of source texts. It has previously been proposed that this may have to do with text comprehension, but other factors, such as the students' subjective theories about what may be copied, may also play a role. It is possible that the students considered the definitions to be general knowledge rather than the intellectual property of the authors.

Patchwritten sentences are characterised by lexical or syntactic alterations to source text strings or a combination of both. Usually, this is achieved by employing a D/A/S paraphrasing strategy (discussed above; see also Keck 2010), i.e. by integrating strings from a source text sentence to which individual words had been added or left out to avoid copying longer strings, as in (6.66). This type of patchwriting at the micro level has also been termed localised patchwriting and distinguished from global patchwriting, which operates at the level of ideas (Abasi & Akbari 2008: 271).<sup>64</sup> Extreme cases of patchwriting, or near copies, in which the student only replaced or inserted individual words between copied strings (see examples in 6.67, copied strings are underlined) are quite rare, unlike in the studies by Keck (2006; 2014). Most patchwritten sentences in the corpus consist of copied passages that are strung together from the source text and to which the student contributed relatively little themselves (see example 6.68, copied strings are underlined).

(6.66 ) *Source text excerpt (copied strings underlined):*

It will therefore be interesting to see whether there are major differences between the eco-words found in the BNC and those in the OED, which may reveal differences of language in use and language as an institutionalised collection of words and expressions. (Kettemann et al. 2003: 138)

*Student paraphrase (copied strings underlined):*

Thus, differences between eco-words in the OED and in the BNC are likely to reveal more general differences of authentic language and institutionalized language. (RR1.G.HB.010)

(6.67 ) Further more, words occurring in dictionaries are filtered, which means that not all words that have been found have been entered and that the words and expressions go through a process of human selection in the filtering and editing process. (RR1.G.HB.024)

To do so, they conduct a contrastive analysis of the morpheme eco in the Oxford English Dictionary (OED) and the British National Corpus (BNC). (RR1.G.HB.010)

(6.68 ) The study is concerned with the morpheme eco and it examines its meanings both in the Oxford English Dictionary (OED), the largest dictionary of the English language, which documents entries of various authors and in the British National Corpus (BNC), which constitutes the largest collection of empirical data which is carefully annotated, e.g. all words starting with the morpheme eco and additional details on this morpheme can be found easily. (RR1.G.HB.022)

Such examples showcase the complexity of an analysis of textual borrowing. Though the student who produced example (6.68) re-used several strings that should be regarded as necessary borrowing, the resulting sentence is very close, arguably too close to the source text. Analyses of textual borrowing should thus take into account both the proportion of copied strings per sentence as well as their content.

The analysis of the corpus data brought to light that different aspects of intertextuality may have been mastered to different extents. This is evident from examples such as RR1.G.HB.092. This learner copied extensively from the source text, with 24% of their text consisting of copied strings of three or more words. However, the learner carefully signalled the act of intertextual writing by using a range of reporting structures and words and clearly attributing the cited content to the authors of the original study. The source text is thus properly acknowledged throughout, yet the reading report relies very closely on the original wording. This is reminiscent of the students in Pecorari's (2003) and Abasi and Akbari's (2008) studies, who acknowledged their sources in the text, had in fact mostly listed them in their bibliographies, but still copied extensively from them without marking these passages as direct quotes. They clearly did not intend to deceive their lecturers, which is especially clear from Pecorari's students, who were even eager to supply her with their sources. Such cases show that the students are keen to do it right, but their behaviours are still problematic insofar as it might be difficult

<sup>64</sup> Global patchwriting was not considered here as it appears to be a phenomenon that plays a role in writing tasks that require students to construct an argument in reference to multiple source texts, which is not a requirement of reading reports.

for a teacher to discern the student's own writing ability. The heavy amount of copying could also be viewed as plagiarism despite the attributions. Such cases emphasise how important it is to give students comprehensive advice and exercises for source-based academic writing, especially for how to avoid extensive copying.

While some students generally struggled to formulate paraphrases using their own words, other students' texts show that copying from the source text does not preclude the ability to paraphrase effectively. That patchwriting is not necessarily caused by insufficient paraphrasing skills is evident from sentences in the data in which elaborate paraphrase is combined with unaltered copied passages. In example (6.69), the student is capable of paraphrasing a passage with minimal lexical overlap using original vocabulary and syntactic structures, only to then insert a copied passage of eight words. It is questionable that insufficient vocabulary or text comprehension are the reasons for this practice. Similarly, in reading report RR1.G.HB.092 there are sentences that are very close to the source text in terms of lexis, but other sentences that synthesise information from across a section with very little lexical overlap. This indicates that a range of factors may be at play in some students' writing behaviour.

(6.69 ) They chose to conduct a contrastive rather than an additive analysis to be able to focus on the actual differences between the two data sources. (RR1.G.HB.023)

Some textual overlap between the reading reports and the source text is possibly incidental, especially when students' paraphrases of one source text section contain strings from another. The paraphrase in the example (6.70), which is based on section 3 of the source text, contains a string from the abstract. It is conceivable that the student was not aware they had read this exact string in the source text and retrieved it from their memory. This cannot be reconstructed using corpus data alone. Process-based approaches are likely to shed light on such instances of potential accidental re-use.

(6.70 ) The OED on the other hand is a summary of words and definitions chosen by lexicographers. (RR1.G.HB.090)

Several reading reports in the corpus show relatively little lexical overlap, at least as far as strings of three or more words are concerned (Figure 13). Avoidance of copying does, however, not necessarily go hand in hand with effective paraphrasing. Some students' insufficient command of academic vocabulary and spelling becomes obvious in their attempts to paraphrase without re-using source text vocabulary; see example (6.71). As proposed above, minimal textual overlap may indicate insufficient understanding of the source text or the task description and can also have a negative effect on the quality of a paraphrase.

(6.71 ) As a next step we have an explanation that eco comes from ecology and the word is been explained as well as its six mayor meanings are revealed. (RR1.G.HB.014)

It is interesting to note that there appears to be a correspondence between genre awareness on the part of the student writers and their textual borrowing. There are two texts in the corpus that are not in line with the task requirements. One mostly comments on the structure of the source text without conveying its content, while the other contains the student's own interpretation of the data presented in the source text. Both texts are also characterised by relatively little textual borrowing. These two students' misunderstanding of the task clearly had an effect on their textual borrowing because their texts have a goal that is different from a prototypical reading report. Though it is beyond the scope of this thesis to explore this aspect in more detail, it points to the genre-specific role of and need for textual borrowing.

In sum, while too many copied strings may be a sign of the writer's deficient paraphrasing skills, too little textual overlap may be due to insufficient understanding of the task requirements and result in questionable forms of intertextuality. Overall, it appears that different types of textual borrowing can be distinguished, confirming that a nuanced view of patchwriting in L2 academic writing is needed (Abasi & Akbari 2008;

Pecorari 2003). This view should also take into account the students' reasons for patchwriting. It is generally acknowledged in second language writing research that borrowing from the source text can be a strategy of appropriating expert discourse and learning genre-appropriate ways of expressing oneself (e.g. Campbell 1990; Howard 1995; Keck 2014). Such strategic re-use of source text vocabulary is discussed in the next section.

### 6.3.4.2 The source text as a language resource

In order to summarise Kettemann et al.'s (2003) research article, the students had to incorporate certain expressions from this source text into their writing. Example (6.72) illustrates the quality of lexical overlap between the source text and a particular student's paraphrase. The underlined words in this example occur throughout the source text and most are not linked to a specific excerpt, unlike in patchwritten sentences. Though the student re-used several individual words from the source, none of this textual borrowing is illegitimate. Instead, it creates an impression of a well-founded understanding of the source text and its core academic and technical vocabulary.

- (6.72 ) Regarding the OED, they describe what institutionalisation of a lexeme is, and that the filtering of words that is an outcome of institutionalisation is a core difference between a corpus, where everything that occurs in language can be recorded, and a dictionary, where only socio-culturally accepted words and meanings are listed. [...]

This causes gaps for meanings that are used widely today but are not included by the makers of the OED for different reasons. (RR1.G.HB.015)

The L2 writers re-use both academic expressions and technical terminology related to the topic in their paraphrases. The re-use of linguistic terminology is illustrated in example (6.73). Though it is impossible to know which of those terms the student was already familiar with before reading the article, the density of overlapping terminology in this example suggests that the student borrowed these expert expressions to improve their style in a similar way to the student in Stapleton's (2010) study.

- (6.73 ) While the latter contains information on token frequency, offers unfiltered results and co-texts, the former features detailed information about history and meaning of institutionalized lexems. (RR1.G.HB.012)

There is also overlap in terms of general and academic expressions, which has been noted in previous studies (e.g. Keck 2006, 2014; Gebriel & Plakans 2016). The vocabulary of the source text and that of the student's text often coincide in the use of research-related academic vocabulary. In example (6.74), the student uses two research nouns which also occur in the source text. In such cases, it is unclear whether the student consciously adopted these words from the source text or was using topic-related vocabulary from their memory, though students have attested to re-using source text vocabulary to assist and elevate their academic writing (see e.g. Leki & Carson 1997; Ruiz-Funes 1999). Certain words seem to be particularly attractive for re-use in paraphrases. For example, eight paraphrases in the corpus contain the phrase *widely accepted*, and the word *nevertheless* is found in five paraphrases of the same source text sentence in which it occurs. Apparently, certain words are more salient than others for writers of paraphrases and function as a kind of 'intertextual teddy bear' (compare Hasselgren's (1994) concept of lexical teddy bears in L2 writing). While this observation is only impressionistic, it provides an interesting point of departure for further research.

- (6.74 ) The findings are being compared by the researcher, bearing in mind the different nature of the two sources. (RR1.G.HB.011)

L2 writers probably use the source text for inspiration with regard to useful expressions throughout their reading reports, not merely when paraphrasing individual sentences. In example (6.75), the student possibly re-used expressions not only from the sentence they were paraphrasing but also from a parallel sentence in the conclusion.

(6.75 ) *Source text excerpts:*

Corpus linguists have a tendency to regard the corpus [...] as the most valuable kind of data in the study of language, sometimes forgetting that other, different forms of data are equally valuable. (Kettemann et al. 2003: 135)

[...] [A]nyone undertaking lexical studies should not ignore the vast offerings of corpus data. (Kettemann et al. 2003: 145)

*Student paraphrase:*

Simultaneously the author mentions that because of the high recognition of corpus data other different kinds of data, which are at least equally valuable as corpus data, are being ignored very often. (RR1.G.HB.011)

The re-use of source text vocabulary is naturally not always an effective strategy, as the boundaries between appropriate and inappropriate textual overlap are fuzzy. The sentence in (6.76) re-uses vocabulary from several source text sentences and does not contain longer copied strings, but due to its brevity, it is obvious that the student contributed little lexical material of their own. Because of the resulting extent of lexical overlap, the paraphrase is arguably too reminiscent of the source text. Such examples illustrate that the re-use of source text vocabulary can quickly veer into patchwriting, as discussed above.

(6.76 ) *Source text excerpt:*

The arguably more interesting difference between corpora and dictionaries, however, is that words in the latter are filtered, which means that not all words that have been found have been entered. [...] Institutionalization thus means that a lexeme has a widely accepted intersubjective status. [...] (Kettemann et al. 2003: 137)

*Student paraphrase:*

Due to the process of institutionalisation, some lexemes which are widely accepted become entered. (RR1.G.HB.092)

Textual overlap may be a sign of patchwriting in reading reports, but it is important to distinguish between appropriate and inappropriate textual borrowing. It is conceivable that the expertly selection and integration of core source text vocabulary or “general links” (Keck 2006) contributes to, if not determines the intertextual quality of effective paraphrases in L2 writing. Process approaches are needed to gauge the ways and extent to which students rely on the vocabulary and phraseology of their sources during the composition process.

## 6.4 Interim summary: Corpus study

The corpus analysis has brought to light several interesting findings with regard to the use of intertextual strategies in reading reports. The writers used material from across the source text that they mostly arranged in chronological order, following the structure of the source text. Most reading reports begin with a summary or summary paraphrase. In the body of the reading report, the students combine a range of different intertextual strategies. Paraphrasing is far more common than direct quotation and summarisation. The students rely on individual sentences for paraphrasing, but also select ideas from across the text. Direct quotation, which is

generally a peripheral phenomenon in academic writing (Petrić 2012), is comparatively infrequent. The students' strategies of creating intertextuality via paraphrases, direct quotes, and summaries are individually different, and not all texts contain all types of intertextual links. They tend to combine intertextual strategies in the same sentence, for example by embedding direct quotes in paraphrases. Exact copies are very rare. These observations point to the students' advanced academic literacy and a general understanding of expert ways of creating intertextuality.

The task of writing a reading report itself seemed to have a considerable influence on the intertextual strategies employed by these L2 writers. There is little to no evaluation, which can be explained in reference to the task description. Referencing is used sparingly, but this is not necessarily due to a lack of familiarity with citation conventions. It should rather be viewed as a text type-specific use of intertextuality that results from the fact that the reading report is based on a single source text.<sup>65</sup> In addition, its intertextual nature is made explicit in intricate ways, for example through the use of meta-level observations that guide the reader through the text and provide an impression of the source text's structure. Attribution and reporting structures are additionally used in students' texts to signal the act of writing from sources.

A further influence on intertextual strategies is exerted by the source text itself. The students apparently used the source text as a language repository and strategically extracted vocabulary to re-use in their own writing. The analysis showed that textual overlap is unavoidable and desirable to a certain extent, but that there is also some patchwriting that needs to be remedied. The structure of the source text appears to influence the students' writing behaviour. It is certainly no coincidence that many of the direct quotes and exact copies are of the definitions of the meaning of *eco-*. Furthermore, certain source text sentences seem to be preferred for paraphrasing across the learner group. The combination of information from different sections of the source text by some students suggest that though the reading reports are usually structured chronologically, the writing process may not have been equally chronological. Further analysis of students' selection strategies in the writing process is certainly needed. Several inappropriate approaches to intertextuality in the corpus indicate insecurities on the part of the students. Among others, these insecurities relate to questions of what to quote, what to place in quotation marks, and how to paraphrase effectively. Some students also appear to hold misconceptions with regard to the amount of textual borrowing that is appropriate and which linguistic changes are sufficient to move beyond patchwriting. This is evident from some students making slight changes that minimise textual overlap but do not result in an effective paraphrase. Finally, improvements can be made with regard to the strategic use of and stylistic variation in reporting structures to avoid the over-reliance on certain verbs, e.g. *explain*.

It is important to stress that many aspects of intertextuality have already been mastered by these students at the end of their first year of study. These include the syntactic integration of direct quotes, the selection of important information, and structuring this information in the reading report. Overall, the students' citation practices are relatively advanced. Apparently, the reading report is a text type that allows students to fully realise their potential in terms of writing from sources. Nevertheless, there is much variability in the corpus in terms of how academic writing conventions are implemented. More training is required with regard to effective paraphrasing strategies and avoiding patchwriting at this stage of academic literacy development. Some students are not yet aware of how to distinguish others' ideas from their own. Moreover, it is apparent from the data that some students struggled to write effective gist statements. Others based most of their paraphrases on individual sentences. It is difficult to assess whether this strategy should be avoided. It would be interesting to investigate if there is a correlation between students' grades and their preferences for gist statements or paraphrases that are based on individual sentences.

The comparison of students' reading reports to the source text indicates that the students are driven by a desire to minimise textual overlap, but it only sheds light onto the final outcome of the textual operations the students performed. The corpus study cannot give answers to the questions of how exactly the students chose passages for and direct quotation, paraphrasing, and summarisation, of how they rephrased source text material,

<sup>65</sup> See also T. A. Hyland's (2009) study for students reporting that they omitted references because of the assessment format.

and of why they copied certain passages. Despite deepening our understanding of the ways in which intertextual strategies are applied and combined in academic writing tasks, corpus studies in general cannot offer insights into the processes that lead to the manifestations of intertextuality in the products.

To triangulate the corpus-based approach, a process study of L2 writers composing reading reports was conducted. Its purpose is to investigate the micro-level processes L2 writers complete in order to create intertextuality in a source-based academic writing task. The process study aims to investigate paraphrasing and summarising as well as quoting and copying from source text excerpt. It is hoped to provide evidence for how and from where students select source text material and how they rephrase these passages during writing. It is also expected to offer interesting complementary insights into deletion, addition, and substitution strategies as well as morphological and syntactic changes that are made in the course of paraphrasing. It will also be possible to gain additional insights into students' motivations for using direct quotes and their preference for quoting the definitions from the source text by Kettemann et al. (2003). The process study will also provide an analysis of students' approaches to making intertextuality explicit by means of referencing and attribution and their choice of reporting structures. A further focus of the process study is on students' strategies of selecting source vocabulary for re-use in their reading reports, especially strategic approaches to this, and their explanations for these. The process study is presented in the following section.

## 7 Process study

This process-oriented study complements the corpus study presented above by shedding light on the intertextual writing processes of L2 writers in disciplinary assignments. It complements the corpus study by providing new insights into the phenomena observed there and offers the opportunity to triangulate this method. The experimental part of the process study consists of six first-year students of English writing a reading report in an electronic environment, which was recorded using screen recording software (see detailed overview in section 7.4.1). The recording was then used as a prompt for a stimulated recall procedure in a retrospective interview (section 7.4.2). The study is based on two main data sources, namely screen recordings of the students' writing processes and transcripts from the retrospective interviews with stimulated recall.<sup>66</sup> The study design and methodological decisions are presented in the next section, followed by a description of the pilot study and its implications for the study design in section 7.2. Section 7.3 provides information about the study participants. A detailed account of the procedure appears in section 7.4, followed by two sections which detail the annotation and analysis of the screen recordings (7.5) and the transcription and annotation of the interviews and (7.6), respectively. The results of the process study are presented and discussed in section 7.7.

### 7.1 Study design and methodological decisions

This study investigates the processes of source-based writing in an entirely electronic environment. The participants were asked to write a reading report on a computer using word processing software and accessed the source text and the instructions only on the screen. It has been reported that some students prefer to use paper to draft their assignments, especially in the planning stage, but most prefer writing on a computer. The students in Bailey and Withers' (2018) study viewed the use of materials on the same screen as one advantage of writing on a computer. For this reason, the use of an electronic environment was seen as representing a writing setting that was familiar – if not preferred – by student writers. Screen recording software was used to capture the students' entire writing process. This type of software records all on-screen actions, for example writing, note-taking, (re-)reading the source text, copying or copy-pasting, and editing. The actions that constitute the writing process are documented in the order and duration in which they take place and can subsequently be coded and characterised (Solé et al. 2013). The recordings can later be transformed into learning resources for students or teacher development materials (Bailey & Withers 2018: 177).

This approach was selected because the video created by the screen recording software allows the researcher to observe and analyse individual parts of the writing process on the micro-level and the order of their occurrence. Keystroke logging was initially considered as an alternative to screen recordings because it produces data which can be analysed statistically<sup>67</sup> (Leijten & Van Waes 2013). A computer programme running in the background records the writer's use of the keyboard and their mouse activity as well as their location and duration. Actions such as cutting, pasting, and deleting are also recorded. All activities are then output to a file, usually in XML. The data produced by keystroke logging allow the researcher to reconstruct all the actions that took place in the process of writing in minute detail (K. Hyland 2016). The data can then be used to evaluate the amount of time and the order of certain processes that take place during writing.<sup>68</sup>

66 Additional data is available in the form of the written products as well as an experiment log filled in by the students while they worked on the writing task, but this data is only used to complement the main data sources.

67 Screen recordings can also be analysed statistically, but only if they are manually annotated (see Chan 2011).

68 See e.g. Breuer (2014) for a study using keystroke logging to explore writing fluency and revision processes in L2 academic writing.

Keystroke logging has previously been used in research into intertextuality in studies of Master's students and experts writing short texts from multiple digital sources (Leijten & Van Waes 2013; Leijten et al. 2019) and of argumentative essays written for language tests (Chan 2017). Its advantage is that it is minimally obtrusive and accurately records every action of the writer with a precise time stamp so that individual differences in writing processes can be minutely captured. However, in another study of L2 writing processes (Zhao 2011), the data produced by keystroke logging to investigate writers' revision behaviour in creative writing proved too complex and an analysis of this data too time-consuming. In the end, the data were only used to facilitate transcriptions of think-aloud protocols. Keystroke logging captures each individual action the writer takes, i.e. each letter that is written or deleted, each mouse movement, and each pause that is taken. Intertextual strategies, however, which are at the focus of this study, are usually manifested in a combination of different actions, e.g. reading, writing, copying, and editing.

As a result, keystroke logging data would have been too detailed for the purpose of this study as it would have necessitated additional analysis of the data to determine and combine the different actions that constitute an instance of intertextual writing. Furthermore, keystroke logging neither records which parts of the source text the writer interacts with nor how they integrate information from it into their writing (Leijten et al. 2019). For the purpose of answering the research questions underlying this study, it was therefore considered more effective and pertinent to record the process using screen recording software, which visualises the writing process in a more user-friendly way (Bailey & Withers 2018).

It has been argued that in studies of writing processes, it should be attempted to "capture processes in a way which best reflects how authors compose in real-life situations that appreciate the context, time and technology as they are presently used" (Stapleton 2010: 305). In a naturalistic writing process, L2 writers make use of a variety of language- and content-related resources, e.g. dictionaries, thesauruses, news websites, and academic publications (see Choi 2016). Different students may have very different preferences as to the circumstances of their writing, i.e. when, in how many sessions, where, and how they would approach a task such as writing a reading report. In a truly naturalistic setting, the participants would be allowed to complete the task in their preferred setting, with the hardware of their choice, and with all the resources that are available to them in real-life contexts. The source text itself also introduces a number of variables, as students may read a printed hard copy or a PDF on screen. This alone might influence the ways in which the students refer to the source text in a writing task. Similarly, the source texts provided and the texts produced in experimental settings are typically shorter than in real-life scenarios (McCulloch 2013). Typically, students will read a number of academic research articles and textbook chapters for a source-based real-life assignment, e.g. a research paper, which influences the writing process.

However, just as classrooms are "messy contexts" (Dovey 2010: 46) with a variety of variables at play, so are the settings in which students may write their assignments. In such environments, it is difficult for the researcher to control variables such as the time and resources used for writing, which is why few researchers have attempted to capture writing processes in highly authentic settings.<sup>69</sup> For this reason, a standardised experimental setting was designed in which all participants wrote the reading report in an electronic environment under the same conditions, but had some freedom in terms of how to arrange the windows on the screen, which word processor to use, and when to take breaks. The source text variable was controlled for by instructing students to use an editable PDF on screen and not permitting the use of printed and other off-screen resources. An academic article of typical length was selected as the source text, which participants prepared at home. The length of the reading report was set to 600 words, which is shorter than a research paper, but typical for this task type in the students' degree programme.

Though screen recordings have the advantage of facilitating direct observation of the composition process, they do not reveal the writer's thought processes. Introspection can give access to information about the

---

69 For an exception see Choi 2016, who trained students to record their writing at home using video recordings, keystroke logging, and screen recordings.

writing process unavailable from observational methods such as screen recordings (Mackey & Gass 2013; see also Gass & Mackey 2000). It can thus complement observational data with information about the learners' cognitive processes and perspectives, for example as regards the types of knowledge that students access during the task and why they employ certain strategies at certain points of the writing process (but perhaps not at others). In this case, introspection can help uncover what the learners know about source use and what motivates them to use sources in a certain way. It can bring to light the learners' subjective theories and the origin of their knowledge about writing from sources (compare Gass & Mackey 2000). It can also elucidate similarities and differences between learners, especially in cases where the written product is similar.

Several introspective methods were considered to complement the experimental task. A common instrument in process studies is the think-aloud protocol, also referred to as a concurrent verbal protocol or a verbal report, which is typically used for introspection into problem-solving behaviour (Mackey & Gass 2013: 77ff.). This method requires the participants to vocalise their thought processes and internal processing. Usually, the participants are trained on a different problem before they are asked to think aloud while completing the task of interest. Participants are sometimes reminded to voice their thoughts during the procedure (see e.g. Plakans 2008; Solé et al. 2013). Think-aloud has been employed in a number of process-oriented studies into L2 writing to explore the processes of reading-into-writing tasks used in language tests for English (e.g. Gebril & Plakans 2009; Plakans 2008; Plakans & Gebril 2012). It has been shown to be useful in collecting information on what participants are thinking while writing, generating data which is not available from other methods.

A common issue with think-aloud is that it requires training and that even training does not guarantee that a participant is able to carry out the task (Gass & Mackey 2000). Furthermore, concurrent verbal protocols are problematic in high-stakes situations because they potentially distract participants from the task (Sasaki 2000). They require the students to speak while writing and to focus simultaneously on writing and on verbalising their actions. This may put participants under additional stress and make it more difficult to concentrate (McCulloch 2013). The products and processes of writing while thinking aloud are likely to be affected by the method because "any technique which imposes constraints on cognitive processes [...] is bound to influence the outcome" (Stapleton 2010: 305). Writing the reading reports for this study required the students to read and write, two processes which are cognitively demanding even without having to speak at the same time. For this reason, think-aloud was regarded as unsuitable for collecting introspective data in the present study.

Other process-oriented studies of L2 writing have used written process logs as an alternative to think-aloud protocols (see e.g. Heine et al. 2014; Li 2013; Stapleton 2010). These written logs, which are also known as integrated problem decision reports (IPDR; Heine et al. 2014), are completed by the students either while they are working on the task or afterwards. Usually process logs are used to record the different processes that take place during individual writing sessions and the time spent on each in as much detail as possible. The logs may serve to track vocabulary learning, tools used, sources consulted, issues encountered, etc. Study participants may be provided with guiding questions on how to fill in the logs (see Li 2013 for an example). Process logs can be produced in several different formats, e.g. as a narrative, by taking notes by hand or in an electronic document, or by tracking changes in the document (Heine et al. 2014).

Unfortunately, process logs, too, require additional cognitive resources. Stapleton (2010), for example, observed that the student in his case study wrote more in the log than in her paper. This suggests that working on two concurrent writing tasks – an assignment and a process log – has an impact on the writing process itself. The quality of the assignment may be influenced by the cognitive demands of recording all composing behaviours simultaneously. Conversely, the quality of the process log may also suffer because the students' focus is on the task and because certain processes and thoughts are difficult to describe. The accuracy of descriptions is difficult to assess, especially since some processes may be described in more detail than others and some may be over- or understated by the writer. Process questionnaires are an alternative to process logs. They are conducted after writing and thus have the advantage of not posing an additional cognitive burden during writing, though they

may be unreliable because writers do not necessarily report their source use strategies accurately, especially those that may be considered inappropriate (see Gebril & Plakans 2009: 64f.).

Based on the considerations outlined above, this study used a combination of introspective methods in order to complement the screen recordings. A retrospective interview with stimulated recall served to facilitate the interpretation of the process data by eliciting the students' verbalisation of their thought processes when watching the screen recording. A set of open questions in the interview focused on various aspects of the process of writing from sources, including the reasons for using direct quotes, the use of paraphrasing strategies, the reuse of source text vocabulary, and the role of the source text. They were designed based on findings from research into L2 writing processes (see section 4.3) and findings from the pilot studies (Wiemeyer 2017a, b, c, 2018, 2019). In addition, an experiment log was given to the students to fill in during the writing task and to complete afterwards, which was to be used as an additional prompt for the stimulated recall procedure.

Stimulated recall is an introspective method used to explore the strategies the participants used in a particular task and the ideas, motivations, decisions, and problems associated with the event. In stimulated recall, the learners are asked to introspect based on a prompt, for example a video or a text that they have produced. The purpose of the stimulus is to reactivate and support the participants' recollection of what they were thinking when working on the respective task so that they can report their thought processes accurately (Mackey & Gass 2013: 79). This ensures that the participants' comments relate to the specific event and not their general strategies and attitudes (Gass & Mackey 2000). The prompt is assumed to enhance the participant's access to memory structures that were created during the event. Gass and Mackey (2000: 4) argue that the fact that a stimulus is used makes stimulated recall more reliable than unstimulated self-reports, because thought processes from the past are difficult to access when the subject has to rely solely on their memory without some sort of aid to the recollection.

The reliability of the recall is increased if the stimulus is strong and allows the participant to retrieve an exact memory of the situation. In order for the recall to be as accurate as possible, it is furthermore important that the data are collected as soon as possible after the event, ideally within 48 hours, so that the participants still have access to the event in their short-term memory (Gass & Mackey 2000; Ruiz-Funes 1999). A short interval between the event and the stimulated recall also decreases the likelihood of recall interference from the long-term memory. If too much time elapses between the event and the recall, the recall becomes increasingly unreliable:

As the event becomes more distant in time and memory, there is a greater chance that participants may say what they think the researcher wants them to or may create a plausible explanation for themselves because the event is less sharply focused in their memories. (Gass & Mackey 2000: 54)

Usually, stimulated recall is accompanied by minimal instructions and the participants are given as much freedom as possible to initiate the recall and select the episodes they want to relate (Mackey & Gass 2013: 78f.). The instructions have to be formulated carefully and provided by the researcher in such a way that they "help participants provide recall comments without challenging their preconceived notions of appropriateness and without leading them" (Gass & Mackey 2000: 61). Stimulated recall procedures are not concurrent to the writing process and thus cannot capture the thought processes during writing. However, unlike think-aloud protocols, they do not distract from the writing task.

According to Gass and Mackey (2000), the coded protocols from stimulated recall procedures are particularly useful for identifying trends, explaining variation in a particular group, and uncovering the reasons behind unexpected results. Stimulated recall can also bring to light extra-experimental factors that influenced the participants' behaviour and strategies (Mackey & Gass 2013). In this study, using a relatively structured stimulated recall procedure that includes general interview questions has the additional advantage of allowing the researcher to ask questions about the students' source use strategies and thus focus the recall on specific aspects of interest. This is not possible in other designs, e.g. using think-aloud protocols or unstructured

stimulated recall. These two methods may produce data that are not conducive to the research question. However, if the researcher selects the focus of the recall, participants are more susceptible to researcher interference. The quality of the recall can be increased by permitting both the researcher and the student to initiate the recall by selecting episodes for comment (Gass & Mackey 2000), which is the approach taken in this study.

In L2 writing research, stimulated recall has been employed in studies of composing processes both cross-sectionally (Bosher 1998; Rose 1984) and longitudinally (Choi 2016; Sasaki 2004). It has also provided insights into writer's cognitive activities and their language awareness (Lindgren 2005; Lindgren & Sullivan 2003) as well as writing fluency, planning, and pausing behaviour (Sasaki 2000). In studies of source-based writing, paraphrasing strategies (McInnis 2009) and reading-into-writing processes (Ruiz-Furries 1999) as well as the cognitive validity of integrated writing tasks in language testing (Chan 2011) have been analysed on the basis of data from stimulated recall protocols. Prompts used for the stimulated recall were video recordings (Bosher 1998; Rose 1984; Sasaki 2000, 2004), keystroke logs (Lindgren 2005; Lindgren & Sullivan 2003; McInnis 2009) or a combination of the two (Choi 2016). Screen recordings (Chan 2011) and process logs (Ruiz-Funes 1999) have also been employed as stimuli.

The present process study was carefully designed so as to fulfil the three conditions that warrant reliability in stimulated recall procedures (see Lindgren 2005). The first condition to be met is a relatively short interval between the writing session and the recall, in this case a maximum of 48 hours. Second, the writing session is to be recreated in some way, which was achieved by replaying the screen recording of the students' writing process from the experiment. Third, the recall is to be structured by the use of carefully chosen prompts that are provided if necessary. In this study, aspects of interest (AOIs) identified by the researcher in each screen recording before the interview served as stimuli for the recall (see further discussion in section 7.4.2). The recording was considered a strong stimulus (Mackey & Gass 2013) because the students were able to review the process and did not have to rely on their memory. In addition, Sullivan and Lindgren (2002) have conveyed that participants enjoy watching a recording of their writing process and find this procedure useful.

The combination of screen recordings and stimulated recall is highly suited to investigations of intertextual writing processes because these methods offer valuable insights into both the processes that constitute writing from sources in disciplinary assignments and the cognitive processes and strategies employed by the students. Petrić (2015) has recently called for studies which place their focus on the context of L2 writers' source use by asking participants to comment on specific instances of intertextuality in their own texts. Retrospective interviews with stimulated recall elicit statements that are anchored in concrete examples of intertextuality and expand on the insights provided by screen recordings. They thus yield data that are both specific and contextualised and are suited to shedding light on the multiple dimensions of source use.

The potential of text-based interviews in L2 writing research has been vividly illustrated by studies into alleged cases of plagiarism showing that students were in fact willing to disclose their sources and had no intention to deceive (e.g. Pecorari 2003; Shi 2012). Interview and stimulated recall data can be a beneficial resource for explaining L2 writers' intertextual strategies and textual borrowing practices (Shi 2004), especially if combined with textual and process analyses. In addition, interviews with students about their source use have been suggested to have pedagogical value for the interviewees themselves because they help them refine their source use strategies (T. A. Hyland 2009).

As yet, this combination of methods has not been employed to study intertextuality in L2 writing. That combining screen recordings with stimulated recall is a worthwhile endeavour was shown by Chan (2011) in the only study in an L2 writing context that has used these methods. Chan used screen recordings of students taking a high-stakes test of academic English and the students' comments from the recall to evaluate the cognitive validity of test items. The stimulated recall took place immediately after the writing task. Chan (2011: 18) contends that this non-intrusive methodology can be useful for investigating cognitive processes of writing. Because the software runs in the background, screen recordings allow the researcher to collect real-time data of

the writing process without distracting writers from it. They provide visual data of the writing process unavailable from keystroke logging. This is thus deemed to be a worthwhile approach so as to gain insights that are complementary to existing process studies of source-based writing that have used different methods. Table 8 gives an overview of the data collection instruments. The process study was piloted in May 2017. The pilot and the changes to the study design it precipitated are presented in section 7.2.

Table 8: Instruments used and data collected in process study.

Instrument	Data collected
Experiment	Screen recording Written product (reading report) Experiment log + participant metadata from learner questionnaire + result of language proficiency test
Retrospective interview	Stimulated recall data based on screen recording (and experiment logs) Responses to general interview questions

## 7.2 Pilot study

In order to ensure that the research design was technically and methodologically sound, a pilot study was conducted at the University of Bremen for both the experiment and the retrospective interview. Its aims were to ensure that the instructions were clear, to determine the time required to complete the task, and to assess the suitability of the source text. Initially, a test run of the experiment and the interview was completed in April 2017 with one student to test the technical aspects and general setup of the study. The student, ‘Nadja’<sup>70</sup>, was sent an e-mail with a source text – a research article on a sociolinguistic focus-group study of Māori English by Jeanette King (1999) – along with instructions and a short introduction to the topic. She was given one week to prepare. Her questions and comments on the consent form and the task description were used to fine-tune their wording for the pilot study. The student was instructed not to use any additional content-related resources during writing, e.g. other articles on the topic or Wikipedia. She was, however, allowed to use language resources, e.g. online dictionaries, thesauruses, and corpora. The writing task was completed on a computer so that the entire writing process could be recorded on screen.

The writing process was successfully recorded using the free software *oCam* (OhSoft 2017). Nadja was asked to complete a log with issues and questions during writing, but did not experience any problems. It took her a total of 1 hour and 35 minutes to complete the reading report. On this basis, the task design was considered appropriate for an experimental setting in terms of duration. During the entire writing process, Nadja consulted an online dictionary only once. For one other language-related issue, she searched the source text for help. One week after the experiment, Nadja participated in a retrospective interview with stimulated recall. The test interview lasted 55 minutes, which was within the target maximum length of one hour. After the test run, some of the interview questions were reworded and two additional questions regarding lexical strategies were added. In order to gauge whether Nadja’s behaviour was representative of that of other students, it was decided to compare groups of students that were allowed to use language resources with groups that were not. The goal was to make the task as naturalistic as possible, which would entail permitting digital language resources such as online dictionaries and thesauruses. However, because it was feared that the students would make excessive use

<sup>70</sup> All student names are pseudonyms.

of this option, which would affect the way they restated the source text, it was decided to observe this variable in the pilot study.

The pilot study was conducted in May 2017, before the corpus data had been analysed. An open call for participants (see Appendix 3) was circulated via e-mail in an introductory linguistics class. It contained general information, a registration link, and a learner questionnaire (see Appendix 5) to be submitted by future participants within two days. Twelve students of English in their second semester participated in the pilot. One week before the session, each group of participants was sent the respective source text as an editable PDF along with a general description of the experiment, instructions for reading the source text, and a short introduction to the topic, including definitions for key terminology (see Appendix 4).

The participants were subdivided into five groups so as to determine a source text for the process study and decide whether or not to allow the use of additional resources. Two versions of the instructions for writing a reading report were created (see Appendices 11 and 12 for instructions for writing a reading report with and without language resources). The groups with access to language resources were not allowed to use printed materials, but only electronic resources which could be captured by the screen recording software (see also Bailey & Withers 2018). In addition, all students were given a choice of two different word-processing programmes, namely *Microsoft Word* and *LibreOffice Writer*. Coincidentally, all twelve participants of the pilot study identified as female. As gender was not a variable of interest, this was considered irrelevant to the validity of the pilot.

The source text for the process study was to be selected from those source texts for which reading reports were available in the CALE so as to generate process data which was maximally comparable to the existing corpus data. Plakans and Gebril (2013: 228) have asserted that “the source materials used in an integrated task clearly affect the way L2 writers approach the task and how they include these materials in their writing”. For this reason, the source text had to be carefully selected for the task in question and its expected outcome. In accordance with the study design, the source text had to be an authentic, published academic research article. Initially, only two source texts (King 1999; Kjellmer 2003) were to be piloted. However, the students in the first session commented that King’s (1999) article was easy to read, but too long, and Kjellmer’s (2003) was too technical. In addition, the reading reports based on King 1999 in the CALE had been written by second-year, not first-year students. For these reasons, a third source text (Kettemann et al. 2003) was piloted that had also been used as the basis for reading reports in the CALE and could be directly compared. The text contained an abstract and would offer an interesting opportunity to observe how this would be used by students in the writing process. This text also matched the topics that the students participating in the process study were currently taught in their introductory linguistics course. An overview of all five groups of participants is provided in Table 9.

Table 9: Overview of pilot study groups.

Group	Participants (pseudonyms)	Source text	Use of language resources
A	Bea, Bianca, Katharina	King 1999	Yes
B	Charlotte, Paula, Svenja	King 1999	No
C	Sarah, Vera	Kjellmer 2003	Yes
D	Anna, Cynthia	Kjellmer 2003	No
E	Miriam, Lena	Kettemann et al. 2003	Yes

The three source texts differed in several respects, e.g. length, topic, research method, use of figures, and type-token ratio (see Table 10). As all three had previously been used in reading reports in linguistics classes at the same university, they were all considered appropriate for the learner group in terms of difficulty and topic.

Table 10: Overview of piloted source texts.

Text	Research area	Topic	Method	No. of words <sup>71</sup>	Type/token ratio	Pages of text	No. of figures and tables	Lexical density
King 1999	Socio-linguistics/ varieties of English	Functions and aspects of use of Maori English	Focus group discussion	6,261	0.2	17	0	0.55
Kjellmer 2003	Semantics	Near-synonymy in adverbs	Corpus analysis	1,067	0.33	8	4	0.56
Kettemann et al. 2003	Word-formation	Productivity and institutionalisation of novel morphemes	Comparative analysis of corpus and dictionary data	4,719	0.22	12	6	0.55

Neither the task of writing a reading report itself nor the corresponding task description were piloted for this study because both had been in use as a disciplinary assignment in English linguistics courses at the University of Bremen for several years. This task thus represented an assignment that students enrolled in the B.A. programme were likely to encounter in the course of their studies.

At the beginning of the piloting session, the students were given general instructions (Appendix 8), signed the consent form (Appendix 9), and were provided with an experiment log (Appendix 10) for noting down problems and questions during the completion of the task. After they had read the instructions, the researcher briefly sat with each student to give them time to ask questions.<sup>72</sup> Each student was then given access to their desktop folder with the documents for the study, the screen recording was started, and they commenced working on the task. When the students had completed their text, the recording software was turned off and the students' experiment blogs were briefly discussed. They were then asked if they were willing to participate in a voluntary interview to take place several days later. Finally, they were asked to complete the *Oxford Quick Placement Test* (OQPT; Syndicate 2001) to assess their language proficiency.

Technical issues encountered by the students could be resolved with the help of a lab assistant. The students positively evaluated the explanations and instructions provided via e-mail and at the beginning of the experiment, which they found easy to understand. Several students reported that because they were used to working on a Mac at home, it took them some time to adjust to the different keyboard and software. This issue had not been anticipated and a sentence was added to the call for participants explaining that they needed to be familiar with working on a Windows PC. Other general comments related to the task and the setting. One student suggested providing printed dictionaries, not just online resources. Because the use of books would have necessitated videotaping of the writing process, however, this could not be considered.

As it was difficult to estimate to which extent the students relied on vocabulary they found in dictionaries when paraphrasing, the use of additional resources was tested in order to decide whether to look at intertextual strategies 'in isolation' as they occur when the students have to rely on their own vocabulary or as they occur in a more naturalistic setting. A superficial analysis of the two groups' writing processes revealed that they were very similar. Those students who were permitted to consult dictionaries, thesauruses, and other language-related tools did so scarcely during writing. The resources they consulted were online dictionaries (both monolingual English and bilingual German-English), online thesauruses, and *Microsoft Word* tools such as a thesaurus and a German-English dictionary. The students thus appeared to be similarly "digi-literate" as those in Bailey and Withers'

71 The number of words (excluding tables, figures, and references), type/token ratio, and lexical density were calculated using the Compleat Lexical Tutor's *VocabProfile* (<http://www.lextutor.ca/vp/eng/>).

72 Only two students had questions prior to writing, both of which concerned the formatting of the reading report. They were advised to approach the task as they would if it were a real-life assignment.

(2018: 182) study. Only one student made extensive use of language resources. Another student reported struggling with the task because she could not use language resources. In consideration of these observations, it was decided to allow the use of language resources in the process study.

The students who had been given the text by Kjellmer (2003) finished writing their reading report after an average of 67 minutes. The students working on King's (1999) and Kettemann et al.'s (2003) research articles took longer, with an average of 88 minutes for each text (see Appendix 6 for an overview of the pilot participants' writing time, use of resources, and OQPT results). Based on these findings and the availability of reading reports in the CALE, the research article by Kettemann et al. (2003) was selected as the source text for the process study.

Seven participants volunteered to take part in a retrospective interview. Four of those students had written their reading report on King (1999), one on Kjellmer (2003), and two on Kettemann et al. (2003). Before the interview, each screen recording was analysed by the researcher to identify Aspects of Interest (AOIs) to be used as cues for the stimulated recall (see Appendix 18). Despite the fact that all the participants and the interviewer were native speakers of German, English was chosen as the medium of the interviews. This was because all participants were students of English and thus proficient in the language. English is also the general language of instruction at the department. Furthermore, since both the reading and the writing had taken place in English, it seemed more natural to interview the students in English so as to avoid having to switch between languages when talking about their reading reports or source text excerpts.

It is acknowledged that conducting stimulated recall in an L2 may influence the quality of the data. For example, it has previously been argued that the complexity of answers could be reduced in the L2 (McCulloch 2013). Mackey, Gass, and McDonough (2000) observed that the EFL learners in their study produced considerably fewer words per recall and were more likely not to comment than the native-speaker participants. As pointed out by Gass and Mackey (2000), it is possible that a limited knowledge of the L2 may cause participants to omit those thoughts that they are not able to verbalise. However, because the OQPT results confirmed the relatively high proficiency of the participants (see Appendix 6), this was considered unlikely in the present study. Piloting confirmed that completing the recall in English was within the linguistic competence of students in this cohort.

The first interview brought to light that it was not sufficiently clear to the student which aspects of the writing process she was supposed to comment on, so the instructions given at the beginning of the interview were expanded. Based on the observation that some students did not comment on their process unless asked and sometimes gave very short answers, the instructions were also amended to include a request for speaking as much as possible. Nevertheless, the pilot study showed that the students did not require training to be able to provide recalls.

A recurrent issue in the interviews was related to the questions regarding the students' use of source text vocabulary. Apparently, their fear of committing plagiarism sometimes meant that they denied re-using source text vocabulary despite having done so. To assure the students that plagiarism was not of interest here, the respective questions were reworded carefully to assure them that language re-use was a normal phenomenon in intertextual writing. Several questions were also collapsed to reduce the total number used in the interview. Two questions were removed after piloting, namely one about the students' thoughts on the appropriate way to complete the task, and their suggestions for teaching. The first aspect was usually discussed when the students broached the issue of plagiarism, the latter emerged as too speculative. The final interview questions appear in Appendix 17.

Following the completion of the pilot study and adjustments to the experiment design, participants were recruited for the main study. The recruitment procedure and participant characteristics are described in the following section.

### 7.3 Study participants

The process study was conducted at the University of Bremen in June 2017. The participants were recruited following the same procedure as in the pilot study (see section 7.2). Twenty students were originally recruited, of which seventeen actually participated in the experiment. Three students had to be removed from the sample after they had completed the experiment.<sup>73</sup> The students were recruited from an introductory linguistics module for which they had to gain experiment credits and were enrolled in one of three parallel sections of the seminar “Introduction to English Linguistics 2: Research Methods”. The course is an integral component of a compulsory introductory module in English linguistics which all B.A. students enrolled in a degree programme in English have to take. In order to complete the module, each student has to earn a total of three experiment credits worth 30 minutes each which can be earned in up to three linguistic experiments offered at the department. The experiment and interview conducted for this study were offered in exchange for three experiment credits; students had to complete both in order to be awarded the credits.

Several calls for participants (see Appendix 3) were circulated via the university’s e-learning platform. Each call was sent out nine days prior to the respective experiment session. The text of the call was standardised and contained information about the researcher, the structure and time frame of the experiment and the interview, and details on registration. The recipients were also told that they would be sent a source text in another e-mail upon registration which they would have to read and which would form the basis of a writing task in the experiment. It was not revealed to the students at that point that the writing process would be recorded or that the study focused on intertextual strategies. The description of the writing task was intentionally kept vague so that participants would be unable to pre-write the reading report.

Potential participants had to register for the experiment and an interview slot online and submit a learner questionnaire. The learner questionnaire (see Appendix 5) was closely based on the questionnaire used for data collection for the CALE (Callies & Zaytseva 2013). It elicited general autobiographic data as well as detailed information about the students’ language learning biographies, e.g. foreign languages learnt, amount of time spent abroad, amount of time generally spent speaking English, etc. One additional question (item 8) was introduced to collect information about the students’ previous academic writing experience in English studies. The question was included because writing expertise has been found to have an impact on the way L2 student writers approach writing tasks and the quality of their writing (Bosher 1998; Petrić 2012; Shi 2012). Studies have shown that the more advanced the L2 proficiency and writing expertise are, the more varied are students’ writing strategies. It has been argued that writing expertise may determine writing strategies as well as organisation and quality of the text more centrally than proficiency in the L2 (see discussion in Plakans 2008). For the design of this question, several faculty members had been asked to compile a list of text types that were typically written in the first year of the B.A. programme. These text types were listed as options in item 8. An ‘other’ option was provided so that the participants could list additional text types.

The *Oxford Quick Placement Tests* (OQPT; Syndicate 2001) completed by the students (see section 7.4.1) showed that their English language proficiency ranged from ‘lower intermediate’ (39 points) to ‘very advanced’ (57 points) with an average of 48 out of 60 points ( $SD = 7.62$ ). This attests to considerable variation in the participant group, which is reflective of the heterogeneity of student cohorts enrolled in English studies. In order to control for language proficiency, which has previously been found to influence source text use (Cumming et al. 2005; Gebril & Plakans 2009; Gebril & Plakans 2012), only participants whose OQPT level was “Advanced” were selected for the study (see results in Appendix 7). External assessment was deemed necessary because it has been suggested that source use may have an influence on students’ grades (Gebril & Plakans 2016; Plakans & Gebril 2013; Petrić 2012). An assessment of language proficiency based on the students’ reading reports was therefore not a viable option, especially since they were to be composed in an experimental setting.

<sup>73</sup> One student explained in the interview that she had understood the instructions to mean that she was not allowed to use the source text, which rendered her writing process and product unnatural, the second submitted an incomplete OQPT test so that her language proficiency could not be assessed, and the third was not enrolled in a B.A. but in a Master’s programme.

The final sample consisted of six participants. Sample sizes in studies of L2 writers' composing processes are generally relatively small. They range from as few as one and three participants in longitudinal mixed-methods studies (Choi 2016; Stapleton 2010) to 23 participants in a study using only think-aloud protocols (Cumming 1989, cited in Sasaki 2000). Considering the mixed-methods approach taken here, which used a combination of screen recordings, experiment logs, and retrospective interviews with stimulated recall, a cohort of six was considered a sufficiently large, but manageable sample for the purpose of this study.

All participants were first-year students in their second semester of study. Four identified as female, two as male. To protect their identities, each participant was assigned a pseudonym and a learner code. At the time of data collection, all were enrolled in the B.A. English-Speaking Cultures in a double Bachelor's programme, and one student had opted for the English language teaching option. The students were born between 1994 and 1998 and all are native speakers of German. One indicated Arabic as a second L1. It was ensured that all students had received their primary and secondary school education in Germany to pre-empt cultural influences on the writing process and recall behaviour. The students had been learning English for 11 to 14 years, one of which was at university. All had learnt at least one other foreign language, namely French (5), Spanish (3), Italian (2), Latin (1), and Arabic (1). The majority (4) named English as their favourite foreign language. Only two of the six students had spent an extended period of time in an English-speaking country, namely six to twelve months as au-pairs in Great Britain and Ireland. None had completed any of their tertiary education in an English-speaking country.

The students had all written summaries and essays prior to the experiment. Four had already written a reading report. Only one student had had to write a research paper and another a review. Overall, the students indicated that they had experience in writing two to four different academic text types in English. Appendix 7 contains the metadata and OQPT results for the six study participants. A detailed description of the experiment and retrospective interviews is given in the following section.

## 7.4 Data collection procedure

The process data were collected in two different settings. First, the participants were asked to write a reading report in an experimental setting while their process was recorded using screen recording software. They also filled in an experiment log during the writing session. Second, retrospective interviews were conducted with the participants one to two days after the experiment. The interviews combined a stimulated recall procedure based on the recordings of the writing process with general questions about source use. Sections 7.4.1 and 7.4.2 provide a detailed description of the data collection procedures for the experiment and the retrospective interviews respectively.

### 7.4.1 Screen recording

The experiment was conducted in five sessions over the course of three weeks in June and July of 2017. Two to five students took part in the experiment per session. One week prior to a session, the participants were sent an e-mail with the source text attached as an editable PDF. They were informed that they would be required to write a text based on the source text in the experiment. The students were not told about the exact nature of the writing task to prevent them from pre-writing the text at home, which could result in an unnatural writing process in the experiment. The e-mail also informed them that having read the source text was a prerequisite for participation in the experiment (see Appendix 5). The participants were thus able to read the article in a setting of

their choice and at their own pace using their preferred reading strategies. It also gave them time to look up words and re-read if they felt the need (see McInnis 2009 for a similar approach).<sup>74</sup>

The source text was the study on word-formation by Kettemann et al. (2003) which had also been used in the corpus study (see section 6) and tested in the pilot study. The authors used the morpheme *eco-* as an example to compare corpus and dictionary data in terms of their usefulness for lexical studies. The paper was not edited for length, vocabulary or content because students of English are generally expected to read published academic research articles in their original state. The source text used here is thus considerably longer and more complex than those used in previous studies (see McCulloch 2013). However, in linguistics courses at the University of Bremen, students regularly read research articles and handbook chapters, and reading reports are typically based on such academic texts. A shortened, non-academic source text would have constituted a relatively unauthentic basis for writing a reading report and would have negatively impacted the comparability to the corpus data.<sup>75</sup>

All participants had received instruction on the use and purpose of corpora and dictionaries in linguistic research as well as their differences, which were at the core of the study by Kettemann et al. (2003). They had all previously completed an introductory class in English linguistics and were familiar with the basic principles of word-formation and linguistic terminology related to the source text topic. This also made this group exactly comparable to the group which had written the texts in the corpus in terms of their previous linguistic education. A short introductory text was provided in the e-mail to clarify the context of the study and key terminology so as to prevent misunderstandings (see Appendix 4).

At the beginning of the experiment session, each student was assigned a computer and the researcher read the general instructions out loud (see Appendix 8). The students were then given time to read the experiment logs (see Appendix 10) which had been placed on their desks prior to the experiment. The experiment logs consisted of three general questions regarding 1) their experience of reading the source text, 2) the clarity of the instructions, and 3) their experience of writing the reading report. In addition, a table was provided in which the students were asked to record anything they noticed about the way they worked with the source text, e.g. decisions they made or problems they encountered. The table also offered space to note down the participants' solutions and suggestions. The students then signed the consent form (see Appendix 9) and read the task description for writing a reading report (see Appendix 11).

After their remaining questions had been answered, each student was given access to their computer. To ensure uniform experiment conditions, all the documents that they needed to complete the task were provided to them electronically in one folder on the desktop: an editable PDF of Kettemann et al.'s (2003) article, a PDF of the task description, and a choice of either a *Microsoft Word* document or a *LibreOffice* document (see Chan 2017 for a similar procedure). Both the *LibreOffice* and the *Word* document were minimally formatted (12 pt. Times New Roman, justified) and the language was set to English. The pseudonym assigned to the student served as the file name for both. The document of their choice was opened and the students were told that they were allowed to re-arrange the windows as they liked. Finally, the screen recording software *oCam* (OhSoft 2017) was set to record and the students started working on the task.

The students were allowed to take breaks at any time during the experiment. In such cases, the screen recording was paused and resumed upon their return. At regular intervals, the students were asked to save their documents. When they indicated to the researcher that they had finished writing, they were asked to proofread their texts and to complete the experiment logs. Afterwards, they were asked to step outside with the researcher and to summarise what they had written on the experiment log. The researcher took notes and asked for clarifications where needed. The students then completed the *Oxford Quick Placement Test* (Syndicate 2001).

74 As the participants prepared the text at home, the processes of reading and planning were not analysed in this study, but see e.g. Plakans (2008) for a study taking the pre-writing processes into account.

75 Other researchers who studied the reading-into-writing processes of L2 writers (e.g. Gebril & Plakans 2009; McInnis 2009; Plakans 2008) have given their participants edited texts from published, but not necessarily academic sources. Because the students in those studies were required to read the texts during the experiment, they were relatively short. Plakans and Gebril (2012), for example, used two source texts of less than 300 words.

## 7.4.2 Retrospective interviews with stimulated recall

After the experiment, each student participated in a retrospective interview in which they were asked to recall their writing process. The purpose of the interview was to explore the students' thought processes in order to gain insights into the problems they had encountered and their strategies of overcoming them and thus provide a more complete picture of the intertextual strategies of L2 student writers as well as differences between them.

The hour-long interview used a structured stimulated recall methodology that combined recall prompts from the screen recordings with questions on source use. This procedure was intended to elicit comments on the process of writing a reading report as well as on source-based writing more generally. The employment of a stimulated recall methodology stipulated that the interviews be conducted shortly after the writing session (within 48 hours) to prevent recall interference when participants have to rely on their long-term memory (see Gass & Mackey 2000; Mackey & Gass 2013; Ruiz-Funes 1999). The interview was usually scheduled for the day following the experiment, in exceptional cases for two days later. 24 to 48 hours were regarded an acceptable amount of intervening time vis à vis the preparation time for each interview. This interval still warranted a high reliability of the recalls, especially because it is assumed that the majority of memory structures are eliminated immediately following the event so that recalls which are conducted three hours after the event are likely to be as reliable as later recalls (Cohen 1987; Bloom 1954, both cited in Gass & Mackey 2000). An immediate recall after the writing session was deemed too cognitively demanding. Furthermore, a choice was made to analyse the recordings prior to the interview to be able to ask specific questions on the source use strategies observed, which necessitated an interval.

In preparation for each interview, the researcher read the student's reading report and marked passages that were considered interesting to observe during the stimulated recall. Each text was also analysed using *Wcopyfind 4.1.5* (Bloomfield 2016), an open-source plagiarism detection software which cross-checks the students' with the source text to identify identical strings of words. The texts were checked for strings of two words or longer<sup>76</sup>. Passages which were copied from the source text but were not terminology or fixed expressions were highlighted in the students' texts. The purpose of this was to identify instances of copying from the source text – intentional or unintentional – which were not immediately apparent from the screen recording as well as to gain a general idea of their reliance on source text lexis to be able to ask students about this in the interview. This analysis also revealed that all six reading reports were in line with the task description and thus generally resembled those analysed for the corpus study. The reading reports written in the experiment were nonetheless not analysed in the same way as those in the corpus study, which had been composed in a naturalistic setting, because it was likely that the setting and time constraint had an effect on the product. However, they provided important cues for the stimulated recall of the writing process, which was of course at the core of the process study.

The screen recording of the student's writing process was subsequently analysed to identify aspects of interest (AOIs) in terms of source use which were to serve as prompts during the stimulated recall session (see Tyler 1995 for a similar procedure). The time stamps for each of these AOIs were recorded along with a short description. They were used as cues for the stimulated recall (see example in Figure 15 and a list of all AOIs by study participant in Appendix 19). The preparation for each interview took approximately 1 to 1.5 hours. The instructions, AOIs, and interview questions were printed on colour-coded cards and laid out on the desk in front of the computer screen. This made it possible to flexibly decide on the order of questions and place those aside that had already been answered.

The researcher and the student sat side by side facing the computer screen (see Rose 1984 for a similar procedure). Before the microphone was turned on, the students were asked to sign an interview consent form (see Appendix 13). At the beginning of the interview, a brief standardised introduction and instructions were read

<sup>76</sup> This setting was chosen because *Wcopyfind 4.1.5* did not reliably identify strings of three words or longer and often subdivided them into shorter strings. This would have necessitated additional manual comparison of matches between the student's and the source text to determine the actual length of the copied passage.

to the student to explain the procedure and what the focus of the interview would be (see Appendix 14). The interviews and stimulated recall procedures were conducted in English, which proved unproblematic given the participants' high proficiency in English. The participants did not receive training in stimulated recall procedures.

<p><i>Miriam</i></p> <p>00:30 – 02:00 heading w/ copy-paste, copying</p> <p>07:00 – 11:15 reads ST, PP from headings, text</p> <p>12:00 – 13:05 scrolling, meta about structure</p> <p>13:30 – 15:20 PP with DQ ([...], ref)</p> <p>16:00 – 18:00 synonym search for ST term</p> <p>18:00 – 21:45 close PP with synonym repl., summarisation</p> <p>22:40 – 27:20 PP with restructuring, DQ, reference</p> <p>29:10 – 32:20 marks ST, checks toolbox, PP with copying</p> <p>41:30 – 43:10 PP from memory, then ST</p> <p>48:25 – 51:25 DQ of defs, looks up reporting verb</p> <p>1.07:50 – 1.12:35 PP from Conclusion, dict use</p> <p>1.24:30 – 1.25:00 searches text for words</p>
--

Figure 15: Example of a flashcard with aspects of interest for stimulated recall.

Each interview began with three intentionally simple open questions (Appendix 15) in order to put the participants at ease and allow them to settle into the interview situation. The first introductory question was intended to break the ice by prompting the students to explain what the writing session had been like for them. The second question enquired about how they had prepared the source text at home. The purpose of this question was to establish which strategies the students had used, e.g. whether they had taken notes or highlighted sections in the source text, and how thoroughly they had prepared the text. The third question elicited the students' own perception of their overall writing process by asking them to describe how they wrote the reading report in as much detail as possible. The goal was to gain insights into the students' own perspective of their writing process that could later be compared to the analysis of the screen recordings.

Once these three questions had been answered, the interviewer gave additional instructions for the stimulated recall procedure (Appendix 16) and encouraged the students to comment on their own process whenever they saw or remembered something they found noteworthy. Giving the participants the chance to voice their thoughts on sequences they recalled as important reduces researcher interference (Mackey & Gass 2013: 78) and increases the completeness of the recall data. Then, the recording was played and the student and the researcher watched the entire recording of the writing session together following an approach by Heine and Knorr (2015). For reasons of time, the recording was played at a faster speed, usually three times the actual speed. This speed made it possible to watch the entire recording within an hour while still easily being able to observe the movements on the screen.

Unless the beginning of the student's writing process was unique in some respect, the participants were initially given time to adjust to watching the screen recording and to volunteer their own comments. This approach was taken because the student in the test run had explained that not being asked questions right away made it easier to get used to watching the recording and recalling the writing session. It was attempted to avoid discussions as well as extended responses on the part of the interviewer so the student's comments in the recall would not be influenced (see Gass & Mackey 2000). The researcher did not intervene if the student recalled something that was not of interest to the present study. While watching the recording, the student was encouraged to comment on the AOIs previously identified. The student was prompted to explain what they had

done in that particular instant by asking questions such as “Could you tell me what you were doing here?” and “Why did you decide to do X here?”. Following Shi (2008), a conscious choice was made not to mention the term ‘plagiarism’ unless the students did so first in order to avoid bias in their answers. Sometimes the researcher briefly summarised what had happened in the scene being watched and prior to it to provide a frame for comments. During the recall of an AOI, the video was occasionally slowed down to the original speed, rewound to re-watch a sequence or paused to avoid distraction and later resumed. Between questions, there was always time to continue watching and give the students time for their own comments.

At appropriate times, the stimulated recall was interwoven with general questions about the students’ recourse to the source text and intertextual strategies to gather general information on tendencies in their behaviour and to identify trends in the group. The interview questions were only introduced once the participant had familiarised themselves with the stimulated recall procedure. Because they required the participants to abstract from the writing process to a certain extent (cf. Gass & Mackey 2000: 35), the questions were posed at appropriate times during the interview, for example when there was a pause or when the student mentioned one of the intertextual phenomena covered by the questions. The interview questions were grouped into three categories and were informed by the pilot study and previous research. The first category, intertextual strategies, is comprised of questions eliciting the students’ motivations to use manifestations of intertextuality such as direct quotes, paraphrases, and references. Their purpose was to identify the students’ existing knowledge and subjective theories about when, why, and how these types of intertextuality should be used. Questions in the second category focused on lexical strategies based on the observation made in previous research that L2 writers tend to strategically re-use source text vocabulary (e.g. Gebril & Plakans 2012; Wiemeyer 2017b). These questions were intended to elicit comments on how and why the students selected source text vocabulary for re-use, if at all. Finally, the third category of questions was aimed at collecting information about students’ recourse to the source text, i.e. what the students looked for when re-reading the source text and how they generally decided which source text material to include into their reading reports. There was no particular order to the questions in the interviews.

It was attempted to keep the conversation flowing as naturally as possible and to touch upon a range of aspects of intertextuality observable in the recording. If the student commented on an aspect of intertextuality that had not yet been discussed, further questions were asked if necessary. At times notes were taken and returned to at a later point to avoid interrupting the flow of conversation. The students’ comments in their experiment logs were also discussed in the interview, especially if they explicitly referred to issues of intertextuality. At the end of the video and before the microphone was turned off, the students were given the opportunity to bring forward aspects for discussion that had not yet been mentioned.

## 7.5 Coding and analysis of process data

The screen recordings were coded using the software *ATLAS.ti* (version 8.2.34.0; ATLAS.ti Scientific Software Development GmbH 2018), which has been used in previous process studies (see McCulloch 2013) and lends itself to the task because of its video coding feature. The coding procedure was developed on the basis of two screen recordings which were not used as data for this study, each of which was first reviewed in full and then segmented into non-overlapping segments.

Since the focus of the study was on the writing stage, this part of the recording was subdivided into small segments. As in the corpus study, the sentence was used as the basic unit for segmentation. This meant that the sequence between two full stops or, in the case of the first sentence of a student’s document, between a capital letter indicating a new sentence and a full stop was considered as one process segment. If the student ended the sentence with a full stop but then went back to make changes to the same sentence, this was still considered to be

part of the same segment. If a sentence was left unfinished and the student began working on an existing or a new sentence, this was treated as a new segment. The point at which the student then returned to the original sentence marked the beginning of a third segment.

This approach ensured that there was no overlap between segments, an issue observed in previous process studies (see Stapleton 2010). A ‘quotation’<sup>77</sup> was created in *ATLAS.ti* for each process segment. Segments were also created for pre- and post-writing stages. Unlike the writing stage, pre- and post-writing activities were not divided into smaller segments, but analysed globally.

In order to categorise segments in the writing stage, codes for ‘activities’ were developed in a combination of deductive and inductive elements based on an explorative analysis of two separate screen recordings (see also Chan 2011). A first set of codes was deduced from the intertextual strategies observed in the corpus study, e.g. Paraphrasing and Direct quotation. New codes were created based on the analysis of the segments in the two separate screen recordings. These activities were then subsumed under higher-level categories to be used for the primary categorisation of process segments (see Choi 2016). The writing processes of the study participants appeared to follow a tripartite structure, as previously observed in other studies (Choi 2016). For this reason, three different types of activities were distinguished in reference to the different stages of the writing process (cf. Choi 2016): pre-writing activities (\*BW), intertextual writing activities (\*IW), and post-writing activities (\*XW).

A further code for pauses (\*ZZ) was introduced for segments of at least thirty seconds without discernible action, i.e. no on-screen movements.<sup>78</sup> Five intertextual writing activities (\*IW) were distinguished in order to accurately describe the type of activity in a segment (see Figure 16).<sup>79</sup> Two of these codes, namely \*IW - SENTENCE LEFT UNFINISHED and \*IW - CONTINUING UNFINISHED SENTENCE, were used to indicate that the student aborted writing a sentence or continued working on a previously unfinished one. One sentence can comprise several of these activities. Each segment of the screen recording was assigned at least one activity code. Double coding allowed for capturing the co-occurrence of several activities, e.g. Direct quotation and Paraphrasing.

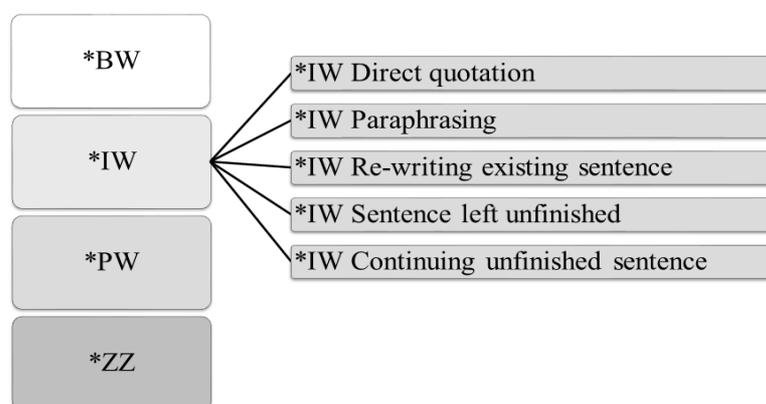


Figure 16: Hierarchy of activity codes for categorisation of process segments.

In the course of coding, it became apparent that most of the categories used for the corpus data could not be applied to the process segments. The distinction of summary, summary paraphrase and meta-level observation from paraphrase proved somewhat problematic due to editing processes in the course of which the character of a sentence was oftentimes changed. Similarly, exact copies and exact copies embedded in paraphrases were

<sup>77</sup> Note that this is the term used in the software for a segment and does not refer to a direct quotation.

<sup>78</sup> All six writers frequently paused throughout the writing processes, but due to issues in inter-annotator reliability, the annotation of pauses within segments was discarded. For studies of pausing behaviour in L2 writers' writing processes see Miller (2005), Miller, Lindgren and Sullivan (2008), and Lee (2002).

<sup>79</sup> Asterisks were used in the codes for activities to distinguish them from the codes used for acts, discussed on page 136, and to indicate their higher order in the hierarchy.

typically modified in the process so that the textual overlap was reduced. Sentences containing new content were not observed. For these reasons, the category \*IW Paraphrasing was used for segments in which the student rephrased material from the source text, even if they contained elements of meta-level observation or summary or an exact copy. In-depth analyses of the micro-level processes in these segments were conducted to compensate for the reduced functional distinction of activities. An overview of codes for activities and their descriptions appears in Table 11.

The coding scheme does not include codes for emergent pre-writing and post-writing activities because it was decided to focus on a deep analysis of the writing stage. The writing stage is the stage which is most closely oriented towards the source text and thus the most relevant for studying the students' intertextual strategies. A detailed segmentation based on sentences and subsequent categorisation of micro-level processes was only performed for the writing stage in each screen recording. Segments at the beginning of the process comprising activities such as writing titles, re-reading before writing, highlighting passages in the source text, and note-taking were categorised using the code \*BW PRE-WRITING without further subcategorisation. Segments at the end of the process consisting of activities such as determining the word count and editing (e.g. corrections of spelling mistakes and grammatical errors, syntactic restructuring, deletions, etc.) were coded as \*XW POST-WRITING. Pre- and post-writing activities were analysed holistically.

Table 11: Codes for activities used to categorise process segments.

Type of activity	Code description	
<b>*BW Pre-writing activity</b>	This code is used for segments which occur before the beginning of the first sentence. Activities that are performed in these segments include writing titles, re-reading before writing, note-taking, highlighting/adding comments in the source text, adjusting settings of word processor, and any other activity that occurs before the first sentence is begun.	
<b>*IW Intertextual writing activity</b>	This code is used for segments in which the student produces new text based on a source text. The activities in such a segment do not serve to prepare the writing task or to edit and format the completed text.	
	<i>Activities in this category</i>	<i>Code description</i>
	*IW Direct quotation	Used for sentences containing passages in single or double quotation marks, irrespective of their content.
	*IW Paraphrasing	Used for segments in which the writer paraphrases the source text by rephrasing textual material, for example directly from the source text, from memory, or from notes or source text material that had previously been copied into the document.
	*IW Re-writing existing sentence	Used for segments in which the writer makes changes to a sentence that was previously completed by a full stop. Such changes may include cutting and pasting the sentence to a different location.
	*IW Sentence left unfinished	Used for sentences which the writer leaves unfinished. This means that the sentence is not ended by a full stop when the writer moves on to work on a different sentence.
	*IW Continuing unfinished sentence	Used when the writer continues working on a sentence that they had previously left unfinished.
<b>*XW Post-writing activity</b>	This code is used for segments at the end of the writing process in which there are only editing and re-writing processes of existing sentences (e.g. corrections of spelling mistakes and grammatical errors, restructuring, deletions, replacement of individual words etc.), but in which no new sentences are written. Characteristically, the student edits the entire text at this stage, often sentence by sentence.	
<b>*ZZ Pause</b>	This code is used for segments of at least 30 seconds of inactivity. No discernible action takes place in these segments, i.e. no cursor movement or typing.	

It soon became evident that there was a hierarchy of processes within each \*IW segment, e.g. that they entailed a combination of micro-level processes such as re-reading the source text and one's own text; writing, deleting, and rephrasing; and adding words to existing strings. As a consequence, an additional level of analysis accounting for these micro-level processes was required in order to capture the students' intertextual writing processes. Following an approach taken by McCulloch (2013) in the analysis of think-aloud protocols, each segment was thus analysed for 'acts'. McCulloch defined an act as "the smallest meaningful event or type of behaviour" (2013: 139) in the writing process.

The acts in the present study were developed inductively in a bottom-up approach (see also Choi 2016; McCulloch 2013) by observing the micro-level processes that took place within each segment, again using the two screen recordings that were not analysed for the study itself. Such acts included re-reading (of the source text, the instructions, the students' own text or notes), deleting and rephrasing, deleting word(s) without replacement, adding word(s) to an existing string, substituting synonyms, and using language resources such as dictionaries and thesauruses. Codes were created, refined, merged, and subdivided in the process of preliminary coding to ensure that each code was meaningful, accurately described an observed micro-level process, and could be usefully distinguished from other codes (see Keck 2010; McCulloch 2013 for similar approaches). Two types of acts emerged from this procedure, namely writing acts (W) and non-writing acts (NW). The distinction was simply based on the absence of writing (in the form of observable typing) in the non-writing acts.

An initial set of writing and non-writing acts was developed based on the micro-level processes observed in the initial analysis of the two separate screen recordings. An additional code was added to the set whenever a new type of act emerged during a writer's process in the subsequent analysis of the process data. In order to identify emergent patterns of source text use (see Plakans & Gebriel 2012), the acts constituting each activity (i.e. each segment) were recorded using the comment function for 'quotations' in *ATLAS.ti*. The goal was to describe the micro-level processes of intertextual writing in as much detail as possible. In this step of the coding procedure, the acts that occurred in the segment under scrutiny were listed in chronological order (see example in Figure 17).

This approach ensured that an accurate description of the chronological sequence of acts in each process segment could be given rather than only an overview of the different acts of which an activity is comprised. In this way, the recursiveness of the writing process on the micro-level and the repetition of acts could be captured. Once the sequence of acts had been recorded, the segment was coded with a corresponding code for the respective writing (W) or non-writing (NW) act. For the sake of economy, some acts were subsumed under the same code. For example, the acts CUT-PASTE FROM OWN TEXT, COPY-PASTE FROM OWN TEXT, CUTS STRING FROM OWN TEXT, and PASTES STRING FROM OWN TEXT were all subsumed under the code NW - COPY-PASTE / CUT-PASTE FROM OWN TEXT.

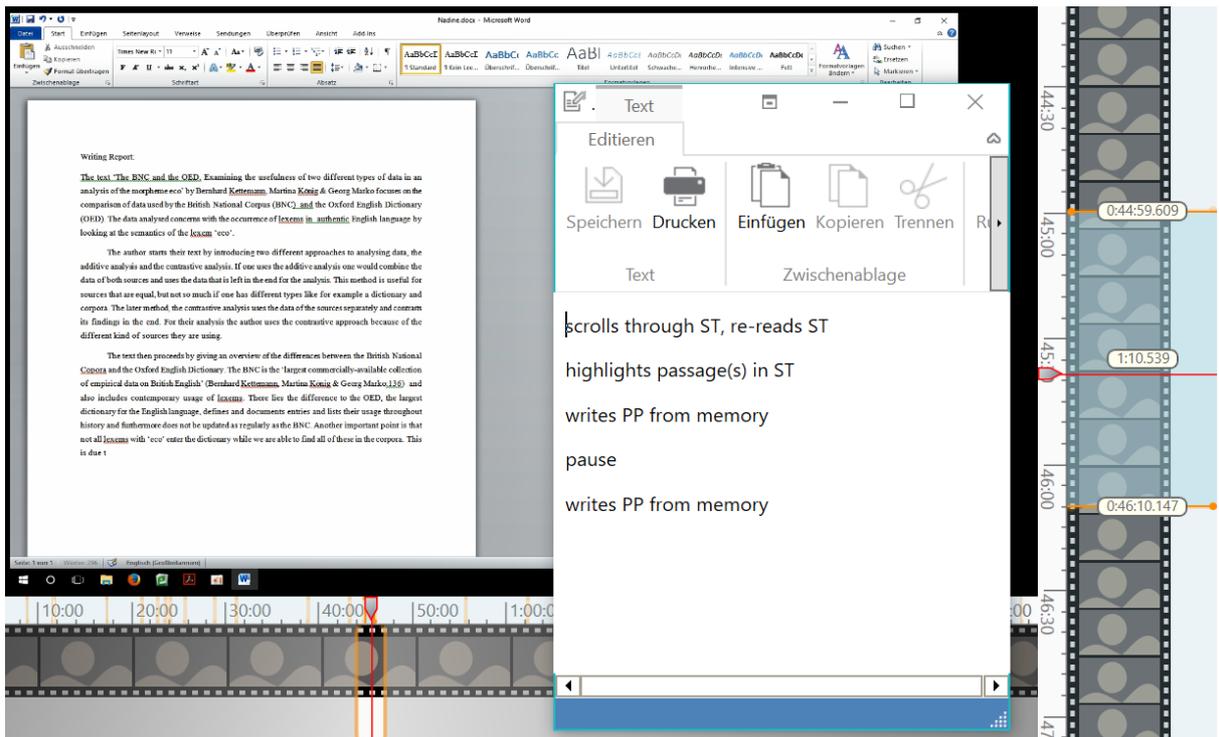


Figure 17: Use of the comment function of *ATLAS.ti* to describe the writing process in each segment chronologically.

A total of 35 different acts were identified during the analysis of the six writers' writing stages. Fifteen of these were classified as writing acts and eighteen as non-writing acts (see Table 12 for an overview of acts and corresponding codes). Once the description of the acts in an \*IW segment had been completed, each segment was coded for each of the acts that had been observed using the corresponding code. The code was used to allow for an identification of segments in which a specific act occurred and their quantification. The process is visualised in Figure 18. Several codes for \*IW activities related to paraphrasing were created, as shown in Figure 19. This approach made it possible to identify segments in which several approaches to paraphrasing occurred, for example paraphrasing from the source text and from the students' notes. The screenshots from *ATLAS.ti* in Figures 20 and 21 show the selection of codes for a segment and a segment that has already been coded.

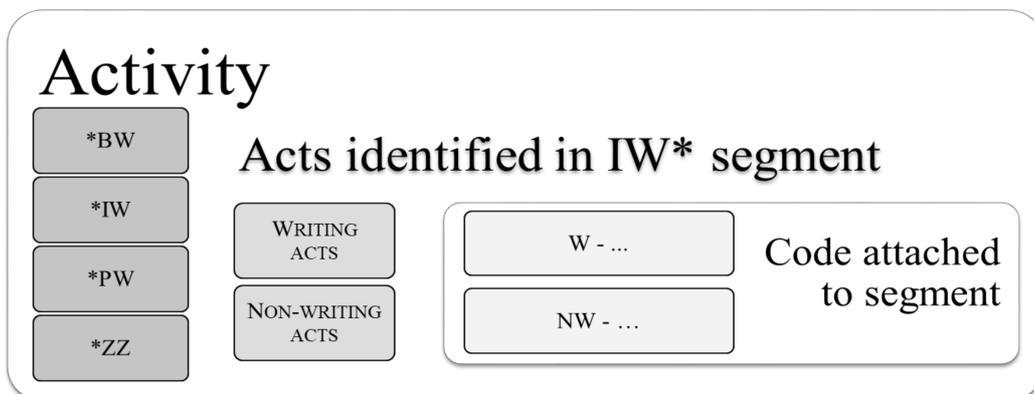


Figure 18: Overview of coding sequence. Once the sequence of acts has been described for an \*IW segment, it is then coded with the corresponding codes for writing and non-writing acts.

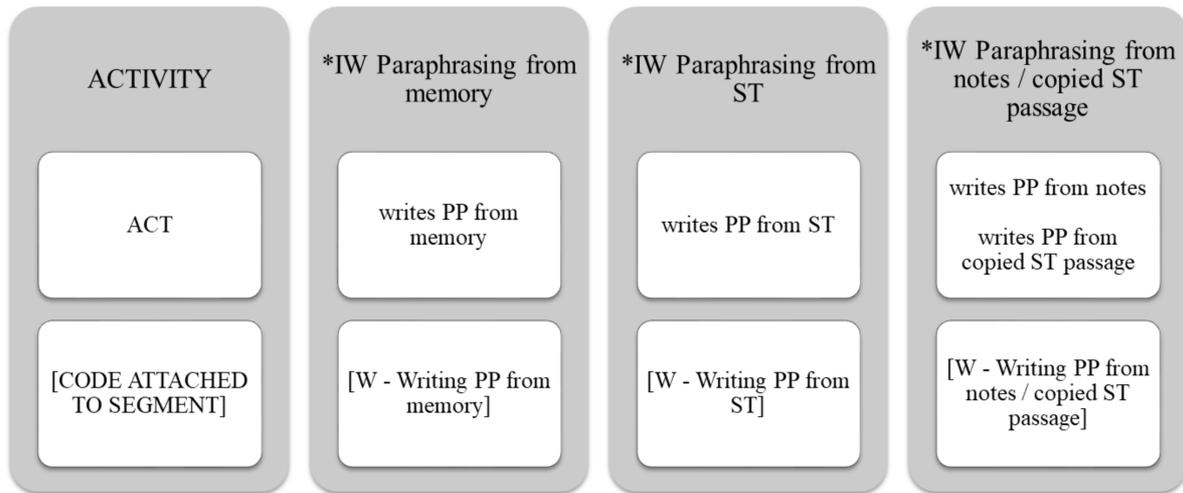


Figure 19: Activities and acts related to paraphrasing.

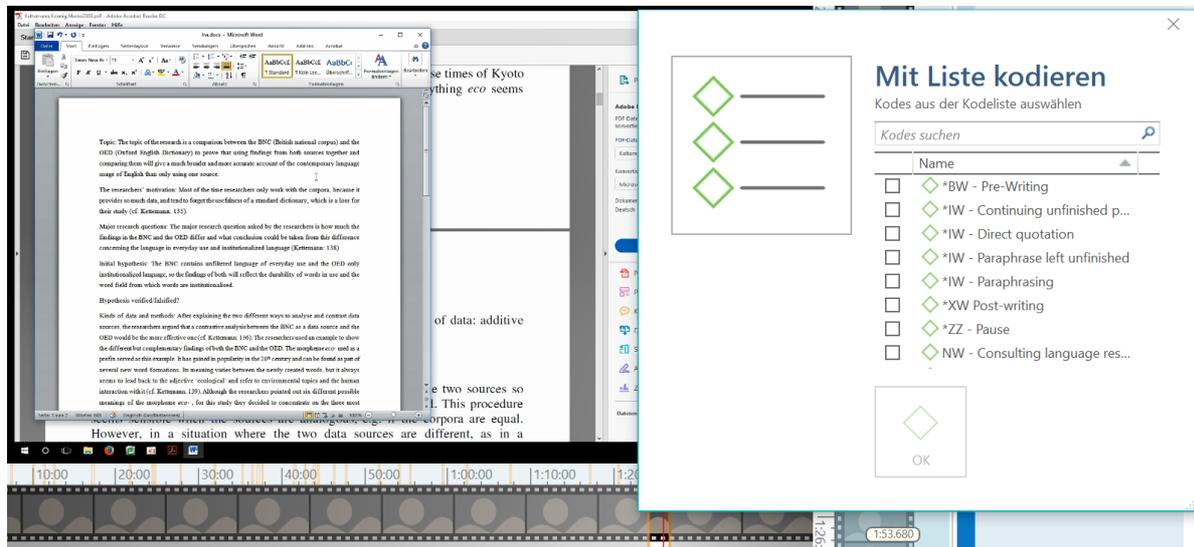


Figure 20: Selecting codes for acts occurring in a segment from a list of codes in *ATLAS.ti*.

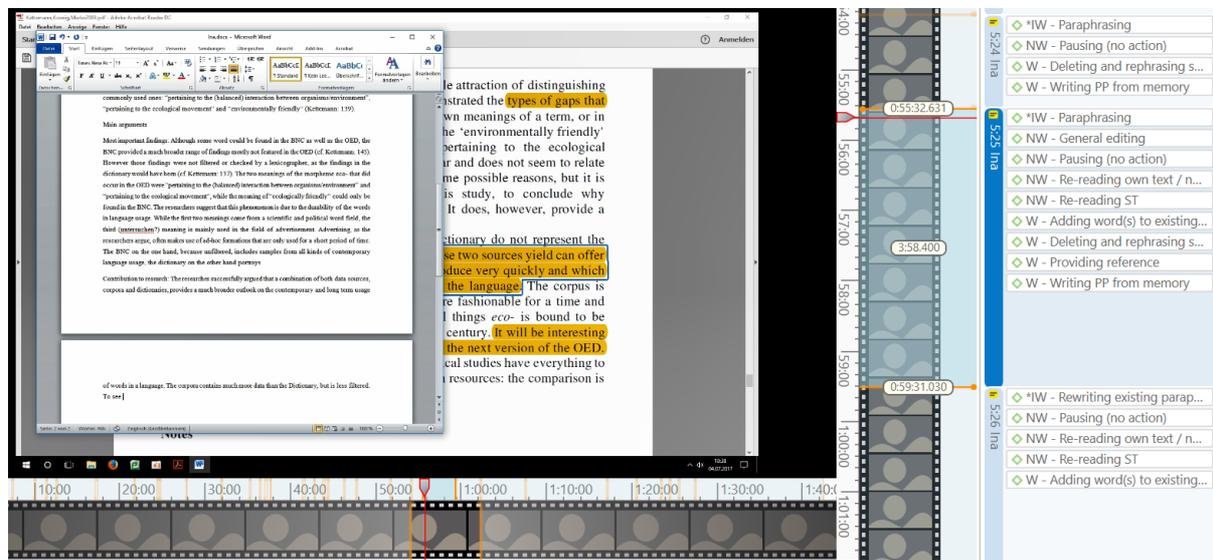


Figure 21: View of segments coded for acts in *ATLAS.ti*.

Once the video recordings had been segmented, 29 segments were randomly selected for review by a second annotator to ensure the reliability of the categories for acts and a consistent approach. The researcher and a student assistant independently annotated these segments for the sequence of acts, i.e. they described the actions on the screen using the categories for acts established earlier. Initially, an inter-annotator agreement of 65% was reached. Pauses were identified as the major source of these disagreements. They had initially been included as an act; a pause had been defined as a sequence of five seconds of more with no action or movement on the screen. Despite this narrow definition, there was a large discrepancy between the annotators in terms of noticing and annotating for pauses. Pauses were thus removed from the set of acts and no longer considered in the annotation of segments.

It was furthermore evident that two provisional acts, namely RE-READS ST and SCROLLS THROUGH ST, RE-READS ST, were not sufficiently demarcated, which resulted in repeated inconsistencies between the two annotators. It was thus decided to collapse the two provisional acts into RE-READS / SCROLLS THROUGH ST so as to reduce the likelihood of errors. The acts WRITES PP FROM NOTES and WRITES PP FROM COPIED ST PASSAGE were combined into WRITES PP FROM NOTES / COPIED ST PASSAGE for the same reason. Initially, codes that referred to students leaving sentences unfinished and returning to these sentences had been included in the set of acts, but as these led to misunderstandings and were already coded on the activity level, the corresponding codes were discarded.

After these changes had been implemented, inter-annotator reliability reached 79%. This was judged to be sufficiently high to ensure consistency of annotation. The remaining disagreements mostly occurred when one annotator noticed a deletion or a general editing act interceding with a paraphrasing act that the other annotator had not noticed. These cases were consequently discussed and the definitions for acts refined. Once these issues had been resolved, the existing annotations were reviewed and corrected and the researcher proceeded to annotate and code the remaining segments. Table 12 contains the final list of acts annotated for in the screen recordings with definitions. One week after the coding of all six screen recordings had been completed, the comments as well as the codes for activities and acts assigned to each segment in each screen recording were reviewed for consistency.

In addition to the coding of the screen recordings, an impressionistic description of each writer's process was recorded in a spreadsheet as a summary of the analysis. First, the length of each writing session was recorded. Then, a qualitative description of the setup, i.e. the way the documents were arranged on the screen, and of the pre- and post-writing stages was given for each process. Most importantly, the writer's general approach to the task and to working with the source text was described based on the screen recording. The aim of this was to identify patterns of behaviour in the writing processes to determine whether there are certain types of writers in such writing tasks.

The coding procedure is visualised in Figure 22 and can be summarised as follows. In the first step, the recording was segmented and a 'quotation' was created for each segment. These segments were subsequently coded for type of activity in a second step. In the third step, a comment was added to the 'quotation' in which the writing and non-writing acts found in this segment were recorded in chronological order. Fourth, each process segment was coded for the identified writing and non-writing acts. In the fifth and final step, each recording was reviewed to identify the general approach to the source-based writing task taken by each student to this source-based writing task.

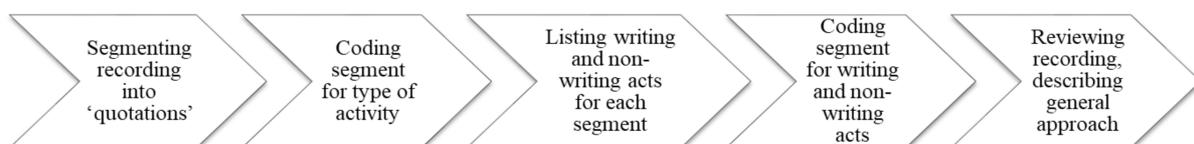


Figure 22: Coding procedure.

Upon completion of the annotation, the results of the coding were quantified using the analysis function of *ATLAS.ti* and exported into *Excel* for further analysis. In addition, the comments containing the annotation for acts were extracted and each writer's approach to writing intertextually was analysed qualitatively with the aim of identifying patterns in the overall processes. Special attention was paid to paraphrasing and direct quotation as the central intertextual strategies. An additional focus was placed on attribution, reporting, documentation, and the re-use of source text material.

Table 12: Overview of acts and corresponding codes.

Type of act	Act	Corresponding segment code	Definition
Writing act (W)	writes PP from memory	W - Writing PP from memory	Used when the writer rephrases the source text while the source text is not visible on screen or not visible enough to be readable and there are no notes or copied source text passages in their document.
	writes PP from ST	W - Writing PP from ST	Used when the writer rephrases the source text while it is visible on screen.
	writes PP from notes / copied ST passage	W - Writing PP from notes / copied ST passage	Used when the writer rephrases material from their own document, either notes they had taken earlier or a source text passage that they had copied into their document.
	adds word(s) to existing string	W - Adding word(s) to existing string	Used when one or more words are added to a string that has already been written, e.g. in the middle of an existing sentence. If words are added at the end of a sentence, this act can only be used if the sentence was previously completed.
	deletes and rephrases	W - Deleting and rephrasing string of paraphrase	Used when the writer deletes (part of) a word or a string and replaces it with a new string immediately in the course of writing. Applied even if only individual letters are deleted and replaced. This act is to be distinguished from 'deletes word(s) without replacement' (no new words) and 'synonym substitution' (word from an existing string is replaced by a (near-)synonym).
	inserts placeholder	W - Inserting placeholder	Used when the writer inserts words or symbols that function as a placeholder and can clearly be identified as elements that are to be replaced later in the process.
	inserts quotation mark(s)	W - Inserting quotation marks	Used when the writer inserts one or more single or double quotation marks.
	manual copying from notes / copied ST passage	W - Manual copying from notes / copied ST passage	Used when the writer manually copies a passage from their notes or from source text passages that have been inserted into their own document. Manual copying implies that the writer types an identical sequence without using the processor's copy-paste functions.
	manual copying from ST	W - Manual copying from ST	Used when the writer manually copies, i.e. types a passage from the source text. Manual copying implies that the writer types an identical sequence without using the processor's copy-paste functions.
	provides reference	W - Providing reference	Used when a bibliographical reference to a publication is inserted, e.g. in parentheses, a footnote etc.
	replaces DQ with PP	W - Replacing DQ with PP	Used when a direct quote, i.e. a passage set in quotation marks, is deleted and replaced by a paraphrase.
	replaces placeholder	W - Replacing placeholder	Used when a previously inserted placeholder is replaced in some way, for example by text.
replaces PP with DQ	W - Replacing PP with DQ	Used when a paraphrase is deleted and replaced by a direct quote (set in quotation marks).	

Type of act	Act	Corresponding segment code	Definition
Writing act (W)	synonym substitution	W – Synonym substitution	Used when a single word is deleted from an existing string and replaced with a word of a related or identical meaning, i.e. a synonym or a near-synonym. Distinguished from ‘deletes and rephrases’ in that synonym substitution is not implemented in the flow of writing but at a later time.
	writes notes	W - Writing notes	Used when the writer takes notes in their document, e.g. bullet points based on source text material or the instructions.
	writes reporting structure	W - Writing reporting structure	Used when the writer composes a passage that highlights the fact that the content of the sentence is reported. Usually, this is achieved via matrix structures (with or without a reporting verb) and complex prepositions, e.g. <i>the author states that, it has been claimed that, a central assumption of the study is that or according to X.</i>
Non-writing act (NW)	consults language resource	NW - Consulting language resource	Used when the writer consults an online dictionary, thesaurus, language forum, Word language tool etc. for reference.
	copy-paste / cut-paste from notes / copied ST passage cuts notes / copied ST passage pastes notes / copied ST passage	NW - Copy-paste / cut-paste from notes / copied ST passage	Used when the writer cuts or copies a passage from their notes or a source text passage they have previously copied into their document and pastes it into their text.
	cut-paste from own text copy-paste from own text cuts string from own text pastes string from own text	NW - Copy-paste / cut-paste from own text	Used when the writer cuts or copies a passage from their own text and pastes it into their text, usually at a different location.
	copy-paste from ST copies from ST pastes from ST	NW - Copy-paste from ST	Used when the writer cuts or copies a passage from the source text and pastes it into their reading report.
	deletes entire sentence / fragment	NW - Deleting entire sentence (fragment)	Used when the writer deletes an entire sentence or an entire sentence fragment (i.e. a sentence that has not been finished yet). This act is only annotated for if the sentence is completely removed from the text without any remains.
	deletes notes / copied ST passage	NW - Deleting notes / copied ST passage	Used when the writer deletes their notes or a source text passage that they have previously copied into their document.
	deletes word(s) without replacement	NW - Deleting word(s) without replacement	Used when one or more words are deleted from the writer’s text and are not replaced by new textual material.
	general editing	NW - General editing	Used for editing and formatting acts such as corrections of spelling, grammar errors, and punctuation; insertion or deletion of line breaks, spaces, and quotation marks, etc.

<b>Type of act</b>	<b>Act</b>	<b>Corresponding segment code</b>	<b>Definition</b>
Non-writing act (NW)	highlights passage(s) in notes / copied ST passage	NW - Highlighting passage(s) in own text / notes	Used when the writer highlights a passage in their notes or a copied source text passage using a highlighting function of their Word processor (e.g. changing the background colour or the font colour).
	highlights passage(s) in ST	NW - Highlighting passage(s) in ST	Used when the writer uses the highlighting function to highlight a passage in the source text.
	re-arranges windows	NW - Re-arranging windows	Used when the writer changes the setup of the windows on the screen, e.g. by moving them to a new position and/or changing their size.
	re-reads instructions	NW - Re-reading instructions	Used when the writer reads the instructions (discernible e.g. through cursor movements over the instructions window).
	re-reads own text / notes	NW - Re-reading own text / notes	Used when the writer reads their own text (discernible e.g. through cursor movements over the text).
	scrolls through ST / re-reads ST	NW - Re-reading ST	Used when the writer reads the source text (discernible e.g. through cursor movements over the source text).
	searches own text	NW - Searching own text	Used when the writer uses the word processor's search function to search their own text.
	searches ST	NW - Searching ST	Used when the writer uses the search function of the PDF reader to search the source text.

## 7.6 Transcription, coding, and analysis of stimulated recall and interview data

The interviews were transcribed by three student assistants using *easytranscript* (version 2.50.5; e-werkzeug 2015). Because the focus of the analysis would be the interviews' content, a broad orthographic transcription system was used. Repetitions, self-corrections, and false starts were smoothed over and grave grammatical errors were corrected. Laughter was indicated as (laughter), short pauses as (.), long pauses as (...), and capital letters were used to indicate emphasis. Sequences in which the interviewer and the student watched the recording were marked as (observation of recording). A <timestamp> tag was inserted before parts of the recording in which the interviewer mentioned the timestamp of the screen recording that was currently being viewed. To prepare the transcripts for analysis, leading questions and students' speculations were excluded.

The transcripts were then divided into segments based on their content to account for emergent themes in the data and a surface categorisation was implemented using *ATLAS.ti* (see Boshier 1998; Plakans 2008 for a similar approach). To focus on specific areas of interest identified in the transcripts (see Gass & Mackey 2000: 101ff.), broad surface categories were created that correspond to intertextual acts such as direct quotation, paraphrasing, and referencing. Other categories were based on common categories of AOIs, such as lexical strategies and synonym substitution. Synonym substitution was highlighted separately despite being a lexical strategy because it had previously been identified as a major strategy in paraphrasing (see Keck 2010) and because it had been a common AOI in the screen recordings. The category 'General process' was introduced to categorise the answers the students had given to the one of the introductory interview questions in which they had been asked to describe their general approach to the task and how they went about writing the reading report.<sup>80</sup> The following are the final surface categories used for the classification of interview sections:

- Paraphrasing
- Direct quotation
- Copying from the source text / plagiarism
- Attribution
- Referencing
- Lexical strategies (finding the right words, re-use of source-text material)
- Synonym substitution
- Re-reading source text
- Note-taking
- General process (introductory question about process)
- Other

The goal of using these broad categories was to identify sections in the interviews to be analysed which could later be related to the analysis of the screen recordings. It also facilitated excluding sections from the interview that did not relate to the research questions<sup>81</sup>. The coding of stimulated recall data is subject to a high level of interpretation, which can be problematic in view of the deep involvement of the researcher (Gass & Mackey 2000: 64). For this reason, the student assistant who acted as a second annotator in the corpus study was instructed to categorise the retrospective data independently to ensure objectivity. Following Boshier (1998), important themes, strategies, and issues of source-based writing discussed by the L2 writers were then identified within each surface category and connected to the observations made in the analysis of the writing process (see discussion in section 7.7).

80 This code was only used for answers to the third introductory question in the interview, namely "Could you please explain to me in as much detail as possible how you wrote the reading report in the experiment?" (see Appendix 14). It was hoped that this would allow for insights into the students' own perception of their general source-based writing process, which could then be compared to the process data obtained from the screen recordings.

81 This included answers to two of the introductory interview questions and to those concerning students' use of additional resources (unless clearly related to paraphrasing) as well as other sections in the interviews in which the conversation focused on topics that were not relevant to the research questions.

## 7.7 Results and discussion

The students spent between 1 hour and 29 minutes and 2 hours and 10 minutes on writing their reading report. As in previous studies into L2 source-based writing (see e.g. Choi 2016; Leijten et al. 2019), each student's writing process is clearly divided into three stages. In all but one writers' processes, the majority of the time was spent on the writing stage. Both a distinct pre-writing stage with a focus on preparation and a distinct post-writing stage with a focus on editing can be observed in each writing process. In all writers' processes, there was a period at the beginning in which the students prepared for their task in various ways. The beginning of the writing stage was determined as the point at which the student wrote the first word of their reading report. The students also spent time editing and reviewing their text at the end of the process. The change from writing to post-writing was always accompanied by a clear change in behaviour. For example, once they had finished writing the students changed the layout on the screen, zoomed out or moved to the beginning of the text. They then re-read and edited the text, usually sentence by sentence.

A total of 201 segments were distinguished in the six screen recordings. The number of segments per screen recording ranges from 29 to 38, including longer pre- and post-writing segments. This corresponds to between 15 and 25 segments per hour, with an average of 19.3 segments per hour (see Table 13).<sup>82</sup> The divergent numbers and differences in writing time can be explained by different approaches to source-based writing, as discussed below. That the majority of segments are categorised as \*IW – Intertextual Writing is a result of the sentence-based segmentation of the writing stage.

Table 13: Segments and activities per writing process.

Student pseudonym	Total no. of segments	*BW - Pre-Writing	*IW – Intertextual writing	*XW - Post-writing	*ZZ - Pause
Ben	29	1	26	1	1
Nadine	30	1	28	1	0
Tara	33	1	30	1	1
Eva	37	1	33	2	1
Ina	34	1	32	1	0
Arne	38	1	35	1	1
TOTAL	201	6	184	7	4
PERCENTAGE	100%	3.0%	91.5%	3.5%	2.0%

Figure 23 shows that the proportions of pre-writing, writing, and post-writing stage are relatively uniform, despite variation in the length of the writing processes. The ratios observed here are comparable to that of the advanced L2 student in the case study by Leijten et al. (2019) and is thus likely to be typical of relatively proficient L2 writers. The individual stages do vary in length, however. While some students' processes comprised relatively extensive pre- and post-writing stages, other students spent relatively little time on these activities. Tara's process stands out because her pre-writing stage is longer than her writing stage (see further discussion in section 7.7.1). An overview of the length of all three writing stages per student is provided in Table 14.

<sup>82</sup> As described in section 7.5, pre- and post-writing stages were not divided into smaller segments, so that there was typically only one segment coded as \*BW – Pre-Writing and one coded as \*XW – Post-writing per process. In Eva's process, there are two post-writing segments, disrupted by a longer pause. Overall, there are only four pauses that are longer than 30 seconds (\*ZZ), each of which occurred in a different writer's process. Most of these pauses were apparently due to the fact that the writers were filling in their experiment logs, as they explained in the retrospective interviews. This shows that the writers otherwise worked continuously on the task, which may have been different outside of an experimental setting.

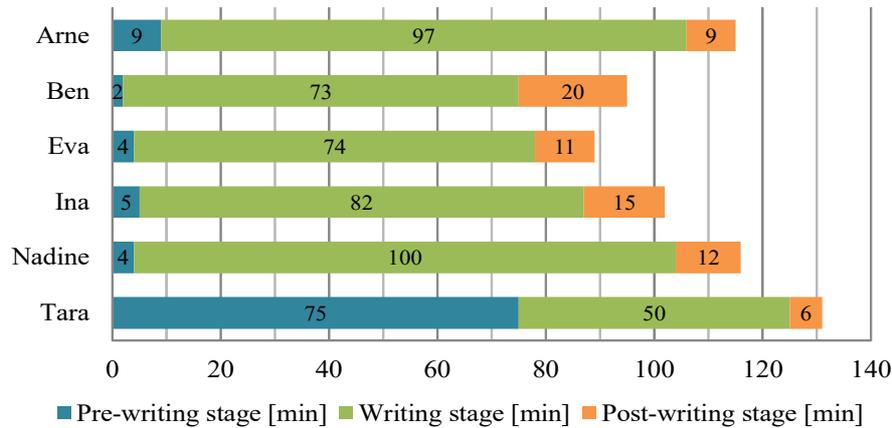


Figure 23: Length of each student's writing process in minutes.

Table 14: Length of recording and of pre-writing, writing, and post-writing stages per student [hh:mm:ss].

Student pseudonym	Total length of recording	Pre-writing stage	Writing stage	Post-writing stage
Arne	01:54:51	00:08:41	01:37:25	00:08:45
Ben	01:35:37	00:02:29	01:12:52	00:20:16
Eva	01:28:58	00:04:22	01:13:47	00:10:49
Ina	01:42:34	00:05:24	01:22:07	00:15:03
Nadine	01:55:15	00:03:40	01:39:49	00:11:46
Tara	02:10:28	01:14:44	00:49:44	00:05:59
MEDIAN	01:48:42	00:04:53	01:13:47	00:11:18

The variation in length is – at least to a certain extent – due to the writers' different approaches to planning and editing (see sections 7.7.1 and 7.7.3). It is likely that the length of the pre- and post-writing stages and the types of processes which constitute them were influenced by the students' preparation at home, by the task prompt, and by the quasi-experimental setting in which they wrote their reading reports. Consequently, the length of stages is not necessarily representative of tasks completed in an uncontrolled setting. For this reason and because this study focuses on the processes of writing based on a source text, an in-depth analysis of pre- and post-writing processes based on these data was not carried out. In lieu of this, short impressionistic descriptions of the writers' pre- and post-writing activities are provided in sections 7.7.1 and 7.7.3. A detailed analysis of the writing stage is presented in section 7.7.2.

### 7.7.1 The pre-writing stage

Before the students began writing, they each engaged in a preparatory stage which commenced immediately after the start of the session. At the beginning of this pre-writing stage (\*BW), they arranged the windows on their screens.<sup>83</sup> Three different setups were observed among the six participants. Eva and Arne arranged the windows so that they were able to see their own text and the source text side-by-side (Figure 24).

Nadine and Ina worked with overlapping windows so that only one was visible at a time (Figure 25). They switched back and forth between the windows in the process, bringing to the foreground their own text, the source text, the instructions or the browser for the use of language resources as needed. Ina occasionally re-arranged the windows so that the source text and her own text appeared side-by-side on the screen, but she did not abide by this setup for longer periods in her writing process. She explained that this was because looking at one document at a time allowed her to focus more easily on its content, while viewing both allowed her to compare her own text to the source text (see excerpt 7.1).

(7.1 ) INA: [...] I find it easier to read the things properly when they were full screen because then I concentrated on one of the things. But in some cases I wanted to compare: what did I write, what did the researchers write. Is that all right or is something missing or is something wrong. That is why I put them next to one another. (Ina's interview, 00:46:24-7 – 00:46:57-8)

A similar third approach was taken by Ben and Tara. Ben first arranged his windows to a side-by-side setup, but later in the process maximised the windows so he was only able to view one at a time. Tara used a side-by-side setup during the pre-writing stage, in which she copied passages from the source text into her document. Once she had completed this part of the process, she changed to overlapping windows and worked predominantly with the text document.

The setup chosen by the students is reflective of their respective paraphrasing behaviour. For example, Arne wrote all paraphrases directly from the source text as it was visible at all times. He did not copy source text passages over to his text for paraphrasing. Tara, on the other hand, paraphrased from source text passages she had copied into her document (see section 7.7.2.4 for further discussion of paraphrasing strategies).

The pre-writing stages varied in length, from 00:02:29 to 01:14:44 (see Table 14). The differences in length are a reflection of the different pre-writing activities in which the writers engaged during this first stage of their writing processes. Five writers' pre-writing stages were under ten minutes in length. Ina and Arne began by arranging their windows. Both then re-read the instructions and either took notes or copy-pasted parts of the instructions into their own document. They also both re-read parts of the source text. Ina additionally highlighted passages in the source text, while Arne arranged the questions he had copied from the reading report instructions in his document and then highlighted central passages in these questions. He also spent some time formatting the copied passages. Eva, Nadine, and Ben, all of whom spent less than five minutes for pre-writing, mostly devoted this time to re-reading the source text. Nadine and Eva additionally highlighted passages in the source text and Ben reviewed the instructions.

---

83 Note that some of the writers re-arranged the windows on their screen repeatedly during the writing process. Overall, there were seventeen instances of the act 're-arranges windows', which occurred in twelve segments. Nadine and Eva did not re-arrange their windows; Nadine worked with overlapping windows throughout and Eva maintained her side-by-side setup.

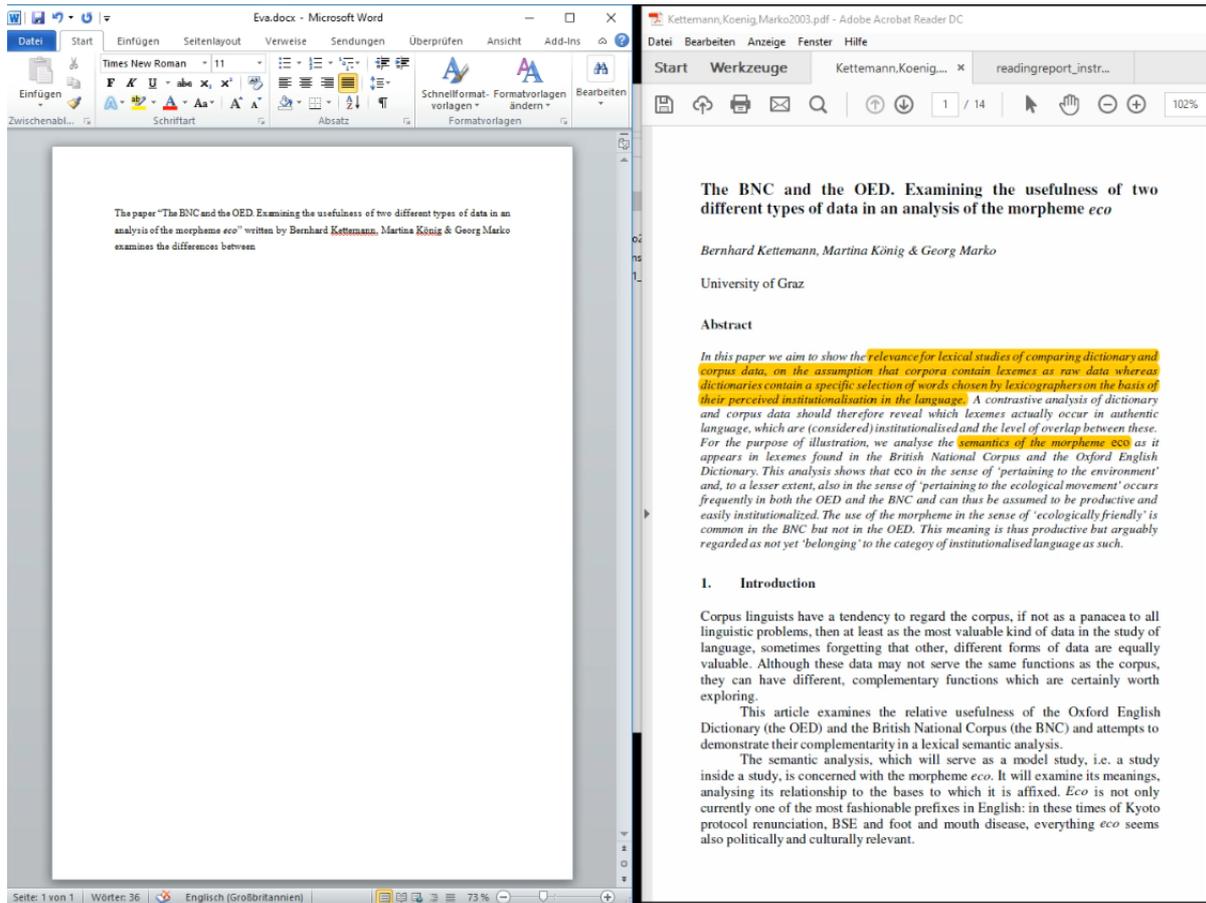


Figure 24: View of documents being used side-by-side in Eva's writing process.

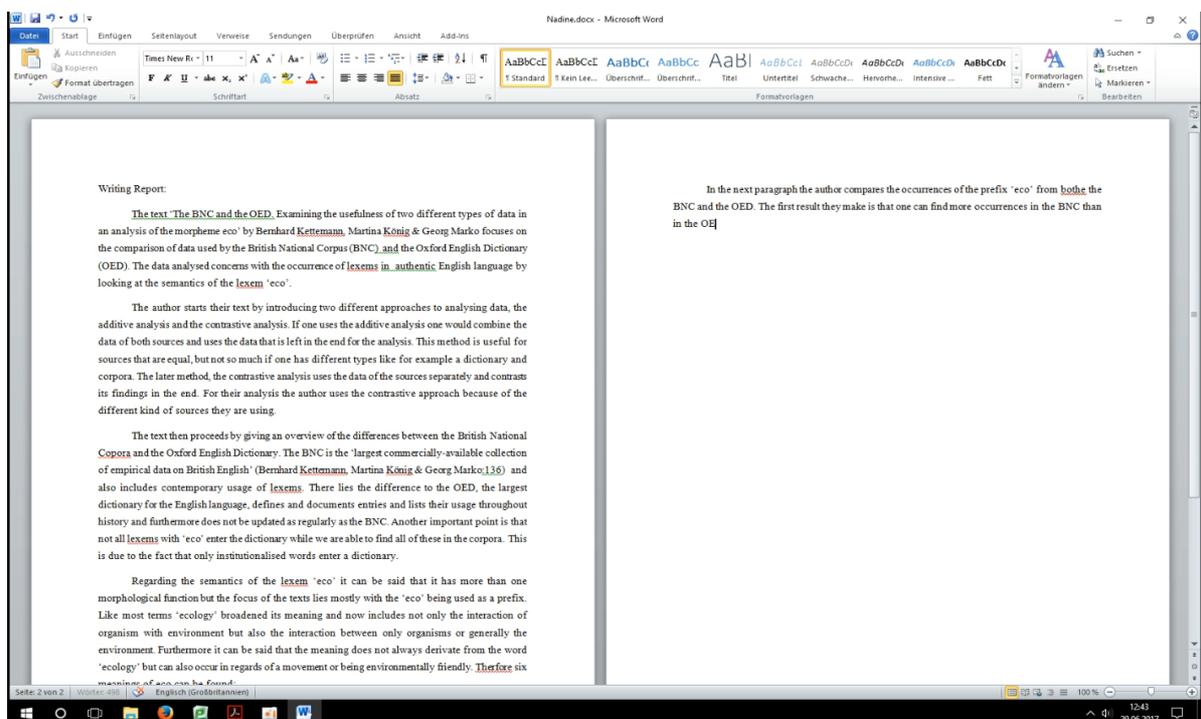


Figure 25: View of only one document being visible at a time in Nadine's writing process.

Tara stood out from the group as the only student with an extensive pre-writing stage. She spent an hour and fourteen minutes on this stage, more than half of her total writing time. During this time, she first arranged the windows and wrote the authors' names and the title of the source text into her document. She then re-read parts of the source text. She spent the majority of her pre-writing stage reviewing the entire source text and copying passages from it to her document. Tara did not place these copied excerpts in quotation marks, but she did mark ellipsis and provided a reference (page number) at the end of each excerpt. These preparatory processes are similar to those of students in the studies by Stapleton (2010) and Li (2013), showing that copying from source texts in the course of summarising is a typical strategy for L2 writers.

In previous studies of intertextual writing, it was observed that L2 writers usually start their writing process by reading the instructions, trying to understand the task, and planning their text (Chan 2017; Plakans 2010).<sup>84</sup> In this as in previous studies (Plakans 2010), this is achieved in a circular process of reading the task description, the source text, and again the task description, which is sometimes extended by other micro-level processes such as note-taking and highlighting. Most of the students spent the majority of the initial phase on reading, which was also observed in a previous case study (Leijten et al. 2019). The observations of the pre-writing stage made for this source-based writing task are in line with Flower and Hayes' (1981) assumption that planning constitutes the initial phase of the writing process (see Shaw & Weir 2007 for further discussion).

The pre-writing processes of the students in this study are comprised of the same processes as in previous process studies (Chan 2017; Choi 2016; Plakans 2010), though they are somewhat reduced, both in terms of time and intensity. The writers had been given time to read the instructions before the beginning of the recording and had already read the source text before coming to the actual writing session, which is likely to be the reason for these less extensive pre-writing stages. The case of Tara, however, shows that despite the general tendencies confirmed by this study, individuals may differ remarkably in the way they shape the pre-writing process.<sup>85</sup>

## 7.7.2 The writing stage

The writing stage of the students' processes, which follows the pre-writing stage, is at the focus of the present study. It lasted between 50 minutes and 1 hours and 40 minutes and was usually the longest stage of the three. 184 of all process segments analysed are devoted to intertextual writing (184 \*IW segments; 91.5%). This second stage of the writing process consists of between 26 and 35 segments per student, with an average of 30.7 segments. Since the sentence was used as the basic unit for identifying segments, this means that each student composed between 26 and 35 sentences and sentence fragments during the writing stage. Due to deletions and other syntactic operations during the process, this number diverges from the total number of sentences in the final product (see Table 15).<sup>86</sup>

### 7.7.2.1 Linearity of the writing stage

On the macro-level, the approaches to the writing stage observed in this study can be broadly grouped into linear and non-linear approaches. As proposed by Zimmermann (2000), writing processes may display linear

<sup>84</sup> See also Choi (2016) for a more extensive analysis of L2 writers' planning processes in real-life term papers written over the course of several weeks.

<sup>85</sup> It must be noted here that it is possible that Tara had not read or prepared the source text in as much detail as the other study participants and this was the reason for her lengthy pre-writing process. It is, of course, also possible that her approach is simply different from the other students'.

<sup>86</sup> The results show a discrepancy between the number of sentences in the final text and the number of segments devoted to writing intertextual links in the process. The total number of all sentences in the six reading reports produced in the experimental study is 134, while the total number of segments coded as \*IW in the process data is 184. Some sentences and sentence fragments that the writers composed in the writing process were either deleted or merged with other sentences later in the process; others were edited several times during the process and thus formed the basis of more than one segment.

structures despite their general recursiveness. Linear approaches were taken by Nadine, Ben, and Arne. Their approaches were linear in the sense that they generally wrote one sentence after the other in a chronological order, with noticeable similarity between the individual acts that occurred during the writing of a sentence.

Unlike these three students, Tara, Ina, and Eva did not take a strictly linear approach to the writing stage. While they still wrote many of the sentences in chronological order, Tara and Ina tended to jump back and forth in their texts. At several points in her writing stage, Tara began writing a sentence, then deleted the fragment completely and continued writing at a different position in the text. Ina initially did not write a coherent text but rather accumulated sentences into categories that she had drawn from the instructions during the pre-writing stage. As a result, she also displayed a relatively non-linear approach to writing. Instead of completing a sentence she had started, she frequently turned her attention to a previously written sentence, rewrote it, and then returned to the fragment. She only turned her collection of sentences into a coherent text by connecting sentences and restructuring her text during the post-writing stage.

Tara's and Ina's processes, which are based on notes from the source text, also refute the claim that note-taking leads to linear approaches (Kirkland & Saunders 1991: 115). Finally, Eva's non-linear approach was manifested in her tendency to insert new paraphrases between existing ones. She also cut sentences from her own text to paste them at a different position, thereby restructuring the reading report. Sometimes she conjoined existing paraphrases, but also deleted some. In the processes of Ina and Tara, the non-linearity in the writing stage coincides with the fact that they both based their reading report on excerpts – either from the source text or the instructions – that they had copied into their document in the pre-writing stage.

It has been proposed that linear approaches characterise the writing process of less proficient writers because they have to devote more resources to text production (Flower et al. 1986). Since the texts produced in this study were not graded, it is difficult to assess whether Nadine, Ben, and Arne are in fact less proficient writers than the other three students despite their comparable proficiency in English. It should be noted that this assumption has been called into question by a study of two intertextual writing processes (Chan 2017) showing that the student with the more linear writing process scored higher than the one with a more recursive process. Relating linearity to the distinction between knowledge-telling and knowledge-transforming writing processes introduced by Bereiter and Scardamalia (1987), Leijten et al. (2019: 558) have argued that linear approaches in source-based writing are a sign of pure knowledge telling where supposedly “the information from the source texts fails to be integrated into a whole”.

In this study, the writing task itself strictly does not require students to perform knowledge transformation. This appears to have been overlooked thus far as a reason for linear approaches. All six reading reports produced in this experimental study integrate ideas from the source text and match the task description. Knowledge transforming takes place irrespective of the linearity of the approach as sentences are rewritten and edited throughout the writing process. Based on these findings, it must be assumed that the connection between the linearity of the writing process and the quality of the product is as straightforward as previously suspected, at least in source-based writing.

The observations made here suggest that the recursiveness of the writing process is a matter of degree, with individuals leaning towards more or less recursive approaches. All six processes in this study contain recursive elements throughout, yet the actualisation of recursiveness is quite different. Overall, it appears that L2 writers approach the source-based writing task either chronologically, resulting in a relatively linear approach in which editing and restructuring are left to the post-writing stage, or they take an approach that integrates restructuring into the writing stage in a more pronounced way and is thus relatively non-linear. It is conceivable that this is a matter of preference or convenience rather than proficiency, though further research is needed to test the relationship between these aspects.

The intertextual nature of the writing task plays an influential role in the linearity observed in the processes. The analysis of the screen recordings provides additional insights related to an assumption derived

from the corpus study, namely that L2 writers follow the structure of the source text in structuring their own text, an observation also made by Chan (2011). The chronological sequence observed in the products is reflected in the writing process of the L2 writers. That the structure of the source text served as orientation was confirmed by Eva, Nadine, and Tara in the interviews (see excerpts 7.2, 7.3, and 7.4). These excerpts, which were responses to one of the introductory interview questions, confirm that these three writers made a conscious decision to use the structure provided by the source text as a template for their reading reports.

- (7.2) EVA: At first, I kind of read through the main parts of the text again. [...] And then I basically used the structure of the text to do my reading report. Like I started with the introduction and then with the different paragraphs. (Eva's interview, 00:01:55-0 – 00:02:28-3)
- (7.3) NADINE: [...] I started re-reading the text and marking or highlighting the text. And then I went through every passage and when I wrote the introduction I just concentrated on the first few paragraphs and then I went on like this. So I had like the structure of the text in my writing report as well. [...] (Nadine's interview, 00:01:45-7 – 00:02:15-6)
- (7.4) TARA: [...] I had to re-structure my own report so I tried to make it similar to the structure of the research paper but because I was explaining different things it was a bit difficult to actually re-structure everything [...]. (Tara's interview, 00:02:42-3-00:02:47-7)

However, not all students followed the structure of the source text, as some rather relied on the guiding questions from the reading report instructions (see Appendix 11). These instructions roughly suggested the IMRD structure observed in the reading reports analysed for the corpus study. Arne explained that he “just followed the order of the questions” (Arne's interview, 00:02:39-9 – 00:03:59-2), and Ben “always looked again at the questions in order to get a well-structured text” (Ben's interview, 00:02:34-5 – 00:03:22-0). Ina read the questions while trying to make sense of the task at the beginning of the experiment. She “wrote parts of [the questions] down and then [...] tried to find the passages in the text again that dealt with those parts” (Ina's interview, 00:02:21-2 – 00:03:38-8). This approach helped her structure her process as well as her reading report. These testimonies show that the source text alone does not explain the structure of the students' reading reports. Instead, the task description largely determined the structure for three of the six writers and has to be taken into consideration as a variable influencing both the writing process and consequently the product.<sup>87</sup>

The next section takes a closer look at the writing stage and the micro-level activities of which it is comprised.

### 7.7.2.2 Activities constituting the writing stage

Five different activities were distinguished in the analysis of the writing stage (see Table 15). Unsurprisingly, as in the corpus study, the most frequent type of activity is paraphrasing (\*IW - PARAPHRASING). 154 segments, that is 84% of all \*IW activities, are instances of paraphrasing<sup>88</sup>. Ten sentences (5.4%) were left unfinished. In nine<sup>89</sup> segments (4.9%), the writer returned to a previously unfinished sentence and continued writing it.

32 intertextual writing segments (17.9%) were coded both as \*IW - PARAPHRASING and as \*IW - DIRECT QUOTATION; there were no instances of \*IW - DIRECT QUOTATION occurring on its own. This shows that in this

87 See also Petrić & Harwood (2013) for a study of the influence of the task representation on L2 writers' citation practices in different kinds of written assignments, and Chan (2017) for a discussion of the influence of the task prompt in source-based writing.

88 As elaborated in section 7.5, such segments may lead to the production of other types of intertextual links such as meta-level observations and summaries, but a broad categorisation was applied here as changes to the sentences in the process made a distinction between these categories from the corpus study difficult.

89 The discrepancy can be explained by one paraphrase in Ina's process that she left unfinished and later deleted.

data set, in contrast to the corpus study, direct quotes were always embedded in paraphrases. There were no non-embedded or free-standing direct quotes. Overall, direct quotation was much more frequent in the process study than in the corpus study, in which 1% of sentences were direct quotes and 6.9% were direct quotes embedded in a paraphrase (see section 6.3). It is important to bear in mind that some direct quotes were deleted in the process and thus did not appear in the final version of the text produced in the process study. Since there was variability in the corpus study with regard to the use of direct quotes, it is nevertheless conceivable that the students who took part in the experiment simply had a preference for direct quotation (see further discussion in section 7.7.2.4). It is also possible that the experiment design and the fact that the source text was available on the screen had an influence on the students' behaviour.

The third most frequent type of activity is \*IW - RE-WRITING EXISTING SENTENCE, which occurs in 28 segments (15% of \*IW activities). In this activity, the L2 writers rewrote or restructured paraphrases they had completed earlier. This bears evidence to the recursiveness of the writing process in general, which is widely acknowledged in the literature (see e.g. Flower & Hayes 1981), and the recursiveness of the intertextual writing process in particular (see also Ruiz-Funes 1999). Despite the linearity of some L2 writers' approaches, all students returned to previously composed paraphrases at later points in their processes to make changes to wording and/or syntax or to otherwise improve upon them. Their approach to writing comprises the same processes of monitoring, elaborating, structuring, and planning observed in L1 writing (Stein 1990, cited in Ruiz-Funes 1999: 47).

Table 15: Intertextual writing activities.<sup>90</sup>

Pseudonym	Total no. of segments coded as *IW	*IW - Paraphrasing	*IW - Direct quotation	*IW - Re-writing existing sentence	*IW - Sentence left unfinished	*IW - Continuing unfinished sentence	No. of sentences in product
Arne	35	32	4	2	0	0	28
Ben	26	19	4	6	3	3	16
Eva	33	29	8	4	1	1	26
Ina	32	23	4	9	4	3	19
Nadine	28	25	9	3	0	0	22
Tara	30	26	4	4	2	2	23
TOTAL	184	154	32	28	10	9	134
PERCENTAGE	100%	83.7%	17.4%	15.2%	5.4%	4.9%	

Within the different activities, the L2 writers performed a variety of writing and non-writing-acts at the micro-level. Individual differences emerged with respect to both the kind and the number of acts, as outlined in the next section.

### 7.7.2.3 Writing acts and non-writing acts in the writing stage

The students' intertextual writing activities are composed of a total of 1694 acts at the micro-level, of which 947 (55.9%) are writing acts and 747 (44.1%) are non-writing acts. The acts in these two categories provide an insight into the general make-up of the students' writing processes and are discussed in turn below. Figures 26 and 27 show writing and non-writing acts by frequency; a detailed overview of these acts and their distribution across the six students' processes is found in Appendices 20, 21, and 22.

<sup>90</sup> The percentages exceed 100% because segments can be comprised of more than one activity, e.g. quoting directly and paraphrasing.

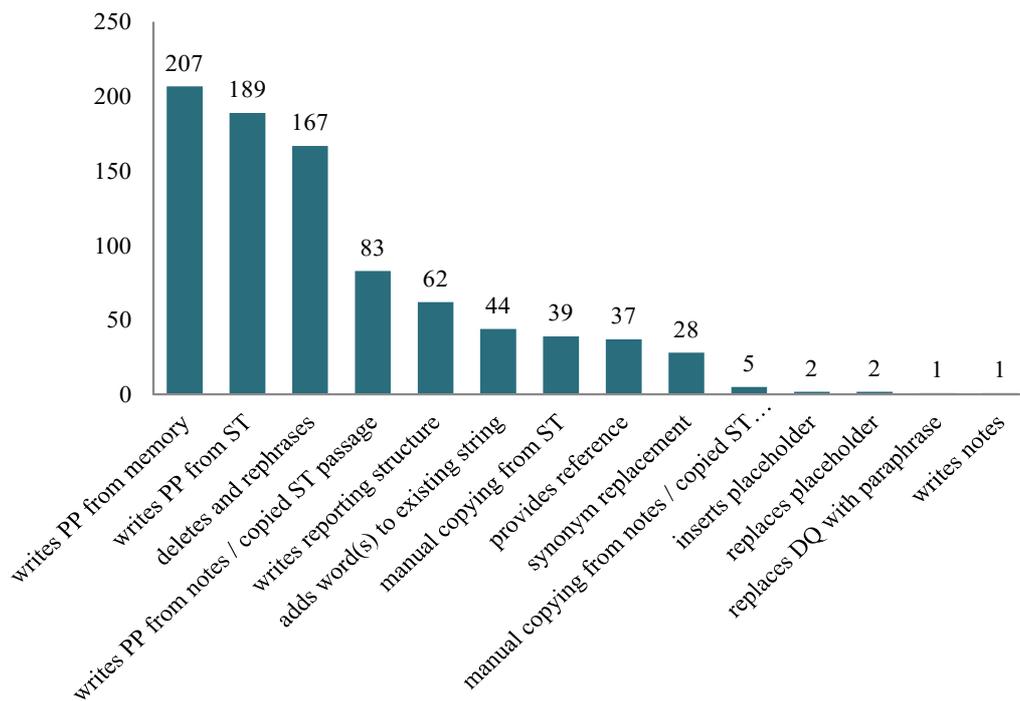


Figure 26: Writing acts – total number of instances in the experimental data.

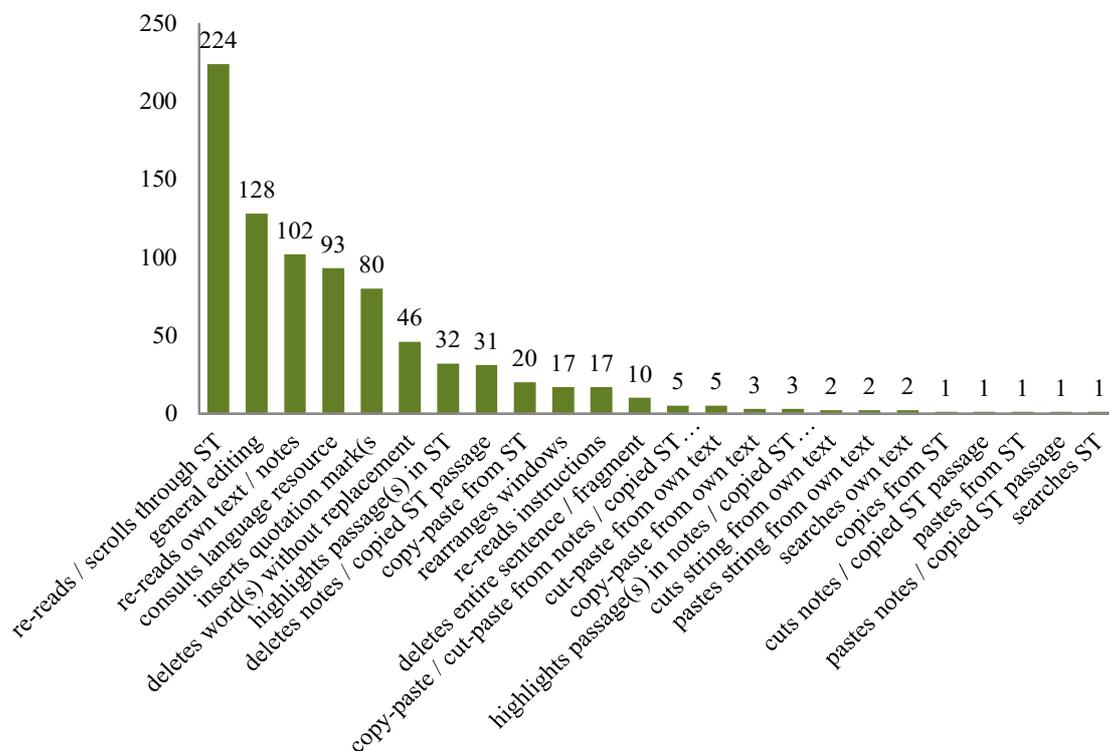


Figure 27: Non-writing acts – total number of instances in the experimental data.

The analysis shows that there is a core set of nine acts that are present in all six writers' processes (Figure 28). The occurrences of these nine acts comprise between 41 and 60% of all acts in the writers' processes. Among them are the writing acts ADDS WORD(S) TO EXISTING STRING, SYNONYM REPLACEMENT, and WRITES REPORTING STRUCTURE. They also include two writing acts related to deletion, namely DELETES AND REPHRASES and DELETES WORD(S) WITHOUT REPLACEMENT. In addition, all writers' processes share the non-writing acts RE-READS / SCROLLS THROUGH ST and RE-READS OWN TEXT / NOTES as well as GENERAL EDITING and INSERTS QUOTATION MARK(S). The most common of these core acts is RE-READS / SCROLLS THROUGH ST (224 instances in total), which is evidence of the significance of source text interaction in intertextual writing processes.

It is striking that the remaining acts are related to the central intertextual strategies of paraphrasing and direct quotation as well as to writing reporting structures and editing. It appears that these are indispensable activities and they are likely to be found in any process of source-based writing. The fact that writing paraphrase is not one of these core acts is the result of the coding scheme which takes into consideration the students' different approaches to paraphrasing and interaction with the source text. An overview of the absolute numbers of core acts per L2 writer's process appears in Appendix 23.

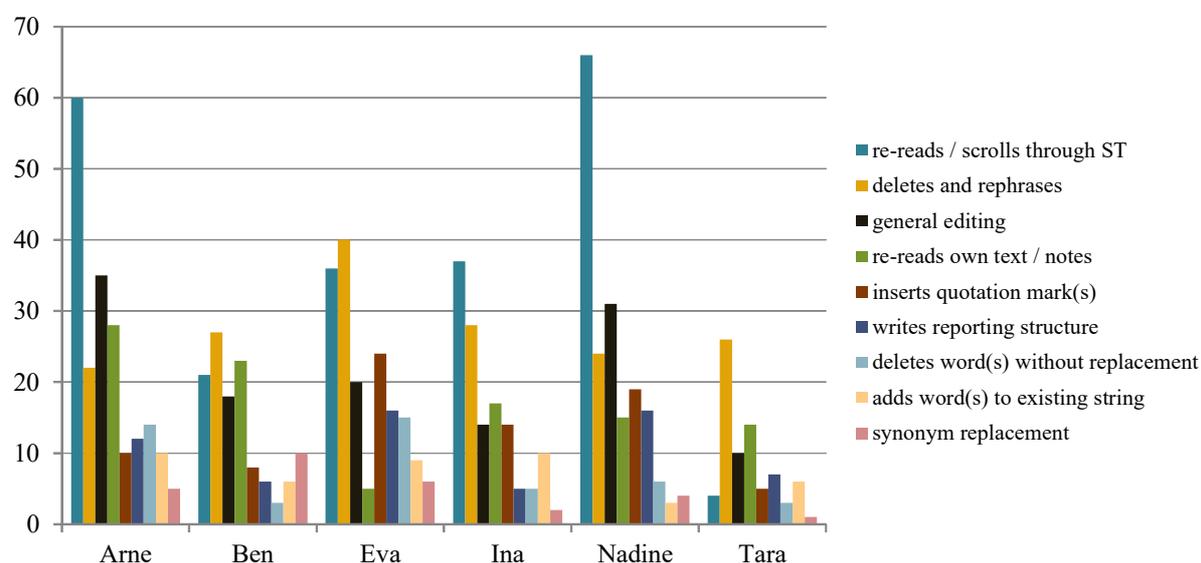


Figure 28: No. of instances of core acts in each L2 writer's process.

Writing paraphrase – from memory, from the source text, or from notes / copied source text passages – is by far the most frequent type of act in the students' writing processes. It occurs in 154 out of 184 \*IW segments (84%). Taken together, there are 479 instances of paraphrase writing, constituting 28.3% of all acts. There are notable differences in the distribution of these three writing acts across the students' writing processes (see section 7.7.2.4).

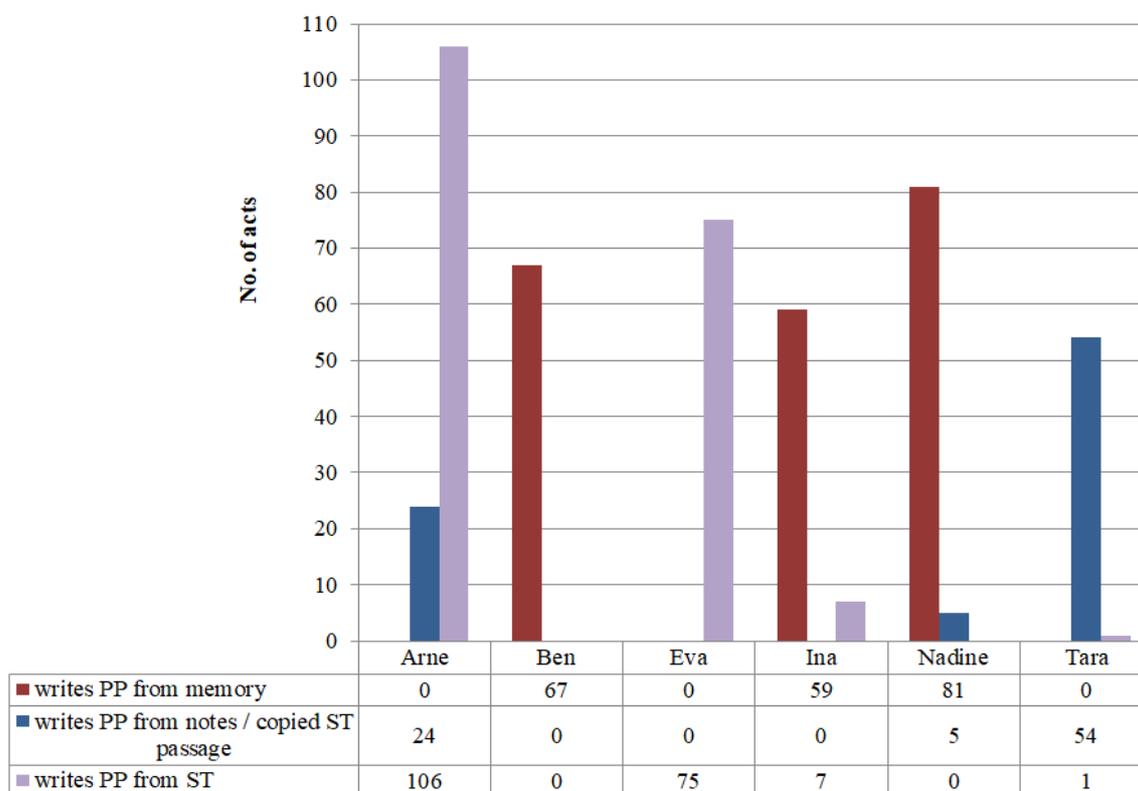


Figure 29: Overview of paraphrasing-related writing acts per student.

As Figure 29 shows, the most frequent writing act is WRITES PP FROM MEMORY, which was performed 207 times (12.2% of all acts). This act occurs in 67 segments in four students' processes, highlighting the fact that the majority of sentences were written without the source text visible. There are 189 instances of the act WRITES PP FROM ST in 59 segments (11.2% of all acts), making it the second most frequent writing act. This act was also employed by four different students. The least frequent type of paraphrasing act is WRITES PP FROM NOTES / COPIED ST PASSAGE, which the students used 83 times (4.9%). Together, this act occurred in only 40 segments because only three students used notes and copied source text passages as scaffolding when writing their reading reports.

It is interesting to note that Ben paraphrased exclusively from memory and Eva wrote all her paraphrases while being able to see the source text on screen. In contrast, Tara, Ina, Nadine, and Arne used different approaches for different paraphrases. This explains why five segments (2.7% of \*IW) contain two or more paraphrasing acts. This was the case for example when a student first wrote a paraphrase from their memory, but then re-arranged the windows on the screen so that they could see the source text and continued paraphrasing from the source text. Nevertheless, each student had a clear preference for one of the four writing acts: Tara wrote the majority of her paraphrases from copied source text passages; Ben, Ina, and Nadine wrote mostly from memory, and Eva and Arne directly paraphrased the source text for the majority of their paraphrases. A detailed analysis of the students' paraphrasing processes appears in section 7.7.2.4.

Aside from the paraphrasing acts, another common writing act was DELETES AND REPHRASES (167 instances across 96 segments; between 14 and 21 segments per process). It occurred in roughly equal numbers across all six writers' processes. While deleting and rephrasing is a common occurrence in any writing process, it often served intertextual purposes here. Passages were deleted and rephrased to convey the source text content using different words and avoid textual overlap. Other writing acts that are closely tied to paraphrasing were also

relatively frequent, namely ADDS WORD(S) TO EXISTING STRING (44 instances in 39 segments) and SYNONYM SUBSTITUTION (28 instances in 25 segments), both of which were employed by all six writers (see further discussion of paraphrasing-related acts in section 7.7.2.4). INSERTS QUOTATION MARK(S) was also relatively frequent (80 instances in 32 segments). Direct quotation is quite frequent in the data, but this remarkably high number of instances is due to the fact that each quotation mark was coded individually. Eighty instances, thus, are roughly equivalent to forty strings in quotation marks (see further discussion of the processes of direct quotation in section 7.7.2.4).

The act of writing a reporting structure was relatively common among the writing acts. There were 62 instances of WRITES REPORTING STRUCTURE in 46 segments from all six writers' processes. Referencing was less common, with 37 instances in thirty segments from four writers' processes. However, while the writing of reporting structures occurred in all writers' processes and the act WRITES REPORTING STRUCTURES was found in four to twelve segments per process, PROVIDES REFERENCE was used extensively by only one student, Tara (18 segments). Referencing was observed in only one to eight segments by three other students. The two remaining students did not provide any references in the process of writing the reading report. These findings are in line with those from the corpus study, in which reporting structures were also very common, while documentation was rare (see section 6.3.3).

MANUAL COPYING FROM ST occurred 39 times in 25 segments, and all but one writer (Tara) made use of this act. It must be emphasised that much of the manual copying served the purpose of creating a direct quote (see section 7.7.2.4). The same can be said for MANUAL COPYING FROM NOTES / COPIED ST PASSAGE, which occurred five times, once in a segment of Nadine's writing process and four times in segments of Tara's process. This is in contrast to a study by Li (2013), in which first-year students copied excerpts into their texts without rephrasing them. In some cases, Li's students provided references for such copied passages, but they were apparently unaware that copying is not usually acceptable in academic writing. Their questionable ideas regarding copying clearly influenced their writing processes. It appears that the writers in this study were aware of academic conventions and the need to acknowledge source text material in their texts. It is perhaps for this reason that copy-pasting from the source text – which may increase the risk of inadvertently retaining copied passages – was rare in their writing processes except for cases of direct quotation.

The most frequent non-writing act is re-reading the source text (224; 13.2% of acts). Students repeatedly re-read the source text when composing an intertextual link, with varying levels of source text access throughout the process. Source text access also varied in the process of a proficient student in the study by Leijten et al. (2019). In her process, reading was common during the initial phase, but only occurred occasionally in the remainder of the process. Interestingly, re-reading was not a very frequent strategy of novice writers in the study by Sasaki (2000), but it increased with proficiency and was most common in the experts' processes. It is possible that these L2 writers' perpetual use of the source text is an indicator of their advanced intertextual writing competence, but it may have been amplified by the experimental setup of the study.

Irrespective of whether or not they had taken notes or copied excerpts from the source text, all six L2 writers re-read the source text during paraphrasing, showing that recourse to the source text is indispensable for intertextual writing and constitutes a central component of the intertextual writing process.<sup>91</sup> All students accessed the source text at different points throughout the writing stage. They all sometimes scrolled back and forth in the text, usually to find a specific part of the text to paraphrase. Tara also scrolled through the text to remind herself of its length so that she was able to balance her text by basing it on different parts of the source text (excerpt 7.5).

---

91 Because some students had the source text visible on the screen aside their word processing document, only those re-reading processes that were accompanied by mouse movements across the source text and scrolling could be directly observed. The entirety of re-reading phases could only have been captured with the additional use of eye-tracking software. It is thus likely that a considerable number of pauses can be attributed to re-reading the source text in those parts of the writing process in which the student were able to see the source text on their screens.

- (7.5) TARA: I think I just looked how long the paper is to make sure that I don't like fill up my entire summary with the first couple of paragraphs and then leave out everything else [...]. I was just making sure that I don't fill up everything with explanations and leave out the analysis itself and the conclusion, to balance. (Tara's interview, 00:13:32-4 – 00:14:33-7)

The students generally reviewed the source text chronologically, referring to the abstract and the introduction first and then progressively moving on to the conclusion via the sections in which the study was presented (see e.g. Ben's interview, 00:19:48-6 – 00:20:13-4). In the interviews, several of the L2 writers indicated this as their strategy (see excerpts 7.2, 7.3, and 7.4), confirming the assumptions made in the corpus study. In addition, unsurprisingly, the L2 writers – at least sometimes – re-read the source text to check that their reading report accurately reported its content (see excerpt 7.6; see also Ben's interview, 00:04:38-0 – 00:05:24-4).

- (7.6) INA: [...] I think from working with the text in general you get a feeling for the topic. So you basically notice in your summary when you read it through if something is completely off or does not fit any more. Yeah. So that is kind of also how I check if everything was sensible that I wrote. [...] (Ina's interview, 00:16:10-1 – 00:17:22-5)

A recursive element in the students' writing processes was the combination of re-reading one's own text and general editing. This combination is well documented in the literature as a feature of the writing process (see e.g. Flower & Hayes 1981: 374). GENERAL EDITING, an act comprising the correction of spelling, punctuation, and grammar errors as well as formatting, e.g. inserting line breaks, occurred throughout all writers' processes (128 instances in 75 segments; see Appendix 20). The writers re-read their own texts or notes in 67 segments, with 102 instances of the act RE-READS OWN TEXT / NOTES overall. It must be borne in mind, however, that some re-reading may not have been observable due to a lack of mouse movement. It is likely that during many of the passages without action that occurred while only the student's own text was visible on the screen, they were in fact re-reading it. Figure 27 shows that the act RE-READS OWN TEXT / NOTES is slightly less common than GENERAL EDITING. This is because editing sometimes takes place immediately after a writer has finished writing a passage and is not necessarily preceded by reading. In addition, general editing is sometimes interrupted by the L2 writer consulting a language resource and then resumed, resulting in two or more GENERAL EDITING acts. The occurrence of these two acts during the writing stage further confirms the recursiveness of the writing process, which entails the re-occurrence of editing processes throughout.

Another common non-writing act is deleting. The act DELETES WORD(S) WITHOUT REPLACEMENT occurs 46 times in 34 segments. In 32 instances, notes or copied source text passages are deleted (19 segments), and in the remaining ten instances, an L2 writer deletes an entire sentence or an entire sentence fragment (10 segments). The act DELETES WORD(S) WITHOUT REPLACEMENT is found in the writing processes of all six participants and is evidently a common strategy in paraphrasing, but not all deleted entire sentences or sentence fragments. Obviously, only those students who had taken notes or copied source text passages into their document were able to perform the act DELETES NOTES / COPIED ST PASSAGE, so this act is only found in the processes of Nadine, Tara, and Arne.

Some writers continued engaging in processes which typically occur in the pre-writing stage. This is apparent from the fact that both HIGHLIGHTS PASSAGE(S) IN ST (32 instances in 22 segments) and COPY-PASTE FROM ST (22 instances in 8 segments) occur in the writing stage. Again, this finding can be ascribed to the recursiveness of the (intertextual) writing process, but also suggests that the transitions between the phases of the writing processes are fluid and there is overlap in terms of micro-level processes. Some copy-pasting from the source text served as scaffolding and orientation for paraphrasing, but other instances formed part of direct quotation processes, see section 7.7.2.4.

All writers except for Tara consulted language resources, usually online dictionaries and thesauruses. This indicates that their writing processes were similar to those by students observed in non-experimental settings (Choi 2016). Four of the writers in this study referred to language resources only occasionally to look up words or find alternative expressions during paraphrasing (see also section 7.7.2.4).<sup>92</sup> Unlike in a study by Bailey and Withers (2018), *Microsoft Word* language tools were largely neglected by these students. NW - CONSULTING LANGUAGE RESOURCE occurred in 49 segments, comprising 93 instances of this act overall. 65 of these instances occurred in segments of Arne's writing process, who made extensive use of online dictionaries when writing paraphrases despite his advanced language proficiency. It is clear from the students' use of bilingual dictionaries that these micro-level processes are mostly related to the fact that the students were writing in their L2, though L1 writers naturally also use dictionaries and thesauruses when writing. In addition to general issues of finding the right words that occur in any writing process, language learners' may additionally find themselves searching for words as a result of their deficits in the L2 (see Krings 1989 for a description of L2 problems in the writing process). Such issues are of course easily remedied by accessing language resources.

While the other writers usually accessed language resources to find the English equivalent for an individual German word or a synonym, Arne appeared to use language resources systematically to help him rephrase the source text. Arne confirmed that his intertextual approach was characterised by a heavy reliance on language resources. He stated that he usually knew the words he looked up, but that the visual aspect of seeing a list of possible translations helped him choose the right one for the context. He described this as a "thinking process", but expressed uncertainty as to whether this was a useful approach. He mentioned that his writing was quite different in exam situations in which he was not allowed to use a dictionary (00:32:11-5 – 00:33:15-2). Because previous research has suggested that the use of language resources is not necessary to produce high-quality paraphrases (Bailey & Withers 2018), it would be interesting to investigate in future studies whether an extensive use of language resources actually makes a difference in individual students' grades, especially at higher levels of language proficiency.

Certain acts were observed only in certain writer's processes, but not in the others'. These acts, which occurred less than five times in total, included the writing acts INSERTS PLACEHOLDER and REPLACES PLACEHOLDER, which only occurred in Ina's and Arne's processes (see section 7.7.2.4 for further discussion). Two non-writing acts, SEARCHES OWN TEXT (1 instance) and SEARCHES ST (2 instances), both of which describe acts that involved the use of the programme's search function, were only used by Arne. He used the search function towards the end of writing his reading report. WRITES NOTES also only occurred only once and only in Arne's process. This is because the other L2 writers who made use of notes prepared them during the pre-writing stage and then did not repeat this act during the writing stage.

Other acts that were very rare across all processes included REPLACES DQ WITH PARAPHRASE, HIGHLIGHTS PASSAGE(S) IN NOTES / COPIED ST PASSAGE (which was only done by Arne in three instances during two segments), and several acts of copying, cutting, and pasting (see section 7.7.2.4). Such low-frequency acts point towards individual approaches to the process of writing from sources. The low frequency of REPLACES DQ WITH PARAPHRASE, which appears counter-intuitive in the light of the copy-pasting observed in some processes, suggests that the students either paraphrased copy-pasted excerpts right away or settled on a direct quote. This act may be more frequent in other text types or in other writers' processes. The processes of creating intertextual links are explored in more detail in the following section.

#### 7.7.2.4 Intertextual writing processes

This section contains an analysis of the intertextual writing processes observed in the screen recordings. First, the processes of paraphrasing are examined. Further analyses focus on the processes of direct quotation

<sup>92</sup> See Bailey & Withers (2018) for a study of the use of software tools in the writing processes of L2 writers. Interestingly, in their study external resources were used only occasionally and most participants relied on the features of *Microsoft Word*.

and those of reporting and referencing. Finally, the occurrence of textual overlap and the strategic re-use of source text material are analysed.

### ***Processes of paraphrasing***

The analysis of the screen recordings brought to light that each L2 writer has a certain general approach to paraphrasing that is repeated – with some variation – throughout the writing stage. The processes of paraphrasing were to a large part determined by the setup the writers had chosen on their screens. Ben, Ina, and Nadine did not have their source texts visible on the screen during writing. Because they had to switch between windows to access the source text, their paraphrasing processes were generally quite similar. In all three of these writers' processes, writing paraphrase was alternated with re-reading the source text, an approach that has also been observed in Chan's (2017) study of the writing process in a reading-into-writing task. Despite this general commonality, certain individual preferences for the combination and succession of paraphrasing-related acts emerged.

Ben wrote his paraphrases mostly from memory without much editing or rephrasing. His process of writing a paraphrase consisted predominantly of writing, pausing, and re-reading. He usually either began by re-reading the source text and then paraphrasing from memory or by writing a short passage, for example a reporting structure, then consulting the source text, and then paraphrasing from memory. As he switched back and forth between the two windows throughout, the source text was his constant companion during paraphrasing, except when he was re-writing existing sentences. Aside from the source text, Ben also frequently re-read the instructions, usually at the beginning of a segment. During long sequences of his paraphrasing processes, he focused on writing, during which he paused frequently (e.g. segments 1:11, 1:14, 1:16, and 1:23). Towards the end of a segment comprising paraphrasing, he typically re-read his sentence and made changes, e.g. by adding words, replacing synonyms, and editing. He often deleted and rephrased; this act occurred in the majority of his paraphrasing processes.

As Ben approached the end of his writing stage, he performed the acts of re-reading the source text and general editing more frequently. After five paraphrases, he began re-writing his paraphrases, then alternated between writing new paraphrases and re-writing existing sentences. Ben's approach to the re-writing of existing paraphrases usually entailed synonym substitution; sometimes he added words and edited his sentences. There was only one case of extensive re-writing in which Ben deleted words, cut and pasted a string from his own text to change the order of constituents, added words, and edited the paraphrase. Ben was one of the writers who tended to leave sentences unfinished, usually to re-write another.

Ina also wrote her paraphrases mostly from memory. During the pre-writing phase, she had created headings from the guiding questions provided in the reading report instructions under which she now placed her paraphrases. She usually began a paraphrasing process by re-reading the source text. She sometimes highlighted passages in it, then began paraphrasing from memory. She re-read the source text from time to time while writing, but wrote some paraphrases from memory without any recourse to the source text (e.g. 5:2, 5:5, 5:20, and 5:21). Like Ben, Ina also tended to leave sentences unfinished to re-write existing ones or to begin writing a new one. She then returned to unfinished sentences several segments later. General editing occurred throughout Ina's paraphrasing processes; she also deleted and rephrased in some segments, but not in others. The re-writing of existing paraphrases featured relatively frequently in her process, especially at the end of the writing stage. When she re-wrote existing paraphrases, this usually entailed re-reading them and adding words to them. In several instances, she also returned to a paraphrase to add a reference.

Much like Ina's, Nadine's general approach to paraphrasing consisted of re-reading, highlighting, and then paraphrasing from memory. She usually began a segment of paraphrasing by re-reading the source text. In some segments, she highlighted passages in the PDF file. She then switched to her own text to write a

paraphrase. Usually, she later returned to re-reading the source text before continuing her paraphrase. Nadine switched back and forth between windows more frequently than Ina and Ben did. She wrote a part of a paraphrase, then re-read the source text, then wrote more, while Ina and Ben generally re-read only once or twice per segment or not at all. Some of Nadine's paraphrasing processes consisted almost exclusively of writing and pausing (e.g. 2:6). She also engaged in extensive general editing throughout, i.e. she often wrote a short passage and then corrected spelling errors. Deleting and rephrasing tended to occur towards the end of her paraphrasing processes. Unlike Ina and Ben, Nadine rewrote existing paraphrases relatively infrequently. In three instances of re-writing, she re-read her own text and then edited it, also making additional changes (deleting and rephrasing, synonym substitution, adding words). While Nadine initially only wrote paraphrases from memory, later in the process she began copy-pasting passages from the source text into her document, which she then used as the basis for paraphrasing. Towards the end of the writing stage, she returned to her usual strategy of writing paraphrases from memory with intermittent recourse to the source text.

The remaining three writers' paraphrasing processes were more varied, though certain similarities also emerged. Tara was the only writer who consistently paraphrased from source text excerpts she had copied into her document during the pre-writing stage. Her process was not linear, and it has previously been observed that adults' processes of turning notes, which are usually non-linear, into linear text are a matter of complex transformations (Bereiter & Scardamalia 1987: 15). A process segment usually ended with her providing a reference (e.g. segments 3:27 to 3:32). She deleted the copied excerpts after she had used them for paraphrasing or restructured them for later use. Tara also copy-pasted material from the copied excerpts into her paraphrases. Although she mostly relied on the excerpts during paraphrasing, Tara did consult the source text at times. When providing references at the end of a segment, she used the references she had added to her copied source text excerpts during the pre-writing stage. She frequently deleted and rephrased in the process of paraphrasing and sometimes also typed sentence beginnings which she subsequently deleted. Two other paraphrasing strategies – adding words to existing strings and synonym substitution – were used occasionally throughout, but there was very little general editing in her paraphrasing processes. Tara began re-writing existing sentences from very early on in the process and, like Ben and Ina, she occasionally left a sentence unfinished to re-write another.

Eva and Arne both had the source text visible at all times, meaning that they had direct access to it during paraphrasing. Arne frequently spent time on re-reading the source text in the process of paraphrasing. During pre-writing, he had copied excerpts from the instructions, in which he had also highlighted important sections. He used these for orientation during writing. It was striking that Arne combined writing paraphrase with extensive use of language resources and for this purpose switched back and forth between his document and the browser repeatedly throughout the writing process. Despite his heavy reliance on language resources, he also engaged in paraphrasing-related acts such as deleting and rephrasing, synonym substitution, and deleting without replacement. Like in Ben's process, these acts tend to occur towards the end of a segment. In the second half of the process, Arne switched to a different strategy and copy-pasted source text passages into his document. This approach was comparable to that taken by Tara during pre-writing. Arne used these passages as orientation for paraphrasing, but did not modify them directly. When he had rephrased their content, he deleted these excerpts. He only copied from them for direct quotation, not for paraphrasing. There were only two cases in which Arne re-wrote existing paraphrases. In one of these segments (quotation 6:13), Arne added words to an existing paraphrase; in the other (quotation 6:33), he thoroughly re-wrote the paraphrase, but then deleted it and replaced it with a new paraphrase.

Eva's paraphrasing processes consisted mostly of writing paraphrase from the source text interspersed with re-reading. During her many pauses, she was likely also re-reading the source text. Similarly to Arne and Ben, she consulted the instructions at times while paraphrasing. In other respects, however, her process was more varied: during some paraphrasing processes she deleted and rephrased strings, during others she added words, in yet others she deleted one or more words without replacement. Deleting and rephrasing appeared to become more common towards the end of her writing stage. Overall, Eva's process was less linear than Arne's as she

sometimes inserted new paraphrases between existing ones or copy-pasted her own paraphrases to a new place in the document. She also tended to delete (parts of) her own paraphrases and to conjoin sentences. Like Arne, Eva consulted language resources throughout the writing stage. She used them much less extensively and often during general editing, suggesting that these resources fulfilled a different function in her process than in Arne's. In some of her paraphrasing processes, she did use language resources to prepare synonym substitution, however. Like Tara, Eva began re-writing existing sentences at a relatively early stage, but only did so occasionally. There was only one case in which a paraphrase was left unfinished, but, as mentioned before, Eva restructured her paraphrases in other ways, for example by conjoining them.

As evident from the analysis of the individual paraphrasing processes, a range of acts were observed that appeared to be closely tied to the L2 writers' goal of rephrasing the source text. DELETES AND REPHRASES was common across all writers' processes. This act occurred in between 14 and 21 segments per process and in 52% of segments overall. Typically, deleting and rephrasing was observed in the flow of writing, i.e. the student wrote a passage, then deleted a part and subsequently continued typing, thereby replacing the deleted string. This approach has been observed previously in a study by Chan (2017) and is clearly related to the students' desire to reword source text material and minimise textual overlap. Deleting and rephrasing frequently occurred at the beginning of a sentence. In such instances, the student began writing, but then apparently changed their mind about the sentence beginning. They then deleted what they had written and began writing the sentence anew. In segment 5:30, Ina began her sentence with the phrase "Often researchers", then deleted it and replaced it with "Most of the time researchers".

This type of 'false start' occurred in the writing processes of all six participants. It appears that L2 writers plan the structure of their sentences right at the beginning and at least some of these deletions occur because different options are considered. Similarly, when adding words to existing strings, the students sometimes began writing, then deleted the new string and wrote a different string. In the same citation (5:30), Ina later added a subordinate clause, which she began as "which is a shame", only to replace it with "which is a loss for their study", thereby opting for a more formal expression. As in Chan's (2017) study, some students deleted more extensively than others, and only some deleted entire sentences or sentence fragments in their processes.

Synonym substitution occurred in all six writers' processes and was found in one to eight segments, with a total of 28 instances. In those cases in which synonym substitution was preceded by the consultation of language resources, the selection of an appropriate synonym seemed to be systematic and selective, unlike in the study by Bailey and Withers (2018), possibly due to the more developed academic literacy and/or language proficiency of the L2 writers in the present study. The majority of substituted words were nouns or noun phrases (10) and verbs (9). Most replaced verbs were related to research and reporting. Replacement of adjectives (3), prepositions (2), numbers (2), subordinators (1), and adverbs (1) also occurred. The range of word classes represented among the substituted words suggests that synonym substitution is not exclusively used to minimise similarity to the source text. Replacement of synonyms often appears to be prompted by stylistic concerns as well as considerations of the context and the semantic fit of a word.

This assumption was confirmed in the interviews, in which the writers described being led primarily by stylistic motivations when replacing words with synonyms. For example, when asked about synonym substitution, Ben explained having done so because he had remembered the content of the source text and felt that the new word expressed it more appropriately. He felt obliged to use synonyms not because he was eager to minimise similarity between his and the source text but because he wanted to achieve sophistication in his writing style. Similarly, Eva used a thesaurus to find synonyms in one process segment, which she explained was "not really about the content, more about the sound of the word" because she thought there may be a word that was more suited to the respective context and her intended meaning. In another segment, she consulted a thesaurus because she was afraid the word she had used might be too informal. However, unlike Ben, she also used synonym substitution as a paraphrasing strategy to avoid copying (Eva's interview, 00:08:38-7 – 00:11:04-9). The reading report instructions clearly played a role in this decision (see excerpt 7.7).

- (7.7) EVA: [...] I think there was ‘examining’ or something. And that is used in the text as well. And so I decided/ Because in your reading report instructions there was written that we should not copy anything from the text. So I decided to use as much of my own words as possible. And in case of ‘examining’ it is easy to change it to a different word with the same meaning basically. So I did that. (Eva’s interview, 00:09:41-8 – 00:10:19-5)

While she replaced *examine* in one sentence because it occurred in the source text, she did use it in a later one, explaining that she strove to be “as diverse as possible” in her expressions (Eva’s interview, 00:10:19-5 – 00:11:04-9). Stylistic motivations thus overrode her ambition to use words that were different from those of the source text. A similar statement was made by Arne, who also did not want to use the verb *examine* because it occurred in the source text. In replacing it, he considered the meanings of several reporting verbs and first replaced it with *compare* before settling on *explore* because he considered its meaning to be general enough for his purposes (Arne’s interview, 00:25:16-8 – 00:25:58-8).

Nadine decided against synonym substitution and re-used an expression from the source text because this seemed easier to her than to search for a synonym that may not be as suitable. She also stated that she sometimes replaced her own words in her text with synonyms from the source texts that she remembered during writing (example 7.8), an approach that stands in contrast to those of the other L2 writers. Their statements demonstrate that the L2 writers did not use synonym substitution solely to minimise textual overlap. This strategy was often related to concerns of style rather than plagiarism.

- (7.8) NADINE: [...] I phrase the sentence in my head and look if it fits it perfectly and if not then I have a look at the text again and look if maybe they have used a word which is fitting best. [...] I will write the sentence myself and if I know that in the text there was a word that fits perfectly, then I use it. (Nadine’s interview, 00:06:37-4– 00:07:17-3)

During coding, it was observed that some of the instances of DELETES AND REPHRASES also comprised synonym substitution, i.e. an L2 writer deleted a passage and used a synonym when rephrasing it. This entailed not necessarily a replacement of an individual word, but rather the ad-hoc reformulation of source text passages. In many such cases, a word was deleted before it was written in full, so that one would have had to guess to identify the original word and to establish whether the new word was in fact a synonym. For this reason, these were not considered to be cases of synonym substitution. Nevertheless, this finding is still relevant as it shows that unlike suggested by the corpus data, synonym substitution may in fact be the result of two different processes – the replacement of a word with a synonym in the flow of writing and the replacement in a separate act after a string had been written.

Synonym substitution also occurred in the post-writing stage when students edited their texts, though there were only very few such instances. It was evident that the students usually did not refer to the source text in such situations. Their synonym substitution in the post-writing stage seemed to be guided by stylistic rather than intertextual considerations. Nevertheless, this observation is interesting insofar as it points to the continuation of the recursiveness in the writing process into the final stage of writing, even if it is not as pronounced and systematic as in the writing stage.

The analysis of synonym substitution and the corresponding comments by the writers who participated in the process study provides a more nuanced view when compared to the findings of corpus-based studies that have assumed synonym substitution to be a grammatical strategy used in paraphrasing (e.g. Keck 2010). It shows that paraphrasing is a process of writing, deleting, and rephrasing which may result in the substitution of words. The presence of synonyms is thus not necessarily the result of a conscious decision by the writer to replace an individual word, but rather of a general motivation to use one’s own words and minimise similarity to the source text in the writing process. Isolated processes of synonym substitution often serve stylistic purposes, not intertextual ones.

Inserting a placeholder in the writing process is a strategy a writer can employ when they are (momentarily) unable to produce the expression they are looking for or need further information to fill a gap in their text. Interruptions of the writing process due to language issues are typical of writing processes in an L2 (Krings 1989: 397). Placeholders apparently helps L2 writers cope with such vocabulary gaps. They can include symbols, blank spaces, and (nonsense) words from the writer's L1 or L2 (see Colyar 2009 for a description of the author's own placeholder use). The insertion of placeholders as a paraphrasing strategy is used by Ina and Arne, each of whom employed it once in their respective processes. In Ina's case (quotation 5:20), the act of inserting a placeholder was preceded by a pause of more than fifty seconds. Though looking for an adjective, she then proceeded by inserting a verb in German followed by a question mark in parentheses. She only returned to the placeholder twelve segments later (quotation 5:32). There was again a pause – perhaps because she searched for the right word – before she finally replaced the placeholder with an English adjective. She did not consult a dictionary.

In the interview, Ina explained that she was unsure of the word in English and used the placeholder because she did not want to interrupt her flow of writing. This strategy allowed her to write down her ideas without being distracted by a missing word, though she sometimes looked up words immediately. She stated that she had originally intended to look up the missing word, but later simply replaced it by one that came to mind. She claimed to use this strategy frequently to not lose her train of thought over a missing word (Ina's interview, 00:31:11-8 – 00:32:08-3). Arne, on the other hand, inserted and replaced the placeholder in the same process segment. Before inserting the placeholder for a noun, he spent several minutes consulting language resources to find a word. When several searches still did not produce any satisfying result, Arne re-opened the source text, which was followed by a pause in which he was probably reading. He then inserted three dots (...) as a placeholder into his text and continued writing. Five words later, he again paused, then performed another search for a suitable German-English dictionary, and then replaced the three dots with a noun.

In the interview, Arne also explained that he was unable to retrieve a suitable word from his memory at the time. As he had the entire phrase in mind, he did not want to pause to find an expression. He later disclosed that he was looking for a noun that was synonymous to a specific one used in the source text which he did not want to re-use (Arne's interview, 00:22:25-6 – 00:23:47-3). In his case, it was the desire to avoid textual overlap, not a vocabulary gap, that prompted the use of a placeholder. This shows that the use of placeholders can be paraphrasing-related and not caused by a lack of linguistic resources. Thus, the use of placeholders is a processing-related, cognitive strategy. It can serve the purpose of avoiding a break-down of the flow of writing, not only in general, as suggested by Colyar (2009), but also in intertextual contexts.

The writers in this study were cautious about copying from the source text and, with the exception of Ben, cited concerns over plagiarism as a major reason. Tara emphasised that she was worried of plagiarism “all the time” during the experiment. She was mostly afraid of accidentally copying something because she still had it in her mind (excerpt 7.9). She also voiced a fear that the risk of accidentally re-using source text material would increase if she had to write longer texts (Tara's interview, 00:28:18-0 – 00:30:07-5). For the other students, the fear of plagiarism seemed to be less pronounced. Eva stated that she was not worried about plagiarising, but bore it in mind during writing. She explained trying to paraphrase the term *institutionalisation* from the source text because she was afraid of committing plagiarism and felt it was necessary to paraphrase this expression. She also wanted to avoid quoting directly because she thought this to be inappropriate in a reading report (Eva's interview, 00:20:51-8 – 00:22:16-0).

- (7.9) TARA: [...] Sometimes you are so sure that you came up with that sentence but [...] actually you read it in like a book two weeks ago and you might end up using it and then in the end it's like it's not yours, you forgot to quote, you forgot to mention the source. [...] (Tara's interview, 00:28:56-3 – 00:28:56-8)

Arne only considered plagiarism when deciding whether or not to use direct quotes, pointing towards his insecurity regarding what is appropriate and inappropriate re-use of source text material. At the same time, he stated that because of the repeated warnings of plagiarism in his courses, he was used to paraphrasing. He also admitted that he sometimes stayed quite close to the source text but generally tried to use his own words. Like Eva, he was not worried about committing plagiarism because he had never had any issues with it. He pointed out that he nonetheless paid attention to this issue (Arne's interview, 00:10:53-6 – 00:12:10-0; 0:17:17-8 – 00:17:50-0). Ina also explained that she always had plagiarism in the back of her mind when writing from sources. She gave the fact that she was reporting someone else's words as the reason why she found paraphrasing difficult. Even if she provided a reference and thus signalled the intertextual nature of the excerpt to the reader, "it is still difficult to use other words that are as efficient as the ones already in the text" (Ina's interview, 00:18:55-0 – 00:20:15-1). Her personal solution to this issue was re-writing excerpts completely (excerpt 7.10).

(7.10 ) INA: It helps when you completely, COMPLETELY rephrase it. Not just exchange words but maybe combine it with another idea or just TRY to grasp the main idea and then write a passage on that main idea. And not just take a sentence and try to rephrase that one sentence because that often does not really work properly. (Ina's interview, 00:20:18-4 – 00:20:47-7)

These testimonials show that the writers are cautious about copying so as to avoid plagiarism, but are not generally afraid of committing plagiarism because they are confident in their ability to paraphrase sufficiently well. They also confirm that these writers are aware of academic conventions and of strategies of avoiding inappropriate textual overlap.

In sum, the writers approached paraphrasing in different ways. These differences are evident from their use of the source text and their employment of other paraphrasing-related acts, e.g. deletion and synonym substitution. Despite these differences, similar motivations and conceptualisations of paraphrasing and of plagiarism were revealed in the interviews. Copying played only a minor role in these students' paraphrasing processes. Where it was employed, it was often used consciously and the students weighed concerns of plagiarism against other considerations, e.g. those of style. Their comments made it apparent that they reflected their paraphrasing practices in the light of warnings of plagiarism and were generally aware of the potential implications of extensive textual borrowing.

### ***Processes of direct quotation***

Direct quotation occurred in all six writers' processes. 32 segments were coded as \*IW - DIRECT QUOTATION overall, with between three and nine such segments per student. All of the direct quotes observed in the students' writing processes were embedded in paraphrases; there were no instances of direct quotes embedded in reporting structures. This stands in contrast to the corpus study and is probably due to personal preferences. It may have been influenced by the experimental setting as it is possible that reporting structures were felt to be more difficult and were thus avoided in this task. Language proficiency may have also played a role, as this factor was tightly controlled in the process study, but not in the corpus study.

The process of direct quotation was relatively uniform across this group of writers. The typical approach was to insert quotation marks, then to manually input the material to be quoted, and to insert another set of quotation marks after the quoted passage. It was rare for a student to manually copy a passage from the source text and then add quotation marks later (but see e.g. Nadine, segment 2:21). There were also instances in which the L2 writers initially forgot the second set of quotation marks and added them later (see Tara, segment 3:24; Nadine, segment 2:17). Within the activities comprising direct quotation, the act INSERTS QUOTATION MARK(S) occurred 80 times. Even if divided by two – since quotation marks are inserted on both sides of direct quotes – this still equates to more than 32, which means that some segments contained more than one instance of direct quotation.

Such combined quotes were also observed in the corpus study. In some instances, the L2 writers wrote direct quotes which they later deleted (see discussion below). The only instance of a direct quote being replaced with a paraphrase was found in Arne's process. The reverse process, replacing a paraphrase with a direct quote, did not occur in the data, but this had been observed in one of the screen recordings used for the development of the coding scheme, so it is a possible act. There was also one direct quote in Arne's process that he had copied completely with quotation marks from a source text excerpt he had previously pasted into his reading report (segment 6:34).

It emerged from the corpus study that many direct quotes contained terminology and definitions from the source text. The reasons for quoting these linguistic units are thus a point of interest in the process study. The majority of direct quotes produced in the process study consist exclusively of terminology. This is a surprising finding since terminology constitutes only 20% of direct quotes in the corpus study (see section 6.3.2.3). The experimental setting, the students' language proficiency or their personal preference may have prompted this difference, as the texts of individual writers in the corpus also contained more quoted terminology. The students in the process study usually added quotation marks around terminology immediately and keyed them in confidently in a fluid movement. Hence, they did not stop to deliberate whether the use of quotation marks was appropriate with terminology. The L2 writers in the process study, like those whose texts were analysed in the corpus study, typically quoted the term *eco*, which was the morpheme investigated in the source text, and its derivatives, e.g. *ecology*. As in the corpus study, this seems to stem from the students' insecurity regarding academic conventions and regarding what to place in quotation marks, which has also been observed in other studies (see e.g. Petrić 2012).

Although quoting definitions from the source text was considerably more frequent in the corpus study, this was also observed in the process study. One motivation for quoting the definitions seemed to be that the students found them difficult to paraphrase. Tara decided against quoting all six definitions since the authors only discussed three (Tara's interview, 00:18:21-9 – 00:18:54-2). In her reading report, she embedded these three definitions into her paraphrases by the use of single quotation marks because she was unsure of how to paraphrase them (Tara's interview, 00:49:31-6 – 00:50:03-5). Ina also admitted finding it difficult to rephrase the definitions because they seemed to be general in nature. She surmised that the definitions were not formulated by the researchers but possibly taken from a dictionary and had hence "existed before" (Ina's interview, 00:22:45-2 – 00:23:26-0).

She refrained from paraphrasing them because she considered them "their [the authors'] definitions" and felt she should maintain the original wording. According to her, rephrasing would have been difficult because they were "so precise", so it "did not really make sense to do so" (Eva's interview, 00:33:08-7 – 00:33:32-5). Arne articulated his motivation more clearly. First, he attributed having quoted the definitions to laziness, but then explained that he was afraid of changing the meaning (excerpt 7.11). A similar explanation was given by Tara, who emphasised that she generally avoided quoting directly (excerpt 7.12).

- (7.11 ) ARNE: I think this was actually just laziness [...]. And I excused myself with/ because I thought if I change this, which is put in quotation marks in the original as well, if I change THIS, the meaning would change and I would need to give specific information. [...] Because it is facts and if I might change this definition of the meaning, it would get a different notion. So I just thought, well, I am not going to change that. I am just including it in my own document. [...] (Arne's interview, 00:42:10-7 – 00:45:21-2)
- (7.12 ) TARA: I tried to avoid quotes as much as possible. The only time I used quotes was, I think, three times when I talked about the different meanings of the word *eco* because I was not sure how to paraphrase that. And I don't know, like honestly, how do you paraphrase *environmentally friendly*? (laughter) (Tara's interview, 00:27:32-9 – 00:27:39-2)

This makes apparent that the writers considered the definitions to be different from other quoted passages. Because of this perceived difference, the students assumed that the wording of the definitions could not or should not be changed, which was the reason why they opted for direct quotation. The interview data thus confirmed the interpretation of the students' preferences given in the corpus study.

The students' perception of terminology and definitions as different also influenced their choices regarding referencing. Eva quoted the definitions because they were also found in section headings and did not consider providing a reference because, as she explained, she tends to only give them for "direct quotes from the text" (Eva's interview, 00:32:20-6 – 00:32:43-0). This confirms that she, too, considered the meanings to be different from other quoted material. She admitted that for the definitions she had simply not thought about referencing. In fact, she stated that she often forgot references for terms and that she had also forgotten them here, thereby putting the quoted definitions in one category with terminology. Eva explained not having given references for the term *eco* because it appeared throughout the source text, so she assumed quotation marks were not required. It is likely that the scarcity of references in these L2 writers' processes is in part caused by their tendency to quote terminology, which they did not consider to require a reference.

This process study also lends support to the assumption that L2 writers' insecurities regarding the use of quotation marks lead to atypical uses (Petrić 2012; Verheijen 2015). Eva, for example, explained that she had placed *eco* in quotation marks because it had been set in italics in the source text. It apparently did not occur to her to use italics in her own text. Similarly, Ina changed the quotation marks around *ecological* from double to single ones. She explained that she felt that *ecological* was "more a phrase but not a direct quote" and that it referred to a concept. She added that she would have used double quotation marks if it had been a direct quote because "it is other people's words" (Ina's interview, 00:21:06-7 – 00:21:35-7). This points towards an intuition on Ina's part with regard to the distinction between terminology and quoted passages and the different ways of marking such passages in one's text.

It appears that uncertainty regarding quotation marks stems from the range of ways of indicating terminology in a text rather than from insecurity regarding how to signal direct quotes. Ina correctly identified the authors of the source text as the originators of direct quotes, but not of the terminology they used. Hence, she used different quotation marks for terminology and quoted passages, a practice that is unlikely to occur in published academic writing but nevertheless serves to distinguish terminology from direct quotes. Tara, who used single quotation marks, first explained this in reference to aesthetics as she felt double quotation marks looked "bulky". She then conceded that she would have used double quotation marks in another context, but that she had decided on single quotation marks for the definitions because she struggled to paraphrase them (Tara's interview, 00:49:31-6 – 00:50:03-5). As suggested above, the writers' insecurity with regard to citation conventions and the demarcation of terminology from other source text material is a probable cause of inconsistent or unusual uses of quotation marks, e.g. with terminology and definitions. While such insecurities can easily be addressed in teaching, these students had apparently not received sufficient assistance. In contrast to Petrić's (2012) study, the findings presented here suggest that the use of quotation marks with terminology is not the result of insecurities regarding direct quotation.

A noteworthy finding is that the L2 writers used both copy-pasting and manual copying in the process of creating direct quotes. Nadine copy-pasted the title, which she placed in quotation marks, in one segment (2:2). Arne, who had copy-pasted source text excerpts into his document during pre-writing, also used copy-pasting to create direct quotes in his writing process (segments 6:5, 6:34). In two other segments, however, he also inserted material manually in the process of writing a direct quote. Both Nadine and Tara, though writing from source text passages they had copied into their respective documents, also manually copied from the source text when creating a direct quote (segments 2:17, 3:19, 3:24, 3:26). Manual copying is a likely explanation for errors in direct quotes, which are of course even more likely to occur if the source text is not available in electronic format. This manual approach may be the explanation for the errors found in direct quotes in the corpus study.

Much more surprising, however, is the finding that not all students used the source text when quoting directly. In fact, only Eva and Arne consistently had the source text displayed on the screen when quoting directly, and Tara based her direct quotes on source text excerpts she had copied into her document. Ben, Nadine, and Ina, however, did not have the source text visible on screen when quoting directly. Nadine and Ben checked the source text and then reproduced the quoted elements from memory. They did not always confirm the correctness of their direct quote by reading the source text afterwards. In the interview, Nadine confirmed that she memorised passages from which she created direct quotes (00:16:57-8 – 00:17:01-7). While the majority of passages quoted from memory were terminology, Ina also relied on her memory when quoting definitions (segments 5:10, 5:16, and 5:18). Though she re-read the source text before quoting, she did not check her quoted definitions against the source text afterwards. Such observations can also explain errors in direct quotes of L2 writers (see Verheijen 2015). Novice writers should thus be reminded to compare their direct quotes with the source text to confirm their correctness, especially when ‘quoting from memory’.

Despite their insecurities concerning the formal characteristics of direct quotes, the writers’ reasons for using direct quotation resembled those postulated for expert writing. According to K. Hyland (2004: 26), expert writers are only likely to use direct quotes if these are the most suitable way to introduce an idea into their text. Nadine justified her use of a direct quotation by stating that “the sentence was perfect”, which was the reason why she “did not want to paraphrase it because it just made the most sense to use it in whole” (Nadine’s interview, 00:17:07-1- 00:17:52-4). She disclosed that she had indeed considered paraphrasing it, but that the sentence sounded better as a direct quote. Ina, who only quoted definitions (see further discussion below), asserted that she would have used another direct quote if she had found another passage that she “could not have rephrased in any way possible or that would not have sounded as well” (Ina’s interview, 00:44:02-1-00:44:44-6). This shows that stylistic considerations similar to expert writers’ play a role in L2 writers’ use of direct quotation. These findings corroborate those by Petrić (2012), in whose study the L2 writers also explained their decisions to quote by their desire to vividly express an idea and achieve stylistic variety. Unlike in a study by McCulloch (2012: 60), the direct quotes created by the students in this study served a purpose. Though this finding may be connected to the text type reading report, it is also indicative of the students’ advanced academic literacy.

Previous research has shown that L2 writers sometimes prefer direct quotation because it is easier to achieve and bears less risk of inadvertent plagiarism than paraphrasing (e.g. Hirvela & Du 2013). In this study, there was evidence of L2 writers writing paraphrases instead of direct quotes for stylistic reasons. Nadine explained having paraphrased some of the definitions from the source text in one of her sentences because she did not want to use too much space by explaining each one. She did quote all six meanings in the following sentence. Since she included an explanation of a semantic change the morpheme *eco-* had undergone, the combination of paraphrase and direct quote allowed her to provide the reader with additional information without being redundant (see Nadine’s interview, 00:26:59-7 – 00:28:00-7). This finding points towards varying motivations for the use of direct quotation, possibly induced by students’ writing proficiency.

Some of the writers consciously decided against the use of direct quotes in their reading reports. Teachers’ attitudes paired with writers’ conceptualisations of the text type reading report play a role in such decisions. Ben did not use any direct quotes, Arne wrote one direct quote and then replaced it with a paraphrase, and Eva and Ina only quoted terminology and the definitions from the source text. Ben explained having consciously avoided direct quotation because he did not like it. He attributed this to his former teacher having told him that direct quotes are “not that appropriate” in summaries (Ben’s interview, 00:29:54-8 – 00:30:34-3). A similar motivation emerged from the interview with Eva, who placed passages in quotation marks in eight segments. She stated that she generally avoids the use of direct quotes, especially if she is unsure if this is appropriate for the text type, because some of her teachers did not like direct quotes. As a result, she did not really think about direct quotation other than with terminology such as *eco*, which she decided to quote “because you kind of have to” (Eva’s interview, 00:22:21-9 – 00:22:35-6). Ina exclusively quoted the definitions and conceded that she was unsure

whether direct quotation was allowed in a reading report. She found this difficult to decide because there are “SO many different opinions on what you are allowed to do in summaries and what not”. For this reason she opted for paraphrasing, which she generally tried to do wherever possible (Ina’s interview, 00:44:02-1 – 00:44:44-6).

These findings further emphasise the considerable influence of teacher advice on students’ intertextual strategies, at least in some students. They mirrors the findings of Davis (2013), in whose study a postgraduate student reported having avoided direct quotation to please his tutor. These L2 writers’ lack of concrete knowledge of the purpose of direct quotation and their eagerness to meet their teachers’ preferences clearly affected their attitudes towards and use of direct quotation. For some L2 writers, the advice of both secondary school and university teachers appears to be a crucial factor in their understanding of direct quotation in the early stages of their higher education.

In addition to teacher advice, the fact that they were writing a reading report for the first time played a role in the students’ quoting behaviour. As mentioned above, Arne decided against keeping a quote he had already written because of his interpretation of the reading report instructions, thereby foregrounding genre-related considerations. He also admitted to feeling unsure of what to quote when asked why he had copied a string of words from the source text into his text (excerpt 7.13).

(7.13 ) ARNE: [...] Because it is again a specific term that I did not want to change. Those are parts of which I am not sure whether this is already a quotation, whether I would need to mark it as a quotation or not. It is difficult to decide here. Would you mark it as a quotation or not? (Arne’s interview, 00:53:09-9 – 00:53:38-1)

His question at the end suggests that he was unsure whether his decision against quotation marks was appropriate, though this insecurity may have been aroused by the interview situation. He claimed that time pressure played a role and that he might have changed the copied string when re-reading his text a few days later had it been a graded assignment (Arne’s interview, 00:53:42-1 – 00:54:25-5). Similarly, Tara did not use direct quotes because she was unsure if this was expected in a reading report. For this reason, she relied on what she had been taught about summaries. She shared Ina’s awareness of different conceptualisations of intertextuality by teachers (compare the studies by Crocker & Shaw 2002; Roig 2001; Shi 2012) and voiced her frustration with teachers who did not follow the rules themselves or invented their own. She also emphasised that she was afraid of facing accusations of plagiarism if she did not mark material as quoted that she considered to be general knowledge (excerpt 7.14). For this reason, she used copied source text excerpts so that she could directly compare her paraphrases to the original sentences. It is clear from her interview that she felt she did not have enough teacher guidance with regard to the expected forms of intertextuality in a reading report.

(7.14 ) TARA: [...] my professors, they do not allow any quotes in your summary because it’s a summary and you should basically just explain what the main thesis of the text is and like what are the argumentations. But there are also other professors who want you to quote from the text. [...] I was also unsure of the quotations because it’s already really confusing for a normal summary because there are a lot of rules but some people don’t like to stick to the rules and make up their own rules (laughter) and again for the reading report, I don’t know what the rules are [...] maybe in a reading report you HAVE to use quotation (laughter) I’m not sure so that’s why I asked. [...] So those are like dangerous areas where you would say everyone says that, should I say that as well and not put it in a quote and maybe risk that someone who reads my texts says that it’s plagiarism, or should I try to paraphrase it? [...] (Tara’s interview, 00:25:02-3 – 00:28:18-0)

Overall, the analysis of direct quotation in the process and interview data shows that the writers in this study approached this intertextual strategy in very similar ways, despite different preferences with regard to the use of the source text. The assumed insecurities described in corpus studies are confirmed by these writers, but the process analysis also confirms that L2 writers at this stage of their studies are proficient users of direct quotes. Their statements disclose expert-like motivations for quoting, though teachers’ advice and lack of

knowledge of genre continue to play a major role in students' decisions with respect to the verbatim re-use of other authors' words.

### *Processes of attribution, reporting, and referencing*

The L2 writers' awareness of the academic convention of attribution in source-based writing is reflected in their writing processes. They used attribution, reporting structures, and referencing to signal the intertextual nature of their texts. Much like their peers whose texts were analysed for the corpus study (see section 6.3.3), all six participants made use of reporting structures, though to different extents (see Figure 30). The number of segments that contain processes of writing a reporting structure range between 4 and 12, with a total of 62 instances of WRITES REPORTING STRUCTURE overall. Nadine, in whose process the writing of reporting structures occurred in 12 out of 39 \*IW segments, made the most extensive use of this way of making intertextuality explicit.

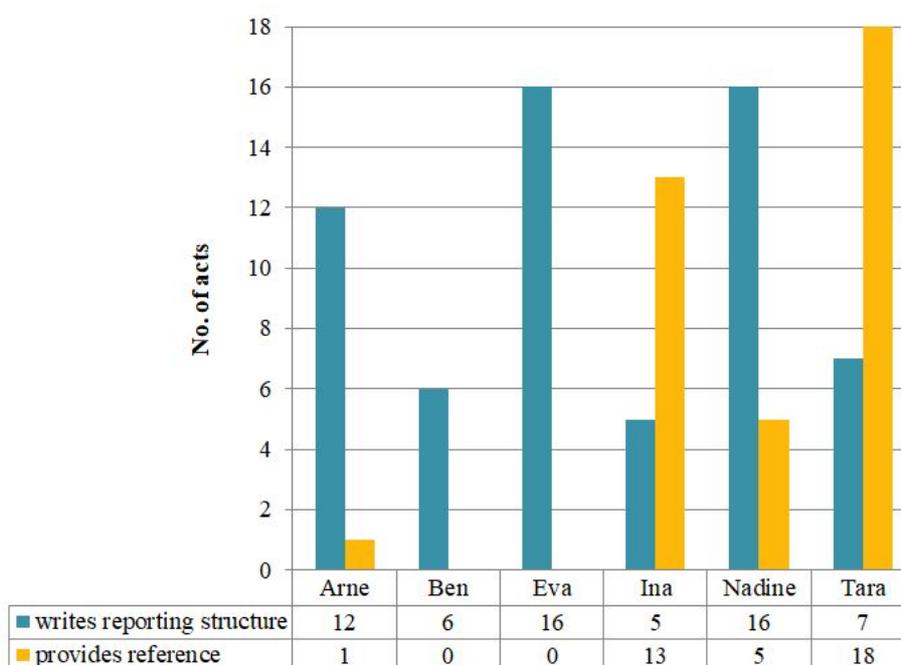


Figure 30: Acts related to reporting structures and documentation of sources in each student's process.

Reporting structures are usually written in a fluid motion as part of the process of writing a sentence and mostly occur at the beginning of sentences, as in the corpus study. Sometimes, a pause after the writing of the reporting structure indicates that the student stopped to consider how to formulate the reported statement. There are no instances in which a reporting structure was added to a sentence at a later point, showing that writing reporting structures has been manifested as an integral component of these L2 writers' processes of producing intertextual links. Reporting verbs are frequently subject to synonym substitution throughout the writing process, which is reflective of the importance the L2 writers attach to these words. The stylistic variety observed in the corpus study is likely to be a result of such substitution processes caused by the students' desire to use a range of reporting verbs and structures.

In the interviews, the L2 writers gave very different reasons as to why they included the authors' names in their text and used reporting structures. Their motivations are usually in line with academic conventions. Four of the L2 writers articulately explained their choice of mentioning the authors in their text in reference to their

intention of showing that the ideas and opinions cited in the text are not their own (excerpts 7.15, 7.16, 7.17). These statements support the findings and interpretation of the corpus study and of previous studies of attribution (e.g. Wiemeyer 2017b, 2017c).

- (7.15 ) TARA: Because I use that sometimes to remind the reader that this is not my own opinion but the opinions of others and I'm just simply telling you what they are thinking or what they did [...] which is why sometimes I said *according to the authors* or *Kettemann, König, and Marko*. (Tara's interview, 00:41:05-3 – 00:41:13-0)
- (7.16 ) ARNE: [...] And well *according to the authors*/ I just tend to include that in my text from time to time when I am writing a summary or something similar. Just to remind the reader that it is not my thoughts I am presenting but someone else or in this case several authors. So, well, I always want to include that but not overuse it. [...] (Arne's interview, 00:34:38-3 – 00:35:40-1)
- (7.17 ) INA: I often try when I write summaries to mention the authors or the text where it came from and/ I mean it was the authors or the researchers decision to use one of the two different analysis or ways to not analyse. So I tried to put away the attention from myself as writing the summary to the TEXT I have read and that I am taking everything from. So that it is not MY idea [...]. (Ina's interview, 00:26:37-2 – 00:27:43-2)

Another reason given by Eva for the use of the authors' names was to create objectivity by attributing a statement to the authors, which is why she decided to insert the reporting structure *according to* (excerpt 7.18). When asked about an instance in which she mentioned the authors in her text, Tara pointed out that author reference was something that was required by one of her professors, even in oral presentations (excerpt 7.19).

- (7.18 ) EVA: Because I wanted to be objective and not/ If I did not do that I thought that it would sound as if that would be my own opinion. Or it would be the truth or the fact. And so I decided to add that. (Eva's interview, 00:27:09-9 – 00:27:29-9)
- (7.19 ) INTERVIEWER: So you mention the authors again here.

TARA: Uhum. Just dropping it in there. (laughter) Making sure that people know that it's not mine. This is also something that one of my other professors likes to point out whenever we have presentations [...] she always tells us make sure that you mention the author a couple of times so people who listen to you will know that it's not your idea but someone else's idea, you're simply just presenting it. (Tara's interview, 00:48:18-7 – 00:48:20-0)

This is interesting insofar as apparently not all students used attribution consciously to create intertextuality. Ben used the phrase *the author* to create coherence when starting a new paragraph, he referred to it as a "good transition signal" (Ben's interview, 00:16:12-1 – 00:16:41-0). Nadine used a phrase containing *the authors* for stylistic reasons; she explained that she "just needed a beginning of a new paragraph" and found that one "most fitting" (Nadine's interview, 00:09:04-4 – 00:09:15-1). When prompted about why she used the authors' names in her text again later, she asserted that this "gives you a broader variety of sentence structures" (Nadine's interview, 00:13:35-8 – 00:13:48-0). Eva used *according to the authors* because she felt that the cited information "sounded like a fact" (Eva's interview, 00:37:09-8- 00:37:37-1). However, she was unable to explain why she had decided to delete this reporting structure later in the process. Evidently, these students held insecurities about why to include reporting structures and used attribution for stylistic rather than intertextual purposes. It is possible that both of these observations cumulatively explain the variance of reporting verbs and structures observed in the corpus study (see section 6.3.3).

Four out of six students provided references (Figure 30). In six cases, these followed direct quotes, the others were given at the end of paraphrases. In the process of providing a reference, Ina, Nadine, and Arne

searched the source text for page numbers (e.g. segments 2:19, 2:24, 5:10, 5:25, 6:5), while Tara, who had already included references in the excerpts she had copied from the source text at the pre-writing stage, simply copied these into her reading report. When combined with paraphrasing, the act PROVIDES REFERENCE predominantly occurred in the last quarter of a segment, as references were usually given at the end of a sentence (e.g. segments 3:27-3:32 by Tara, 5:11, 5:14, 5:15 by Ina), or rarely during re-writing of a paraphrase (e.g. segments 3:8, 3:21). As in the corpus study (see section 6.3.3), many references contained only page numbers. The inclusion of the authors' names in the reference was rare (but see e.g. segment 2:12). As in the corpus study, this should be attributed to the students' awareness of the fact that the source text on which the reading report was based was known to the reader.

Those students that did document their source use provided references regularly in their writing processes, but they cited diverging motivations for doing so. The most diligent student with regard to referencing was Tara, who created references in 18 out of 31 segments. After she had added a reference to a sentence, she deleted the copied excerpt on which this sentence was based. Providing references is thus an integral part of her intertextual writing process. This is reflected in her final product, which contains eighteen references in eighteen sentences. In her interview, she explained that she provided references so that an interested reader would be able to find the respective passages in the source text (Tara's interview, 00:07:48-8 – 00:07:50-2). Ina and Nadine provided references in eight and three segments, respectively. Ina decided to use referencing despite not being sure whether it was required in a reading report. Her decision was prompted by the fact that she had attended a lecture on plagiarism in her first semester and she had repeatedly been warned of committing plagiarism in her other seminars (see excerpt 7.20).

(7.20 ) INA: [...] And people are saying never, never, ever do plagiarism and even if it is unconsciously. But just write down where you got your ideas from and then everything is alright. So basically that is what I did. [...] (Ina's interview, 00:17:45-4-0:18:46-3)

This illustrates that Ina's use of references was inspired first and foremost by concerns over plagiarism. She stated that it did not initially occur to her to provide references, but when it did, she added references to parts of her reading report she had already completed. In her references, Ina used *cf.* with paraphrases, but not with direct quotes. She once added *cf.* to a reference later in the process, explaining that she "remembered again that for direct quotes you give the name of the text and the page number but for indirect quotes you give 'confer'" (Ina's interview, 00:23:43-9 – 00:24:21-2). When asked why she had not provided references for all passages in quotation marks, Ina pointed out the fact that she had quoted terminology and stated that she did not use a reference because she had already provided one for earlier direct quotations of the same terminology. Despite the apparent misconception with respect to what should be placed in quotation marks, Ina's statements suggest that her choices regarding the use of references were not arbitrary. It is likely that diverging practices across disciplines and a lack of concrete teacher advice guided her decisions.

Nadine explained having used references because the words cited were "not my own words and I used [them] directly from the text" and, as she emphasised, she had to provide a reference in such a case (Nadine's interview, 00:29:21-7-00:29:40-4). In her reading report, references accompanied three paraphrases and one direct quote. Nadine used a full reference only the first time she provided one and gave only page numbers in the other references. She explained that she had been taught in school that if citing only one text in an assignment, one had to give only one full reference and that page numbers would suffice for the following ones (Nadine's interview, 00:29:45-7-00:30:03-5). She was unsure of why she provided a reference in one case, assuming it might have been because it was "too close to the text to just use it without giving a reference" (Nadine's interview, 00:34:08-2-00:34:22-0). She also admitted that she was unsure of how to reference properly and explained that she would have consulted a style guide under different circumstances (see excerpt 7.21).

(7.21 ) NADINE: I was actually not quite sure how to reference a quote. And I was like, okay, just do it this way. I was not sure. Usually I would look that up. I was not sure if we were allowed to do that.  
(Nadine's interview, 00:17:59-3 – 00:18:13-9)

Arne used referencing only in one segment, accompanying a direct quote. The reference contained only the authors' names and a page number, a citation style he explained he had learned in his literature courses. In the same segment, he replaced the direct quote with a paraphrase and deleted the reference so that no references remained in the final product. When asked about this in the interview, he explained that while he was thinking about how to reference the direct quote, he started wondering whether he was actually allowed to use direct quotes in the reading report at all. He checked the instructions and, upon finding the passage instructing the writers to use their own words, decided to replace the direct quote.

The L2 writers' interview statements show that they are aware of the necessity of acknowledging one's sources, though it is also evident that some are still unsure of why and when to attribute content to the authors and to use reporting structures. The different citation styles used in different disciplines at the same faculty were also a cause of confusion. Several writers explained there were unsure of whether references were required for this particular text type. There was also insecurity about the appropriate format of references. It is interesting that previous studies have often criticised a lack of stylistic variety in students' use of reporting verbs and structures, yet these students mostly base their choices on stylistic considerations and disregard the intertextual functions of these structures. These functions are clearly central to source-based writing tasks and need to be discussed by teachers. It may also be helpful to make expectations regarding referencing explicit in task descriptions in first- and second-year university courses. Some of the differences between the L2 writers, despite their similar progress in their study programmes, could be explained by the fact that some of their teachers paid close attention to the compliance with academic standards and made these explicit, while others were more lenient. This shows that teacher guidance is crucial in L2 writers' attainment of certain features of intertextual writing such as the strategic use of reporting structures.

### ***Textual borrowing, copying, and strategic re-use of source text material***

The process analysis revealed that students copy from the source text both manually and by using the computer's copy-paste functions. Copying processes often serve to prepare direct quotes, but passages are also transferred into the students' documents to provide easy access to source text material and thus scaffold their paraphrasing processes. Textual material from the source text is re-used during paraphrasing processes even if students are cautious about copying and textual overlap. This of course includes desirable forms of textual overlap, e.g. terminology. In the analysed processes, copy-pasting and cut-pasting acts form part of 24 segments. Altogether, there are 41 instances of copying, cutting, and pasting. Though they are generally quite rare, such micro-level processes are performed by five of the six writers (see Figure 31).

The acts subsumed under the code NW - COPY-PASTE FROM ST (COPY-PASTE FROM ST, COPIES FROM ST, PASTES FROM ST) were used by Nadine and Arne and occurred in eight segments (22 instances in total). These acts were thus slightly more frequent than those coded as NW - COPY-PASTE / CUT PASTE FROM OWN TEXT (22 instances), of which CUT-PASTE FROM OWN TEXT is the most frequent act in the data (5 instances). The acts subsumed under the code NW - COPY-PASTE / CUT-PASTE FROM NOTES / COPIED ST PASSAGE only occurred in Arne's and Tara's writing processes (7 instances in 7 segments). It was the least common code related to copying because the other students simply had not taken notes or copied excerpts from the source text into their document.

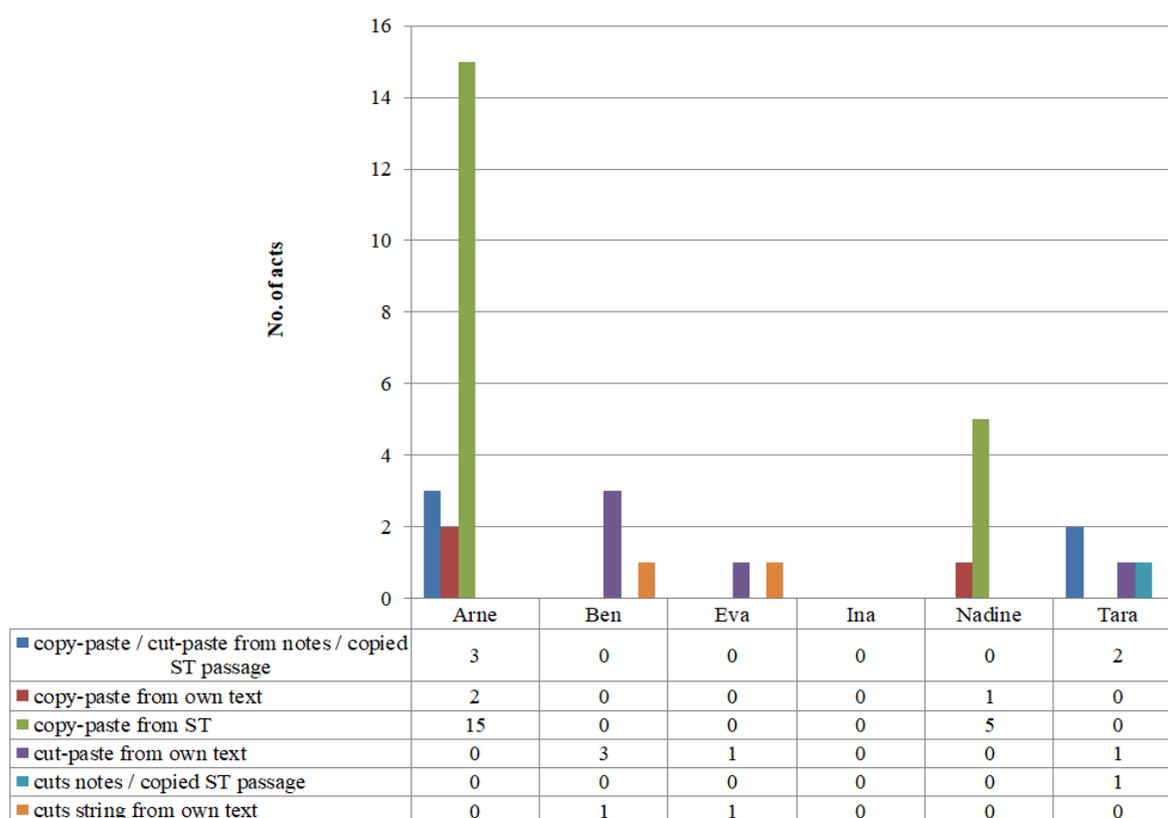


Figure 31: Acts comprising copy-pasting across study participants.

Unlike the other students, who typically employed the computer's copy-paste function to insert source text excerpts into their documents, Tara manually copied from the source text. In her view, this prevented her from copying too much as well as copying unimportant passages (Tara's interview, 00:17:03-7 – 00:17:28-1). Such individual preferences illustrate that L2 writers consciously select certain strategies of source use to avoid plagiarism. Tara's approach to copying is similar to that of the subject in Stapleton's (2010) study, who also copied from her sources both manually and by copy-pasting. The other students in this study also copied excerpts from the source text into their document, which they then either modified directly or used as orientation for their paraphrases. This confirms that copying and the subsequent restructuring and rephrasing of copied excerpts can be major source use strategies. Apparently, this is true not just of real-life source-based writing tasks, but also of experimental settings, irrespective of whether the sources were accessed in an electronic format. The similarities between Tara and the other writers in this study and Stapleton's student suggest that these findings possibly extend to student writers in general.

In the interviews, it became apparent that at least some copy-pasting served to avoid having to switch back and forth between the writer's document and the source text (see e.g. Nadine's interview, 00:24:28-9 – 00:24:51-5, 00:32:26-9 – 00:32:40-4). Arne confirmed that this was a strategy he commonly used in summary writing; see example (7.22). His explanation suggests that the copied passages provide scaffolding in the writing process. Selecting specific excerpts from the source text serves to pre-select important sections so as to make paraphrasing easier by "putting one step in between" (Arne's interview, 00:47:00-4 – 00:48:43-8).

(7.22) ARNE: [...] I just copy the main parts, or the main information, in my own document so I do not need to/ So I can see more specifically what is/ what I think is IMPORTANT and I can just leave out the other parts. This is actually how/ When I write summaries I am just reading the whole text and I am

copying my own/ these most important parts in a document. And I do not even look at the original text I just look at those quotations [copied source-text passages, LW] and then I am writing my own. And sometimes I am already paraphrasing. So these notes are what I use for writing the summary. [...]  
(Arne's interview, 00:47:00-4 – 00:48:22-6)

The students purposefully selected passages to copy, for example because they regarded them as well-phrased. Nadine copied a passage because “it was so nicely done in the source text” – despite the fact that she was unsure of how to reference it (Nadine's interview, 00:27:01-7 – 00:27:29-8). Tara explained that she relies mostly on topic sentences as they contained important information in a condensed format (excerpt 7.23), corroborating the findings of the corpus study (see section 6.3.2.1). She then carefully selects material for paraphrasing from the copied excerpts so as to include only relevant material and avoid redundancy, which sometimes also means that she will condense the content of several copied sentences into one sentence of her own (see Tara's interview, 00:07:50-2 - 00:08:49-1).

(7.23 ) TARA: [I]n the case of this research paper, the first sentence of the paragraph tended to be an introduction or like it kind of gave you the main topic of the paragraph and that's basically what I need for my summary, so I tended to use that. (Tara's interview, 00:11:01-6 – 00:11:11-3)

The use of copy-pasting or cut-pasting for the purpose of changing the order of constituents in a paraphrase can be observed in only one segment across the screen recordings. This is a notable observation in the light of the corpus study presented above and Keck's (2010) study of grammatical operations in paraphrasing. It strongly suggests that changing the order of constituents is not an editing strategy in paraphrasing. The differences in syntactic structure between original text and paraphrase are thus the result of online planning and syntactic operations in the process of rephrasing, at least at this stage of L2 writing proficiency.

Several writers asserted that they had avoided copying from the source text. Ben cited teacher advice as his motivation. When queried about why he had tried to avoid copying, Ben suggested that copying a sentence or its structure “would not be scientific writing”, which is why a writer should use their own words. He also stated that he learnt in school that copying should be avoided and sentences should be re-written “because it is more functional to improve your writing and your language” (Ben's interview, 00:08:54-2 – 00:09:58-4). That teacher expectations motivate students' use of intertextuality has been shown in other studies of L2 writing (see e.g. Davis 2013; Petrić & Harwood 2013). Because the adherence to their expectations is appreciated by lecturers, Ben's strategy potentially leads to higher grades, as the case study of a successful student writer by Petrić and Harwood (2013) suggests.

The strategic re-use of source text material was not analysed in detail in the experimental data, but it was discussed with the participants in the interviews. The responses reveal that there is unmistakably some strategic re-use of source text material, though the students have different attitudes towards and reasons for textual borrowing. Several students cited stylistic considerations as the reason for having re-used vocabulary from the source text. Ina, for instance, explained that this approach helped her make the text “sound nice” and avoid repeating the same words (Ina's interview, 00:30:11-0 – 00:31:05-5). She made reference to an academic writing class in which she had been explicitly instructed to vary her word choice. Nadine asserted that she did not scan the source text for useful expressions but that she did re-use words she had read if she remembered them during writing. She also stated that she tried to think of the perfect word for a certain context and consulted the source text if she could not think of one herself (Nadine's interview, 00:06:55-1 – 00:07:17-3). It is clear from these statements that some textual overlap in these students' texts is intentional.

That the students were very cautious about what they copied from the source text – even for acceptable uses – was noticeable in the discussion of re-use of source text expressions. Ben asserted that though he sometimes looked through the source text for useful expressions, he would always feel as if he were copying, a feeling that he described as “weird” (Ben's interview, 00:16:59-2 – 00:17:16-8). He felt differently about

terminology because of the specific meanings of such words. He also made reference to the frequency of certain terms in the source text, apparently using this as a criterion for establishing whether they should be considered to be terminology. Using the expressions *corpus data* and *corpus* as examples, he explained that he did not know if alternative expressions existed, and that this prompted him to use them in his reading report (Ben's interview, 00:17:16-8 – 00:18:18-4).

The specificity of a certain expression was also cited by Tara and Arne as a criterion for deciding whether it was terminology and could thus be used in the reading report. Tara explained that terminology had to be re-used because of the specificity of meanings and that replacement of such terms would lead to vagueness, which she wanted to avoid (Tara's interview, 00:34:39-3 – 00:34:51-3). She opined that "there are words that you can't really change because it will change the meaning and others where you can easily replace them with something else but the meaning doesn't change" (Tara's interview, 00:34:39-3 – 00:34:51-3). Arne stated that he could not exchange terms for other words, unlike reporting verbs, which both he and Tara gave as an example for words that could easily be replaced (Arne's interview, 00:19:13-7 – 00:20:15-2). When reviewing a particular segment in which he had retained the source text wording, Arne said that he had prioritised conveying the correct meaning over rephrasing the passage more extensively (Arne's interview, 00:19:13-7 – 00:20:15-2). In reference to a later segment, Arne explained retaining a string of three words because he thought "it was just too good to change it", but explained that he may have changed it if it had been a real-life assignment and he had been given more time to complete it (Arne's interview, 00:53:42-1 – 00:54:32-4). The students' statements document their awareness of the necessity to adopt certain words and phrases from the source text in order to transmit its meanings accurately, which was also evident from the analysis of textual borrowing in the corpus study.

Source text expressions are also re-used for quite unexpected reasons, for example laziness. Tara confessed to an inclination towards adopting words from the research article because she was too lazy to consult a dictionary (Tara's interview, 00:39:21-4 – 00:39:32-5, 00:50:17-8 – 00:51:03-03). In spite of this, she openly expressed her insecurity with regard to the extent of re-use that was still appropriate. Her strategy of combating this self-identified deficit was to rephrase as thoroughly as possible (see 7.24). She also insinuated that with increased writing expertise it should be easier for her to decide whether textual borrowing is appropriate. Tara's statement implies that she expects to improve her ability to confidently select words for re-use through writing, not through being taught. This is interesting insofar as Pohl's (2007) study shows that student writers' intertextual competence grows over time even without explicit instruction.

(7.24 ) TARA: I'm not sure to what degree I can do that [re-use source text vocabulary; LW] without risking it being the same sentence as they used, so again I try to avoid that because I don't know I think I'm not as familiar with the whole process of re-paraphrasing and rephrasing so I try to make it as different as possible from the original text but maybe like once I get used to it once I wrote maybe like a couple more reading reports or couple more summaries it might be easier for me to determine where I can use what. (Tara's interview, 00:36:10-6 – 00:36:24-6)

The interviews generally reveal a certain reluctance regarding the re-use of source text expressions. When detailing how they went about finding the right words to paraphrase, the students alluded to the fact that they did not use certain words because they occurred in the source text. Arne stated that he did not consciously search the source text for useful expressions as he usually tried to use different words so as to make his paraphrase less similar to the original. Because of the repeated warnings of plagiarism across different courses, he did not even consider searching the source text for useful words (Arne's interview, 00:10:30-6 – 00:11:20-6). When Ina was asked about a specific word she had re-used, she was surprised to learn that it occurred in the source text, despite the fact that in her process she had spent a considerable amount of time trying to replace it with a synonym. She surmised that she may have copied it unintentionally because she still had the word and its context in mind from reading the source text (Ina's interview, 00:47:20-9 – 00:49:51-4). It is possible, of course, that she was unable to recall the actual motivation for the synonym search. Nevertheless, these exchanges from the interview are in line

with findings from previous studies showing that students are often worried about re-using source text wordings because of the association between textual overlap and plagiarism.

In discussions of copying during the interviews, the students' retrospective reports reveal that they were not necessarily aware of textual overlap between theirs and the source text (see e.g. Nadine's interview, 00:24:51-5 – 00:25:45-4). Ina's and Tara's comments disclose an awareness of the fact that textual borrowing can be unintentional but may still be perceived by some as plagiarism. Tara's solution to this was to copy sentences from the source text into her document to which she could then compare her paraphrases to ensure that she really used her own words (Tara's interview, 00:28:56-8 – 00:29:27-0). In contrast, Ina avoided copying passages from the source text to then rephrase them because she did not want to commit unintentional plagiarism. She also expressed a fear that if she re-used individual words and phrases, a reader might notice this and consider them plagiarised. She felt it was important to be able to write eloquently on her own. She did not want to use the source text for inspiration to improve her style because it could have made her text too similar to the source text. Instead, if a concept was explained well in the source text, she would rather quote from the text (Ina's interview, 00:18:46-3 – 00:20:47-7, 00:52:50-1 – 00:54:20-5).

In sum, the creation and avoidance of textual overlap is a feature common to all L2 writers' processes and mentioned in all six interviews. Copying is used for scaffolding and to facilitate the transfer of source text ideas, not so much to provide a basis for paraphrasing processes. Because it allows for a direct comparison between the source text sentence and the paraphrase, copying can actually be a way for students to avoid textual overlap. While some students report searching the source text for words that they could use in their own writing, for example to instil stylistic variety or to compensate for laziness, others are cautious about strategic re-use due to warnings of plagiarism. Apparently, the use of the source text for scaffolding language-related aspects in the writing process was less pronounced than in previous studies (Leki & Carson 1997; Plakans & Gebril 2012). It may play a more significant role as language support in testing contexts. It is likely that the language resources generally fulfilled this function in the present study, and the students are perhaps less reliant on the source text because of their advanced language proficiency. Though some unintentional textual borrowing occurred, it is clear that these L2 writers generally reflected on their re-use of source text material and weighed their lexical choices against the risk of committing unintentional plagiarism. The following section is devoted to the post-writing stage, which takes place at the end of the writing process.

### 7.7.3 The post-writing stage

Each student engaged in a post-writing stage which lasted between five and sixteen minutes (see Figure 23 and Table 14). Despite the differences in length, the post-writing stages of all six writers are relatively similar. Tara, Eva, Arne, and Nadine spent the majority of this time re-reading and editing their text. Tara, Nadine, and Arne shortened and rephrased some sentences to meet the required word count. They also consulted language resources in the editing process. In contrast, Eva's changes were restricted to synonym substitution and formatting. She and Arne used the post-writing stage to delete the copied excerpts from the instructions during post-writing and provide a word count. Ina rewrote her text more thoroughly than her peers by substituting words for near-synonyms, integrating her notes into her sentences, and adding material to them. She also created a title for her reading report. Ben, whose post-writing stage was the longest at over sixteen minutes, was the only one in the group who re-read the source text. He spent the majority of the post-writing stage on this activity. After having reviewed the source text, he re-read and edited his text for spelling and grammar. For this, he consulted the internal language resources of the word processor, i.e. the synonym finder and the dictionary. All six students' post-writing stages clearly served to touch up the text and primarily consisted of editing processes. The fact that the majority of students did not consult the source text during this stage suggests that these

processes are not intertextual in nature. It is probably that the post-writing stage is thus similar to that of independent writing tasks.

In previous studies (e.g. Plakans 2008), the post-writing stage has sometimes been interpreted as a part of the writing stage. In all of the screen recordings made for this study, there is a visible change in the students' writing behaviour as they began working on the post-writing stage, for example by re-arranging windows and scrolling to the beginning of the text to read it from top to bottom. This suggests that post-writing is a distinct phase, as others have proposed (e.g. Choi 2016), though a more in-depth comparison of micro-level processes that take place in the two stages was beyond the scope of this study.

## 7.8 Interim summary: Process study

All of the L2 writers' processes can be clearly divided into a pre-writing, a writing, and a post-writing stage. There are clear transitions between these stages, confirming a tripartite structure of intertextual writing processes, at least as concerns this level of writing proficiency and the given text type. During pre-writing, the students prepared for their writing task by re-reading the source text, highlighting and extracting passages, and taking notes. Highlighted and extracted passages were used in similar ways later in the writing stage, which is why these acts can be seen as highly similar. The post-writing stage at the end of the process was mostly spent on editing the written text, deleting notes, and formatting. Occasionally, the students referred back to the source text during this stage.

Structurally, the intertextual writing processes were determined by two aspects. First, the students' pre-writing strategies had an effect on their writing process. For those students who took notes or copied excerpts from the source text or the instructions into their documents during the pre-writing stage, these excerpts provided scaffolding for composing the reading report throughout the writing process and were integral to processes of paraphrasing. For others, highlighted sequences from the source text served as orientation and consequently influenced their writing process. Second, the setup of the windows on the screen and how the students worked with the source text shaped their writing processes. Whether or not the source text was visible on the screen determined whether the students paraphrased from memory or directly from the text. The students' intertextual writing processes thus differed depending on whether the source text was permanently visible, whether it was accessed at different points throughout the process of paraphrasing but hidden out of sight otherwise, and whether excerpts from it had been copied into the writer's document which were then paraphrased.

Previous studies have also described different approaches to working with the source text in the writing process. Plakans (2008) observed individual differences in the way students read a source text: Some read the text all at once and then quickly begin planning, others' processes are more interactive and consist of strategies such as summarising and extracting useful phrases from the source text. In general, individual differences have emerged more strongly in reading-into-writing than in writing-only tasks. Plakans explained this in reference to the participants' previous writing experience as students who had reported an interest in writing and experience in academic and other writing showed more source text interaction than less experienced writers. The results from the present study are similar to hers. As the participants in this study had comparable writing experience, it is unlikely that this prompted them to approach the task in different ways. It appears that teachers' advice and individual preferences play a role here, which was also confirmed in the interviews. Different task representations may have also led to different processes and different approaches to paraphrasing (compare the studies by Plakans 2010; Plakans & Gebriel 2012).

Overall, all writers' processes are recursive and integrate writing with re-reading and a range of micro-level processes. The recursiveness of writing processes is well documented in the literature (see section 4.3). Apparently, writing recursively is also a conscious strategy these students employ to monitor the quality of their

university assignments. When asked about their thoughts on the experimental setting of the study, several students explained that they would usually write in several sessions to give themselves the opportunity to think about the text and look at it again later with fresh eyes (see e.g. Arne's interview, 00:53:42-1 – 00:54:25-5). Similarly, the student in Ruiz-Funes' (1999) study explained finding it helpful to view writing as a process and to divide it into stages and return to the document later. This gave her a different perspective on her own text and made it easier to notice mistakes. These findings suggest that the recursiveness of writing processes is not incidental but that student writers make a conscious decision to write recursively to improve the quality of their text not only within one session, but across several sessions. This bears important implications for the interpretation of studies which require students to write in one session. Previous research has identified linear approaches and little time spent on revision as typical features of poor writers' texts (see Ruiz-Funes 1999: 47). Future studies may look at whether such features prevail in less constrained settings in which students are able to choose for themselves in how many sessions they work on their text.

As widely recognised in the literature and confirmed by this study, the process of source-based L2 writing, as other writing processes, is recursive and hierarchically organised into macro-level and micro-level processes. Paraphrasing processes, for example, were shown to be comprised of several acts on the micro-level which are repeated and combined in similar ways throughout the writing process. Nevertheless, it is evident that the writing process is also linear in some aspects, as suggested by Zimmermann (2000). Certain differences in the L2 writers' approach to writing the reading report are observed in the study. Two groups emerge with respect to the linearity of the writing stage. One group of L2 writers wrote one sentence after the other in a relatively linear approach, using the chronology of the source text for guidance. The other group was less linear in their approach and jumped from one sentence to another, leaving some unfinished and returning to others to rework them. Sentences were composed in a relatively similar linear succession of acts by students from both groups. The source-based writing process as observed in this study thus combines linear with recursive elements. Knowledge-transforming, which is understood here in Bereiter and Scardamalia's (1987) terms as the reworking of the ideas in a text through various rounds of revision, rephrasing, and addition, takes place in linear and non-linear approaches.

The present L2 writers' writing processes all contain the same activities with similar frequencies. Paraphrasing segments are by far the most frequent, but all students also quoted directly. The re-writing of existing sentences occurs in all L2 writers' writing stages – usually precluded by the writer's abandonment of another sentence, which is consequently left unfinished and later resumed. It can be assumed that this is a general feature of L2 writers' processes when writing from sources. Three different approaches to paraphrasing emerge from the process analysis. The first approach encompasses re-reading the source text and then paraphrasing from memory. This approach was observed in Ben's, Nadine's, and Ina's writing stage. Ina additionally used the instructions as scaffolding. The second approach consists of taking notes or copying excerpts from the source text into the document and then paraphrasing those with little recourse to the source text. This approach was taken by Tara. The third and final paraphrasing approach entails a process of paraphrasing directly from the source text (Arne and Eva), relying on the instructions to provide scaffolding to a certain extent. The analysis also suggests a modular make-up of the writing process in the sense that if a writer chooses a particular approach to writing a summary, this has an effect on the processes of paraphrasing. This means that writers who approach the task of summarising a source text in a similar way also have similar paraphrasing processes. For example, Ina's and Arne's processes are similar because they both copied from the instructions to create scaffolding for their text, despite the fact that they used the source text in different ways and relied on language resource to different extents.

Apart from students' recourse to the source text, however, there is a striking similarity between paraphrasing processes, with acts such as ADDS WORD(S) TO EXISTING STRING and SYNONYM SUBSTITUTION occurring in relatively similar numbers across processes. The students also combined acts in similar ways. Deleting and rephrasing acts occur in all writers' paraphrasing processes, irrespective of the way they used the

source text, consulted language resources, and prepared their writing stage, e.g. by copying excerpts or taking notes. In addition, all writers copied some words and phrases from the source text. This is central to any act of paraphrasing, as suggested by earlier corpus studies (e.g. Keck 2006, 2014; Wiemeyer 2017b). It is clear that these acts are integral parts of the process of paraphrasing. They do not form a separate, consecutive process after having written a sentence in which the writer attempts to minimise the overlap between their paraphrase and the source text, contrary to what may be expected from corpus analyses (see e.g. Keck 2010). Though some writers re-write some of their paraphrases, much of the minimisation of similarity takes place in the course of the initial process of writing a sentence that paraphrases source text material.

The process of writing a direct quote was also quite uniform across all writers' processes. Nevertheless, as in the corpus study, there are individual differences with regard to which material was quoted. Terminology was the most common type of material placed in quotation marks, but the L2 writers also quoted the definitions from the source text. Furthermore, they chose different ways of including quoted material into their text, e.g. manual copying and copying 'from memory'. Though it may seem to be the easiest approach, copy-pasting was not a preferred approach. The process analysis and the interviews show that students at this level of writing expertise hold insecurities with regard to what to quote and which type of quotation marks to use, confirming the interpretations of the corpus data. Those that were unsure of whether quoting was allowed in a reading report often followed teachers' advice or 'rules' they had learnt in the past.

Similar observations can be made with regard to referencing and the use of reporting structures and attribution, where students' stylistic considerations sometimes override those related to citation conventions. It is clear from the interviews that the L2 writers are generally aware of the need to highlight cited material as such, but that they are also dependent on their teachers' advice. For this reason, attribution is sometimes interpreted by these writers as a way to achieve a sophisticated style of writing rather than of acknowledging the origin of ideas. Reporting structures are produced in writing processes along with the rest of a sentence and thus have a close relationship to the intertextual links they accompany. References, on the other hand, are occasionally added later to a sentence, but generally used as sparingly as in the corpus study. Reporting verbs frequently undergo synonym substitution, suggesting that the students pay special attention to these expressions.

Finally, the discussion of textual borrowing and strategic re-use of source text material reveals fear of plagiarism among the study participants, but also conscious choices with regard to the amount and quality of material that is integrated into the writers' reading reports. Even though they reported having considered whether the re-use of source text material would be appropriate in a given context, there are also instances in which the students inadvertently re-used words, indicating that some textual borrowing occurs because the students unconsciously memorise source text passages and include them into their texts. Some students actually commented in the interviews that the textual overlap was accidental, though these explanations may have been influenced by their understanding of plagiarism. The analysis nevertheless shows that copy-pasting in the writing process does not necessarily lead to inappropriate textual borrowing as long as the students are aware of the risks of copying and equipped with sufficient paraphrasing skills (see also Li 2013).

In the next section, the findings of the corpus and the process study are compared and discussed in detail. A central focus of this discussion are the complementary, converging, and diverging findings regarding the three research questions. The usefulness of this mixed-methods approach and implications for second language writing research as well as for teaching are also addressed, followed by considerations concerning the limitations of this study and suggestions for future research.

## 8 Discussion

The present study endeavours to provide a fully-fledged analysis of source-based L2 writing by combining corpus data with process data in the form of screen recordings and retrospective interviews. In the following, the central results from both the corpus and the process study will be scrutinised in the light of the three research questions. While interpretations of individual intermediate results were already provided for each study individually, this section will provide interpretations based on the combined results of both studies regarding key aspects of interest. It also highlights the study's major contributions to L2 writing research and reflects the insights from the process study in reference to existing models of the writing process.

The present study is based on three research questions, the first of which is concerned with the kinds of intertextual strategies L2 student writers use and how they are combined, both in the process and the product. The study shows that L2 student writers use an impressive range of intertextual strategies in their interaction with the source text and in the integration of source text material and ideas. Students' writing stages differ in length and in the number of segments as a result. Ten different types of intertextual links were identified in the corpus. They are characterised by different degrees of reliance on the source text. Across students' texts, paraphrases are the most common type, but each student has an individual approach to combining intertextual links and structuring the text. Most of these approaches are in line with the task description for reading reports and disciplinary citation conventions. The students in the process study preferred paraphrases over direct quotes for stylistic reasons and sometimes decided against direct quotation, for example because of teachers' advice or their conceptualisation of the text type reading report.

There is remarkable variation in both parts of the study with regard to the features of intertextual links, for example the amount of textual borrowing; the use of various means of attribution, documentation, and reporting; and the students' inclination to comment on the meta-level, interpret the source text, and add information. The study thus emphasises the individual nature of source-based writing and the diversity of approaches. That intertextuality is both an individual and a social phenomenon was evident from the interviews, in which the students often cited teachers and academic conventions as their influence for decisions regarding intertextuality. The study accordingly also enhances our understanding of students' processes of social alignment with respect to academic writing skills.

The source text is a constant companion for students throughout the writing process. Interaction with the source text is in fact the most common of all acts in the writing stage. This finding corresponds to the importance of source text interaction observed by Plakans and Gebril (2012). Most recourse to the source text in the present study falls into Plakans and Gebril's category of academic writing and thinking processes, whereas the language support function is much less common. It is evident that the students generally follow a chronological approach when extracting information from the source text (see also Chan 2011; Plakans & Gebril 2012). They either use the structure of the source text or the guiding questions from the reading report instructions as the basis for structuring their own text.

A certain linear pattern appears in several writers' processes in the screen recordings. The students with a linear approach typically interact with the source text page by page and write one intertextual link after another. Those with a non-linear approach jump back and forth in their text and restructure it regularly. Even though linearity has previously been claimed to occur in less proficient writers who do not achieve knowledge transformation and fail to integrate information (e.g. Leijten et al. 2019: 558), this does not appear accurate in the light of the language and writing proficiency of these writers. It is more likely a result of the task description, which elicits a knowledge-telling approach. It may also be a result of the writing task, which does not necessarily require knowledge transformation.

The degree of linearity appears to be a matter of preference rather than writing expertise. In the words of Bereiter and Scardamalia (1987: 11), “a knowledge-transforming approach to writing can be found even among people who have no particular talent for or commitment to writing, some of whom would even be judged to be bad writers by literary standards”. Irrespective of the linearity of the process, students sometimes leave sentences unfinished and return to them later or delete them. This is telling of the recursive interaction of micro-level processes in source-based writing that is assumed to be an inherent feature of any writing process and has been observed in previous L2 writing studies (Flower & Hayes 1981; Hayes & Flower 1980; Sullivan & Lindgren 2002).

Based on the findings of this study it appears reasonable to distinguish between linearity and recursiveness on the macro- and on the micro-level of writing processes. It was observed in this and previous studies that there are certain stages, referred to here as pre-writing, writing, and post-writing, that are completed in a relatively linear succession, though there are not always clear boundaries between them. Within these stages, certain macro-level processes are completed in a specific, fixed order as well (see also Zimmermann 2000). The source-based writing process is nevertheless clearly recursive on the micro-level. Subprocesses are regularly repeated and frequently overlap or even interrupt each other. One micro-level process may occur in only a specific stage or in all stages (see also Choi 2016). Micro-level processes that characterise the writing stage may also occur in the pre-writing and post-writing stages. Editing is a suitable example as it may be a subprocess of each of the three stages. Clearly, linear and recursive processes co-occur in L2 students’ writing processes and generally appear to be tendencies rather than strictly definable categories.

There are different approaches to setting up the documents during the pre-writing stage and to highlighting or extracting passages from the source text taken by the writers. These are reflected in the students’ processes of paraphrasing. Students continue their individual approaches to paraphrasing throughout the writing process. Similarities between students arise from a similar setup on the screen and approach to accessing the source text. Paraphrases may be written with the source text visible, from copied excerpts, from the students’ notes or from memory. The presence or absence of the source text influences source text interaction in the construction of paraphrases. While some paraphrases are created directly from copied excerpts, others are paraphrased on the basis of what the student remembers from the source text. In many process segments, the source text is hidden from view during writing, which may be the cause of some accidental textual borrowing. Not all paraphrases are compared to the source text once they are completed. It is likely that misrepresentation of source text content is linked to setups in which the source text is not visible or simply not used for direct comparison.

L2 writers perform a variety of acts to create a paraphrase. Paraphrasing always entails writing and re-reading the source text – sometimes before and sometimes during writing – as well as deletions both during writing and after passages are completed. The students also replace individual words, for example with synonyms, they insert placeholders when they encounter a vocabulary gap or they change the order of constituents. Paraphrases are often modified again later in the writing stage and edited in the pre-writing stage. The study confirms findings from previous research in that the majority of these writers’ paraphrases are based on individual sentences (Hirvela & Du 2013; Howard, Serviss & Rodrigue 2010; Shi 2004, 2008). Certain sections are preferred for selection of material, for example the introduction and the conclusion, and paraphrases are often based on topic sentences.

The selection of individual sentences is also observable in students’ writing processes. It occurs despite the fact that the source text used in this study is significantly longer than those used in other studies (e.g. Keck 2006; 2014). This suggests that the length of the source text, which in this case was a 12-page research article, does not lead students to change this strategy. The process study allows for more detailed insights into this preference and shows that there is a process of identifying these sentences prior to paraphrasing that is common to all six student writers. The students’ reliance on topic sentences is apparently not strategic, as only one student

described looking specifically for topic sentences to paraphrase. It is more likely that the informational content leads students to first highlight or extract and then paraphrase topic sentences, as suggested by Sherrard (1986).

The paraphrases in the corpus study appear to be the result of two broad approaches to rephrasing source text material. The first is the adaptation of source text excerpts, in which a source text sentence functions as a template. The approach consists of the application of D/A/S strategies, of morphological and/or syntactic adaptation or of syntagmatic adaptation to this sentence to reduce similarity with the source text. The second is re-writing source excerpts, which entails thorough modification and amalgamation of source text material so that there is little to no lexical and grammatical similarity. Based on the observations from the corpus study regarding the students' paraphrases, it was assumed that in the first approach students use sentences from the source text as templates or even copy and then modify them using the three mentioned strategies. This was indeed observed in the process study, but it was a relatively rare phenomenon. In the majority of cases, the students considered highlighted or copied passages during paraphrasing processes, but did not modify them directly. They in fact employed them to compare their paraphrases to the original with the goal of avoiding textual overlap.<sup>93</sup>

Nevertheless, the process study illuminates that the D/A/S strategies are a central component of intertextual writing processes. Paraphrases are typically created in turn by writing, deleting, and more writing, but are additionally altered throughout the process by acts such as adding word(s), deleting strings, and substituting synonyms. These are core acts that occur in all students' processes irrespective of whether they paraphrase from the source text, from memory, from notes or from copied excerpts. The students frequently rephrase sentences by recursively performing these micro-level operations in various combinations during writing. Synonym substitution as a micro-level process is often performed for stylistic reasons once a sentence is completed. However, individual words as well as longer phrases are commonly deleted and usually replaced in the flow of writing. Students also tend to add one or more words to the passages they have already written.

The study cannot confirm the assumption that students copy excerpts into their document and then modify them morphologically and syntactically. However, this first detailed analysis of intertextual micro-level processes shows that the D/A/S strategies posited based on corpus data in this and previous studies (see Keck 2010) do indeed occur in the writing process. While corpus data only allow for a comparison between a paraphrase and the source text, this study offers a complementary view by making the paraphrasing process visible and documenting the sequence of text creation and modification at the micro-level. It thus provides us with a much more comprehensive understanding of the thus far relatively sparsely investigated micro-level processes of source-based writing.

In both studies, the students combined material from several sentences and from across paragraphs to create gist statements. These have not previously been discussed in L2 writing research, potentially because many corpus-based studies have used relatively short source texts. Gist statements are likely a result of both the text type and the length of the source text, which require students to condense information more substantially than in summaries of short texts. They are also a sign of an advanced academic literacy as gist statements require knowledge transforming and interpretation of source text statements. Another important finding of the corpus study is that many summaries are in fact based directly on summarising sentences of the source text. These were categorised as summary paraphrases. Summaries and summary paraphrases are generally rare, but primarily occur only at the beginning of the reading report and contain the title of the research article.

In the process study, it was not possible to discern differences in the processes leading to paraphrases, summaries, and summary paraphrases, suggesting that the same micro-level processes result in these different but closely related manifestations of intertextuality. It is clear from the present study that direct quotation constitutes a distinct process from paraphrasing. Though there are different manifestations of intertextuality in the form of summaries, summary paraphrases, and meta-level observations in the products, the only discernible

---

93 It may also be the case that only certain students rely directly on source text sentences as templates. In the few patchwritten sentences in the corpus, it was evident that D/A/S strategies had been applied, so it is possible that source excerpt adaptation is rather a strategy of less proficient writers with a tendency towards patchwriting.

differences in the processes of creating different intertextual links are between sentences that contain direct quotes and sentences that do not. It seems that the processes of writing summaries, summary paraphrases, meta-level observations are closely related to processes of writing paraphrases despite the different features in the writing product. It would be interesting for future studies to take a bottom-up approach by distinguishing a range of intertextual links in the product first and then analysing the micro-level processes and source interactions that led to their creation to identify possible differences.

Generally, the analysis of segments from the writing stage revealed that rephrasing and quoting processes are very similar throughout a writer's process. Much of the rephrasing for paraphrases is apparently done online in the process under consideration of the respective source text excerpt(s). The two approaches observed in the corpus study, adaptation and re-writing of source excerpts, are apparently the result of similar processes and mostly depend on the extent of the student's transformations. Direct quotes are usually created by inserting a quotation mark, then manually copying or copy-pasting the passage to be quoted into the text, and then inserting a second quotation mark. Quotation marks are only very rarely added after copying. The students' preference towards accessing the source text influences not only their paraphrases but also their direct quotes. A surprising finding of the process study is that these are sometimes created from memory. Three students employed this strategy, and neither of them always consulted the source text to confirm that they had transferred the passage correctly.

All direct quotes in the corpus and the screen recordings are embedded in the text. The majority occur within paraphrases and there are some combined quotes. While direct quotation is relatively infrequent in the corpus, it occurs in all students' reading reports in the process study, making it a more dominant strategy in this group. Direct quotes are often used without a reference. They are sometimes deleted in the process, but it is very rare for a student to replace a direct quote with a paraphrase. In the present study, students often quoted the definitions from the source text, sometimes several in the same sentence. In the interviews, the students described the definitions as difficult to paraphrase. One student explained that she refrained from placing the definitions in quotation marks because they were general knowledge and she did not consider the authors to be the originators. The students shared a feeling that the definitions were different from other sections of the source text. Such subjective theories probably also explain the decisions of their peers from the corpus study.

As in previous studies (e.g. Petrić 2012), terminology is frequently placed in quotation marks, especially in the process study. The process of inserting quotation marks around terminology is fluent and not interrupted by pauses, which indicates a purposeful behaviour. It appears that the students choose one of two options: using quotation marks for all terminology and definitions from the source text, or not using quotation marks for either. Both approaches can potentially lead to undesirable results, stressing the importance of providing students with guidelines regarding the use of quotation marks as well as the conventional integration of expert terminology, definitions, and technical explanations from source texts. There are two major differences with respect to direct quotes between the corpus and the process study, namely that direct quotes are generally more frequent in the process study and more commonly contain terminology. As the process study was controlled for language proficiency, these results suggest that there is a relationship between students' proficiency level and the frequency of direct quotes, especially those containing terminology.

Terminology has previously been found to be especially common in high-graded students' texts, which generally contained more and shorter direct quotes than those of their lower-rated peers (Petrić 2012). It is likely that the relatively proficient students in the process study also had more advanced academic writing skills than some students from the corpus study and that this the reason for the discrepancies regarding the use of quotation marks with terminology. The students' interview comments further corroborate this assumption. They cite expert-like motivations for using direct quotes, e.g. that the source text passage was very well phrased and could not be rephrased. Nevertheless, several students said that they were unsure whether direct quotation was acceptable in a reading report. This should be addressed in teaching, for example by specifying whether direct quotes are appropriate in a specific text type.

Another noteworthy finding is that there is a category of intertextual link in the data that serves to describe the source text on the meta-level and is thus not a paraphrase. Meta-level observations make explicit that the reading report is based on a source text. They are not based on sentences of the source text and contain new propositions. They sometimes constitute the majority of intertextual links in a reading report, which results in a superficial summary that does not provide information on the authors' research and arguments and is not intended by the task description. Meta-level observations may be a compensation strategy by students with text comprehension issues. Exact copying is only found in very few texts. The few exact copies that were found were mostly embedded in paraphrases. It was also interesting to observe that students sometimes add material of their own to the reading report, both in entire sentences and within paraphrases. This kind of knowledge display is quite rare, probably because it is not intended in a reading report. It is nevertheless notable as it shows how student writers connect their own knowledge with that of the source text. As the addition of new content was not observed in the process study, they may be features of texts written by less proficient students whose academic literacy is less advanced.

The process study provides further insights into the difficulties novice writers face when trying to integrate complex passages into their academic texts (see e.g. Swales 2014; Wiemeyer 2017b, 2017c, 2019). There are some cases of misrepresentation of content in the corpus, for example because of unsuitable synonyms or apparent misunderstandings. In the corpus study, a significant number of patchwritten sentences is based on the definitions, which were apparently difficult for the learners to integrate into their reading reports. It is clear that textual overlap with the definitions is not the result of insufficient paraphrasing but rather of an insecurity regarding whether or not the definitions would have to be marked as direct quotes. Several students confirmed in the interviews that they chose to quote the definitions because they did not know how to paraphrase them. These insights stress that it is important to consider the nature of source text passages that are reproduced in patchwritten sentences, because there can be a variety of reasons for patchwriting. Overall, it must be asserted that in both studies students displayed an eagerness to minimise similarity to the source text and employ their own vocabulary for rephrasing, which has also been emphasised in other research (e.g. McInnis 2009).

In sum, L2 student writers employ a range of intertextual strategies and combine a variety of intertextual links in individual ways. The present study sheds light on both the strategies of integrating words and ideas into their own text as well as on the strategies of structuring their process and their text in consideration of the source text. The students' writing processes are influenced by the presence of the source text, which they constantly access while writing. The most important strategies are paraphrasing and direct quotation, which are created in processes that are similar across the cohort. D/A/S strategies are observed in the written products and also in micro-level processes. The students use meta-level observations to structure their reading reports and sometimes add new information to them. Summaries are regularly based on individual source text sentences with a summarising function. Unlike in earlier studies, exact copying is rare, emphasising the students' advanced academic literacy.

The second research question addressed the ways in which L2 student writers document, attribute, and report information and material from the source text. 44% of all intertextual links in the corpus of reading reports contain attribution, documentation, and/or reporting structures, demonstrating how relevant these means of acknowledging sources are in integrated writing tasks. They are also observable in the process study. The students interviewed in the process study named their intention to acknowledge the origin of ideas in their texts as their motivation for including attribution, reporting structures, and references, confirming the assumption from the corpus study. Other motivations include the desire to create objectivity by attributing content to the authors, which is an expert-like behaviour, but also meeting teachers' expectations.

In the writing process, attribution and reporting structures are produced in fluent sequences and display a close relationship to the intertextual links in which they occur. Attribution is present in all reading reports. The writers of the corpus texts included a range of targets of attribution, which creates stylistic variety. In the vast majority of cases, the authors themselves are acknowledged, but there is also attribution to the type of

publication, to the type of research, and to the section of the research article. Moreover, indirect attribution via passive structures is observed in several reading reports. This means of acknowledging the author indirectly has been overlooked in other studies, probably because the use of the passive is a general feature of academic writing, but it is clearly an important linguistic phenomenon for marking content as cited and signalling objectivity. Surprisingly, the students' interview statements reveal that attribution, especially in reporting structures, is often used for stylistic rather than intertextual reasons, confirming an assumption from previous research (Bloch 2010). In fact, some students were seemingly unaware of the intertextual functions of phrases such as *according to the authors* and used them solely to vary their wording. There is therefore a different aspect to the stylistic variation achieved through attribution in the corpus study, namely that some students are not aware of the intertextual functions of attribution and reporting structures. It is thus important for teachers to pay attention not just to stylistic options, but also point out the role of attribution and reporting in indicating to the reader which parts of the text are cited.

Much attribution occurs in reporting structures, which accompany all types of intertextual links except new content. There is an average of seven reporting structures per text in the corpus, with at least one reporting structure per student. Reporting structures also occur in all six students' writing processes, often with the majority of sentences. They are always written in the same segment as the rest of the sentences and never added later, unlike references. Most of these structures contain a reporting verb, all remaining structures are relatively infrequent. The verbs employed by the students are quite varied. Certain verbs are preferred with certain structures, yet there is no apparent 'overuse', contrary to the claim that this is an issue of L2 writers (e.g. Manan & Noor 2014; Verheijen 2015). It is relatively uncommon for a student in the present study to use a reporting verb more than twice.

As a matter of fact, the process study revealed that reporting verbs are frequently subject to synonym substitution, indicating that L2 writers strive towards variation in reporting structures. Evaluative verbs are rare, which is an expected observation given the task description. Despite some unorthodox reporting structures, it is clear that the students consciously select a range of verbs and vary their choices throughout their texts. It is stressed here that the assumptions made in studies that cumulate the frequency of reporting verbs across texts are unlikely to apply to all the individual texts in a corpus, and it is possible that previous studies have misrepresented the reporting verb use of L2 writers by generalising across cohorts.

Documentation is relatively marginal in the corpus and in the screen recordings. It is employed by few students and sometimes consists only of page numbers, which is likely a result of the text type and the students' task representation.<sup>94</sup> Reading reports are based on an individual source text that is known to the reader, which may cause students to omit references. As in expert writing, references are relatively common with direct quotes, but also occur with paraphrases and summaries, especially at the beginning of reading reports. In the process study, the students explained that they document their source use so as to eliminate the risk of being accused of plagiarism. They provide references to acknowledge the origin of cited ideas, but also as a service to the reader. References are sometimes added later in the process to sentences that have already been completed. This indicates that they have a different status than the other forms of acknowledging one's sources; that is, the connection between the text and references is less immediate than for attribution and reporting structures.

The students made it transparent in the interviews that they follow certain implicit rules when providing references. For example, they explained that they do not consider references necessary for terminology, which may explain their scarcity despite the use of direct quotes. The comments of the students also disclose certain insecurities that are assumed to have influenced the students in the corpus study. These insecurities regard the correct formatting of references and the exact requirements for the text type reading report. Teachers can help

---

94 Previous research has shown that the task description has an impact on the students' texts, especially on their use of sources, because it guides them towards what is expected in the final product and identifies those skills on which their assessment will be based (Petrić & Harwood 2013; Wette 2017). Though no evidence of this emerged in the interviews, varying task representations may nonetheless have influenced the students' source use (see also Plakans 2010; Plakans & Gebrel 2012: 31).

students overcome such insecurities by clarifying their expectations regarding the format of references and why and where they are necessary.

It is clear that the students in both studies are aware of the need to foreground the intertextual nature of their sentences and also exploit the various ways of doing so for their stylistic potential. Attribution is realised in manifold ways across all students' texts. A large number of sentences contains some kind of attribution, mostly to the authors and in reporting structures. A wide range of reporting structures is attested, including some creative and somewhat atypical uses. The findings of the present give rise to methodological implications for L2 writing research. It is crucial to take different forms of acknowledgement and their combination into consideration when analysing intertextual strategies in order to fully capture students' approaches to source-based writing. Furthermore, there is no evidence that individual students repeatedly rely on the same verbs. Instead, students pay close attention to varying their choice of reporting verb. Some students in the process study, however, use reporting structures and attribution only for stylistic reasons and are apparently unaware of their intertextual functions, emphasising a need for awareness-raising activities in university courses. References were used scarcely by most writers, which is likely a consequence of the students' task representation, but there are also insecurities regarding the correct formatting and when documentation is required. Overall, the students acknowledge the source text in their reading reports to different degrees and most have an understanding of the means of doing so, though the exact requirements and intertextual functions are sometimes a cause of insecurities.

The third research question focuses on textual borrowing and copying of source text material as well as the strategic re-use of vocabulary. To my knowledge, it is the first study investigating copying and textual borrowing in the L2 writing process, including copied excerpts that cannot be traced in the product. It addresses a significant desideratum of L2 writing research (Leijten et al. 2019: 577). Incidental observations in previous studies have suggested that the source text serves as a language resource (e.g. Gebril & Plakans 2016; Plakans & Gebril 2012; Pecorari 2015), which is why the present study investigated this form of source text interaction. The two studies show that copying and textual borrowing are integral features of the intertextual writing process, which is also reflected in the product. During the writing process, L2 writers frequently insert passages from the source text into their documents by copying them manually or using the computer's copy-paste functions.

The present study reveals that copying passages fulfils a variety of functions in the writing process. First and foremost, these passages are required for creating direct quotes, which are often copied manually from the source text into the student's text. In addition, copied excerpts can serve as scaffolding because they allow the student to structure their reading reports and to paraphrase source text material directly in their document. Sometimes they form the basis for paraphrases when copied excerpts are modified and turned into new sentences, though this is not a frequent strategy. Students occasionally transfer strategically selected expressions for stylistic reasons or because they cannot find a suitable synonym. Copying is evidently a valuable tool in the writing process and exploited by student writers to facilitate source-based writing. Yet some textual overlap in the corpus appears to be incidental. The unconscious transfer of topic-related expressions and remembered phrases also occurs during the writing process and is confirmed by students in the interviews.

All writers in both studies included copied passages into their texts, and copied strings of 3+ words constitute a median of 17% of the words of a reading report. A central finding is that this textual borrowing includes source text terminology and compounds that are necessary to accurately report the content of the research article. This finding is important insofar as it proves that students are not necessarily patchwriting even if there are longer overlapping strings, which has not always been considered in previous studies (e.g. Shi 2004). Similar observations have been made by Cumming et al. (2005), who noted that the more successful students in their study copied common and ordinary expressions that were not clearly definable as textual borrowing. Keck's (2006) paraphrase taxonomy also makes a distinction between unique links and general links, the latter of which is comprised of common expressions. Some overlapping terms cannot be avoided in source-based writing without altering the original meaning, others are simply typical of academic writing. This observation has direct

implications for L2 writing research and teaching, in which textual overlap is sometimes identified via software and treated uniformly. It is important for researchers and teachers to distinguish desirable textual overlap from transgressive forms and to consider both the frequency and the content of copied strings when judging students' texts. This also entails educating students about the difference so they are not driven to reword phrases that would have been acceptable to copy.

Across the corpus and the screen recordings, students' strategies are characterised by varying degrees of reliance on the source text. Micro-level processes of copy-pasting and cut-pasting are rare overall. Some L2 writers copy sentences and passages into their texts before they write their paraphrases, while others limit themselves to copying passages for direct quotes and a few cautiously selected words and phrases. The corpus study revealed that less than 10% or more than 20% textual overlap of strings of 3+ words may be an indicator of problematic source use. More than 20% textual overlap is usually a sign that the text is very close to the source text, while less than 10% indicates that the reading report summarises the source text only very superficially, possibly because the task was misunderstood. Such issues are symptomatic of students' insecurities regarding how to write from sources in particular text types and how much textual overlap is acceptable, which are amply documented in the literature (see e.g. Shi 2004, 2012; Pecorari 2003; see also section 3.2). It has previously been asserted that "more does not necessarily equal better" in terms of source material use (Plakans & Gebril 2012: 30), which is confirmed in the present study. It is furthermore argued here that less does not necessarily equal better either, because a certain amount of textual overlap is necessary to accurately reproduce ideas when writing from sources.

An important finding of the process study is that copied passages are a negligible source of textual borrowing. Several students in this study regularly copy passages from the source text into their documents, but use them only as inspiration for their paraphrases. They purposefully select important sentences for copying. The excerpts are deleted once the students have completed their paraphrases. Copying therefore provides direct access to central source text excerpts and thus scaffolds the paraphrasing process, in terms of language (see also Leki & Carson 1997: 56; Plakans & Gebril 2012: 29) and of creating intertextuality and avoiding plagiarism. In contrast, some students copy only for direct quotes and not for paraphrases. Some even consciously avoid any form of copying because they are afraid of committing plagiarism or because they try to follow their teachers' advice.

These findings emphasise that copying is either employed as a strategy with much care or avoided. The students in this study largely make conscious decisions about copying from the source text which are guided by their knowledge of academic writing conventions. Concerns over plagiarism are always in the foreground, and the students take care to re-write sentences using their own words. The present results thus extend our knowledge of patchwriting as a learning strategy by insights into the use of copying as a scaffolding strategy that actually serves to avoid textual overlap. It is likely that these findings extend to other students with a similar level of writing expertise, but there is of course evidence in the corpus that not all students are successful in their attempts of avoiding unnecessary textual overlap.<sup>95</sup>

There are cases of patchwriting, near copies, and copied passages with only syntactic changes in the corpus, but these are quite rare. Some paraphrases have also been so vigorously rewritten that they misrepresent the source text or seem far removed from it. Extensive textual overlap mostly occurs in sentences that contain the definitions of *eco-* provided in the source text without quotation marks. Difficulties in understanding source text passages, especially technical ones, have previously been identified as a cause of patchwriting and unusual forms

---

95 It must be noted that the language proficiency of the students in the corpus study is more diverse than in the process study and that their average proficiency is likely to be lower. This possibly has an influence on the observed paraphrasing strategies. Previous studies have shown that language proficiency correlates with some, but not with all aspects of source-based writing. Gebril and Plakans (2012) found no significant correlation between students' grades and the amount of copied strings of 3+ words, but that more proficient students are significantly less likely to adopt lexis and syntax from the source text. It may of course be the case that despite having spent the same amount of time at university, the students in this study have very different writing experience and aptitude at academic writing. Thus, the paraphrasing processes of students in the same course may actually be more diverse and may be more reflective of the variety of approaches in the corpus study.

of direct quotation (Hirvela & Du 2013; Howard, Serviss & Rodrigue 2010; Wiemeyer 2019). They are likely to have played a role in the copying of the definitions in this study. It is also possible that L2 writers tend to consider certain parts of source texts to be general knowledge and thus regard quotation marks as superfluous, as one of the writers in the process study explained. Neither excessive textual borrowing nor near copies were observed in the process study, which considered students with a relatively high language proficiency. In the light of previous research into the reasons for illicit copying (see section 3.1), it can thus be assumed that the issues observed in some of the reading reports in the corpus are also connected to limited writing experience and language issues. Nevertheless, as some students mix elaborate paraphrases with patchwritten sentences, it is clear that various factors are at play and that only detailed analyses of individual students' texts can provide a complete picture.

In the corpus study, there is some textual overlap with individual words from the source text that were assumed to have been borrowed strategically. Certain words stand out because they occur in several students' texts. Among them are general terms as well as linguistic terminology. The borrowed words are selected from all sections of the source text, but not all are re-used effectively. Contrary to expectation, only a few of the interviewed students explained that they used the source text for language support. Some refrained from doing so out of fear of plagiarism. Because they are aware of the expectation that copying be avoided, re-using words often feels strange to them unless it is terminology. Of course, such explanations may also be distorted by the students' understanding of copying as inappropriate.

Despite being relatively uncommon, the strategic use of the source text for language support described by some of the students is very similar to that observed by Plakans and Gebril (2012: 31). It helps the students find expert terminology as well as useful expressions. The process study shows that students' use of the source text as a vocabulary resource is driven by an eagerness to vary their word choices. L2 writers may re-use individual words because they prefer an expression from the source text over their own words, because they cannot substitute a word, or because of laziness. Unlike in Plakans and Gebril's study, the students did not use the source text as a replacement for a dictionary, probably because the use of language resources was allowed. The interviews made it clear that the students were very careful to select only words for re-use that could not be considered as plagiarised, making it unlikely that this lexical strategy has a significant effect on students' vocabulary. Nevertheless, this study provides evidence for the assumption that some re-use of source text material is strategic, though it appears to be a marginal phenomenon in disciplinary contexts.

The identity of individual words between the source text and a student's reading report is apparently sometimes the result of an inability to find a suitable synonym or of unintentional copying of phrases the students remember from reading the source text. Many identical words are closely related to the topic and so the transfer is incidental rather than strategic. Some of the students are aware of the fact that copying can be accidental and take measures to avoid it by rephrasing thoroughly. Despite earlier findings that the language support function is especially pronounced in undergraduate content courses (Leki & Carson 1997), it appears that this is dependent on the task, the source text, and the students' proficiency in English and academic writing. Language resources seem to be a more relevant source for vocabulary for these students, and much overlap of individual words concerns topic-specific terms. Thus, while the corpus data seemed to indicate that there is a certain amount of intentional re-use, it must be conceded that only some students purposefully consider the source text in their lexical choices beyond topic-related terminology.

To summarise, copying and textual borrowing occur throughout L2 writers' processes, but are not necessarily connected to textual overlap in the product. Copied excerpts often function as scaffolding and are deleted in the course of the writing process without any copied material remaining in the students' paraphrases. This is because the students compare their paraphrases to these copied passages and take great care to rephrase them as thoroughly as possible. Some of the textual overlap observed in the product is apparently the result of accidental copying or of a student's inability to rephrase an expression. It is clear that some textual overlap results from students' insecurity regarding what needs to be placed in quotation marks, especially concerning

terms and definitions from the source text. There is desirable and unavoidable textual overlap, e.g. of terminology and topic-related compounds. Some students use the source text for language support and intentionally adopt expressions from it, for example to increase the lexical variation in their reading reports. The students in this study employ varied textual borrowing practices with academic conventions firmly in mind. Despite some questionable textual borrowing, many L2 writers have apparently internalised the need to avoid extended copied passages at this level of academic writing expertise.

The study also offers insights into source-based writing and the underlying processes beyond the three research questions. The general approach towards writing from a source text taken by the students in this study mirrors the general source use observed by Plakans and Gebriel (2012) in their study of an undergraduate L2 reading-to-write task. These strategies are thus likely to be generalisable across different populations of novice academic writers of English both in testing and in disciplinary contexts. There are three distinct stages in all six writers' processes, with the writing stage usually being the longest. The boundaries are relatively clear, unlike in Choi's (2016) study, perhaps because the writing task was completed in one session.<sup>96</sup> The assumption that post-writing is part of the writing stage (e.g. Plakans 2008) is contradicted by these writers' processes in which there is always a clear transition, for example a change of the setup on the screen. It is clear from this study that the post-writing stage is dedicated to language-related editing that does not serve to minimise similarity to the source text. Syntactic restructuring and intertextual acts are not usually found in this stage, in which the source text is usually hidden from view. The assumption that intertextual writing takes place during the writing stage is thus proven to be true at least for this group of students. There is much variation in students' processes with respect to the length of the pre-writing and post-writing stages. These are apparently reflective of different approaches towards preparing the writing stage and editing the text at the end, but more detailed analyses are needed.

Both the corpus and the process study have emphasised that there are certain patterns across (at least some) students' processes and texts. Their writing processes are recursive, and yet there are always elements of linearity, for example in following the chronology of the source text. The pre-writing stage to a certain degree influences the processes that occur during the writing stage, while the writing stage has an influence on the processes that take place during the post-writing stage. For example, if the writer takes notes during the pre-writing stage, the composition process is likely determined by these notes during the writing stage, and if a lot of editing takes place during the writing stage, not as much editing is performed during post-writing. Because similarities between student writers have thus far been investigated mostly in the context of (unintentional) plagiarism and language assessment, the present study provides new insights into proficient source use in the process and in the product, especially regarding the micro-level processes that lead to intertextuality, and makes an important contribution to L2 writing research.

The corpus and the process study produced converging results regarding many aspects of source use that have thus far not been investigated in detail. They show that there are certain shared features both in students' writing processes and in completed reading reports. Among these are the chronological approach to summarising the source text, the predominance of paraphrases, and the similar processes that lead to the creation of intertextual links. In general, behaviours from the process are observable in the product and the final product discloses information about the process. An example is the copying and quoting of the definitions from the source text. Both studies also found that the students used certain sentences from the source text as islands to land on when writing their reading reports. Some source text sentences are more frequently paraphrased than others, for example topic sentences from the beginning of paragraphs. Thus, while the present study confirms many of the results of earlier studies and thus adds to our knowledge of L2 writing in general, it also offers

---

<sup>96</sup> The time limit imposed on the study participants has to be considered when interpreting its results. Writing a research paper is likely to take students hours, even if they do not put much effort into the assignment. Stapleton (2013) argues that writing a research paper over the course of several weeks using a variety of electronic and printed resources and written in different locations, usually on a computer, is a much more likely scenario. Though a reading report is a different text type, the three-hour session likely affected the participants' composition process.

tantalising new insights into source-based writing in L2 English regarding the selection of source text material as well as the strategies of integrating and rephrasing excerpts.

As intended in the study design, the results of the corpus and the process study provide complementary insights regarding the central aspects of intertextuality investigated in this dissertation. The corpus analysis allows for a deeper understanding of the linguistic features of intertextual links, while the process study discloses information about students' interaction with the source text and manipulations of their own text. Textual borrowing, for example, is a general feature of the reading reports in the corpus study and there are copied strings of more than ten words, suggesting that copying may be a major strategy of undergraduate students. In the process study, however, the students copied relatively seldom and most rephrasing was completed online. Even though some textual overlap found in the corpus study appears to be intentional, the process study does not confirm that strategic borrowing and language support functions are common among L2 writers. Another example of complementary results are reporting verbs, which are remarkably varied in the corpus study, pointing towards students' intertextual awareness. Yet the process study shows that though these verbs are often subject to synonym replacement and students take much care to create stylistic variation, they are often unaware of the intertextual functions of reporting structures. The methods were effectively combined for the purpose of triangulation: The screen recordings and interviews, which were completed after the corpus study, complemented the corpus data by allowing the researcher to directly observe source-based writing processes and gauge students' motivations and intentions.

The present study also provides further general insights into source-based writing processes in electronic environments. It shows that source-based writing processes consist of writing and non-writing acts at the micro-level. A large variety of such micro-level processes are repeated and intertwined throughout the writing process. Their combination and succession is crucially influenced by the presence of the source text and the writer's approach to extracting information and material from it. Some micro-level processes are specific to the writing stage, yet several acts are apparently elements of more than one of the three writing stages, for example general editing, adding words, and re-reading the source text. This highlights the fact that recursiveness exists both within and across the stages of the writing process. It is clear that processes of writing reading reports are recursive and consist of several stages and that the model by Hayes and Flower (1980) is thus applicable in its basic assumptions. These include the presence of three stages in source-based writing processes, which has also been confirmed in previous studies (see Choi 2016). There is also a certain linearity to the students' approach, as previously described by Zimmermann (2000).

While the existing process models are applicable to the processes observed in this study in general terms, they are usually based on investigations of independent writing tasks and do not account for interactions with source texts. Hirvela's (2004) assertion still stands that there is currently no model that comprehensively captures the connections between reading and writing in the L2 and that may be used as a starting point for investigations into the ways in which these connections are investigated and taught. When comparing the multitude of micro-level processes to those described in models of the L2 writing process (Krings 1989; Zimmermann 2000), it is striking that the use of a source texts adds remarkably to the list of possible acts. While the micro-level processes described by Krings and Zimmermann occur in the present study, source-based processes are generally much more complex. They consist of a larger range of subprocesses related to formulation as well as to interactions with the source text, the task description, and language resources. The L2 problems and solutions postulated in the existing models are also observed in this study, but are relatively minor in comparison, either because of the advanced language proficiency of the writers or because source use requires considerable cognitive resources and thus overshadows language-related concerns, or both.

It is clear that the process of writing from a source text contains many of the elements proposed in the models, but they do not account for interactions with and influences of the source text, the conventions of the discourse community, and external resources such as dictionaries, all of which considerably shape the writing processes (and products) in this study. If the model by Hayes and Flower (1980) were to be extended to source-

based writing processes (see Figure 32), the task environment in source-based L2 writing tasks crucially would have to consist not only of the writing assignment and the text produced so far, but also of the source text(s) and external resources, including language tools. The present study shows that these are central pillars of the task environment in that they determine a considerable number of the various micro-level processes performed by the writer. The source text is used as a template for paraphrases and direct quotes, as a guide in terms of structure and content; it is used to check for accuracy of meaning and as a language resource.

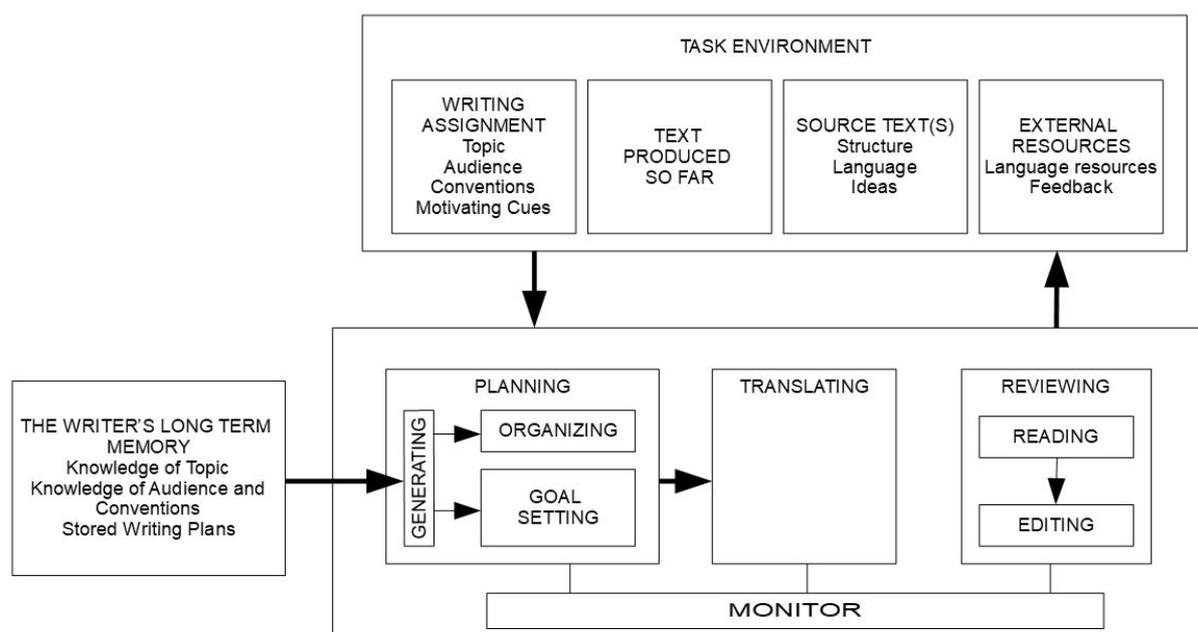


Figure 32: Extension of Hayes and Flower's (1980) model to the source-based writing process, taking into consideration interactions with the source text(s) and external resources as well as the influence of academic conventions.

In addition, it is evident from the interviews that academic conventions, as mediated by teachers and implied in the task description, strongly influence students' decisions, not just regarding source use. They are an important factor in the writing process. According to Hayes and Flower (1980: 12), the writing assignment exerts an influence on the process in terms of the topic, information about the audience, and motivating cues. In source-based and other writing tasks, the writing assignment typically also refers to and implicates the conventions of the respective academic community, for example regarding citation. It contains information about intertextual features of the writing product and the intended source integration. Knowledge about this is additionally stored in the writer's long-term memory and shapes the writing process, which is evident from the students' comments in this study. These aspects should be reflected in a model of the source-based writing process.

It is an important aim of this study to provide an empirical basis for the design of effective writing pedagogy, and many of the findings discussed above are directly relevant to academic teaching. Furthermore, the study also provides important insights regarding mixed-method investigations of source-based writing that should be taken into account in future studies, as discussed in the following sections.

## 8.1 Implications

On the basis of the discussion and interpretation of the central study results, a critical examination of the study's methodology and design is performed in order to gauge their practical applicability. The present study

has implications for L2 writing research as well as for teaching of source-based writing in academic contexts, as elaborated in the following sections.

### 8.1.1 Benefits of mixing methods in research into L2 source-based writing

The mixed-method approach taken in the present study warrants an examination of potential implications of the study's methodological design for future research into source-based writing tasks, for example in academic contexts. A triangulation of methods such as the one performed in this study is still rare in research into source-based writing. This research thus far has usually been based on case studies, e.g. text-based interview studies, and corpus studies. Process-oriented research has mostly been conducted in testing contexts. The present combination of a corpus-based and a process-oriented approach integrates perspectives from two areas of linguistic discourse: learner corpus research and process-oriented L2 writing research. It bridges two often distinct strands of research and gives rise to several implications for studies in this area.

The present study highlights the fact that combining product and process perspectives in a mixed-methods approach has a number of benefits. First, it allows for a direct comparison of the processes and products of the same kind of task. Second, data from the analysis of the processes of a smaller learner group can be studied in depth and correlated with the results from a corpus analysis of a larger data set that is representative of a larger group of learners and thus more generalisable (see Plakans & Gebriel 2012). The process study presented here provides a more nuanced look at the phenomena observed in the corpus study and enhances the results of the study of the product with insights into the creation of certain intertextual phenomena. Stimulated recall procedures provide especially valuable insights because, as Petrić (2012: 105) argues, “[u]ncovering writers’ perspectives on and reasons for citing is especially important when writers are L2 students, because such information may provide pedagogically useful insights into the types of difficulties L2 student writers encounter when writing from sources”. One clear advantage of combining screen recordings with stimulated recall is that these data can bring to light discrepancies between students’ self-reports and their actual writing processes.<sup>97</sup> Of course, the prerequisite for such a triangulation of methods is a comparable setting. In this study, this was achieved by basing the analyses on reading reports using the same source text and written by students at the same stage of their university studies and with comparable writing experience.

The complementary functions of corpus, process, and stimulated recall data are evident in the complementary findings regarding central aspects of source use in the reading reports. For example, the corpus study shows that the majority of direct quotes are not referenced. The influence of the text type and the fact that the lecturer knew the source text were posited as possible explanations, but such findings could also be related to the students’ ignorance towards citation conventions. The process study demonstrates that references are sometimes provided during writing, but frequently added at a later stage. This indicates that students make a conscious decision to insert references, but do not necessarily consider doing so when typing their sentence, even when the source text is readily available for extracting the page number and other relevant information. The interviews revealed that students are indeed influenced by their teachers’ advice, but also have subjective theories about the text type and its characteristics, for example regarding references. Often they are unsure of the applicable citation conventions and develop a suitable approach, for example providing references for direct quotes, but not terminology placed in quotation marks. They make conscious choices based on their understanding of citation conventions, which is sometimes not yet fully developed. All three research methods combined thus allowed for very detailed, in-depth insights into this phenomenon.

Due to the different nature of the data sources, certain aspects of intertextual writing can be more easily addressed using one type of data. As corpus data allows for the quantification of certain phenomena, it is more

---

<sup>97</sup> See also Ranalli, Feng & Chukarev-Hudilainen (2018) for an example of a student who emphasised his revisions in the interview despite very little editing having taken place in his writing process.

suiting to the investigation of the use of reporting verbs, for example. However, other aspects cannot be observed in corpus data. The process data allows insights into synonym substitution with respect to reporting verbs that are not available from the products of writing. Taken together, these different types of data allow the researcher to reconstruct intertextual phenomena in much more detail and from multiple perspectives. Due to the deep analysis of the data, the process study also brought to light incidental findings that go beyond the research questions, for example regarding the use of language resources and general editing.

While mixing methods offers new perspectives for any empirical study, such an approach is especially important in research into source-based L2 writing because it is located in the context of plagiarism or at least refers to the ethics of source use. It is crucial to consider the full picture of students' behaviours and choices because studies of intertextuality often touch upon apparent deficits of student writers. The literature review in section 3 shows that L2 writers' texts may be scrutinised more rigorously than those of their peers. Their practices have been considered in the context of their native language and cultural attitudes and even likened to ignorant or deceitful behaviour. It cannot be emphasised enough that students' comments in this and previous interview studies have shown that they are very much aware of and try to avoid illicit textual overlap. The present process study provides empirical evidence of these attempts. Researchers must consider the full range of intertextual manifestations, from paraphrasing and direct quotation to attribution, documentation, and reporting. By analysing these features of texts in isolation or in only one type of data, researchers risk drawing conclusions that do not do L2 writers justice and may miss an opportunity to identify individual and effective approaches to writing from sources. Multiple perspectives in this area of research are thus a crucial instrument to avoid stigmatisation of L1 and L2 student writers.

The findings of the present study confirm that despite general tendencies, L2 writers' strategies of creating intertextuality are individually different, establishing the need for targeted pedagogical advice instead of generalised warnings of plagiarism. Analyses of disciplinary assignments therefore have important implications for academic writing instruction for learners of English and the potential of enriching academic writing pedagogy.

### **8.1.2 Pedagogical implications**

There are manifold aims to teaching students how to write from sources in academic contexts. Students must understand the purpose of academic writing – and of integrating sources in their texts. In order to do so effectively, they must understand their topic and their sources and reflect on the literature, which requires critical thinking skills. Academic writing entails the ability to employ paraphrasing, direct quotation, referencing, and other intertextual strategies and understand them as more than preventative measures for avoiding plagiarism. Students must learn how to apply these skills in creating intertextuality in line with academic conventions. For this, they need to be familiar with the formal and stylistic aspects of source-based writing. Students' skills and awareness of disciplinary expectations need to be developed in such a way that they can be applied in various writing tasks. It is crucial that L2 writers learn how to develop their own point where this is expected. In order to position themselves towards their sources and integrate them into their argumentation, students must also reflect on their role as writers and budding members of a discourse community.

Considering these complex and multi-layered cognitive requirements, it is not surprising that writing from sources is an integral part of academic writing courses from undergraduate to postgraduate level. Intertextual strategies are taught to facilitate knowledge transformation and help students avoid plagiarism (Hirvela & Du 2013; Wette 2010). Because L2 writers are expected to incorporate ideas from sources into their texts in accordance with academic writing conventions, they are often warned of inappropriate textual borrowing and insufficient referencing (Shaw & Pecorari 2013). However, it has been noted that there is not always sufficient

training of strategies for the effective incorporation of information from source texts (e.g. Solé et al. 2013) and that warnings of plagiarism may lead to anxiety rather than improvement (Abasi & Graves 2008).

It is apparent from this and previous studies that L2 writers are usually aware of the need to avoid plagiarism, but are not always sure how to circumnavigate transgressive forms of intertextuality (see Pecorari 2003). Some students are not yet aware of how to distinguish others' ideas from their own. Some also cling to advice given by teachers and university lecturers and overgeneralise it without critical reflection. They may then be overwhelmed if they are unsure whether this advice is applicable to a new text type. Such issues can be targeted by teaching and raising awareness of effective intertextual writing strategies and comprehensive steps towards avoiding patchwriting and illicit copying. Such combinations of awareness-raising exercises and explicit teaching of strategies have previously proven successful in gradually improving students' source use (Wette 2010).

In order to learn effective paraphrasing, students must be given explanations and exercises to learn what it actually means to 'say it in your own words' and what is considered as effective paraphrasing. In this study, it was apparent that some students rephrased only to minimise similarity to the source text, not because they were trying to write sophisticated paraphrases (see section 6.3.2.1). It is especially vital for teachers' explanations to include the steps towards creating a paraphrase, a summarisation, or a direct quote, not just the necessity to avoid identity with the source text and to achieve grammatical integration (Plakans & Gebril 2013: 227). Teaching "mechanic strategies" (Chan 2017: 25) to avoid plagiarism is not likely to convert into improved writing skills. Such strategies may even stand in the way of effective learning of intertextual strategies (Yamada 2003).

Teachers should bear in mind that L2 writers use patchwriting as a learning strategy of language appropriation, and students should not be deprived of this way of learning to write academically and negotiating complex ideas in their sources (see also Currie 1998; Shi 2004). It is important for teachers to strike a balance here and to draw attention to alternative ways in which students can improve their writing skills. This will enable them to become competent members of the respective academic discourse community. In a similar vein, the present study suggests that the teaching of paraphrasing, summarisation, and direct quotation must include a discussion of unavoidable and desirable overlap as well as the functions of textual borrowing. Students have to know what is accepted in their discipline and what is not. The focus should be on content and semantics of overlapping phrases, not merely on the length of a copied string that may or may not be acceptable. Such an approach could also include the explicit teaching of vocabulary learning strategies using source texts as a language resource.

This study suggests that the nature of paraphrases crucially depends on the type of source text and its properties, for example its length. While even the advanced L2 writers in Keck's (2006) study re-used longer stretches from the source text in their summaries, the students in this study borrowed very little and tended to combine propositions from several sentences into one paraphrase. Clearly, the writing task, the length of the source text, and the fact that it was an academic research paper rather than a non-fiction newspaper article influenced their paraphrasing strategies. These observations have important implications for teachers. Source texts for paraphrasing exercises must be carefully selected because they may require different intertextual approaches. If the aim of a summary task is to prepare students for writing an academic literature review, then the source texts should be academic research articles. This also means that examples that are used in teaching intertextual strategies should stem from the same text type, because the intertextual strategies employed in a research article may not be helpful as orientation for students writing a reading report, and vice versa. Finally, the intertextual requirements of writing a particular text type must be made explicit in the task description.

Paraphrasing activities should focus on both the use for knowledge telling and for knowledge transforming so that L2 student writers are equipped for writing text types that require these skills. This includes raising awareness of the various ways in which source text material can be embedded for the purpose of building an argument while staying faithful to the original meaning. Students also need background knowledge of the topic to understand the source text, to summarise it, and to connect its ideas to their own. Expert writers combine

source text material with their own argument in complex ways. For this reason, Yamada (2003) has proposed that students' ability to think inferentially is essential to their ability to write intertextually and should complement the teaching of paraphrasing skills. If students understand the inferential thought processes that have occurred during the writing of experts' paraphrases, they may be able to gradually adopt these themselves.

A basic prerequisite for effective intertextuality is, of course, a certain aptitude in selecting those excerpts that are central to the summarised text. A writer's reading comprehension and reading strategies largely determine their ability to summarise and accurately integrate source text material (see e.g. Kim 2001; Plakans 2009a; Uludag et al. 2019: 5). Ideas from source texts can only be transferred accurately and connected if students understand them. For this reason, exercises that scaffold the reading of academic research articles are an important preparation for summary writing (see also Hirvela 2016; Gebril & Plakans 2013; Plakans & Gebril 2012) and should be integrated into source-based tasks. As it has been suggested that specific paraphrasing strategies may enhance reading comprehension in inclusive middle school classrooms (Katims & Harris 1997) and undergraduate EFL contexts (Karbalei & Amoli 2011), reading-into-writing exercises may even have an effect in several areas of literacy attainment. Furthermore, drawing L2 writers' attention to their reading processes in their L1 could positively influence the ways in which they form interpretations of ideas presented in source texts written in their L2 and connect them in their own writing (see Yamada 2003).

Strategies of attribution need to be taught explicitly as well and must go beyond formal explanations of the formatting and frequency of references. Teachers should focus on the importance of being able to deduce the origin of a statement. It was evident in the process study that some students understood reporting structures and attribution as means of creating stylistic variety rather than intertextual reference. It would thus be productive for teachers to delineate their purposes first and then discuss stylistic options, including reporting structures and referencing as well as various forms of in-text attribution. Teachers should also verify students' understanding of reporting verbs and their use of reporting structures. In case of an obvious reliance on a restricted set of verbs and structures, teachers can draw attention to alternative structures to introduce more variation. The evaluative potential of reporting verbs should be pointed out in a suitable context, i.e. when writing texts that require the students to take a stance towards the cited sources.

For such measures to be fruitful, L2 writers also require sufficient knowledge of grammar and lexis to be able to paraphrase source text material and embed citations in reporting structures in meaningful ways. Teachers should provide sufficient support, for example by having students search for appropriate expressions in published articles, noting their use and position, and compiling lists of reporting structures (see also Chen et al. 2015). Generally, the results reported here suggest that L2 student writers are aware of reporting practices and that 'overuse' of certain verbs is probably not an issue for the majority at this level of writing expertise. Corrective measures should only be taken in cases of incorrect, unintelligible, or transgressive reporting of source content. Since there is a lot of variation in the ways in which sources are acknowledged and evaluated, students should primarily be made aware of the multitude of means of attribution and reporting. It is furthermore important to raise awareness of register-specific use of reporting verbs. Swales reports that "raising the issue of how we are to account for citations has a valuable consciousness-raising effect on [our] students, both NS and NNS, and starts them towards developing a high-level sense of where and why citational support for their statements may be advisable" (1990: 7).

The present study confirms that source use is a highly individual area of academic writing, which it is why it is so important to take into account individual and contextual variables (see Hirvela 2016) and the students' current level of attainment. Intertextuality can take many forms, even in published writing. Intertextual competence should be considered "a variable and complex, rather than uniformly realized, ability" (Cumming et al. 2016: 51). This is especially important to bear in mind as several studies have shown that not all learners benefit from source-based writing exercises equally and that some do not show any improvement at all (see Chen et al. 2015; Choy & Lee 2012). Depending on their attitudes, their writing styles, and the progress of their writing process, writers may benefit from different writing strategies (Leijten et al. 2019: 577). For instance,

Wette (2017: 56) suggests that once undergraduate writers have attained a basic knowledge of both the language and the technical requirements of citation, they are likely to benefit from strategies to develop their rhetorical and conceptual skills much more than from advice on how to avoid textual overlap. Plakans and Gebril (2013: 228) propose that the micro-level processes of writing from sources should be the focus right at the beginning of students' university education, while higher-order skills could be addressed later. Students should also be given the space to develop their own processes and explore various writing styles in order to become proficient writers. In sum, any teaching of source-based writing should allow for individual approaches and provide exercises that match the individual students' level of academic writing expertise.

Indeed, the two studies presented here show that many aspects of intertextuality, such as the integration of direct quotes and the selection and structuring of important information, have already been mastered by these students at the end of their first year of study, though not always to the same extent. They are also aware of the potential repercussions of plagiarism and seek to avoid it in their writing. Overall, the students' citation practices are relatively similar to those found in expert writing (see e.g. K. Hyland 2002, 2004). Such findings show the relevance of studies of learner writing to identify not only the problems, but also those aspects of writing in the L2 that have been mastered.

Research into the effectiveness of instructing students in writing from sources unanimously shows that students do improve through explicit instruction (see overview in Cumming et al. 2016: 52). Teachers are therefore encouraged to address aspects of source use with specific exercises and comprehensible individual advice instead of issuing general warnings of plagiarism. Since the present study was conducted in Germany, this implication deserves special emphasis because students at German universities are often still expected to acquire academic writing skills by themselves and are rarely given explicit instructions (Breuer 2014; Pohl 2007). If students are not given comprehensive support, they may spend an insufficient amount of time on reading and their intertextual strategies tend to consist of transferring information in very simple ways, sometimes by copying (see e.g. Plakans & Gebril 2012; Solé et al. 2013). It is clear from the students' comments in this and previous studies (e.g. Davis 2013; Petrić & Harwood 2013) that teacher's advice with respect to source use does not fall on deaf ears and that students are eager to meet their teachers' expectations. Teachers can exploit this by making their expectations regarding source use transparent and offering individualised support for improving intertextual writing skills.

Ideally, this should take into account the whole process – from reading through writing to editing. Despite a growing body of research into reading-into-writing processes, explicit teaching of writing processes is rare (see Dovey 2010; Ruiz-Funes 1999: 45 for discussion). Ruiz-Funes (1999: 56) observes that

[w]e often take for granted that our students are well aware of their own reading and writing processes and that they only need our assistance in learning the language. Yet, they do need our guidance in order for them to exploit the full potential of each of these processes and, more importantly, to become fully aware of the function(s) that each of these processes has in the assigned reading-to-write task.

Studying the processes of intertextual writing is the first step towards designing teaching exercises that serve to improve students' understanding and command of the process (see also Ruiz-Funes 1999). In order to improve students' intertextual writing, they need to be made aware of the structure of the writing process and the micro-level processes of which it is comprised, and teachers should be prepared to include process strategies into their teaching. Process-based research has shown that active forms of engagement with the source text, especially those that trigger knowledge processing, are associated with high test scores and little textual overlap. Source text interactions that comprise the selection, organisation, and connection of information have a positive impact on L2 writers' text quality, whereas copying, patchwriting, and the application of templates do not (Yang & Plakans 2012). The way in which students engage with their sources thus determines their success in writing. Process-oriented teaching of source-based writing should thus focus on such organisational skills.

The data analysed for this study reveal that students approach source-based writing tasks in very conscious ways. They have a lot of explicit and implicit knowledge about writing from sources, but are also insecure about certain aspects. The analysis brought to light many individual strengths, issues, and concerns. It is therefore recommended that the teaching of intertextual strategies be based on a needs analysis using authentic data of student writing products and processes. Corpus data can be used as an analytical tool to identify areas of concern, and as a source for positive and negative examples of source use. If teachers build corpora of their own students' writing, manual and automated analyses can provide insights into qualitative aspects of source use, e.g. the nature of textual borrowing, the integration of direct quotes, and the ways in which content is reported and attributed.<sup>98</sup> Such a corpus-based approach to intertextuality would allow teachers to identify issues that concern individuals as well as larger groups of students and consequently to tailor their teaching to their students' needs. In addition, corpora of published research articles can be used to raise students' awareness of how experts create intertextuality, bearing in mind text type-specific differences. Examples from learner corpora can also serve to illustrate effective and inappropriate uses of intertextuality from a context that the students are familiar with, especially once they have mastered the formal aspects of source use. Such exercises may, for example, have the goal of raising students' awareness of what to cite, how to select excerpts of appropriate length for quotation and embedding them in paraphrases, how to paraphrase excerpts of technical nature, and how to integrate direct quotes into the co-text so that they are effective in terms of both style and content.

Once students' strengths and weaknesses have been tentatively identified via corpus analyses, they can be confirmed through recordings of the writing process. It has been established in earlier studies that students may find it difficult to differentiate paraphrased from plagiarised material (Roig 1997) and sometimes overestimate their ability to paraphrase effectively (Choy & Lee 2012). Such issues could be more effectively identified and targeted if the writing process is recorded. The methods used in this study can even be incorporated into teaching units. The findings suggest that watching one's own process can raise students' awareness of their own issues as well as helping teachers identify students' misconceptions regarding academic writing conventions.<sup>99</sup> Watching recordings of their writing processes with students also bears the potential of triggering self-reflection and informing self-assessment (see Ranalli, Feng & Chukarev-Hudilainen 2018: 89f.; Sullivan & Lindgren 2002). The goal should be to create a space for students to ask for what they need in teaching based on issues discovered in their processes and products. Process data could be used as a starting point for discussions between teachers and students. Insights from screen recordings should also be used to praise students for effective strategies.

Aside from knowledge about a variety of linguistic features and writing processes, effective academic writing also presupposes an understanding of the respective discourse community's culture and conventions. It is paramount that source use exercises are meaningful and applied in the context of the disciplinary discourse because "[i]t is by participating in the activity of the community that one can become an insider in the knowledge and conventions of that circle" (Canagarajah 2002: 30). It is important to raise awareness of the role of citations in constructing a persuasive argument in the relevant discipline. Students should be alerted to the disciplinary differences not only in citation frequency, but also in the use of different intertextual manifestations and their positioning in a sentence (see also Okamura 2008a, b). The present study revealed students' insecurities regarding the features of the text type, and such issues can be easily addressed by talking about expectations.

Integrated writing tasks should ideally be situated in authentic discourse, for example research papers that make a real contribution to research. Engaging with and orienting towards the conventions of a discourse community in safe learning spaces may help students developing the knowledge and practices necessary to become full members of the community (Canagarajah 2002). Such exercises may also empower L2 writers by fostering their self-image as academic writers and researchers. The teaching of intertextual strategies should

98 For instance, Hyland (2002) emphasises the need for explicit teaching of reporting verbs as they are crucial in convincing the reader of the writer's membership in the respective research community. According to him, the discipline-specific differences in the selection of reporting verbs require attention in EAP classes. Such teaching can be informed by corpus analyses of expert as well as novice writing.

99 See Ranalli et. al (2018) for related findings regarding the use of process-tracing technologies in L2 writing assessment.

place a strong focus on combining other authors' ideas without being dependent on one's sources. It has been suggested that students can improve their ability of avoiding unintentional plagiarism if they understand authorship and learn how to adopt a more authorial role when writing academically (Elander et al. 2010). Students should be encouraged to explore a variety of possible textual interpretations when working with source texts so that they are able to develop their authorial voice (Angéil-Carter 2000, cited in Yamada 2003). In the present study, the students often reflected on their own role as an academic writer. Such reflections offer a fruitful starting point for discussions of writer's stance and positioning oneself in the discourse community.

The learning of intertextual strategies requires contextualised, goal-oriented teaching, which in turn requires qualified teachers with extensive academic writing expertise in their discipline. Source use exercises should thus be integrated into disciplinary courses or taught by qualified teachers familiar with the conventions of their discipline. Teaching of source use is much more effective if teachers themselves are knowledgeable and experienced in the use of intertextuality in various genres: "The varied academic practices of writing and reading cannot be seen as general skills that can be taught in marginalised university 'Language Centres' by anyone with a reasonable grasp of English and a textbook" (K. Hyland 2000: 145). The interviews conducted in this study bear evidence to students' confusion over varying citation practices across disciplines and contrasting teacher preferences. The teaching of source use practices should therefore include raising L2 writers' awareness of variation in the creation of intertextuality both within and across disciplines.

Furthermore, teachers of disciplinary courses have to move away from basing assessment on intertextuality without making sure that their students are equipped with the means necessary to use sources meaningfully and in line with academic writing conventions. The intended instructional outcomes need to be reflected in the instructional processes: If assessment is going to be based on writing, then students have to be taught how to write. This certainly also presupposes departmental agreement on the intertextual practices and citation styles that are deemed acceptable and the curricular integration of source-based writing for specific text types. This is clear from the statements made by the students in this study pertaining to confusing and conflicting instructions given to them by lecturers of the same department. Their statements suggest that there may be a mismatch between teachers' expectations of what students should be able to do and what they are actually taught. Only if a department's rules concerning acceptable source use are accessible, transparent, and unambiguous do L2 writers stand a chance of internalising and applying them in expected ways. Teachers can do their part by supporting students' learning processes and reconsidering their own advice and guidelines.

In sum, students need explicit teaching and awareness-raising activities in order to improve their intertextual competence. Students should be alerted to their own patchwriting and possible misconceptions about textual borrowing. They must be equipped with easy-to-follow, step-by-step guidelines for writing paraphrases, summaries, and generalisations, and for integrating direct quotes. Once they are able to use these intertextual strategies in acceptable ways, the focus can be shifted to more fine-grained aspects. Students need information on the functions of the various intertextual strategies as well as on attribution and documentation of sources. In order to create texts that are formally and also stylistically effective, they also need advice on how to combine various intertextual manifestations and how frequently to use them. It is pertinent for teachers to bear in mind that intertextual strategies "are not uniform practices reducible to generic advice" (K. Hyland 2000: 145) and to value and encourage individual approaches within the boundaries of acceptability. Teachers of disciplinary courses should view it as their responsibility to teach students the intricacies of source use in their discipline and should use synergies arising from the students' knowledge from other contexts. It is important that teachers instruct their students not just in the correct implication of style guides, but also inform them of the purposes of source use and expert writers' approaches across genres. Corpus and process data can complement and enhance the teaching of source use and help identify areas for improvement both in the processes and the products of source-based academic writing.

## 8.2 Limitations

As with any empirical research, it must be acknowledged that there are certain limitations to the present study of intertextuality in L2 writing. The findings regard a specific text type, namely the reading report. An advantage of reading reports is that they alleviate some of the challenges of complex writing tasks based on multiple source texts. Though there are certain advantages to limiting the writing task to one source text, reading reports pose a similar risk as isolated literature reviews.<sup>100</sup> They require students to report information from a source text without embedding it in an argument. This may have influenced the students' strategies, especially as paraphrasing by using one's own words is an explicit requirement in the task description. The results thus cannot necessarily be generalised to other text types.

The students in this study had a specific linguistic, cultural, and educational background. Their intertextual strategies may not be representative of other cohorts with different L1s, or even of other cohorts with the same L1 but different writing experience. A study by Shi and Dong (2018) suggests that intertextual practises from the students' native language may influence their source use in the L2. This may very well be the case here since all students in this study also have academic writing experience in German. Their prior education about avoiding plagiarism and the text types they have written in school and at university are also likely to have had a bearing on their source use. The group of students was also relatively small to allow for an in-depth analysis of writing processes, but those of other students possibly differ from the ones observed here. The fact that many findings of the present study corroborate previous research suggests that the patterns are also found in other cohorts. Further studies of more writers of the same and of different backgrounds are needed to confirm this.

Though many of the features of L2 student writers' texts have been argued to be a result of their academic writing expertise, not their language proficiency, it is possible that some of the differences between corpus and process data are related to language proficiency, especially when it comes to textual borrowing. While English language proficiency in the process study was externally assessed by means of the *Oxford Quick Placement Test* (Syndicate 2001), it was not possible to take it into account in the corpus study. Furthermore, writing proficiency was not assessed, and the grades for the reading reports in the corpus were not available in the metadata, so it is probable that the groups were not homogeneous in this respect. It is also unclear to what extent the topic of the source text and the participants' prior knowledge acted as confounding variables. It is conceivable that individual academic writing skills and experience account for variation in source use practice, both in the corpus and the process study. The comparison between the two studies must thus be taken with a grain of salt, though there are obvious advantages to a broader-scale corpus analysis.

While the task given in the experiment is representative of assignments in this particular linguistics programme, the artificial setting does not reflect a typical scenario of writing a reading report. The exclusive use of electronic media must also be considered, as students may use printed sources in other settings (see e.g. Choi 2016). Real-life writing performance may be quite different from the response to an ad-hoc writing task (Stapleton 2010: 296). McCulloch (2013: 137) suggests that writing tasks which take place in one sitting elicit "strategies that are more typical of (and more suited to) exam-type writing than those they would apply under more naturalistic conditions, where students read multiple self-selected sources and read, write and reflect over a longer period of time". This is also true for the students participating in the present study and is likely to have affected their writing processes, which may be different in a naturalistic environment.

In their review of literature on the use of verbal report data in SLA research, Gass and Mackey (2000) list possible disadvantages of stimulated recall procedures, for example that cognitive processes are complex, that some processes may be unconscious, and that participants may not report them accurately (see also Stapleton 2010: 305). These limitations certainly have to be borne in mind when considering the interpretations made here. Stimulated recall mediates, but does not provide direct access to the thought processes that are at work during

---

<sup>100</sup> As Dovey (2010) showed, student writers find it more difficult to connect their sources meaningfully if they are not required to create a research gap. In her study, students struggled with the task of writing a literature review that was not embedded in a project with a clearly stated goal.

writing. They also initiate new cognitive processes. As a result, the stimulated recall is likely to contain some interpretation by the students of their own actions. The students may have – intentionally or unintentionally – omitted or exaggerated certain aspects, for example because of their awareness of plagiarism. They also may have forgotten certain details, as they sometimes openly admitted. Students’ awareness of intertextual strategies may be heightened by participation in the stimulated recall (see also Ruiz-Funes 1999: 55-56) so that they ascribe them more importance than they did during writing. It must be acknowledged that, as with any retrospective self-report, the stimulated recall provides an incomplete account of students’ actual motivations and source use strategies: “Even if the researcher is not the student’s instructor and has no influence on grades, young novice writers may still feel hesitant to talk openly with someone on his or her own university campus.” (Li & Casanave 2012: 178; see also McCulloch 2013) Qualitative research always presupposes a certain selectivity and bias. Nevertheless, in the absence of neurological measures, screen recordings and interviews are a useful window to students’ writing process and strategies, and utmost care was taken to minimise bias by triangulating research methods, by employing reliability measures, and by mindful analysis.

### **8.3 Directions for future research**

The results of this study deepen our understanding of intertextuality in L2 academic writing and also offer exciting starting points for future research. While some of the presented findings may be specific to writers with these characteristics, i.e. German L2 writers in their first year at university, the comparison to the existing research confirms that many of the results also apply to larger groups or even novice academic writers in general. Future mixed-methods studies of intertextuality in the writing of other learner cohorts, e.g. with a different native language and educational background, will thus provide us with a more comprehensive understanding of which intertextual strategies are context-specific and which are general features of novice writers’ composing processes and their products. In addition, a comparison of the strategies and micro-level processes of the same writers in parallel source-based writing tasks in their L1 and their L2 represent a tantalising subject for further research.

The process study confirms that source use depends not just on language proficiency and writing experience, but also on the degree to which student writers critically engage with teachers’ advice and the strategies observed in the texts they read. The processes of those students who relied on generic advice differ from those of their peers who created their own conceptualisations of writing from sources and were critical of their own writing. As the texts composed in the process study were not graded, it is impossible to judge the effect of these different attitudes on the quality of the text, but this warrants further research. In general, it would be interesting to expand on existing research from testing contexts (e.g. Plakans & Gebriel 2012, 2013; Uludag et al. 2019) to gauge the correlation of certain micro-level processes with grades in assignments typical of disciplinary contexts. This would include an investigation of which source use is viewed favourably by teachers and prompts them to give high grades. Such studies could also integrate a comparison of the writing processes of high- and low-graded students with teacher interviews to determine the source use strategies that lead to successful and to problematic products and have a direct positive or negative effect on students’ grades. Knowledge about processes and products of source use can also provide information about construct validity in testing contexts (Plakans 2008: 112).

The present study has shown that the structure of the source text has an influence on the structure of the writing product because of students’ chronological selection strategies. It could not be investigated whether linear approaches of interacting with the source text in the process promote similar structures and intertextual manifestations in the student’s finished text and whether recursive approaches lead to different results. It has previously been proposed that recursive approaches to creating and managing the structure of a source-based text and to integrating sources should be explicitly taught in academic writing contexts, especially for large-scale

writing assignments (Dovey 2010), so it is worth exploring how different approaches in the process materialise in student writing. Researchers may also take an anti-chronological approach by analysing the products of real-life writing processes for intertextual strategies first and then observing these strategies in the corresponding screen recordings. This would allow them to identify similarities and differences in the processes leading to certain kinds of intertextual links, e.g. paraphrases, summaries, and summary paraphrases at the micro-level. Such a process-based approach would also help to discern whether the distinction of these intertextual links is useful or whether such phenomena should in fact be considered as different surface manifestations of the same underlying intertextual strategy.

Beyond the recursiveness of micro-level processes, the macro-level processes of the source-based writing process tend to be individually different and may also recur, interrupt each other, and overlap (Choi 2016; Plakans 2009b). In a longitudinal writing process, for example one of writing a research paper, it is likely that there are several pre-writing, writing, and post-writing stages. A writing stage may be followed by a post-writing stage, which in turn is followed by another writing stage, possibly on a different day, and the beginning of a new chapter may later prompt a new pre-writing stage. This would also entail new combinations of macro- and micro-level processes. Future studies, especially longitudinal ones, may look into the linearity and recursiveness of long-term writing projects to identify differences between the writing processes associated with different types of tasks.

As discussed above, such a controlled writing task as a reading report is not representative of other academic text types, e.g. research papers. According to McCulloch 2013: 138), it is possible that “when student writers are engaged in higher-stakes writing and required to use a wide range of self-selected sources, they represent the task differently, taking more time to locate and compare sources”. As a result, the processes and strategies of writing assignments which are based on both previous literature and the student’s own research are likely to be different, because they require writers to transform knowledge and align themselves with the disciplinary discourse (Dovey 2010: 46). The present study provides detailed insights into source use strategies for this particular text type and a fruitful starting point for parallel studies of other text types.

Strictly speaking, the present study only considered a part of the writing process. According to Krings (1992), the writing process begins with becoming aware of the writing task and ends with the finalisation of the writing product. An interesting starting point for research of the entire writing process are students’ reading skills as they are assumed to have at least a mild impact on source use in integrated writing tasks.<sup>101</sup> Because the students were asked to read the source text at home prior to the experiment, it cannot be excluded that their approach to reading and their textual comprehension had an effect on their writing processes which could not be considered here. Their knowledge of the topic is also likely to influence their understanding of the source text and consequently their approach to the writing task. Future studies may address these aspects by integrating data on students’ writing expertise and their strategies of reading the source texts and preparing their writing with analyses of the writing process.

Screen recordings make writing processes visible and the findings of this study are thus complementary to previous research using think-aloud protocols and keystroke logging (e.g. Breuer 2014; Chan 2017; Hirvela & Du 2013; Krings 1989; Leijten & Van Waes 2013; Leijten et al. 2019; Plakans 2008). Our knowledge of source-based writing processes could be further extended by triangulating these process-oriented research methods with eye-tracking. This method is expected to provide additional information about the micro-level processes of source use, especially about sequences without on-screen action, for example by making visible which parts of the source text a student reads before writing a paraphrase and which words they focus on. Such an approach would provide valuable insights into the cognitive processes leading to textual borrowing and the selection of source text material in the writing process.

---

101 See e.g. the correlational study by Asención Delaney (2008) and the qualitative process studies by Esmaeili (2002) and Plakans (2009a).

It would be an interesting endeavour to investigate the entire process of an authentic source-based writing task, from inception to completion, and the role of interaction with the source text therein, to explore whether the processes and strategies observed in the experimental task in this study are representative of real-life source-based writing in the L2. Moreover, research into the composing processes of L2 writers has shown that the time allotted to individual composing behaviours differs in naturalistic settings. This is the case because real-life academic writing tasks require students to generate ideas on what to write, research their topics, and locate and read potential sources, especially if using multiple sources (Choi 2016; Stapleton 2010). Such processes do not occur in the writing of reading reports in an experimental setting and thus could not be considered in the present study.

In general, the time spent on individual micro-level processes and how this relates to the quality of the writing product seem worth exploring, especially if this knowledge is transferred into academic writing pedagogy. Writing processes are also likely to be different if students take breaks, write in their preferred location, and use their own equipment and tools. Studies of real-life writing tasks, ideally combining screen recordings with video recordings and eye tracking, would thus expand our knowledge of source-based writing processes. Future triangulated, mixed-methods studies of source text interactions could broaden the scope of L2 writing research by guaranteeing naturalistic settings in which students select and synthesise a range of source texts and which are thus reflective of the complexity of real-life university assignments.

## 9 Conclusion and outlook

Since the writing process invariably leads to the writing product, studying the processes and products of intertextual writing can reveal which processes lead to certain products, which parts of the process are not manifested in the product, and whether similar processes lead to similar products. Corpora lend themselves to an analysis and quantification of the linguistic features of source-based texts. Observing the process of the same writing task provides complementary insights into the manifestations of source use found in the corpus data. It also sheds light on the actual processes of reformulating and restructuring the source text during paraphrasing. A full picture of the characteristics of source-based student writing can only emerge when both the processes and the products are analysed, because not all actions from the writing process are reflected in the product. Such research allows L2 writing scholars and educators to understand patterns of source use and to explain similarities and differences between individual writers.

The present study employed a mixed-methods approach to investigate source use in L2 writers' reading reports. In the first part of the study, intertextual strategies were analysed in a corpus of reading reports written by German first-year students of English. In the second part, the intertextual writing processes of six L2 writers from the same degree programme were studied on the basis of screen recordings and stimulated recall data. As reflected in the three central research questions, intertextual links, strategies of attribution, documentation, reporting, and (strategic) textual borrowing were analysed in both parts of the study in order to shed light on how students create and combine various intertextual manifestations. The goal of this study was to examine the strategies used by L2 student writers in source-based writing and to identify patterns and markers of intertextuality in writing processes and products.

Taken together, the corpus study and the process study have produced five central results regarding proficient source-based writing in L2 English that pave the way for a new evaluation of students' intertextual strategies. The first result is that L2 writers' reading reports are created by employing a broad range of non-transgressive intertextual strategies. Intertextual links are combined in very individual ways and differ in their closeness to and reliance on the source text. They are comprised of paraphrases of individual sentences, gist statements, and direct quotes as well as meta-level observations and summaries of the entire text, among others. Paraphrases are the most common type of intertextual link and are combined with other manifestations of intertextuality in intricate ways to convey the content of the source text.

The second central finding is that intertextual writing is achieved by a combination of multiple recursive micro-level processes in the writing stage. These include core acts such as adding words, substituting synonyms, deleting, and rephrasing, which are repeated throughout all writers' processes. The operations on the micro-level of paraphrasing processes correspond to the D/A/S strategies identified in the corpus study. Intertextual links are always created in close interaction with the source text, yet differences emerge in the writing process as a result of students' preferred way of accessing it.

Third, L2 writers employ a multitude of means to acknowledge authorship of source text ideas and phrases, not just references. Attribution, documentation, and reporting structures serve to characterise passages as intertextual and are omnipresent in students' texts. Their individual approaches generally result in transparent acknowledgement of sources in accordance with academic citation conventions. The students are aware of the need to avoid plagiarism. They employ manifold intertextual markers to achieve this goal, but also to instil stylistic variety into their writing.

The fourth result, namely that writing processes can be linear, recursive, and knowledge-transforming at the same time, allows for a deeper understanding of the recursiveness of the writing process. L2 writers are inspired by the chronology of the source text when writing reading reports and follow its structure when

extracting information and words from it. They consciously select specific passages as the basis of their intertextual links and their pre-writing processes influence the way in which these are used in the writing stage. Linear and recursive elements co-occur and together result in reading reports that are in line with the task description.

The fifth finding regards the extensively researched subject of textual borrowing. Much textual borrowing in this study concerns source text terminology and other phrases that are essential for a correct representation of source text information, although there is also textual overlap that would be considered problematic by most teachers. The study reveals that the borrowing of individual words from the source text may be accidental and that there are very few instances in which the source text is used strategically for language support. Copying serves as a scaffolding strategy in the writing process and actually serves as a way of avoiding textual overlap for these students.

The present study adopted a positive perspective on students' intertextual strategies by focusing on proficient source use. It revealed many expert-like intertextual manifestations and motivations for source use, pointing towards the advanced academic writing skills of these students. The students in the process study show a well-rounded awareness of plagiarism and academic conventions regarding intertextuality, often relying closely on their teachers' advice. They take great care to avoid textual overlap in their writing processes. In this and in previous studies, apparent cases of transgressive intertextuality are more often than not caused by insufficient writing experience and insecurity regarding conventions. The study thus confirms that source use is primarily a matter of academic literacy. Many of the results of the present study are likely to be generalisable to novice writers irrespective of their native language. It therefore appears most useful to distinguish between novices and experts, leave behind the distinction of writers based on L1 background in research and pedagogy, and focus on the teaching of effective academic writing skills.

The corpus study offers new insights into the linguistic features of integrated writing tasks. It provides a comprehensive quantitative and qualitative analysis of paraphrases, direct quotes, summaries and other intertextual links and how they are combined with references, reporting structures, and attribution. It enhances our understanding of students' individual preferences, for example regarding combinations of intertextual links and reporting structures, of the chronological selection of source text ideas as well as of desirable and undesirable forms of textual borrowing. The combination of screen recordings and stimulated recall in the process study has allowed for the collection of rich data, which in turn has facilitated a detailed and deep analysis of the phenomena under scrutiny. It has brought to light the succession of micro-level processes of source-based composition and the steps students take in the construction of the text as a whole.

Central aspects of L2 writing processes were investigated on the basis of these data, such as source text interaction, lexical and syntactic operations, use of language resources, and editing. This research complements previous qualitative process studies using think-aloud protocols (e.g. McCulloch 2013; McInnis 2009; Plakans 2008) and keystroke logging (e.g. Leijten & Waes 2013; Leijten et al. 2019). By making visible revision processes, deletion, and substitution and eliciting students' comments thereon, the study illuminates that appropriate intertextuality is a central concern of L2 students when writing from sources. The integration of stimulated recall discloses connections between students' writing processes and their motivations, insecurities, and subjective theories. The process and interview data empirically confirm assumptions made in the corpus study, yet each study also highlights specific aspects of source-based writing that cannot be studied in the other. The triangulated design allows for an integration of complementary perspectives on the research questions and thus creates a more complete picture of L2 writers' source use.

This study highlights the potential of pedagogic text types both for teaching source use in disciplinary contexts and for investigating intertextuality in student writing. The study results suggest that there is merit in assigning summaries of individual research articles as practice for writing research papers. Unlike summaries of short non-fiction texts and source-based argumentative essays, reading reports require L2 writers to extract information about the presented study. Such exercises can sharpen students' awareness of features of research

articles by drawing their attention to the IMRD structure. The task description and the length of the source text presuppose an elaborate process of selecting information. As a result, students have to synthesise in a focused and structured way. Reading reports allow students to practise extracting ideas from a research article without the added pressure of having to relate them to other literature, construct an argument, and take a stance. A variety of intertextual strategies can be practised in a controlled environment before moving on to more complex genres. In linguistic studies, the restrictive, focused design of reading reports allows researchers to gather information about areas of literacy development that may otherwise be overlooked. Because students are required to structure the reading report and select source text passages for intertextual links themselves, reading reports have the potential to provide fundamental insights into L2 student writers' academic writing skills.

The present study is innovative in a number of ways. First, it is the first study providing a detailed analysis of the use and combination of intertextual strategies, including the core strategies direct quotation, paraphrasing, and summarisation. It takes into consideration the role of legitimate textual borrowing and the various forms of signalling intertextuality via attribution, reporting structures, and documentation. It is original insofar as it considers all these manifestations of intertextuality together and thus takes into account the interconnected nature of source use and its acknowledgement. Second, this study combines corpus data with process data and introspective data in a proficiency-oriented approach. By combining qualitative and quantitative analyses of a broad range of intertextual phenomena, it thus offers a completely new perspective of L2 source-based writing and students' motivations. Third, it is the first study to shed light on the micro-level processes of writing from sources and fills a long-standing gap in research (Pecorari & Petrić 2014) by providing detailed insights into the operations leading to the integration of ideas and words from a source text into the student writer's text. Based on these insights, an extension of Hayes and Flower's (1980) model of the writing process is proposed that takes source use into account. More importantly, the findings regarding micro-level processes of source use, the linguistic manifestations of intertextuality, and students' underlying motivations for their approaches to writing from sources have direct implications for the teaching of intertextual strategies.

The scope of L2 writing research can be widened by combining analyses of writing processes and products to improve our own teaching of academic writing and tailor advice to students' needs. This affords researchers and educators alike with the opportunity to identify the issues underlying transgressive and unconventional intertextuality, which is a crucial step towards effectively fostering learning (Ranalli et al. 2018: 87). The findings of this study can inform the teaching of source use strategies both in academic writing classes and in disciplinary courses. Combining corpora with process data provides a basis for identifying the kind of pedagogical interventions students need for concrete issues in intertextual writing. This approach can thus be an important step towards well-founded and comprehensive writing research that informs pedagogical practice in all areas of academic writing.

In order to improve L2 student writers' ability to write from sources, we must foster their declarative, procedural, and conditional knowledge of the multiple dimensions of intertextuality and their functions in disciplinary contexts. Curricula need to be improved so that academic writing classes prepare students for the kinds of writing tasks they are expected to work on in their disciplinary classes. Writing from sources is highly individualised and affords the writer with a wide range of approaches to choose from. Consequently, it is important to give advice that is tailored to students' individual needs for a specific text type and that takes into account the multifaceted nature of intertextuality. This also presupposes that writing assignments eventually focus on expert genres. Such tasks allow students to develop communicative competence by engaging with the respective research community and becoming acquainted with its conventions regarding source use and language (Canagarajah 2002: 30).

Two desiderata for writing research emerge from the present study. First, a comprehensive understanding of intertextual writing across the curriculum would be promoted by initiating and connecting research into source-based writing processes and products in different disciplines. This would entail comparisons of the same text type across disciplines, e.g. research papers, as well as research into source use in discipline-specific genres.

Second, our knowledge about source-based writing would be significantly expanded by merging the discourses on disciplinary writing, EAP writing, and testing contexts. Ideally, this would entail bringing together perspectives of academic writing products and processes, in the L1 and the L2, in foreign-language and second-language contexts, and by novices and experts. This liaison of research would permit researchers to develop a fully-fledged description of (source-based) academic writing processes that would also serve as a solid basis for effective academic teaching and assessment task design.

The present study demonstrates the usefulness of combining corpus, process, and interview data. Mixed-methods studies of learner corpus data and process data complement the existing research in the area of intertextuality in L2 writing. They offer a more well-rounded picture of the investigated intertextual phenomena than could be achieved by a single research method because assumptions made on the basis of one type of data can directly be confirmed or disproved using another. By mixing methods, researchers can gain much more detailed insights into students' source use while at the same time reducing the risk of misinterpretation. Screen recordings also offer insights into other aspects of the writing process, such as the use of language resources and pausing, and stimulated recall procedures can be adapted according to the researcher's interest. These complementary methods can thus be employed as a supplement to corpus data in order to answer a wide range of research questions in writing research.

Writing from sources is a complex multidimensional and multimodal process. It requires L2 writers to apply their reading and writing expertise as well as their language skills to create a reader-friendly text that strikes a balance between source text information and their own knowledge and argumentation. This integration process entails many iterations of organising and re-organising source text material, which is concurrently evaluated and rephrased. While performing the demanding task of moving back and forth between the source text and their own text in the writing process, the writer must take measures to acknowledge all sources in accordance with academic writing conventions and to ensure that any intertextual passage faithfully conveys the intended meaning. By providing new data on the ways in which L2 writers achieve this challenging feat, this study provides a valuable contribution to the discourse on intertextuality in L2 writing beyond plagiarism and guidelines for improving the teaching of source-based academic writing.

# References

- Abasi, Ali R. & Akbari, Nahal (2008): "Are we encouraging patchwriting? Reconsidering the role of the pedagogical context in ESL student writers' transgressive intertextuality". *English for Specific Purposes* 27, 267–284.
- Abasi, Ali R.; Akbari, Nahal & Graves, Barbara (2006): "Discourse appropriation, construction of identities, and the complex issue of plagiarism: ESL students writing in graduate school". *Journal of Second Language Writing* 15, 102–117.
- Abasi, Ali R. & Graves, Barbara (2008): "Academic literacy and plagiarism: Conversations with international graduate students and disciplinary professors". *Journal of English for Academic Purposes* 7, 221–233.
- Ädel, Annelie & Garretson, Gregory (2006): "Citation practices across the disciplines: The case of proficient student writing". In Pérez-Llantada Auría, Carmen, Pló Alastrué, Ramón & Neumann, Claus-Peter (eds.), *Academic and Professional Communication in the 21<sup>st</sup> Century: Genres and Rhetoric in the Construction of Disciplinary Knowledge. Proceedings of the 5th International AELFE Conference*. Zaragoza: Prensas Universitarias de Zaragoza. 271–280.
- Asención Delaney, Yuly (2008): "Investigating the reading-to-write construct". *Journal of English for Academic Purposes* 7, 140–150.
- ATLAS.ti Scientific Software Development GmbH (2018): *ATLAS.ti* [computer software]. Available from <http://www.atlasti.com/>.
- Bailey, Carol & Withers, Jodi (2018): "What can screen capture reveal about students' use of software tools when undertaking a paraphrasing task?". *Journal of Academic Writing* 8 (2), 176–190.
- Bazerman, Charles (1988). *Shaping Written Knowledge*. Madison: University of Wisconsin Press.
- Bereiter, Carl & Scardamalia, Marlene (1987): *The Psychology of Written Composition*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Biber, Douglas & Reppen, Randi (1998): "Comparing native and learner perspectives on English grammar: A study of complement clauses". In Granger, Sylviane (ed.), *Learner English on Computer*. London: Longman. 145–158.
- Bloch, Joel (2010): "A concordance-based study of the use of reporting verbs as rhetorical devices in academic papers". *Journal of Writing Research* 2 (2), 219–244.
- Bloch, Joel (2018): "Textual borrowing and intellectual property". In Liontas, John I. & DelliCarpini, Margo (eds.), *The TESOL Encyclopedia of English Language Teaching*. Hoboken, NJ: Blackwell-Wiley. <https://onlinelibrary.wiley.com/doi/book/10.1002/9781118784235>; last accessed 20 June 2020.
- Bloch, Joel & Chi, Lan (1995): "A comparison on the use of citations in Chinese and English academic discourse". In Belcher, Diane & Braine, George (eds.), *Academic Writing in a Second Language: Essays on Research and Pedagogy*. Norwood, NJ: Ablex. 231–274.
- Bloomfield, Lou (2016): "Wcopyfind". *The Plagiarism Resource Site*. <http://plagiarism.bloomfieldmedia.com/wordpress/software/wcopyfind/>; last accessed 26 May 2017.
- Börner, Wolfgang (1987): "Schreiben im Fremdsprachenunterricht: Überlegungen zu einem Modell". In Lörcher, Wolfgang & Schulze, Rainer (eds.), *Perspectives on Language in Performance. Studies in Linguistics, Literary Criticism and Language Teaching and Learning* (Vol. II). Tübingen: Narr, 1336–1349.
- Borg, Erik (2000): "Citation practices in academic writing". In Thompson, Paul (ed.), *Patterns and Perspectives: Insights into EAP Writing Practices*. Reading: University of Reading, 27–45.
- Bosher, Susan (1998): "The composing processes of three Southeast Asian writers at the post-secondary level: An exploratory study". *Journal of Second Language Writing* 7 (2), 205–241.
- Breuer, Esther O. (2014): *First Language versus Foreign Language Fluency, Errors and Revision Processes in Foreign Language Academic Writing*. Doctoral dissertation. Bonn: Rheinische Friedrich-Wilhelms-Universität zu Bonn.

- Callies, Marcus (2016): "Towards corpus literacy in foreign language teacher education: Using corpora to examine the variability of reporting verbs in English". In Kreyer, Rolf; Schaub, Steffen & Güldenring, Barbara Ann (eds.), *Angewandte Linguistik in Schule und Hochschule. Neue Wege für Sprachunterricht und Ausbildung*. Frankfurt/Main: Peter Lang. 391–415.
- Callies, Marcus & Zaytseva, Ekaterina (2013): "The *Corpus of Academic Learner English (CALE)* – A new resource for the assessment of writing proficiency in the academic register". *Dutch Journal of Applied Linguistics* 2 (1), 126–132.
- Calsamiglia, Helena & López Ferrero, Carmen (2003): "Role and position of scientific voices. Reported speech in the media". *Discourse Studies* 5 (2), 147–173.
- Campbell, Cherry (1990): "Writing with others' words: Using background reading text in academic compositions". In Kroll, Barbara (ed.), *Second Language Writing: Research Insights for the Classroom*. Cambridge: Cambridge University Press. 211–230.
- Canagarajah, Suresh (2002): "Multilingual writers and the academic community: towards a critical relationship". *Journal of English for Academic Purposes* 1, 29–44.
- Chan, Sathena H. C. (2011): "Demonstrating the cognitive validity and face validity of PTE Academic Writing items". *Pearson External Research Notes*. <http://pearsonpte.com/research/research-summaries-notes/>; last accessed 18 July 2017.
- Chan, Sathena H. C. (2017): "Using keystroke logging to understand writers' processes on a reading-into-writing test". *Language Testing in Asia* 7 (10). <https://languagetestingasia.springeropen.com/articles/10.1186/s40468-017-0040-5>; last accessed 18 July 2017.
- Chandrasegaran, Antonia (2000): "Cultures in contact in academic writing: students' perceptions of plagiarism". *Asian Journal of English Language Teaching* 10, 91–113.
- Chandrasoma, Ranamukalage; Thompson, Celia & Pennycook, Alistair (2004): "Beyond plagiarism. Transgressive and nontransgressive intertextuality". *Journal of Language, Identity and Education* 3 (3), 171–193.
- Charles, Maggie (2006a): "Phraseological patterns in reporting clauses used in citation: A corpus-based study of theses in two disciplines". *English for Specific Purposes* 25, 310–331.
- Charles, Maggie (2006b): "The construction of stance in reporting clauses: A cross-disciplinary study of theses". *Applied Linguistics* 27 (3), 492–518.
- Chen, Mei-Hua; Huang, Shih-Ting; Chang, Jason S. & Liou, Hsien-Chin (2015): "Developing a corpus-based paraphrase tool to improve EFL learners' writing skills". *Computer Assisted Language Learning* 28 (1), 22–40.
- Cheng, An (2006): "Analyzing and enacting criticism: The case of an L2 graduate learner of academic writing". *Journal of Second Language Writing* 15, 279–306.
- Choi, Yeon Hee (2016): "Writing strategies in the process of L2 computer-mode academic writing with the use of multiple resources". *English Teaching* 71 (3), 3–28.
- Choy, S. Chee & Lee, Mun Yee (2012): "Effects of teaching paraphrasing skills to students learning summary writing in ESL". *Journal of Teaching and Learning* 8 (2), 77–89.
- Clark, Herbert H. & Gerrig, Richard J. (1990): "Quotations as demonstrations". *Language* 66 (4), 764–805.
- Colyar, Julia (2009): "Becoming writing, becoming writers". *Qualitative Inquiry* 15 (2), 421–436.
- Crocker, Jean & Shaw, Philip (2002): "Research student and supervisor evaluation of intertextuality practices". *Hermes Journal of Linguistics* 28, 39–58.
- Cumming, Alister (2002): "Assessing L2 writing: Alternative constructs and ethical dilemmas". *Assessing Writing* 8, 73–83.
- Cumming, Alister; Kantor, Robert; Baba, Kyoko; Erdosy, Usman; Eouanzoui, Keanre & James, Mark (2005): "Differences in written discourse in independent and integrated prototype tasks for next generation TOEFL". *Assessing Writing* 10, 5–43.
- Cumming, Alister; Lai, Conttia & Cho, Hyecheon (2016): "Students' writing from sources for academic purposes: A synthesis of recent research". *Journal of English for Academic Purposes* 23, 47–58.

- Currie, Pat (1998): "Staying out of trouble: Apparent plagiarism and academic survival". *Journal of Second Language Writing* 7, 1–18.
- Davis, Mary (2013): "The development of source use by international postgraduate students." *Journal of English for Academic Purposes* 12 (2), 125–135.
- Davis, Mary & Morley, John (2015): "Phrasal intertextuality: The responses of academics from different disciplines to students' re-use of phrases". *Journal of Second Language Writing* 28, 20–35.
- Dong, Yu Ren (1996): "Learning how to use citations for knowledge transformation: non-native doctoral students' dissertation writing in science". *Research in the Teaching of English* 30, 428–457.
- Dovey, Teresa (2010): "Facilitating writing from sources: A focus on both process and product". *Journal of English for Academic Purposes* 9, 45–60.
- Du, Qian (2013): "What Do You Mean 'In My Own Words'?" *Undergraduate ESL Writers' Paraphrasing Experiences in an Advanced Academic Writing Course*. Doctoral dissertation. Columbus, OH: Ohio State University.
- Dubois, Betty Lou (1988): "Citation in biomedical journal articles". *English for Specific Purposes* 7, 191–193.
- e-werkzeug (2015): *easytranscript* [computer software]. <https://e-werkzeug.eu/index.php/de/>; last accessed in November 2017.
- Egbert, Jesse & Plonsky, Luke (2015): "Success in the abstract: exploring linguistic and stylistic predictors of conference abstract ratings". *Corpora* 10 (3), 291–313.
- Elander, James; Pittam, Gail; Lusher, Joanne; Fox, Pauline & Payne, Nicola (2010): "Evaluation of an intervention to help students avoid unintentional plagiarism by improving their authorial identity". *Assessment & Evaluation in Higher Education* 35 (2), 157–171.
- Erikson, Martin G. & Erlandson, Peter (2014): "A taxonomy of motives to cite". *Social Studies of Science* 44 (4), 625–637.
- Esmaceli, Hameed (2002): "Integrated reading and writing tasks and ESL students' reading and writing performance in an English language test". *Canadian Modern Language Journal* 58 (4), 599–622.
- Fairclough, Norman (1992): "Intertextuality in critical discourse analysis". *Linguistics and Education* 4, 269–293.
- Feilke, Helmuth (2012): "Bildungssprachliche Kompetenzen fördern und entwickeln". *Praxis Deutsch* 233, 4–14.
- Flower, Linda S. & Hayes, John R. (1981): "A cognitive process theory of writing". *College Composition and Communication* 32, 365–387.
- Flower, Linda S.; Hayes, John R.; Carey, Linda; Schriver, Karen & Stratman, James (1986): "Detection, diagnosis, and the strategies of revision". *College Composition and Communication* 37, 16–55.
- Flower, Linda; Stein, Victoria; Ackerman, John; Kanz, Margaret J.; McCormick, Kathleen & Peck, Wayne C. (1990): *Reading-to-Write: Exploring a Cognitive and Social Process*. New York: Oxford University Press.
- Flowerdew, John & Li, Yongyan (2007): "Language re-use among Chinese apprentice scientists writing for publication". *Applied Linguistics* 28 (3), 440–465.
- Galbraith, David & Vedder, Ineke (2019): "Methodological advances in investigating L2 writing processes. Challenges and perspectives". *Studies in Second Language Acquisition* 41, 633–645.
- Gass, Susan M. & Mackey, Alison (2000): *Stimulated Recall Methodology in Second Language Research*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Gebriel, Atta & Plakans, Lia (2009): "Investigating source use, discourse features, and process in integrated writing tests". *Spain Fellow Working Papers in Second or Foreign Language Assessment* 7, 47–84.
- Gebriel, Atta & Plakans, Lia (2016): "Source-based tasks in academic writing assessment: Lexical diversity, textual borrowing and proficiency". *Journal of English for Academic Purposes* 24, 77–88.
- Grabe, William & Zhang, Cui (2013): "Reading and writing together: A critical component of English for academic purposes teaching and learning". *TESOL Journal* 4 (1), 9–24.

- Groom, Nicholas (2000): "Attribution and averral revisited: three perspectives on manifest intertextuality in academic writing". In Thompson, Paul (ed.), *Patterns and Perspectives: Insights into EAP Writing Practice*. Reading: CALS. 14–25.
- Harwood, Nigel (2009): "An interview-based study of the functions of citations in academic writing across two disciplines". *Journal of Pragmatics* 41 (3), 497–518.
- Harwood, Nigel & Petrić, Bojana (2012): "Performance in the citing behaviour of two student writers". *Written Communication* 29, 55–103.
- Hasselgren, Angela (1994): "Lexical teddy bears and advanced learners: a study into the ways Norwegian students cope with English vocabulary". *International Journal of Applied Linguistics* 4 (2), 237–258.
- Hawes, Thomas & Thomas, Sarah (1997): "Tense choices in citations". *Research in the Teaching of English* 31 (3), 393–414.
- Hayes, John & Flower, Linda (1980): "Identifying the organization of writing processes". In Gregg, Lee W. & Steinberg, Erwin R. (eds.), *Cognitive Processes in Writing*. Hillsdale: Erlbaum. 3–30.
- Heine, Carmen & Knorr, Dagmar (2015): "Selbstreflexion akademischen Schreibhandelns anstoßen. Nicht-direktive Gesprächsführung als Haltung des Betreuenden". In Knorr, Dagmar; Lehnen, Katrin & Schindler, Kirsten (eds), *Schreiben im Übergang von Bildungsinstitutionen*. Frankfurt/Main: Peter Lang. 109–131.
- Heine, Carmen; Knorr, Dagmar; Spielmann, Daniel & Engberg, Jan (2014): "New methods of text production process research combined". In Knorr, Dagmar; Heine, Carmen & Engberg, Jan (eds.), *Methods in Writing Process Research*. Frankfurt/Main: Lang, 123–145.
- Hirvela, Alan (2004): *Connecting Reading and Writing in Second Language Writing Instruction*. Ann Arbor, MI: University of Michigan Press.
- Hirvela, Alan (2016): "Academic reading into writing." In Shaw, Philip & Hyland, Ken (eds.), *The Routledge Handbook of English for Academic Purposes*. London: Routledge. 127–138.
- Hirvela, Alan & Du, Qian (2013): "'Why am I paraphrasing?': Undergraduate ESL writers' engagement with source-based academic writing and reading". *Journal of English for Academic Purposes* 12, 87–98.
- Howard, Rebecca Moore (1995): "Plagiarisms, authorships, and the academic death penalty". *College English* 57 (7), 788–806.
- Howard, Rebecca Moore; Serviss, Tricia & Rodrigue, Tanya K. (2010): "Writing from sources, writing from sentences". *Writing and Pedagogy* 2 (2), 177–192.
- Hryniuk, Katarzyna (2016): "The use of citations in research articles written by Polish and English native-speaker writers." In Chodkiewicz, Halina; Steinbrich, Piotr & Krzemińska-Adamek, Malgorzata (eds.), *Working with Text and Around Text in Foreign Language Environments. Second Language Learning and Teaching*. Cham: Springer, 143–157.
- Hu, Guangwei & Wang, Guihua (2014): "Disciplinary and ethnolinguistic influences on citation in research articles". *Journal of English for Academic Purposes* 14, 14–28.
- Hunston, Susan (1995): "A corpus study of some English verbs of attribution". *Functions of Language* 2, 133–158.
- Hyland, Ken (1999): "Academic attribution: citation and the construction of disciplinary knowledge". *Applied Linguistics* 20 (3), 341–367.
- Hyland, Ken (2000): *Disciplinary Discourses: Social Interactions in Academic Writing*. London: Longman.
- Hyland, Ken (2002): "Activity and evaluation. Reporting practices in academic writing". In Flowerdew, John (ed.), *Academic Discourse*. Harlow: Longman. 115–130.
- Hyland, Ken (2004): *Disciplinary Discourses: Social Interactions in Academic Writing*. Ann Arbor, MI: University of Michigan Press.
- Hyland, Ken (2005): "Stance and engagement: a model of interaction in academic discourse". *Discourse Studies* 7 (2), 173–192.
- Hyland, Ken (2008): "Persuasion, interaction and the construction of knowledge: representing self and others in research writing". *International Journal of English Studies* 8 (2), 1–23.
- Hyland, Ken (2016): "Methods and methodologies in second language writing research". *System* 59, 116–125.

- Hyland, Theresa Ann (2009): "Drawing a line in the sand: Identifying the borderzone between self and other in EL1 and EL2 citation practices". *Assessing Writing* 14, 62–74.
- Ishikawa, Shin'ichiro (2016): "Use of *that*-clauses after reporting verbs in Asian learners' speech and writing: Frequency, verb type, and *that*-omission". In Moreno Ortiz, Antonio & Pérez-Hernández, Chantal (eds.), *CILC2016. 8th International Conference on Corpus Linguistics*. 202–215. <https://easychair.org/publications/volume/CILC2016>; last accessed 17 July 2020.
- Jabulani, Sibanda (2014): "An analysis of the language of attribution in university students' academic essays". *South African Journal of Education* 34 (3), 1–10.
- Jacoby, Sally (1987): "References to other researchers in literary research articles". *ELR Journal* 1, 33–78.
- Jakobs, Eva-Maria (2003): "Reproductive writing – writing from sources". *Journal of Pragmatics* 35, 893–906.
- Johns, Ann M. & Mayes, Patricia (1990): "An analysis of summary protocols of university ESL students". *Applied Linguistics* 11 (3), 253–271.
- Johns, Ann M. & Swales, John (2002): "Literacy and disciplinary practices: opening and closing perspectives." *Journal of English for Academic Purposes* 1 (1), 13–28.
- Karbalaei, Alireza & Amoli, Fatemeh Azimi (2011): "The effect of paraphrasing strategy training on the reading comprehension of college students at the undergraduate level". *The Asian EFL Journal* 13 (3), 229–245.
- Katims, David S. & Harris, Sandra (1997): "Improving the reading comprehension of middle school students in inclusive classrooms". *Journal of Adolescent & Adult Literacy* 41 (2), 116–123.
- Keck, Casey (2006): "The use of paraphrase in summary writing: A comparison of L1 and L2 writers". *Journal of Second Language Writing* 15 (4), 261–278.
- Keck, Casey (2010): "How do university students attempt to avoid plagiarism? A grammatical analysis of undergraduate paraphrasing strategies". *Writing & Pedagogy* 2 (2), 193–222.
- Keck, Casey (2014): "Copying, paraphrasing, and academic writing development: A re-examination of L1 and L2 summarization practices". *Journal of Second Language Writing* 25, 4–22.
- Keck, Casey (2015): "Investigating textual borrowing in academic discourse: A need for a corpus-based approach". In Cortes, Viviana & Csomay, Eniko (eds.), *Corpus-based Research in Applied Linguistics: Studies in Honor of Doug Biber*. Amsterdam: John Benjamins. 177–196.
- Kettemann, Bernhard; König, Martina & Marko, Georg (2003): "The BNC and the OED. Examining the usefulness of two different types of data in an analysis of the morpheme *eco*". In Granger, Sylviane & Petch-Tyson, Stephanie (eds.), *Extending the Scope of Corpus-based Research. New Applications, New Challenges*. Amsterdam: Rodopi. 135–148.
- Kim, Sung-Ae (2001): "Characteristics of EFL readers' summary writing: A study with Korean university students". *Foreign Language Annals* 34 (6), 569–581.
- King, Jeanette (1999): "Talking bro: Māori English in the University Setting". *Te Reo* 42, 20–38.
- Kirkland, Margaret R. & Saunders, Mary Anne P. (1991): "Maximising student performance in summary writing: Managing cognitive load". *TESOL Quarterly* 25 (1), 105–121.
- Kjellmer, Göran (2003): "Synonymy and corpus work: on *almost* and *nearly*". *ICAME Journal* 27, 19–127.
- Krings, Hans P. (1989): "Schreiben in der Fremdsprache – Prozessanalysen zum 'vierten skill'" In: Antos, Gerd & Krings, Hans P. (eds.), *Textproduktion. Ein interdisziplinärer Forschungsüberblick*. Tübingen: Niemeyer. 377–436.
- Krings, Hans P. (1992): "Schwarze Spuren auf weißem Grund – Fragen, Methoden und Ergebnisse der empirischen Schreibprozessforschung im Überblick". In: Krings, Hans P. & Antos, Gerd (eds.), *Textproduktion. Neue Wege der Forschung*. Trier: WVT. 45–110.
- Lee, Young-Ju (2002): "A comparison of composing processes and written products in timed-essay tests across paper-and-pencil and computer modes". *Assessing Writing* 8, 135–157.
- Leijten, Mariëlle & Van Waes, Luuk (2013): "Keystroke logging in writing research: Using Inputlog to analyze and visualize writing processes". *Written Communication* 30 (3), 358–392.

- Leijten, Mariëlle; Van Waes, Luuk; Schrijver, Iris; Bernolet, Sarah & Vangehuchten, Lieve (2019): "Mapping master's students' use of external sources in source-based writing in L1 and L2". *Studies in Second Language Acquisition* 41 (3), 555–82.
- Leki, Ilona & Carson, Joan (1997): "'Completely different worlds': EAP and the writing experiences of ESL students in university courses". *TESOL Quarterly* 31 (1), 39–69.
- Li, Yongyan (2013): "Three ESL students writing a policy paper assignment: An activity-analytic perspective". *Journal of English for Academic Purposes* 12, 73–86.
- Li, Yongyan & Casanave, Christine Pearson (2012): "Two first-year students' strategies for writing from sources: Patchwriting or plagiarism?" *Journal of Second Language Writing* 21, 165–180.
- Lim, Jason Miin-Hwa (2012): "How do writers establish research niches? A genre-based investigation into management researchers' rhetorical steps and linguistic mechanisms". *Journal of English for Academic Purposes* 11, 229–245.
- Lindgren, Eva (2005): *Writing and Revising. Didactic and Methodological Implications of Keystroke Logging*. Doctoral dissertation. Umeå: Umeå University.
- Lindgren, Eva & Sullivan Kirk P. H. (2003): "Stimulated recall as a trigger for increasing noticing and language awareness in the L2 writing classroom: A case study of two young female writers". *Language Awareness* 12, 172–186.
- Mackey, Alison & Gass, Susan M. (2013): *Second Language Research: Methodology and Design*. London: Routledge.
- Mackey, Alison; Gass, Susan M. & McDonough, Kim (2000): "How do learners perceive interactional feedback?" *Studies in Second Language Acquisition* 22, 471–497.
- Manan, Nor Azma & Noor, Noorizah Mohr (2014): "Analysis of reporting verbs in Master's theses". *Procedia - Social and Behavioral Sciences* 134, 140–145.
- Mansourizadeh, Kobra & Ahmad, Ummul K. (2011): "Citation practices among non-native expert and novice scientific writers". *Journal of English for Academic Purposes* 10, 152–161.
- Marshall, Stephen & Garry, Maryanne (2006): "NESB and ESB students' attitudes and perceptions of plagiarism". *International Journal for Educational Integrity* 2 (1), 26–37.
- McCulloch, Sharon (2012): "Citations in search of a purpose: Source use and authorial voice in L2 student writing". *International Journal for Educational Integrity* 8 (1), 55–69.
- McCulloch, Sharon (2013): "Investigating the reading-to-write-processes and source use of L2 postgraduate students in real-life academic tasks: An exploratory study". *Journal of English for Academic Purposes* 12, 136–147.
- McInnis, Lara E. (2009): *Analyzing English L1 and L2 Paraphrasing Strategies through Concurrent Verbal Report and Stimulated Recall Protocols*. M.A. thesis. Toronto: University of Toronto.
- Mei, Wu Siew (2007): "The use of engagement resources in high- and low-rated undergraduate geography essays". *Journal of English for Academic Purposes* 6, 254–271.
- Miller, Kristyan S. (2005): "Second language writing research and pedagogy: A role for computer logging?" *Computers and Composition* 22, 297–317.
- Miller, Kristyan S., Lindgren, Eva & Sullivan, Kirk P. H. (2008): "The psycholinguistic dimension in second language writing: Opportunities for research and pedagogy using computer keystroke logging". *TESOL Quarterly* 42 (3), 433–454.
- Neff, JoAnne; Dafouz, Emma; Herrera, Honesto; Martínez, Francisco; Rica, Juan Pedro; Diez, Mercedes; Prieto, Rosa & Sancho, Carmen (2003): "Contrasting learner corpora: the use of modal and reporting verbs in the expression of writer stance". In Granger, Sylviane & Petch-Tyson, Stephanie (eds.), *Extending the Scope of Corpus-Based Research. New Applications, New Challenges*. Amsterdam/New York: Rodopi. 211–230.
- OhSoft (2017): *oCam* [computer software]. Available from <http://ohsoft.net/eng/>.
- Okamura, Akiko (2008a): "Citation forms in scientific texts: Similarities and differences in L1 and L2 professional writing". *Nordic Journal of English Studies* 7 (3), 61–81.
- Okamura, Akiko (2008b): "Use of citation forms in academic texts by writers in the L1 and L2 context". *The Economic Journal of Takasaki City University of Economics* 51 (1), 29–44.

- Oxford Dictionaries (2016): "Punctuation in direct speech". *Oxford University Press*. <http://www.oxforddictionaries.com/words/punctuation-in-direct-speech>; last accessed 4 August 2016.
- Pecorari, Diane (2003): "Good and original: Plagiarism and patchwriting in academic second-language writing". *Journal of Second Language Writing* 12 (4), 317–345.
- Pecorari, Diane (2006): "Visible and occluded citation features in postgraduate second-language writing". *English for Specific Purposes* 25, 4–29.
- Pecorari, Diane (2008): *Academic Writing and Plagiarism: A Linguistic Analysis*. London: Continuum.
- Pecorari, Diane (2015): "Plagiarism in second language writing: Is it time to close the case?" *Journal of Second Language Writing* 30, 94–99.
- Pecorari, Diane (2016): "Writing from sources, plagiarism and textual borrowing". In Manchón, Rosa & Matsuda, Paul Kei (eds.), *Handbook of Second and Foreign Language Writing*. Berlin: de Gruyter. 329–348.
- Pecorari, Diane & Petrić, Bojana (2014): "Plagiarism in second-language writing". *Language Teaching* 47 (3), 269–302.
- Pecorari, Diane & Shaw, Philip (2012): "Types of student intertextuality and faculty attitudes". *Journal of Second Language Writing* 21, 149–164.
- Pennycook, Alastair (1996): "Borrowing others' words: Text, ownership, memory, and plagiarism". *TESOL Quarterly* 30 (2), 201–230.
- Petrić, Bojana (2007): "Rhetorical functions of citations in high- and low-rated master's theses". *Journal of English for Academic Purposes* 6, 238–253.
- Petrić, Bojana (2012): "Legitimate textual borrowing: Direct quotation in L2 student writing". *Journal of Second Language Writing* 21 (2), 102–117.
- Petrić, Bojana (2015): "What next for research on plagiarism? Continuing the dialogue". *Journal of Second Language Writing* 30, 107–108.
- Petrić, Bojana, & Harwood, Nigel (2013): "Task requirements, task representation, and self-reported citation functions: An exploratory study of a successful L2 student's writing". *Journal of English for Academic Purposes* 12 (2), 110–124.
- Pickard, Valerie (1995): "Citing previous writers: What can we say instead of 'say'?" *Hong Kong Papers in Linguistics and Language Teaching* 18, 89–102.
- Plakans, Lia (2008): "Comparing composing processes in writing-only and reading-to-write test tasks". *Assessing Writing* 13, 111–129.
- Plakans, Lia (2009a): "The role of reading strategies in integrated L2 writing tasks". *Journal of English for Academic Purposes* 8, 252–266.
- Plakans, Lia (2009b): "Discourse synthesis in integrated second language writing assessment". *Language Testing* 26 (4) 561–587.
- Plakans, Lia (2010): "Independent vs. integrated writing tasks: A comparison of task representation". *TESOL Quarterly* 44 (1), 185–194.
- Plakans, Lia & Gebril, Atta (2012): "A close investigation into source use in integrated second language writing tasks". *Assessing Writing* 17, 18–34.
- Plakans, Lia & Gebril, Atta (2013): "Using multiple texts in an integrated writing assessment: Source text use as a predictor of score". *Journal of Second Language Writing* 22, 217–230.
- Plakans, Lia; Liao, Jui-Teng & Wang, Fang (2019): "'I should summarise this whole paragraph': Shared processes of reading and writing in iterative integrated assessment tasks". *Assessing Writing* 40, 14–26.
- Pohl, Torsten (2007): *Studien zur Ontogenese wissenschaftlichen Schreibens*. Tübingen: Niemeyer.
- Polio, Charlene (2012): "How to research second language writing". In Mackey, Alison & Gass, Susan M. (eds.), *Research Methods in Second Language Acquisition: A Practical Guide*. Chichester, UK: Blackwell-Wiley. 139–157.
- Polio, Charlene & Shi, Ling (2012): "Perceptions and beliefs about textual appropriation and source use in second language writing". *Journal of Second Language Writing* 21 (2), 95–101.

- Purdue Online Writing Lab (2016): "Research and Citation Resources". *Purdue University*. <https://owl.english.purdue.edu/owl/section/2/>; last accessed 4 August 2016.
- Ranalli, Jim; Feng, Hui-Hsien & Chukharev-Hudilainen, Evgeny (2018): "Exploring the potential of process-tracing technologies to support assessment for learning of L2 writing". *Assessing Writing* 36, 77–89.
- Ridley, Diana (2012): *The Literature Review. A Step-by-Step Guide for Students*. London: SAGE.
- Roig, Miguel (1997): "Can graduate students determine whether text has been plagiarized?" *The Psychological Record* 47, 113–122.
- Roig, Miguel (2001): "Plagiarism and paraphrasing criteria of college and university professors". *Ethics and Behavior* 11 (3), 307–323.
- Rose, Mike (1984): *Writer's Block*. Carbondale: Southern Illinois University Press.
- Rowley-Jolivet, Elizabeth & Carter-Thomas, Shirley (2014): "Citation practices of expert French writers of English: Issues of attribution and stance". In Lyda, Andrzej & Warchał, Krystyna (eds.), *Occupying Niches: Interculturality, Cross-culturality and A-culturality in Academic Research*. Cham: Springer. 17–34.
- Ruiz-Funes, Marcela (1999): "The process of reading-to-write used by a skilled Spanish-as-a-foreign-language student: A case study". *Foreign Language Annals* 32, 45–62.
- Salager-Meyer, Françoise (1999): "Contentiousness in written medical English discourse: a diachronic study (1810–1995)". *Text* 19, 371–398.
- Samraj, Betty (2013): "Form and function of citations in discussion sections of master's theses and research articles". *Journal of English for Academic Purposes* 12 (2), 299–310.
- Sasaki, Miyuki (2000): "Toward an empirical model of EFL writing processes: An exploratory study". *Journal of Second Language Writing* 9, 259–291.
- Sasaki, Miyuki (2004): "A multiple-data analysis of the 3.5-year development of EFL student writers". *Language Learning* 54, 525–582.
- Segev-Miller, Rachel (2004): "Writing from sources: The effect of explicit instruction on college students' processes and products". *L1 – Educational Studies in Language and Literature* 4 (1), 5–33.
- Shaw, Philip & Pecorari, Diane (2013): "Source use in academic writing: An introduction to the special issue". *Journal of English for Academic Purposes* 12 (2), A1–A3.
- Shaw, Stuart & Weir, Cyril J. (2007): *Examining Writing: Research and Practice in Assessing Second Language Writing*. Cambridge: Cambridge University Press.
- Sherrard, Carol (1986): "Summary writing: A topographical study". *Written Communication* 3, 324–343.
- Shi, Ling (2004): "Textual borrowing in second-language writing". *Written Communication* 21 (2), 171–200.
- Shi, Ling (2008): "Textual appropriation and citing behaviours of university undergraduates". *Applied Linguistics* 31 (1), 1–24.
- Shi, Ling (2011): "Common knowledge, learning, and citation practices in university writing". *Research in the Teaching of English* 45 (3), 308–334.
- Shi, Ling (2012): "Rewriting and paraphrasing source texts in second language writing". *Journal of Second Language Writing* 21, 134–148.
- Shi, Ling & Dong, Yanning (2018): "Chinese graduate students paraphrasing in English and Chinese contexts". *Journal of English for Academic Purposes* 34, 46–56
- Shi, Ling; Fazel, Ismaeil & Kowkabi, Nasrin (2018): "Paraphrasing to transform knowledge in advanced graduate student writing". *English for Specific Purposes* 51, 31–44.
- Sinclair, John (1988): "Mirror for a text". *Journal of English and Foreign Languages* 1, 15–44.
- Solé, Isabel; Miras, Mariana; Castells, Núria; Espino, Sandra & Minguela, Marta (2013): "Integrating information: An analysis of the processes involved and the products generated in a written synthesis task". *Written Communication* 30 (1), 63–90.
- Spivey, Nancy Nelson (1990): "Transforming texts: Constructive processes in reading and writing". *Written Communication* 7 (2), 256–287.

- Stapleton, Paul (2010): "Writing in an electronic age: A case study of L2 composing processes". *English for Academic Purposes* 9, 295–307.
- Sullivan, Kirk P. H., & Lindgren, Eva (2002): "Self-assessment in autonomous computer-aided L2 writing". *ELT Journal* 56, 258–266.
- Sun, Yu-Chih (2013): "Do journal authors plagiarize? Using plagiarism detection software to uncover matching text across disciplines". *Journal of English for Academic Purposes* 12, 264–272.
- Sutherland-Smith, Wendy (2005): "Pandora's box: academic perceptions of student plagiarism in writing". *Journal of English for Academic Purposes* 4, 83–95.
- Swales, John M. (1986): "Citation Analysis and Discourse Analysis". *Applied Linguistics* 7 (1), 39–56.
- Swales, John (1990): *Genre Analysis. English in Academic and Research Settings*. Cambridge: Cambridge University Press.
- Swales, John (2004): *Research Genres: Exploration and Applications*. Cambridge: Cambridge University Press.
- Swales, John (2014): "Variation in citation practice in corpus of student biology papers. From parenthetical plonking to intertextual storytelling". *Written Communication* 31 (1), 118–141.
- Syndicate, U. C. L. E. (2001): *Quick Placement Test*. Oxford: Oxford University Press.
- Tadros, Angele (1993): "The pragmatics of text averral and attribution in academic texts". In Hoey, Michael (ed.), *Data, Description, Discourse*. London: Harper Collins. 98–114.
- Thomas, Sarah & Hawes, Thomas P. (1994): "Reporting verbs in medical journal articles". *English for Specific Purposes* 13 (2), 129–148.
- Thompson, Celia H. (2009): "Plagiarism, intertextuality and emergent authorship in university students' academic writing". *PORTAL Journal of Multidisciplinary International Studies* 6 (1), 1–16.
- Thompson, Celia; Morton, Janne & Storch, Neomy (2013): "Where from, wo, why and how? A study of the use of sources by first year L2 university students". *Journal of English for Academic Purposes* 12 (2), 99–109.
- Thompson, Geoff (1990): "Voices in the text: Discourse perspectives on language reports". *Applied Linguistics* 17 (4), 501–530.
- Thompson, Geoff & Ye, Yiyun (1991): "Evaluation in the reporting verbs used in academic papers". *Applied Linguistics* 12, 365–382.
- Thompson, Paul (2000): "Citation practices in PhD theses". In Burnard, Lou & McEnery, Tony (eds.), *Rethinking Language Pedagogy from a Corpus Perspective*. Frankfurt/Main: Peter Lang. 91–102.
- Thompson, Paul (2001): *A Pedagogically-Motivated Corpus-Based Examination of PhD Theses: Macrostructure, Citation Practices and Uses of Modal Verbs*. Doctoral dissertation. Reading: University of Reading.
- Thompson, Paul & Tribble, Chris (2001): "Looking at citations: Using corpora in English for Academic Purposes". *Language Learning & Technology* 5 (3), 91–105.
- Tufescu, Florina (2008): *Oscar Wilde's Plagiarism: The Triumph of Art over Ego*. Dublin: Irish Academic Press.
- Tyler, Andrea (1995): "The coconstruction of cross-cultural miscommunication. Conflicts in perception, negotiation, and enactment of participant role and status". *Studies in Second Language Acquisition* 17, 129–152.
- Uccelli, Paola; Dobbs, Christina L. & Scott, Jessica (2013): "Mastering academic language. Organization and stance in the persuasive writing of high school students". *Written Communication* 30 (1), 36–62.
- Uludag, Pakize; Lindberg, Rachael; McDonough, Kim & Payant, Caroline (2019): "Exploring L2 writers' source text use in an integrated writing assessment". *Journal of Second Language Writing* 46, 1–7.
- Verheijen, Lieke (2015): "The language of quoting in academic writing". *Dutch Journal of Applied Linguistics* 4 (1), 101–121.
- Vieyra, Michelle; Strickland, Denise & Timmerman, Briana (2013): "Patterns in plagiarism and patchwriting in science and engineering graduate students' research proposals". *International Journal for Educational Integrity* 9 (1), 35–49.
- Wang, Wenyu & Wen, Qiufang (2002): "L1 use in the L2 composing process: An exploratory study of 16 Chinese EFL writers". *Journal of Second Language Writing* 11, 225–246.

- Weigle, Sarah C. (2004): "Integrating reading and writing in a competency test for non-native speakers of English". *Assessing Writing* 9 (9), 27–55.
- Weigle, Sarah C. & Parker, Keisha (2012): "Source text borrowing in an integrated reading/writing assessment". *Journal of Second Language Writing* 21 (2), 118–133.
- Wette, Rosemary (2010): "Evaluating student learning in a university-level EAP unit on writing using sources". *Journal of Second Language Writing* 19, 158–177.
- Wette, Rosemary (2017): "Source text use by undergraduate post-novice L2 writers in disciplinary assignments: Progress and ongoing challenges". *Journal of Second Language Writing* 37, 46–58.
- Wiemeyer, Leonie (2015a): "Intertextuality in second language writing: Source use and paraphrasing in academic learner English". Paper presented at the Learner Corpus Research Conference 2015 (LCR), Radboud University, Cuijk/Nijmegen, the Netherlands, 11-13 September 2015.
- Wiemeyer, Leonie (2015b): "Intertextuality in second language writing: Source use and paraphrasing in academic learner English". Paper presented at the Symposium on Second Language Writing 2015 (SSLW), Auckland University of Technology, New Zealand, 19-21 November 2015.
- Wiemeyer, Leonie (2016a): "Direct quotation in second language writing: A corpus-based study of intertextuality in academic learner English". Paper presented at the Teaching and Learner Corpora (TaLC) Conference 2016, University of Gießen, Germany, 6-7 April 2017.
- Wiemeyer, Leonie (2016b): "Direct quotation in L2 academic writing: A corpus-based study". Paper presented at the Symposium on Second Language Writing 2016, Arizona State University, Tempe, AZ, USA, 20-22 October 2016.
- Wiemeyer, Leonie (2017a): "Exploring academic writing competence: Intertextual strategies of L2 writers in linguistics assignments". Paper presented at the Writing Symposium 2017, University of Gießen, Germany, 6-7 April 2017.
- Wiemeyer, Leonie (2017b): "Intertextuality in L2 academic writing: The use of paraphrases by German learners of English". In de Haan, Pieter; van Vuuren, Sanne & de Vries, Rina (eds.), *Language, Learners and Levels: Progression and Variation. Corpora and Language in Use – Proceedings 3*. Louvain-la-Neuve: Presses universitaires de Louvain, 45–68.
- Wiemeyer, Leonie (2017c): "Direct quotes, paraphrases, and summaries in L2 academic assignments". Paper presented at the Learner Corpus Research Conference 2017 (LCR), Eurac Research Centre Bolzano, Italy, 5-7 October 2017.
- Wiemeyer, Leonie (2018): "L2 writers' strategies of paraphrasing, quoting, and textual borrowing in linguistics assignments". Paper presented at ICAME 39, University of Tampere, Finland, 30 May-3 June 2018.
- Wiemeyer, Leonie (2019): "Direct quotation in second language writing. A corpus-based study of intertextuality in academic learner English". In Götz, Sandra & Mukherjee, Joybrato (eds.), *Learner Corpora and Language Teaching*. Amsterdam: John Benjamins. 129–153.
- Yamada, Kyoko (2003): "What prevents ESL/EFL writers from avoiding plagiarism?: Analyses of 10 North-American college websites". *System* 31, 247–258.
- Yang, Hui-Chun & Plakans, Lia (2012): "Second language writers' strategy use and performance on an integrated reading-listening-writing task". *TESOL Quarterly* 46 (1), 80–103.
- Zhao, Yan (2011): *L2 Creative Writers: Identities and Writing Processes*. Doctoral dissertation. Warwick: University of Warwick.
- Zimmermann, Rüdiger (2000): "L2 writing: subprocesses, a model of formulating and empirical findings". *Learning and Instruction* 10, 73–99.

# List of figures

Figure 1: Mixed-methods design.....	4
Figure 2: Distinction of intertextual strategies based on direct and indirect source use.....	14
Figure 3: Strategies of signalling intertextuality in academic writing.....	24
Figure 4: Hayes and Flower’s model of the writing process (author’s own visualisation).....	53
Figure 5: Triangulation of research methods.....	67
Figure 6: Study design process.....	70
Figure 7: Inter-learner variability in the use of intertextual links.....	83
Figure 8: Use of material from different sections of the source text.....	84
Figure 9: Inter-learner variability in selection strategies for paraphrases.....	90
Figure 10: Approaches to paraphrasing. ....	90
Figure 11: Target of author attribution.....	106
Figure 12: Types of reporting structures.....	108
Figure 13: Proportion of copied strings of 3+ words per 100 words.....	111
Figure 14: Word cloud representing copied strings.....	111
Figure 15: Example of a flashcard with aspects of interest for stimulated recall.....	132
Figure 16: Hierarchy of activity codes for categorisation of process segments.....	134
Figure 17: Use of the comment function of ATLAS.ti to describe the writing process chronologically.....	137
Figure 18: Overview of coding sequence.....	138
Figure 19: Activities and acts related to paraphrasing.....	138
Figure 20: Selecting codes for acts occurring in a segment from a list of codes in ATLAS.ti.....	138
Figure 21: View of segments coded for acts in ATLAS.ti.....	138
Figure 22: Coding procedure.....	139
Figure 23: Length of each student’s writing process in minutes.....	146
Figure 24: View of documents being used side-by-side in Eva’s writing process.....	148
Figure 25: View of only one document being visible at a time in Nadine’s writing process.....	148
Figure 26: Writing acts – total number of instances in the experimental data.....	153
Figure 27: Non-writing acts – total number of instances in the experimental data.....	153
Figure 28: No. of instances of core acts in each L2 writer's process.....	154
Figure 29: Overview of paraphrasing-related writing acts per student. ....	155
Figure 30: Acts related to reporting structures and documentation of sources in each student’s process.....	169
Figure 31: Acts comprising copy-pasting across study participants. ....	173
Figure 32: Extension of Hayes and Flower’s (1980) model to the source-based writing process.....	191

## List of tables

Table 1: Overview of major intertextual strategies and their features.....	22
Table 2: Corpus overview.....	73
Table 4: Additional coding categories applied for each type of intertextual link.....	81
Table 5: Overview of quantitative findings.....	82
Table 6: Example of non-chronological structuring of information in the reading report.....	85
Table 7: Most frequent reporting verbs.....	108
Table 8: Instruments used and data collected in process study.....	124
Table 9: Overview of pilot study groups.....	125
Table 10: Overview of piloted source texts.....	126
Table 11: Codes for activities used to categorise process segments.....	135
Table 12: Overview of acts and corresponding codes.....	141
Table 13: Segments and activities per writing process.....	145
Table 14: Length of recording and of pre-writing, writing, and post-writing stages per student [hh:mm:ss].....	146
Table 15: Intertextual writing activities.....	152

# Appendices

Appendix 1: Reporting verbs and their frequency in the corpus.....	220
Appendix 2: Reoccurring copied strings of 3+ words in the corpus by frequency.....	221
Appendix 3: E-mail with general information for potential study participants.....	222
Appendix 4: E-mails with general instructions and short introductory text for writing a reading report.....	223
Appendix 5: Learner questionnaire.....	224
Appendix 6: Pilot study participants.....	225
Appendix 7: Metadata of study participants.....	226
Appendix 8: General instructions for experiment.....	229
Appendix 9: Consent form for experiment.....	230
Appendix 10: Experiment log to be completed by participants during the writing session.....	231
Appendix 11: Instructions for writing a reading report with language resources.....	232
Appendix 12: Instructions for writing a reading report without additional resources.....	233
Appendix 13: Consent form for interview.....	234
Appendix 14: General introduction and instructions for retrospective interviews with stimulated recall.....	235
Appendix 15: Introductory questions for retrospective interviews with stimulated recall.....	235
Appendix 16: Instructions for stimulated recall given orally during the interview.....	235
Appendix 17: Questions for retrospective interviews.....	235
Appendix 18: Aspects of interest for interview by participant (pilot study).....	236
Appendix 19: Aspects of interest for interview by participant.....	239
Appendix 20: Total no. of codings for writing acts (W) and non-writing acts (NW) in the writers' processes....	241
Appendix 21: Overview of instances of acts in each L2 writers' process.....	243
Appendix 22: Absolute number of instances per act per student's writing process.....	244
Appendix 23: Absolute number of instances per core act in L2 writers' processes.....	247

**Appendix 1: Reporting verbs and their frequency in the corpus.**

Rank	Reporting verb	Freq	Rank	Reporting verb	Freq	Rank	Reporting verb	Freq
1.	explain	12	35.	present	2	69.	express their interest in	1
2.	examine	10	36.	proceed (with NP) by	2	70.	figure out	1
3.	point out	9	37.	provide NP with	2	71.	finish with the thought	1
4.	state	8	38.	start (NP) with	2	72.	found out	1
5.	analyse	7	39.	suggest	2	73.	give	1
6.	compare	7	40.	summarise	2	74.	give an overview of	1
7.	deal with	7	41.	take	2	75.	give X as examples	1
8.	show	6	42.	try to find out	2	76.	go into depth for	1
9.	conclude	5	43.	verify	2	77.	go on by	1
10.	use	5	44.	advance the hypothesis	1	78.	highlight	1
11.	claim	4	45.	advise	1	79.	identify	1
12.	describe	4	46.	aim at demonstrating	1	80.	interpret	1
13.	discuss	4	47.	ask	1	81.	list up	1
14.	argue	3	48.	be concerned with	1	82.	make clarify	1
15.	concentrate on	3	49.	choose NP	1	83.	make out	1
16.	find	3	50.	come to the conclusion	1	84.	move on with	1
17.	list	3	51.	come to the first deduction	1	85.	prove	1
18.	mention	3	52.	conclude NP with	1	86.	record	1
19.	reveal	3	53.	continue to discuss	1	87.	say	1
20.	take a (closer) look at	3	54.	continue with	1	88.	select	1
21.	aim to	2	55.	deal with the fact	1	89.	shows	1
22.	apply X to reveal	2	56.	deduct from their findings	1	90.	stress	1
23.	choose to	2	57.	deem as	1	91.	suspect	1
24.	close (NP) with	2	58.	demonstrate	1	92.	test their theory by	1
25.	conduct	2	59.	develop	1	93.	touch on the open question	1
26.	consider	2	60.	disagree with	1	94.	try to answer the research question	1
27.	contrast	2	61.	divide	1	95.	try to convince the reader	1
28.	find out	2	62.	draw out	1	96.	turn the focus	1
29.	focus on	2	63.	draw the conclusion	1	97.	want to examine	1
30.	go on to	2	64.	emphasise	1	98.	want to show the relevance of	1
31.	introduce	2	65.	end NP with	1			
32.	make clear	2	66.	evaluate	1			
33.	make use of	2	67.	explain their interest by mentioning	1			
34.	outline	2	68.	express	1			

**Appendix 2: Reoccurring copied strings of 3+ words in the corpus by frequency.**

Rank	N-Gram	Freq	Rank	N-Gram	Freq
1.	the morpheme eco	23	36.	the semantics of the morpheme eco	3
2.	in the OED	22	37.	the two data	3
3.	in the BNC	17	38.	with the prefix	3
4.	the British National Corpus	14	39.	500,000 lexical items	2
5.	the Oxford English Dictionary	14	40.	a number of	2
6.	eco in the sense of	7	41.	accepted intersubjective status	2
7.	the BNC and the OED	7	42.	additive and contrastive analysis	2
8.	as a prefix	6	43.	all words starting with	2
9.	of the morpheme eco	6	44.	and the BNC	2
10.	on the other hand	6	45.	appear in both sources	2
11.	the meaning of	6	46.	approximately 500,000 lexical items	2
12.	types of data	6	47.	are found in	2
13.	a contrastive analysis	5	48.	as a productive morpheme	2
14.	in a dictionary	5	49.	based on the criterion of institutionalization	2
15.	in the BNC and the OED	5	50.	between affixes and free morphemes	2
16.	sources of data	5	51.	between institutionalized and less widely accepted	2
17.	the two sources	5	52.	both data sources	2
18.	dictionary and corpus	4	53.	British National Corpus	2
19.	dictionary and corpus data	4	54.	collection of empirical data	2
20.	in a corpus	4	55.	collection of empirical data on British English	2
21.	meanings of eco	4	56.	comparing dictionary and corpus data	2
22.	Oxford English Dictionary	4	57.	comparison is essential	2
23.	than the OED	4	58.	constitutes the largest	2
24.	the English language	4	59.	contain a specific selection of words	2
25.	words that are	4	60.	corpora and dictionaries	2
26.	a contrastive analysis of	3	61.	corpora such as	2
27.	chosen by lexicographers	3	62.	data sources for semantic analyses	2
28.	found in the BNC	3	63.	different types of	2
29.	found in the OED	3	64.	different types of data	2
30.	in the dictionary	3	65.	distinguishing between meanings	2
31.	of the OED	3	66.	environmentally friendly	2
32.	pertaining to the ecological movement	3	67.	feature in the OED	2
33.	six meanings of eco	3	68.	few shared words	2
34.	the British National Corpus (the BNC)	3	69.	having to be inferred	2
35.	the semantics of	3	70.	in both sources	2

Rank	N-Gram	Freq	Rank	N-Gram	Freq
71.	in the corpus	2	83.	some information about	2
72.	in the sense of	2	84.	source of eco-words	2
73.	interaction between organisms	2	85.	the BNC is	2
74.	is concerned with the morpheme eco	2	86.	the different nature of the two sources	2
75.	is highly productive	2	87.	the ecological movement	2
76.	lexical semantic analysis	2	88.	the largest dictionary of the English language	2
77.	likely to be	2	89.	the OED and the BNC	2
78.	meaning of the	2	90.	two different types of	2
79.	of comparing dictionary and corpus data	2	91.	very likely to be institutionalised	2
80.	once a word	2	92.	which lexemes actually occur in authentic language	2
81.	related to the	2	93.	with the meaning	2
82.	six possible meanings of eco	2			

### Appendix 3: E-mail with general information for potential study participants.

Dear students,

I am looking for [no.] participants for my experiment on [date] and [date]. I am a PhD student with Prof. Callies and I am conducting the experiment for my dissertation. I hope I can find people who are willing to help me with my study! Please note: You have to be a native speaker of German to participate.

#### What is the experiment about and how does it work?

The experiment consists of two parts:

- 1) In the first part, you will work on a writing task in an experimental setting. This part will take place on [date].
- 2) In the second part, I will interview you about how you wrote the text. The interviews will be conducted on [date].

In order to participate, you have to register for the experiment, which will take place on [date] between [time] and [time], and an interview slot on [date]. You have to register for both to be allowed to participate. I will only award you the credits if you have participated in both sessions. You will receive 3 experiment credits (3 x 0.5 hrs), which is the total number of credits you need for the module. The task will not be graded.

Once you are registered, I will send you a source text that you have to read thoroughly before the experiment. Without reading the text, you cannot participate in the experiment, so it is important that you do it if you want to get credits. We will then meet in room [room no.] on [date], at [time] for the experiment, and again in your selected time slot on [date] for the interview.

Completing the writing task will take around 2 hours. During the test run, some students finished faster, but since people write at different speeds, it could also take you a little longer. The interview will take a maximum of 1 hour.

#### What's in it for you?

You can get all 3 credits in one go. After the interview, you will receive feedback on your text that you can apply to other writing tasks in your degree. So you'll gain writing experience and will learn ways to improve your writing.

If this sounds like something you're willing to do, **please sign up for the experiment AND an interview slot** on my Stud.IP profile: [URL] (I hope the link works; if it doesn't, please search for my profile on Stud.IP.)

In addition, **please fill in the form attached to this e-mail and send it back to me via e-mail by 12 p.m. (noon) on [date]**. Only those people who sign up in Stud.IP AND send me the form by [date] can participate in the experiment.

There is a total of 5 spaces. **Please consider your registration as binding**. This is for my dissertation, so it's very important that I can rely on you. If you cannot make it, please let me know as soon as possible so I can find a replacement. Thank you for your cooperation! If you do not get a space this time, please be patient – there will be another round of experiments soon.

I look forward to working with you! If you have any questions, please don't hesitate to ask.

## Appendix 4: E-mails with general instructions and short introductory text for writing a reading report sent to participants.

### a) Source text King 1999

Dear participants,

I am very thankful and excited that you are supporting me by participating in my experiment!

Please read the attached article thoroughly and make sure you know what the study was about and what its findings were. You'll need this knowledge for a writing assignment in the experiment. You won't need to bring the article to our session, but it is very important that you will have read and understood it before we get started. If you cannot manage, please let me know as soon as possible so I can find a replacement.

Here's a short introduction to the topic:

Maori English is one of the varieties of English used in New Zealand. It has certain linguistic features which are different from Standard New Zealand English. Maori English is spoken predominantly by Maori, i.e. the indigenous people of New Zealand. It is sometimes also used by Pakeha, which is the Maori word for New Zealanders of European descent. Around 15 % of New Zealanders are Maori, some of whom are bilingual and speak Maori as well as English. The speakers of Maori English do not necessarily use it in all contexts, as you'll see in the article.

We will meet on [day] at [time] in [room]. If you have any questions, please do not hesitate to ask.

See you on [day]!

Best,  
Leonie

### b) Source text Kjellmer 2003

Dear participants,

I am very thankful and excited that you are supporting me by participating in my experiment!

Please read the attached article thoroughly and make sure you know what the study was about and what its findings were. You'll need this knowledge for a writing assignment in the experiment. You won't need to bring the article to our session, but it is very important that you will have read and understood it before we get started. If you cannot manage, please let me know as soon as possible so I can find a replacement.

Here's a short introduction to the topic:

An important area of research in semantics is synonymy. Synonyms are two words which have exactly the same meaning. However, some words that appear to have an identical meaning may differ in certain aspects of use. For example, they may be used in different contexts or have different words with which they usually co-occur, i.e. different collocates. Corpora are large electronic databases of authentic texts used in linguistic research. They can help to identify the differences between apparent synonyms, as you'll see in the article.

### c) Source text Kettemann et al. 2003

Dear participants,

I am very thankful and excited that you are supporting me by participating in my experiment!

Please read the attached article thoroughly and make sure you know what the study was about and what its findings were. You'll need this knowledge for a writing assignment in the experiment. You won't need to bring the article to our session, but it is very important that you will have read and understood it before we get started. If you cannot manage, please let me know as soon as possible so I can find a replacement.

Here's a short introduction to the topic:

An important area of research in linguistics is word-formation, i.e. how new words are coined from morphemes of the language. Of special interest are creative word-formation patterns, how productive they are, and whether their formations have become institutionalised. Productivity refers to a morpheme's ability to create new words. Elements that have a low productivity only give rise to a small number of new words, or neologisms, in a certain period of time, whereas highly productive elements can produce hundreds of such neologisms. Some of these words become institutionalised, which means that the majority of speakers of that language knows their meaning and they are incorporated into dictionaries. Linguists use different data to investigate productivity and institutionalisation in word-formation, for example dictionaries and corpora, which are large electronic collections of authentic texts. Both have their advantages and disadvantages, as you'll see in the article.

## Appendix 5: Learner questionnaire.

*Experiment zum wissenschaftlichen Schreiben in der Fremdsprache*

**Fragebogen zum (fremd)sprachlichen Hintergrund**

Kodierung (bitte hier nichts eintragen)

Name: \_\_\_\_\_

Geburtsjahr: \_\_\_\_\_ Geschlecht:  männlich  weiblich  \_\_\_\_\_

Geburtsland: \_\_\_\_\_ Muttersprache: \_\_\_\_\_

Sind Sie bilingual aufgewachsen, d.h. sprechen Sie von Kindheit an mehr als eine Sprache?  ja  nein

In welchem Studiengang studieren Sie Englisch (bitte nur einen ankreuzen)?

B.Ed.  M.Ed.  BA  MA  Andere: \_\_\_\_\_

Fächerkombination bzw. Name des BA/MA-Studiengangs: \_\_\_\_\_ Fachsemester: \_\_\_\_\_

In welcher Sprache findet der universitäre Unterricht im Studiengang statt?  Englisch  Deutsch

1) Welche Fremdsprache(n) haben Sie erlernt bzw. erlernen Sie?

Sprache	Anzahl der Jahre	davon in der Schule	an der Universität	anderswo (z.B. VHS)
a) Englisch				
b)				
c)				

2) Hatten Sie Englisch als Fremdsprache in der Schule durchgehend bis zum Abitur?  ja  nein  
 Falls ja, als Grundkurs oder Leistungskurs?  GK  LK

3) Haben Sie ein Auslandssemester/-jahr oder einen längeren Aufenthalt in einem englischsprachigen Land verbracht?  ja  nein  
 Falls ja, wo, zu welchem Zweck und wie lange?

Land	Zweck	Dauer (in Monaten)
Großbritannien		
Irland		
USA		
Kanada		
anderes englischsprachiges Land:		

4) Wie bewerten Sie Ihre Englischkenntnisse?  
 muttersprachenähnlich  sehr gut  gut  durchschnittlich  eher schlecht  schlecht

5) Wie bewerten Sie Ihre Kenntnisse in den anderen Fremdsprachen? (Bitte ankreuzen)

Sprache	muttersprachenähnlich	sehr gut	gut	durchschnittlich	eher schlecht	schlecht
a)	<input type="checkbox"/>					
b)	<input type="checkbox"/>					
c)	<input type="checkbox"/>					

6) Welche Ihrer erlernten Fremdsprachen gefällt Ihnen am besten? (Nur eine nennen)

7) Welche Ihrer Fremdsprachen sprechen Sie auch privat und nicht nur im Unterricht? (Bitte ankreuzen)

Sprache	fast täglich	einmal in der Woche	einmal im Monat	noch seltener
a) Englisch	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

8) Welche Textsorten haben Sie auf Englisch in den Bereichen Sprach-, Literatur- und Kulturwissenschaft schon geschrieben? (Bitte möglichst alle auflisten!)

Hausarbeit  Essay  Zusammenfassung / Summary  Reading report  Abstract  
 Rezension / Review  Research proposal  Andere: \_\_\_\_\_

**Appendix 6: Pilot study participants.**

<b>Student</b>	<b>Source text</b>	<b>Resources allowed</b>	<b>Writing time in minutes</b>	<b>OOPT score</b>	<b>OOPT level</b>
Bea	King 1999	Yes	85	45	Upper Intermediate
Bianca	King 1999	Yes	102	54	Advanced
Katharina	King 1999	Yes	96	56	Very Advanced
Charlotte	King 1999	No	85	41	Upper Intermediate
Paula	King 1999	No	98	55	Very Advanced
Svenja	King 1999	No	59	53	Advanced
Sarah	Kjellmer 2003	Yes	79	54	Advanced
Vera	Kjellmer 2003	Yes	64	44	Upper Intermediate
Anna	Kjellmer 2003	No	55	43	Upper Intermediate
Cynthia	Kjellmer 2003	No	68	51	Advanced
Lena	Kettemann et al. 2003	Yes	92	45	Upper Intermediate
Miriam	Kettemann et al. 2003	Yes	86	55	Very Advanced

## Appendix 7: Metadata of study participants.

### *Basic information and OQPT results.*

Participant pseudonym	Year of birth	Gender	Country of birth	L1	Study programme	Subjects	Semester (English)	OQPT score	OQPT level
Arne	1997	male	Germany	German	B.A.	English, Spanish	2	54	Advanced
Ben	1997	male	Germany	German	B.A. (teacher training)	English, Politics	2	51	Advanced
Eva	1998	female	Germany	German	B.A.	English, Media and Culture Studies	2	53	Advanced
Ina	1996	female	Germany	German	B.A.	English, Culture Studies	2	53	Advanced
Nadine	1994	female	Germany	German	B.A.	English, Linguistics	2	52	Advanced
Tara	1997	female	Germany	German & Arabic	B.A.	English, Biology	2	53	Advanced

### *Experience abroad and favourite L2.*

Participant pseudonym	Spent period abroad?	Country of stay 1	Purpose of stay 1	Duration of stay 1	Country of stay 2	Purpose of stay 2	Duration of stay 2	Favourite L2
Arne	no							Spanish
Ben	no							English
Eva	no							English
Ina	yes	Ireland	au pair	6	New Zealand	school exchange	3	English
Nadine	yes	Great Britain	au pair	12				English
Tara	no							<i>not specified</i>

### Foreign languages (L2s) (1)

Participant pseudonym	L2.1	L2.1 Years	L2.1 Years at school	L2.1 Years at uni	L2.1 Years elsewhere	L2.1 Self-assessment	L2.1 Use outside class	L2.2	L2.2 Years	L2.2 Years at school	L2.2 Years at university	L2.2 Years elsewhere	L2.2 Self-assessment	L2.2 Use outside class
Arne	English	11	10	1		not specified	once a month	Spanish	5	3	1	1	very good	about once a week
Ben	English	11	10	1		good	about once a week	French	7	7			fair	not specified
Eva	English	12	9	1	2	good	almost every day	French	11	11			rather poor	not specified
Ina	English	13	11	1	1	native-like	almost every day	French	4	4			rather poor	not specified
Nadine	English	14	11	1	2	native-like	almost every day	Italian	3	3			rather poor	not specified
Tara	English	11	10	1		good	almost every day	French	6	6			fair	once a month

### Foreign languages (L2s) (2)

Participant pseudonym	L2.3	L2.3 Years	L2.3 Years at school	L2.3 Years at uni	L2.3 Years elsewhere	L2.3 Self-assessment	L2.3 Use outside class	L2.4	L2.4 Years	L2.4 Years at school	L2.4 Years at university	L2.4 Years elsewhere	L2.4 Self-assessment	L2.4 Use outside class
Arne	Italian	1			1	rather poor	even rarer	Latin	4	4			<i>not specified</i>	<i>not specified</i>
Ben	Spanish	2	2			fair	<i>not specified</i>	---						
Eva	---							---						
Ina	Spanish	3	3			fair	<i>not specified</i>	---						
Nadine	French	1	1			rather poor	<i>not specified</i>	---						
Tara	Arabic	12			1	11	very good	almost every day	---					

***Participants' writing experience in English (no. of texts written).***

<b>Participant pseudonym</b>	<b>Essay</b>	<b>Summary</b>	<b>Review</b>	<b>Reading report</b>	<b>Proposal</b>	<b>Abstract</b>	<b>Research paper</b>	<b>Other</b>	<b>TOTAL</b>
<b>Arne</b>	1	1	1	1					4
<b>Ben</b>	1	1		1			1		4
<b>Eva</b>	1	1		1					3
<b>Ina</b>	1	1							2
<b>Nadine</b>	1	1		1					3
<b>Tara</b>	1	1							2
<b>SUM</b>	12	14	3	6	0	1	1	0	

**Appendix 8: General instructions for experiment.**

Good morning / afternoon everyone,

Thank you very much for participating in my experiment. Today, you are helping me to do the study for my dissertation. That means that you are very important for the success of my research, and I am very happy and grateful that you are here today.

As you know, you will be writing a text today, so please adjust your seats and get comfortable. Please put away the source text and any notes you took. You will be able to use an electronic copy of the source text on the screen. Please also put away all your personal items, including your phone. You do not need anything at your place except for a pen and the things I give you. You can take breaks to eat, go to the bathroom or have a cigarette at any time during the experiment. Please just raise your hand to let me know that you want to take a break and wait for me to come over. While you are writing, please save your document regularly so nothing is lost.

During the experiment, it is important that you do not talk to anyone about your task, neither with the people in the room nor with anyone else via social media or your phone. I'd appreciate your cooperation in that matter. Thank you.

When you have finished the task, please let me know. There are a few more things I'd like to talk to you about before you leave.

I have also given each of you an experiment log. [Show log] Please fill in the log during or after the experiment. Please write down anything that comes to your mind in response to the questions. Please take a moment now to look at the experiment log so you know which questions are on there. [Give time to read]

At your table, you will find a consent form with all the information, and the instructions for your task. I will come over to you to answer your questions and get you started on the experiment, so please let me know when you are done reading. Please do not start writing before I have talked to you.

**Appendix 9: Consent form for experiment.****Experiment – Writing a reading report**

Thank you for participating in this experiment.

The experiment is part of my dissertation, in which I look at how students use source texts when writing in English.

Your task in this session is to write a reading report. A reading report is a kind of summary of the text you prepared at home. The text will be available as a PDF during writing. You will find the instructions on the next page. Before you start, please read the instructions carefully and then complete the task accordingly. There will be time to ask questions.

While you are writing, a programme running in the background will record everything that happens on the screen. There is no video camera or microphone – the only thing that will be recorded is what happens on the screen. No-one else can see your screen while you are writing. Please raise your hand if you would like to take a break or ask a question at any time during the experiment.

This study is confidential. Your participation is voluntary and anonymous. Your name, the recording, and your text will be anonymised and will not be identifiable. The data will be used for research purposes only. It will not be possible to trace them back to your person. If you are interested, a copy of the screen recording will be made available to you.

If you have any questions about the project, please feel free to contact me.

Thank you very much for your time, effort, and thought in participating in this experiment.

Leonie Wiemeyer, PhD candidate

By signing this document I confirm that I have read and understood the instructions and agree to participate in this study.

I consent to having my text and the screen recording used for research purposes.

Participant's name

Date and signature

**Appendix 10: Experiment log to be completed by participants during the writing session.**

Experiment Log

What was your experience of reading the source text? What did you (or didn't you) like about it? Anything that was difficult to understand? Anything that was helpful? Anything that was structured in an unfortunate way?

Was there anything that could have been clearer in the instructions for writing a reading report or the explanations given immediately prior to the experiment? Anything you were unsure about?

What was your experience of writing the reading report? Anything you liked / found easy? Anything that you found difficult or that you weren't sure about?

Please write down anything you notice during the experiment about how you worked with the source text, e.g. ideas that you have, things that you decide (not) to do, things that you don't understand, that you have problems with, technical issues, things you are worried about, etc. If you found a solution or have a suggestions for changes, please also write it down. (continues on the next page)

Time	Issue	Solution / suggestion

...

## Appendix 11: Instructions for writing a reading report with language resources.

### Instructions: Writing a reading report

A reading report is an academic text type that contains the main points, hypotheses, methodology and general line of argumentation of a research paper. A reading report should not exceed **600 words**. Give an **exact** word count, using your word processor. The document has already been formatted for you.

The aims of this task are to learn

- to understand the structure, content and line of argumentation of a scholarly research paper
- to summarize a text coherently, and
- to improve your writing skills generally.



Your reading report should make it clear to an outsider what the paper is about. Ask yourself: Is my report comprehensible for somebody who does not know the paper?

#### How do I write a reading report?

Don't panic if you don't understand a point or a passage in the paper. Read over it again, if you still don't get it, move on – very often such passages become clearer in the light of further explanation. If you don't understand a passage, don't try to summarize it – better leave it out completely. Ideally, your report should take up and answer the following questions:

- What is the topic of the paper?
- Why does the author deal with the topic? In how far is it interesting, or why does it pose problems for the researcher (e.g. open questions or diverging opinions among researchers)
- What is/are the major research question(s) the author wishes to answer?
- Are there any initial hypotheses (if any), and are these verified or falsified in the end?
- What kind(s) of data and methods did the author use (e.g. experiments, corpora)?
- What are the main arguments the author uses to get his/her point across?
- What are the most important findings of the paper, and what is its contribution to research?



**Important:** The report must be written in your own words. Do not take over or 'recycle' whole chunks of the original text. **Do not** summarize stretches that you have not properly understood. Concentrate on the things that you do understand and which you find most interesting.

Please approach the task the way you would approach it if you had to take it for a class and wanted to get a really good grade. Please be ambitious to write a good text. Please also make sure you proofread it at the end, save the document regularly, and only stop writing when you have completed the task to the best of your ability.

You may use the following language-related resources:

- online dictionaries
- thesauruses
- corpora, both online and offline.

You are **not** allowed to use content-related resources, for example

- other (academic) texts or articles
- Wikipedia
- social media
- etc.

Please do not contact anyone during writing, either in person or via social media. Thank you for your cooperation. This document will be available as a PDF for you to consult during writing.

## Appendix 12: Instructions for writing a reading report without additional resources.

### Instructions: Writing a reading report

A reading report is an academic text type that contains the main points, hypotheses, methodology and general line of argumentation of a research paper. A reading report should not exceed **600 words**. Give an **exact** word count, using your word processor. The document has already been formatted for you.

The aims of this task are to learn

- to understand the structure, content and line of argumentation of a scholarly research paper
- to summarize a text coherently, and
- to improve your writing skills generally.



Your reading report should make it clear to an outsider what the paper is about. Ask yourself: Is my report comprehensible for somebody who does not know the paper?

#### How do I write a reading report?

Don't panic if you don't understand a point or a passage in the paper. Read over it again, if you still don't get it, move on – very often such passages become clearer in the light of further explanation. If you don't understand a passage, don't try to summarize it – better leave it out completely. Ideally, your report should take up and answer the following questions:

- What is the topic of the paper?
- Why does the author deal with the topic? In how far is it interesting, or why does it pose problems for the researcher (e.g. open questions or diverging opinions among researchers)
- What is/are the major research question(s) the author wishes to answer?
- Are there any initial hypotheses (if any), and are these verified or falsified in the end?
- What kind(s) of data and methods did the author use (e.g. experiments, corpora)?
- What are the main arguments the author uses to get his/her point across?
- What are the most important findings of the paper, and what is its contribution to research?



**Important:** The report must be written in your own words. Do not take over or 'recycle' whole chunks of the original text. **Do not** summarize stretches that you have not properly understood. Concentrate on the things that you do understand and which you find most interesting.

Please approach the task the way you would approach it if you had to take it for a class and wanted to get a really good grade. Please be ambitious to write a good text. Please also make sure you proofread it at the end, save the document regularly, and only stop writing when you have completed the task to the best of your ability.

You are **not** allowed to use any additional resources, including

- online dictionaries
- thesauruses
- corpora, both online and offline
- other (academic) texts or articles
- Wikipedia
- social media
- etc.

Please only refer to the source text and the task description during writing. Please do not contact anyone during writing, either in person or via social media. Thank you for your cooperation.

This document will be available as a PDF for you to consult during writing.

### Appendix 13: Consent form for interview.

#### Interview – Writing a reading report

Thank you for agreeing to be interviewed. The interview is part of my dissertation, in which I look at how students use source texts when writing in English.

The interview will take approximately 1 hour. It will be held in English. Your participation is voluntary. You have the right to stop the interview or withdraw from the research at any time.

The interview will be recorded and will be transcribed to text in parts or completely. Everything you say in the interview will be treated confidentially. Both the recording and the transcript of the interview will be anonymised and it will not be possible to trace them back to your person. Summaries and direct quotations from your interview may be used in publications resulting from this research; they will also be anonymised using a pseudonym. I will take care to ensure that any information in the interview that could identify you will be anonymised or removed.

The data will be used for research purposes only. If you are interested, a copy of the recording will be made available to you. You can also request a copy of the transcript to ensure that there are no errors.

If you have any questions about the project at any point, please feel free to contact me.

Thank you very much for your time, effort, and thought in participating in this study.

Leonie Wiemeyer, PhD candidate

By signing this document I confirm that I have read and understood the conditions of participation and agree to participate in this study. I understand that participation is voluntary and I can stop the interview at any time. I can request a copy of the transcript of my interview and correct any errors.

I consent to having the recording and the transcript of the interview used for research purposes, for example in research publications. I agree to be quoted directly if a pseudonym is used.

\_\_\_\_\_

Participant's name

\_\_\_\_\_

Date and signature

## Appendix 14: General introduction and instructions for retrospective interviews with stimulated recall.

Thank you again for agreeing to this interview. As a reminder, the interview is being recorded. Both the recording and the transcript from the interview will be completely anonymised and it will not be possible to trace them back to you. Anything you say here will be treated completely confidentially, and the data will be used for research purposes only.

The interview will take between 45 minutes and an hour and we will speak English. You have the right to stop the interview or withdraw from the research at any time. If you do not want to answer a question, that is no problem at all.

During the interview, we will watch the recording of your writing session together and I will ask you to comment on specific passages. I will also ask you some general questions, which I have on these cards. I myself will say very little and mostly ask questions during the interview to give you time and space to explain what you did in the writing task. Please feel free to mention anything that comes to your mind with regard to the writing process. There is no right or wrong. Anything you have to say might be useful for my study. Just to let you know, I might take notes when you mention an aspect that I want to come back to later, so don't be confused if I start writing. Do you have any questions before we start?

## Appendix 15: Introductory questions for retrospective interviews with stimulated recall.

Setting the scene

Okay, let's get started.

So yesterday [two days ago] you wrote a reading report based on the source text about [topic].

What was that like for you?

Preparation

How did you prepare the source text at home?

Process

Could you please explain to me in as much detail as possible how you wrote the reading report in the experiment?

## Appendix 16: Instructions for stimulated recall given orally during the interview.

So let's look at what you wrote together. I will play the recording at a faster speed and ask your comments on certain parts.

If there is anything you'd like to mention to me about what you did during writing and about how you used the source text, I'd love to hear it. I am interested in the way you used the source text when writing and how you worked with it. I would also like to know how you selected important passages, how you re-used words and phrases from the source text, and how you rephrased the content of the source text.

So if you remember or notice something interesting while we watch, you can just start talking. Okay?

## Appendix 17: Questions for retrospective interviews.

Category	Question
Ice breaker	What is it like for you to watch the recording?
Source text use	Which words and phrases did you select from the source text to use in your own text? When you went back to the source text before writing a sentence, what were you looking for?
Intertextual links	Paraphrase When you wrote a sentence that was based on sentences from the source text, how did you do that? How did you select sentences for paraphrase? When paraphrasing, how do you decide which words to keep and for which to find alternative expressions? Direct quotes I noticed you quoted directly from the source text in your reading report. Can you explain why? No direct quotes I noticed you did not quote directly from the source text. Can you explain why?

Category	Question
Documentation and evaluation of sources	References I noticed that you provided [did not provide] references [in parentheses] in your reading report. Could you explain why? Evaluation Why did you decide to provide an evaluation of the source text?
Lexical strategies	Lexical strategies - rephrasing When you write academic texts, how do you go about finding the right words to rephrase something from the source text? Do you look through the source text to find useful expressions? Lexical strategies - terminology I noticed that you sometimes used terminology and key expressions from the source text. Can you explain how you selected them?
Additional resources (pilot study only)	Additional resources allowed You were allowed to use additional resources. How did you use them? No additional resources You were not allowed to use additional resources. How did you deal with situations where you normally would have used a dictionary?
Last words	Thank you again for allowing me to interview you. Before we finish, is there anything you would like to mention that we haven't talked about?

### Appendix 18: Aspects of interest for interview by participant (pilot study).

Pseudonym	Time stamp	Aspect of interest
Nadja	20:40	copy-paste with paraphrase and DQ; use of 'reflect on' which also occurs in the text
	30:30	paraphrase from copied text
	31:35	copy-paste to first paragraph for direct quote
	51:00	paraphrase from memory?
	56:00	two quotes in one sentence, no interpretation
	59:00	change of reporting verb
	1.04:30	quote attributed to student but taken from author comments
	1.08:40	change of 'Maori English' to 'ME' in quote
Vera	1.27:30	search for 'tool' in source text
	05:30 – 08:00	introductory sentence: scrolling between beginning and end of article
	09:40 – 10:00	use of quotation marks for "synonymy"
	11:47 – 12:30	writes and deletes direct quote
	16:00 – 17:30	summary with some patchwriting (initially use of own words, changed to patchwritten text)
	23:20 – 23:45	use of "according to the author"
	23:45 – 24:10	patchwriting from source, then changes made to resemble it even more closely
	24:30 – 24:45	repetition of exact copy "style and text type preference"
	25:40 – 26:05	copying from text after longer pause, followed by paraphrase (ask about paraphrasing here!)
	32:16 – 32:45	evaluation of Kjellmer's overview
	39:23 – 40:25	close paraphrase with copying, then "nouns and" deleted
40:50 – 44:20	paraphrase with corrections and "(according to Kjellmer)" in parentheses	
48:20 – 48:55	paraphrase with change of syntax	
53:30 – 54:30	addition of exactly copied string	
54:35 – 57:08	rephrasing of final sentence, deletion of more specific description of results	
Katharina	03:55 – 05:10	marks passages in instructions
	06:05 – 06:30	marks passages in source text
	24:40 – 29:30	paraphrase of first paragraph with some copying

<b>Pseudonym</b>	<b>Time stamp</b>	<b>Aspect of interest</b>
	32:40 – 32:50	adds reference
	36:40 – 37:20	marks passages
	37:35 – 37:45	“she claims that” → attribution
	40:50 – 41:40	searches source text (for what?)
	44:35 – 46:00	writes paraphrase, then changes it back to exact wording of source text
	49:30 – 50:00	several edits of author attribution
	51:40 – 53:30	bulletpoints with paraphrases
	53:30 – 54:40	looks up “zusammenhalt” in Word dictionary
	55:50 – 57:20	looks up “parameter” in linguee and integrates synonym
	1.14:30 – 1.19:00	paraphrase from several marked passages
	1.17:00 – 1.17:20	use of reporting structure
	1.22: 50 – 1.23:44	paraphrase from one marked sentence with some copying
Paula	11:40 – 18:00	paraphrase from Introduction
	20:13 – 22:20	paraphrase with deletions
	26:50 – 29:18	embedded DQ with reference, then deleted
	30:30 – 31:00	replacing with synonym
	31:00 – 34:10	close paraphrase of sentence explaining structure
	35:50 – 38:23	editing paraphrases
	41:30 – 42:30	insertion of reporting structure
	45:55 – 47:40	comment on participant quote
	58:45 – 59:40	checks synonyms
	1.05:00 – 1.07:05	long DQ with attribution and reference
	1.13:10 – 1.13:40	attribution to focus group in reporting structure
	1.17:20 – 1.20:10	paraphrase from two passages with copying and checking of synonym
	1.24:20 – 1.26:00	copy&paste to DQ, act of quoting made explicit
	1.26:25 – 1.27:20	embedded DQ
	1.28:10 – 1.28:20	repetition of “rituals of encounter”
Bianca	03:15 – 03:44	paraphrase with synonym substitution
	03:50 – 08:30	paraphrase from memory with edits
	16:25 – 18:00	searches text, then edits paraphrases
	20:48 – 24:00	deletes paraphrase, replaces with new, later expands
	26:50 – 27:10	adds reporting structure
	35:25 – 36:50	adds new introductory sentence, checks source text
	49:10 – 52:23	paraphrase, checking source text, delete, new version
	57:20 – 59:40	searches source text, then starts new paragraphs at the bottom
	1.01:10 – 1.02:50	reads Discussion, then continues PP
	1.09:00 – 1.12:20	reads instructions, PP, instructions, PP similar to intro
	1.12:50 – 1.13:35	deletes from intro, pastes again
	1.20:35 – 1.21:45	searches source text, then adds PP
Bea	06:20 – 10:10	notetaking, checking source text, copy-paste
	11:00 – 12:40	paraphrase with copy-paste
	14:50 – 16:46	paraphrase, then deletes notes
	17:27 – 21:05	notes from source text with copy-paste
	21:20 – 23:15	copy-paste, editing notes
	27:50 – 29:50	paraphrase from notes
	30:10 – 30:50	paraphrase with secondary citation

Pseudonym	Time stamp	Aspect of interest
	46:00 – 47:55	copy-paste of heading, meta-level description
	49:00 – 49:45	paraphrase from notes, checks source text
	49:45 – 50:05	re-arranges notes
	52:00 – 52:25	searches dict for 'usage', then replaces in text
	1.00:34 – 1.01:30	copy-paste of Maori word from heading
	1.06:15 – 1.06:20	deletes reference to author
	1.06:30 – 1.09:20	extensive editing
	1.11:25 – 1.12:25	2 paraphrases from 1 sentence with copy-paste
	1.14:00 – 1.14:10	exchange "this reading report" for "I"
	1.18:15 – 1.20:00	insertion of structuring sentences
Svenja	04:00 – 04:30	evaluation "interestingly"
	06:35 – 07:20	searches source text, then summarises
	12:25 – 14:45	paraphrase from Introduction, some copying
	16:00 – 17:45	writes, then searches ST, changes wording
	19:25 – 22:30	adds sentence about methods, searches text, changes wording (→ 'invited' from source text)
	34:20 – 35:00	edits title, adds year in parentheses
	41:40 – 45:10	searches source text, then paraphrases
	46:55 – 49:20	scrolls through ST, then PP
	49:40 – 52:35	paraphrasing and editing
	56:20 – 58:15	evaluation in final sentence
Miriam	00:30 – 02:00	heading w/ copy-paste, copying
	07:00 – 11:15	reads ST, PP from headings, text
	12:00 – 13:05	scrolling, meta about structure
	13:30 – 15:20	PP with DQ ([...], ref)
	16:00 – 18:00	synonym search for ST term
	18:00 – 21:45	close PP with synonym repl., summarisation
	22:40 – 27:20	PP with restructuring, DQ, reference
	29:10 – 32:20	marks ST, checks toolbox, PP with copying
	41:30 – 43:10	PP from memory, then ST
	48:25 – 51:25	DQ of defs, looks up reporting verb
	1.07:50 – 1.12:35	PP from Conclusion, dict use
	1.24:30 – 1.25:00	searches text for words
Lena	02:25 – 03:40	copying without pasting
	05:20 – 08:45	summarising from memory
	09:00 – 12:30	gist passage, some invention
	12:50 – 17:20	checks abstract to PP, checks 'institutionalised'
	25:00 – 31:45	PP from text, rephrasing, looks up word
	33:00 – 34:00	searches ST, then repositions windows
	36:30 – 42:30	re-reading
	50:00 – 1.02:50	PP, looks up words, some copying, 1 error
	1.09:00 – 1.13:30	PP from memory, some deletions
	1.14:50 – 1.18:10	scrolls to conclusion, PP, some copying
	1.27:50 – 1.29:50	close PP, no more summarisation

**Appendix 19: Aspects of interest for interview by participant.**

<b>Pseudonym</b>	<b>Time stamp</b>	<b>Aspect of interest</b>
Arne	01:25 – 07:25	copy-pastes Q from instr., highlights parts
	10:00 – 11:45	scrolls, PP, copy-paste from Abstract
	17:05 – 21:45	copy-paste to notes in italics, marks red
	22:45 – 25:30	RS + DQ + ref by copy-paste, checks instr., deletes
	29:30 – 35:10	PP, extensive synonym search, leaves gap ...
	37:20 – 41:35	PP, some borrowing, synonym check, settles on ST expression, deletes copied notes
	46:05 – 50:50	PP from 2 sentences
	58:10 – 59:20	RS + PP
	1.21:00 – 1.23:05	PP + DQ from defs
	1.27:00 – 1.35:30	copy-paste from Concl., PP from notes, deletion
	1.37:40 – 1.40:30	marks passages in notes, PP, deletion
	1.42:40 – 1.46:10	close PP, synonym search, deletes notes
	Ben	05:25 – 08:55
13:00 – 15:40		PP mostly from memory
18:20 – 18:50		attempts to rearrange windows
24:50 – 25:30		PP, looks up words
26:10 – 32:05		reads Intro, scrolls for first time, PP from Concl.
32:10 – 32:50		PP from Conclusion from memory
45:25 – 50:15		PP, reads Conclusion, edits
50:20 – 52:10		scrolls through ST, reads tables
52:50 – 58:30		adds sentence about tables
1.19:00 – 1.24:30		edits, checks ST
1.27:00 – 1.28:30	looks up 'Untertitel', scrolls	
Eva	01:30 - 04:00	highlights passages in Abstract, Concl.
	08:25 - 09:00	changes PP so it is closer to ST
	11:25 - 20:30	PP, checks instr., looks up synonyms, RS
	21:35 - 25:05	adds PP of highlighted sentence, looks up reporting V
	32:15 - 34:00	repeats reporting V, copy-paste from own text
	34:00 - 38:15	PP, some borrowing
	39:30 - 45:05	scrolls, highlights, PP from those and other sentences
	48:30 - 50:10	PP, deletes, adds RS
	55:50 - 57:05	PP with inference
	59:40 - 1.02:50	scrolls, edits, adds DQ from defs
	1.03:00 - 1.04:00	highlights more in Concl., adds info to sentence
	1.12:50 - 1.15:45	PP with RS, deletes RS, close PP, checks synonyms
	1.17:00 - 1.18:10	scrolls to Abstract, PP
1.25:05 - 1.27:00	changes reporting V, looks up synonyms	
Ina	01:00 – 03:40	notes from instructions
	08:45 – 12:40	highlights passages in Intro, PP from memory
	16:00 – 21:40	highlights, PP, adjusts windows, ref, changes ""
	22:40 – 27:55	highlights meanings, scrolls to headings, PP, corrects misrepresentation, DQ from defs + ref
	29:15 – 32:00	PP with RS + ref
	34:00 – 37:45	PP from Concl., adds refs with 'cf.'
	39:40 – 41:40	PP with DQ, no ref

<b>Pseudonym</b>	<b>Time stamp</b>	<b>Aspect of interest</b>
	44:30 – 46:20	PP from memory, German verb in parentheses
	46:30 – 47:50	PP from memory with RS
	48:20 – 54:15	PP, highlights in Concl., PP from highlighted part
	55:30 – 58:20	PP, edits, settles on ‘institutionalised’, ref
	1.02:30 – 1.07:40	PP from 1 sentence, some borrowing, ref
	1.21:50 – 1.24:40	PP with synonym replacement
Nadine	01:15 – 03:40	highlights passage in Abstract
	09:00 – 11:50	restructures, looks up word, PP, some borrowing
	13:00 – 15:40	highlights passages in Intro, no PP
	15:50 – 17:25	PP with RS from other passage
	17:45 – 23:55	highlighting, PP from paragraph from memory
	24:10 – 25:50	close PP from 1 sentence
	33:00 – 33:30	highlights passages already paraphrased
	35:00 – 39:00	begins PP, marks text, copies, adds ‘+ ref
	48:00 – 54:00	highlights, PP, checks ST, corrects, checks instr.
	57:00 – 1.02:30	copy-paste from ST, PP, deletes, some borrowing
	1.03:40 – 1.10:30	copy-paste from defs, PP, reformatting, ref
	1.21:20 – 1.23:50	copy-paste, RS, PP from copy, deletes, ref
	1.27:45 – 1.29:30	PP with RS, then highlights
	1.32:00 – 1.35:00	checks instr., scrolls to Concl., highlights, PP
	1.41:30 – 1.43:30	PP from 1 sentence, edits
Tara	03:20 – 07:00	copies from Abstract, adds ref, no quotation marks
	09:20 – 10:30	copies to notes, leaves out segments from quote
	14:40 – 17:10	copies topic sentences from subchapters
	18:00 – 20:05	copies, marks ellipsis with ...
	35:50 – 37:00	copies sentence about meanings, skips meanings
	1.14:15 – 1.19:00	minimises ST, starts intro (PP from notes)
	1.20:00 – 1.22:20	looks up terms in ST, PP, deletes copied notes
	1.22:20 – 1.24:25	PP from notes, some patchwriting, changes order, provides ref
	1.30:50 – 1.33:10	PP from notes, reuses word, synonym repl.
	1.38:00 – 1.39:30	RS, patchwriting from notes, ref
	1.43:10 – 1.43:40	rearranges copied excerpts, marks replaced note
	1.49:30 – 1.50:30	close PP from copied notes
	1.55:50 – 1.58:45	PP from notes, misrepresentation

**Appendix 20: Total no. of codings for writing acts (W) and non-writing acts (NW) in the L2 writers' processes.**

Type of act	Code	Writer						No. of coded *IW segments	% of coded *IW segments	No. of instances
		Arne	Ben	Eva	Ina	Nadine	Tara			
W	W - Writing PP from memory	0	19	0	25	23	0	67	36,4%	207
W	W - Writing PP from ST	27	0	29	2	0	1	59	32,1%	189
W	W - Deleting and rephrasing string of paraphrase	14	15	21	16	15	15	96	52,2%	167
W	W - Writing PP from notes / copied ST passage	8	0	0	0	3	29	40	21,7%	83
W	W - Inserting quotation marks	3	4	8	4	9	4	32	17,4%	80
W	W - Writing reporting structure	8	6	11	4	12	5	46	25,0%	62
W	W - Adding word(s) to existing string	10	5	8	8	3	5	39	21,2%	44
W	W - Manual copying from ST	2	4	6	4	9	0	25	13,6%	39
W	W - Providing reference	1	0	0	8	3	18	30	16,3%	37
W	W - Synonym replacement	5	8	5	2	4	1	25	13,6%	28
W	W - Manual copying from notes / copied ST passage	0	0	0	0	1	4	5	2,7%	5
W	W - Inserting placeholder	1	0	0	1	0	0	2	1,1%	2
W	W - Replacing placeholder	1	0	0	1	0	0	2	1,1%	2
W	W - Replacing DQ with PP	1	0	0	0	0	0	1	0,5%	1
W	W - Writing notes	1	0	0	0	0	0	1	0,5%	1
W	W - Replacing PP with DQ	0	0	0	0	0	0	0	0,0%	0
NW	NW - Re-reading ST	23	13	22	19	25	3	105	57,1%	224
NW	NW - General editing	14	9	13	9	19	11	75	40,8%	128
NW	NW - Re-reading own text / notes	17	11	4	14	10	11	67	36,4%	102
NW	NW - Consulting language resource	29	5	11	1	3	0	49	26,6%	93
NW	NW - Deleting word(s) without replacement	8	2	13	4	4	3	34	18,5%	46
NW	NW - Highlighting passage(s) in ST	2	0	3	8	9	0	22	12,0%	32
NW	NW - Deleting notes / copied ST passage	3	0	0	0	3	13	19	10,3%	31
NW	NW - Copy-paste from ST	4	0	0	0	4	0	8	4,3%	22

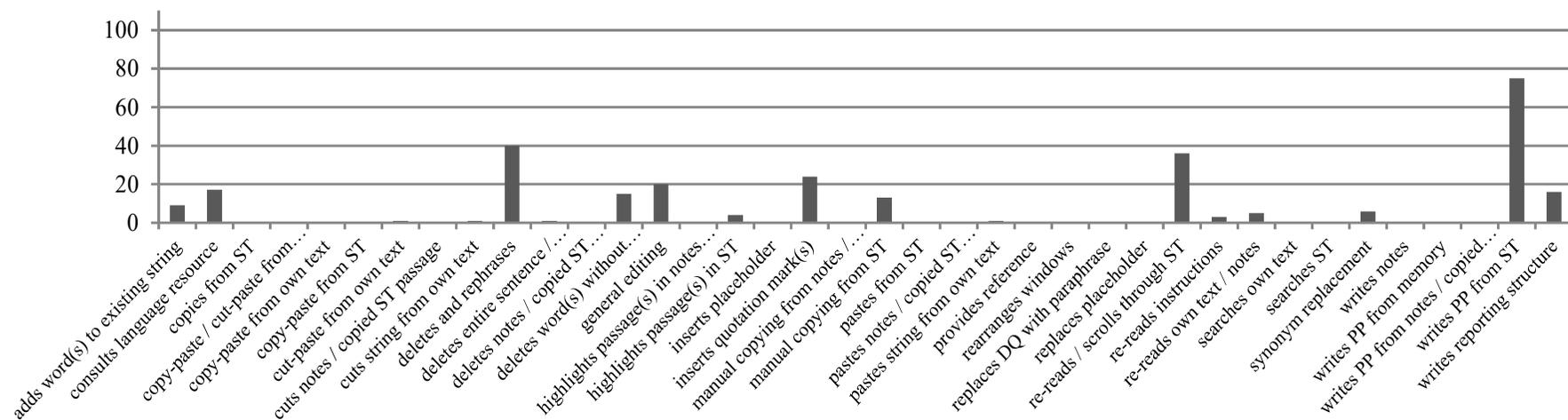
Type of act	Code	Writer						No. of coded *IW segments	% of coded *IW segments	No. of instances
		Arne	Ben	Eva	Ina	Nadine	Tara			
NW	NW - Re-reading instructions	1	6	3	1	3	0	14	7,6%	17
NW	NW - Rearranging windows	2	3	0	6	0	1	12	6,5%	17
NW	NW - Copy-paste / cut-paste from own text	1	3	2	0	1	2	9	4,9%	12
NW	NW - Deleting entire sentence (fragment)	3	2	1	0	1	3	10	5,4%	10
NW	NW - Copy-paste / cut-paste from notes / copied ST passage	4	0	0	0	0	3	7	3,8%	7
NW	NW - Highlighting passage(s) in notes / copied ST passage	2	0	0	0	0	0	2	1,1%	3
NW	NW - Searching own text	2	0	0	0	0	0	2	1,1%	2
NW	NW - Searching ST	1	0	0	0	0	0	1	0,5%	1

**Appendix 21: Overview of instances of acts in each L2 writers' process.**

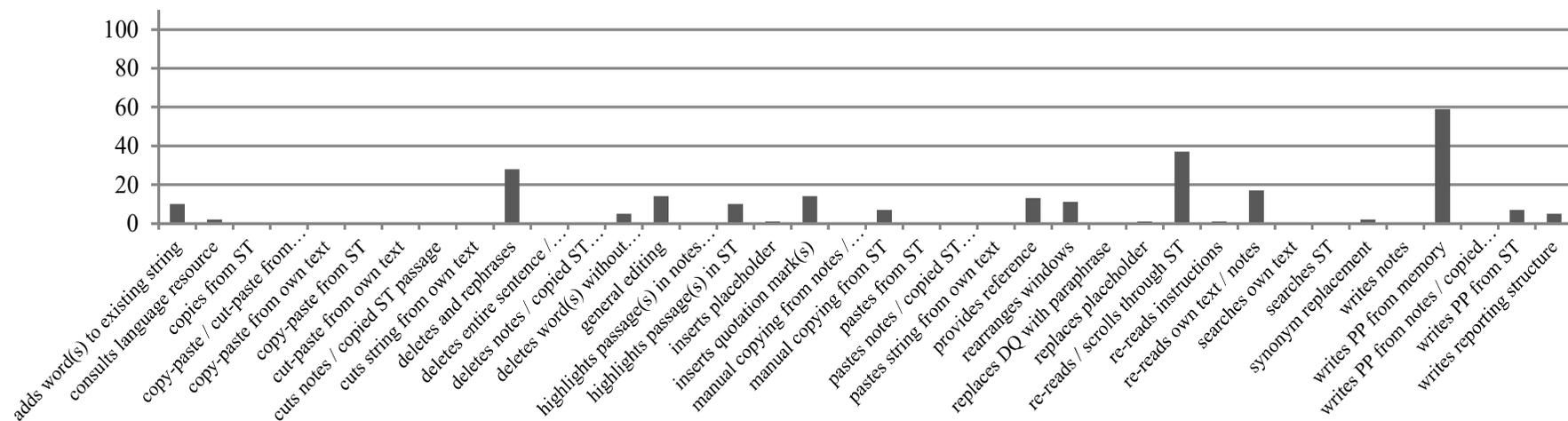
Act	Arne	Ben	Eva	Ina	Nadine	Tara	TOTAL
adds word(s) to existing string	10	6	9	10	3	6	44
consults language resource	65	6	17	2	3	0	93
copies from ST	0	0	0	0	1	0	1
copy-paste / cut-paste from notes / copied ST passage	3	0	0	0	0	2	5
copy-paste from own text	2	0	0	0	1	0	3
copy-paste from ST	15	0	0	0	5	0	20
cut-paste from own text	0	3	1	0	0	1	5
cuts notes / copied ST passage	0	0	0	0	0	1	1
cuts string from own text	0	1	1	0	0	0	2
deletes and rephrases	22	27	40	28	24	26	167
deletes entire sentence / fragment	3	2	1	0	1	3	10
deletes notes / copied ST passage	5	0	0	0	4	22	31
deletes word(s) without replacement	14	3	15	5	6	3	46
general editing	35	18	20	14	31	10	128
highlights passage(s) in notes / copied ST passage	3	0	0	0	0	0	3
highlights passage(s) in ST	2	0	4	10	16	0	32
inserts placeholder	1	0	0	1	0	0	2
inserts quotation mark(s)	10	8	24	14	19	5	80
manual copying from notes / copied ST passage	0	0	0	0	1	4	5
manual copying from ST	4	4	13	7	11	0	39
pastes from ST	0	0	0	0	1	0	1
pastes notes / copied ST passage	0	0	0	0	0	1	1
pastes string from own text	0	1	1	0	0	0	2
provides reference	1	0	0	13	5	18	37
rearranges windows	2	3	0	11	0	1	17
replaces DQ with paraphrase	1	0	0	0	0	0	1
replaces placeholder	1	0	0	1	0	0	2
re-reads / scrolls through ST	60	21	36	37	66	4	224
re-reads instructions	2	8	3	1	3	0	17
re-reads own text / notes	28	23	5	17	15	14	102
searches own text	2	0	0	0	0	0	2
searches ST	1	0	0	0	0	0	1
synonym replacement	5	10	6	2	4	1	28
writes reporting structure	12	6	16	5	16	7	62
<b>SUM</b>	<b>309</b>	<b>150</b>	<b>212</b>	<b>178</b>	<b>236</b>	<b>129</b>	<b>1694</b>



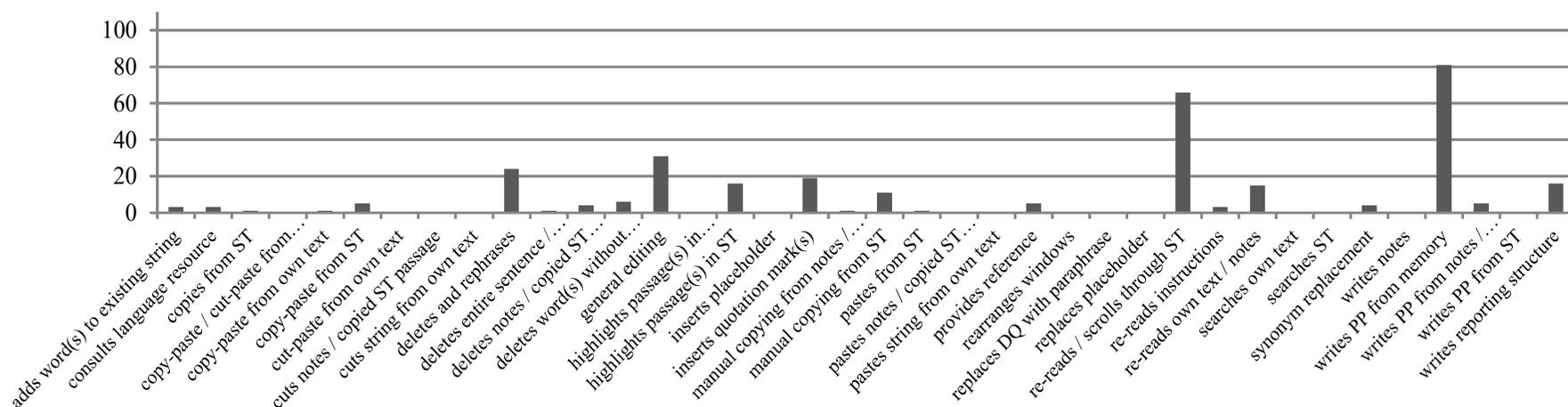
## Eva



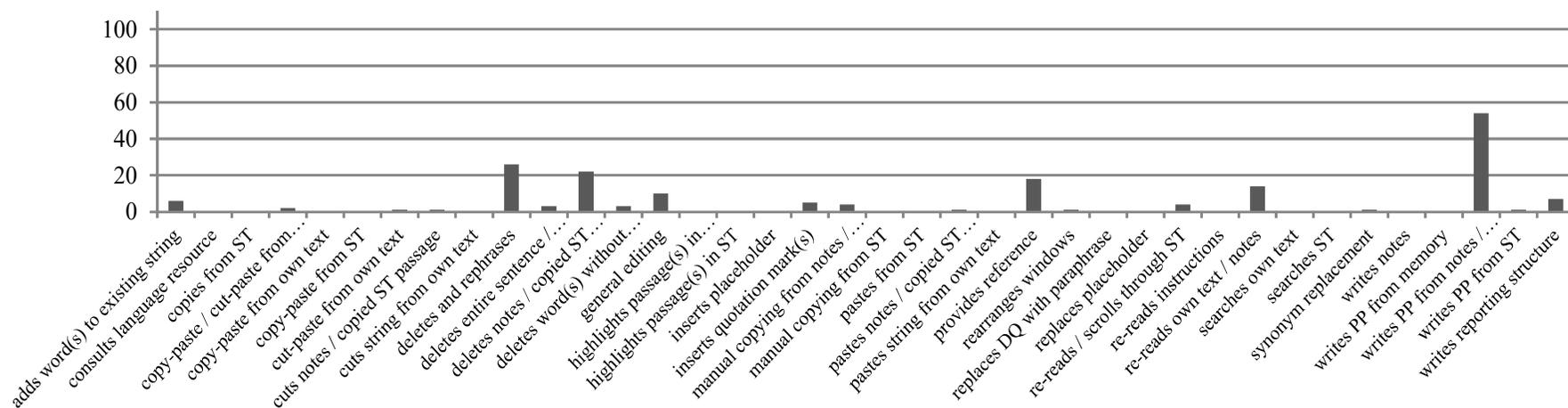
## Ina



## Nadine



## Tara



**Appendix 23: Absolute number of instances per core act in L2 writers' processes.**

Act	Arne	Ben	Eva	Ina	Nadine	Tara	TOTAL
re-reads / scrolls through ST	60	21	36	37	66	4	224
deletes and rephrases	22	27	40	28	24	26	167
general editing	35	18	20	14	31	10	128
re-reads own text / notes	28	23	5	17	15	14	102
inserts quotation mark(s)	10	8	24	14	19	5	80
writes reporting structure	12	6	16	5	16	7	62
deletes word(s) without replacement	14	3	15	5	6	3	46
adds word(s) to existing string	10	6	9	10	3	6	44
synonym replacement	5	10	6	2	4	1	28
SUM	196	122	171	132	184	76	881
% OF ALL ACTS	44.5%	56.2%	59.60%	54.10%	57.10%	41.3%	