Identifying the typicality of student academic writing: a comparative corpus study

Abstract

This study compared a corpus consisting of student theses with a corpus of academic articles at various linguistic levels, in order to single out characteristics that are particular to student academic writing. Unlike earlier research, this study focused on only a single academic discipline – history. It found that very few distinctions existed between the corpora at the lexical level. However, it found marked differences in the dialogic stance taken by the student and expert writers, as well as in their use of secondary sources and the way the authors introduced their research topic.

Student name: Tess Dudink

Student number: 1490974

Supervisor: Dr L. Fonteyn

Date: 09-02-2021

Wordcount (excl. reference list and title page): 17,025

Table of contents

Introduction

English has established itself as the leading language of academia, and over the past decades academic writing courses for non-native speakers have become part of the standard university curriculum all over the world. Academic English proficiency for students has never been more relevant. It is a generally accepted opinion that each academic discipline has its preferred text type and its own specialized vocabulary. Until now, however, student writing has almost solely been researched through corpus studies that consisted of texts from various disciplines. By instead focusing on a single discipline – history – the current study aims to do justice to these specific characteristics, through the comparison of student and expert academic writing.

The approach put forward in this study involves a direct comparison between a corpus consisting of student theses, and one consisting of professional articles, singling out those characteristics that are particular to student academic writing. In order to do justice to the intricacies of academic English, comparisons must happen at various distinct levels. The corpora are first analysed at the word level, using lexical verbs as markers of academic writing proficiency.

The second part of the study applies a dialogical framework to student and expert academic writing, in order to establish how these two groups position their work in relation to other academic material. It also employs a quantitative analysis to gain insight into their interaction with secondary sources. Finally, it combines the results from these inquiries with a qualitative discussion of the way in which the research topic was introduced by these writers, and how this reflects their own vision of their work.

Combined, these measurements will illustrate not only the students' English proficiency and their competence in academic writing , but also their ability to interact with – and position their own work in – a broader academic discourse.

The following chapter will first provide an extensive theoretical framework based on insights from previous research. It will start with a discussion of the key elements of discourse competence that go into writing an academic paper. This includes such aspects as correct grammar and vocabulary, but also the construction of a solid argumentative structure, and the mastery of the appropriate register and text type. This will be followed by a discussion of how an authoritative stance is linked to good academic writing. These sections will in turn be followed by a consideration of the specifics of history writing, as a genre that stands apart from the academic writing of other disciplines. The next part of the paper will discuss one of the great pitfalls of academic writing: the production of overly complex texts. Together these sections should provide an overview of all the elements that are necessary to produce a successful academic paper.

Section 1.2 will make use of a dialogic framework to illustrate how academic texts are placed in a broader discourse, and how writers interact with other academic articles. This will be followed by a brief discussion on plagiarism, which arguably is also a form of interaction with another academic work, albeit a morally questionable one. The final part of the theoretical section of this paper points out some key findings from previous studies regarding the differences between student and expert writing. Specifically, it will argue that verbs and linking adverbials are markers of academic literacy.

The second chapter is devoted to the methodology used in the current study. Chapter 3 will provide the results of its analyses and discuss these findings. The final part of this paper will present the study’s conclusions, as well as its limitations and some suggestions for future research.

1. Theoretical framework

This chapter provides an overview and discussion of relevant previous research. It first considers the characteristics of academic writing as a genre, its purpose and its building blocks. The second part of the chapter is concerned with the interaction between academic papers, specifically how the concepts found in dialogism apply to written texts. The final sections of the chapter will introduce the findings of existing research on the differences between student and expert academic writing, and discuss to what extent these previous conclusions are relatable to the current research groups.

1.1 Academic writing

The following section illustrates how the universality of academic English has inspired research into what constitute the defining characteristics of academic writing. In order to establish guidelines for teaching academic English to students, the elements that are required for writing a paper had to be identified and made concrete. The results of these inquiries are presented in Section 1.1.1. The following Section 1.1.2 will then discuss authority in academic texts. An authoritative stance has been linked to (university) success in academic writing. However, as Section 1.1.2 will illustrate, there is no straightforward linguistic measure for authority. Section 1.2.3 will next provide an in-depth analysis of the features that separate academic history writing from other academic disciplines. This section will thus illustrate why it is necessary to consider academic writing as consisting of numerous sub-genres. This is followed by Section 1.2.4, which will present insights in what is perhaps the greatest weakness of academic writing, namely the fact that it can be complex up to the point that this complexity undermines the content. The primary aim of this section is to make the reader aware that even though most corpus studies – the present one included – will compare student writing to a corpus of professional articles, these expert authors are still fallible. Finally, Section 1.1.5 will give some preliminary conclusions and explain their relevance for the current study.

1.1.1 Academic English

Apart from its status as a lingua franca, English has also established itself as the leading language of academia and science writing (Salazar, 2014). According to Pot and Weideman, the 'output of research' forms the primary concern of all academic discourse, and the preferred format to present research to the world is as an academic article published in a peer-reviewed journal (2015, p.24). At present, the majority of top scientific journals require contributions to be written in English, and this holds true for publications in (nearly) every field (Salazar, 2014). Scientists who are not native speakers of English have had to adapt to this reality, and currently academic English forms part of the standard curriculum for university students globally. In order to establish what would be the best method to teach academic English to students, it was necessary to obtain a better understanding of this discipline and the challenges it offered to non-native speakers in particular. In consequence, various corpora-based studies were employed in the attempt to establish universal guidelines for successful academic writing.

Among the early discoveries of this corpus-based research was the recognition of lexical bundles in papers written by native speakers. The concept of a lexical bundle was introduced in the *Longman grammar of spoken and written English* by Biber et al (1999). Essentially, 'instead of constantly making new combinations of individual words, native speakers often depend on a stock of prefabricated, semi-automatic word chunks'(Salazar, 2014, p.1; see also Biber et al, 1999; Sinclair 1991). Writers less familiar with the English language cannot rely on such a mental library of ready text, making the production of academic papers a much more difficult (and time-consuming) task. A survey taken in 2007 among under- and postgraduate students at a South African university revealed that non-native users of English often lacked the required language proficiency to function successfully at university level (Pot & Weideman, 2015). Indeed, 'a low level of academic language proficiency often results in failure to achieve academic success' (Pot & Weideman, 2015, p.21; see also Weideman, 2003; Van Dyk and Weideman, 2004 for more information on this topic).

Further research into the structure of academic texts also concerned the deliberation of what Pot and Weideman termed the typicality of academic discourse (2015, p.24), in other words, the search for the defining characteristic of academic writing as a distinct discipline. Patterson and Weideman (2013) pointed out the act of making distinctions as the key factor of all successful academic communication. This same view has been expressed by Weideman and Van Dyk, who argued that such distinction-making occurs through analysis, which lies at ‘the core of academic argumentation and discourse’ (2014, p.4). That this is indeed a generally held opinion was confirmed by the results of a questionnaire held at the University of Pretoria, which showed that nearly all academic supervisors held academic argumentation to be the most important writing skill that their students had to master (Butler, 2010).

In order to successfully master the art of academic writing in English, students are required to improve their discourse competence in the language. Discourse competence can be defined briefly as a student’s ability to integrate their knowledge of the language in order to produce a text that is both ‘linguistically accurate and socially appropriate’ (Bruce, 2008, p.1). A more complete definition can be found in the Common European Framework for Reference of Languages, which defines discourse analysis as ‘the ability of the user/learner to arrange sentences in sequence so as to produce coherent stretches of language’ (Council of Europe, 2001, p.123). The definition further implies ‘knowledge and the ability to control’ such elements as natural sequencing, coherence, and register, as well as knowledge of text design (Council of Europe, 2001, p.123). Discourse competence is thus the mastery and application of these and other skills, all of which are required to successfully structure an academic paper.

Precisely because proficiency in academic English has become a determining factor for university success, special assessments have been developed specifically to ascertain students’ academic literacy, using the typicality of academic writing as a directory. Essentially, these tests illustrate the extent to which students have mastered the necessary discourse competence in the English language to produce academic texts. One such an assessment is the Test of Academic Literacy for Postgraduate Students (TALPS), which as its name indicates is a test designed to measure the abilities of (prospective) postgrads to understand and produce academic literature.

The TALPS is aimed at students functioning at the very highest level of university education, and as such the level of proficiency it tests for very closely approaches the quality of professional research. The different aspects that are tested in the TALPS can thus be used to illustrate the different aspects that are important for academic literacy. Table 1 shows the separate sections of the TALPS, along with a summary of the skills tested in each section, based on the descriptions given by Butler (2010) and Pot and Weideman (2015). Combined these can be said to provide a list of the necessary skills to be able to produce and interpret academic literature successfully.

TALPS sections	Skills tested
Section 1: scrambled text	Restructuring a paragraph to form a cohesive whole, while recognising text relations and lexical clues, and interpreting context.
Section 2: visual and graphic literacy	Interpreting graphs and visual information, making calculations and drawing inferences.
Section 3: vocabulary	Proficiency in general academic vocabulary.
Section 4: text types	Recognition of different in text types.
Section 5: comprehension	Classifying information, making inferences, understanding metaphorical language and distinguishing between essential and non-essential information.
Section 6: grammar	E.g. sentence construction, communicative function
Section 7: editing	Recognising and correcting text errors.
Section 8: writing	Applying a formal register, logical structure, and proper referencing .

Table 1: The eight sections of the Test of Academic Literacy for Postgraduate Students (TALPS) with a summary of the skills tested in each section (based on Butler, 2010, pp.6-12; Pot & Weideman, 2015, p.26).

1.1.2 Authority in academic writing

A further key aspect of academic writing is the use of authority. Matsuda and Tardy (2007) recognised an ‘authoritative voice’ as one of five distinct forms of voice in writing (p.236). Although the term voice has frequent connotations with identity and individualism in writing, Prior (2001) demonstrated that voice may also be ‘collective or social’ (p.62). Similarly, Matsuda (2001) stated that ‘voice is the amalgamative effect of the use of discursive and non-discursive features that language users choose, deliberately or otherwise, from socially available yet ever-changing repertoires’ (p.40). Research has shown an authoritative voice to be directly linked to academic success (Tang, 2009).

Hyland (2002) compared pronoun use in academic writing by students and researchers working in range of disciplines. The study found that published authors were more likely to use first person pronouns than students. Hyland (2002) concluded that this correlated with a more authoritative role that was taken on by the expert-writers. However, it should certainly be considered a possibility that in drawing this conclusion, Hyland may have too easily dismissed the role prescriptivism plays in student writing. Students are expressly taught that the use of the passive is a characteristic of good academic writing, and most university students will risk a deduction in their final grade by being too enthusiastic in their use of a first person perspective. It is by no means the case that first person pronoun use is defined solely by the authoritative stance taken by the writer.

A more nuanced look at authoritative writing is provided by Tang (2009), who explored three distinct concepts of authority within academic discourse. In the first, authority is related to the writer’s ability to demonstrate familiarity with the conventions and practices of their chosen discipline. Authority is thus construed as expertise in a given field. The second view links authority to the creation of original content. This is the most literal interpretation of the word ‘author’ in the sense of ‘writer’. The last concept discussed by Tang (2009) followed the argument that authority is not an element of writing in itself, but rather a type of title that is bestowed by the reader.

Although students should ideally aim to master each of the three forms of authority, it is arguably the first view that is of most interest to the discussion of academic writing, as it is this form of authority that is actively taught to (university) students. It is moreover important to highlight that within this view of authority, Tang referred to the specifics of certain disciplines. Rather than a single academic English that suffices for every topic, there are different genres within academics that students must learn to recognize and master (Bruce, 2008; also Tang, 2009).

1.1.3 Academic history writing

Section 1.1.1 has pointed out those aspects of discourse competence that are necessary for successful academic writing. These elements, also tested by the TALPS, have thus far been presented as a rather straightforward list of skills to be mastered in order to produce a correct academic work. However, such an approach may be problematic in light of the fact that, instead of a single academic English, there exist different categories within academic writing (Conrad, 1996). As it turns out, each academic discipline carries with it its own text type and its own specialized vocabulary. As such, although it is true that all of the skills tested by the TALPS are necessary to produce a good academic work, several of them (most notably the use of academic vocabulary and the recognition of correct text types) require very different concrete knowledge depending on a student's chosen discipline. This section will explore the specifics of academic history writing.

The language variation that marks different academic disciplines results from the epistemological traditions of each discipline (Conrad, 1996). The most obvious break in academic subjects is the one between the so-called 'hard sciences' (such as chemistry, biology, medicine, etc.) on the one hand, and on the other hand the 'soft sciences', which include the social sciences, history, and literary studies among other disciplines. It is possible to look at early scientific papers and find a text on philosophy and one on medicine with very similar structures. However, compare current papers on these topics and you will immediately discover that these feature a very different framework indeed. This change has been decidedly carried on by the 'hard' scientists. Humanities writing still retains much of the historical style, whereas the science authors have become more fact-oriented and less elaborately argumentative (Biber & Gray, 2016).

Biber and Gray (2016) considered the language variation of hard science researchers and humanities researchers by comparing works on literature and on biochemistry, with a focus on the complexity of academic discourse and the way in which this complexity is achieved. Biber and Gray's (2016) findings showed that humanities writing has grammatical complexity due to structural elaboration, whereas science writing tends to be more complex as a result of structural compression. Practically, this implies the following similarities and differences: Biber and Gray's found that both the hard and the soft science writers employed a technical vocabulary and nominalizations, each of which contributed to the texts' complexity. Both disciplines were also marked by preferred use of the passive voice. In the humanities the use of attributive adjectives was markedly higher than in the biochemistry papers. The science papers, on the other hand, contained more nouns as noun pre-modifiers, as well as more noun and participle combinations in the function of noun pre-modifiers. Along with a significantly higher use of appositive noun phrases these structures formed part of the complexity of the text as a whole (Biber & Gray, 2016).

The challenge for students is to understand and apply the conventions of their own genre, both in terms of preferred grammatical structures and the ordering of content. As it turns out, history proves a particularly difficult discipline to master, since little agreement exists even among professionals on what constitutes a good paper. Sources of scholarly debate include questions such as 'what sources are trustworthy?', 'which analysis methods are preferred?' and 'what is a significant question to write about?'. These are inquiries that lie at the heart of academic writing, and they are much easier to answer for the 'hard science' disciplines than they are for the humanities or even the social sciences (see Conrad, 1996; and also Novick, 1988 for a discussion on this topic).

As a humanities subject, academic history is argumentational rather than experimental (Conrad, 1996). History is not an evidence based science. This means that new conclusions in this discipline are generally caused by a shift in perspective, and only rarely by the discovery of new material. This is unlike the hard science disciplines, which are propelled forward by experimental research and rely on the new information gained from such studies. Because historians rely primarily on their argumentation to support their viewpoint, the narrative organization of historical texts takes on a particularly important role (Stockton, 1995).

This highlights the importance of the discourse competence mentioned before. It should also be mentioned that academic history is hardly a uniform discipline. The source materials available to a classical historian or to a researcher of 20th century political relations are entirely different both in nature and in quantity. Similarly, a medieval historian and a colonial historian will hardly ever agree on a preferred approach or research method. Even within a single academic discipline – history – there are different genres of academic writing.

1.1.4 Complexity in academic texts

The differences between separate scientific disciplines, discussed in the previous section, have perhaps led to a rather skewed view of academic writing (Conrad, 1996). This is due to the fact that studies have naturally tended to stress the differences of the hard and soft sciences more than their similarities. Nevertheless, similarities between subgenres do exist. One good example of a general characteristic of academic writing – already mentioned in the first section of this chapter - is that academic writing can be very, some would say overly, complex. This has led to the stereotype that professional researchers exhibit a tendency to be more concerned with the linguistic complexity of their text than with its content. In other words, that academic English is ‘deliberately complex and more concerned with impressing readers than communicating ideas’ (Biber & Gray, 2016, p.1).

Biber and Gray (2016) have explored the linguistic features that produce vagueness in texts, or in their own and more forgiving terminology, led to texts being ‘not maximally explicit’ (Biber & Gray, p.14). On the one hand there exist the hard sciences. The feature that, according to Biber and Gray, lies at the hard of academic complexity in these disciplines is the frequent use of embedded phrases. Embedded phrases form part of the structural compression of a text – put very simply the number of words used to express an idea is reduced - with the result of making the texts much harder to interpret for the reader due to ‘a major reduction in explicitness’ (Biber & Gray, 2016, p.18).

Researchers working in the soft sciences, however, tend to err on the side of structural elaboration rather than structural compression. They use the long, winding sentences that typically earn academic writing’s bad reputation. A perhaps counterintuitive finding is that complex sentences – in the linguistic term of a sentence with one or more dependent clauses – are not on their own a good predictor of the overall complexity of a text. In fact, natural speech contains more dependent clauses than academic writing does (Biber & Gray, 2016) The vagueness that these authors are charged with must have some additional cause.

At the lexical level, Robinson argued that it is ‘the use of words without fixed or clear meanings [which forms] a major part of what makes academic writing so terrible’ (Robinson, 2017). This does not necessarily mean, as might be construed, that academic writing is always filled with unusual or highly specialized vocabulary. Well-known – ‘everyday’ – words can be equally vague when context leaves them open to a variety of interpretations. Robinson discusses several extracts to illustrate these instances of vague language. One such extract is given here, in this case taken directly from the source text ‘Towards a Relational Phenomenology of Violence’ by Staudigl in order to preserve the formatting (e.g. cursive) of the original article.

> ‘As I want to show in this article, it is indeed of utmost importance to examine the various faces of violence in their intrinsic *relationality*. To unveil their relational character, I will attempt to substantially broaden the phenomenological concept of sense. By *sense*, I propose not only to examine the immanent accomplishments of the subject’s engagement in and with the world, but, first and foremost, a relation that unfolds in-between the one and the other. Sense, in other words, unfolds in the subject’s relation with those it encounters in this world, who can make this world appear to it, *dys*appear, or, finally, disappear, and accordingly shape its selfunderstanding, self-conception, and agency.’ (Staudigl, 2013, p.44)

As Robinson pointed out, in this quotation even fairly common words such as ‘sense’ or ‘relation’ are hardly likely to evoke the same meaning to any two readers of this text. Whatever its origin, such lack of clarity carries with it serious implications. It allows researchers, especially those working in the soft sciences to publish claims that, through the use of vague language, are not merely open to more than one interpretation, but through the same means withstand the possibility of a clear refutation. This arguably goes against the fundamentals of science-related philosophy as introduced by Popper, that any scientific proclamation must be falsifiable in nature. For the soft sciences this might be interpreted as meaning that academic writing should proffer a clear and unmistakable message, which lends itself to refutation by means of argument by other researchers.

1.1.5 Summary and implications

Several researchers have identified the key element of academic discourse to be the ability to make distinctions through an argumentative structure (Patterson and Weideman, 2013; Pot & Weideman, 2015). Previous research has moreover indicated that discourse competence and the mastery of an authoritative voice are skills that lie at the heart of successful academic writing (Tang, 2009). In the current academic climate, it is necessary for (emerging) researchers to master these aspects of the English language. This can prove particularly challenging for non-native speakers, who do not possess the advantages (e.g. lexical bundles) that native speakers do. For students of the humanities, and of history in particular, academic writing poses additional challenges as they are required to defend their choice of research questions and methods in a field where such decisions are not based on any universally accepted guidelines. Even after manoeuvring all the variables that go into the writing of an academic work, students still run the risk of overcomplicating their texts to the point where they lose explicitness, a pitfall to which many professional writers have succumbed before them.

As a result of this variety of elements that go into any academic text, the possibility to determine the quality of student academic writing based on a predetermined set of characteristics is necessarily limited. Although many elements that are necessary to academic writing have been identified (see the TALPS sections in Table 1 for examples), it can still be difficult to assign a value to these different elements. In addition, some of the aspects that have been linked to good academic writing – most notably an authoritative stance – can be difficult to reduce to linguistic features. It is therefore far more useful to employ a corpus of acclaimed academic writing, and compare this to the writing of students. Any differences found as a result of this method can be used to determine the quality of student writing in comparison to what is currently considered the highest achieved level of academic writing. This is the method that will be employed in the current study.

1.2 Source use in academic writing

The previous sections have dealt with some of the most notable characteristics of academic writing, but have been limited to the original content written by a given author. Of course, an academic paper never stands alone, but rather is embedded in a larger discourse on its topic. As stated in a 2013 article by Fryer, Bech and Andersen, 'the way in which an author engages with and positions him/herself in relation to other voices in the discourse, e.g. with the literature and the putative reader, is an integral part of … research' (p.183). The following section will explore the interactions between academic papers.

In order to facilitate this discussion it is necessary to first establish the distinction between primary and secondary sources. Although these are likely long familiar terms, they do have slightly different definitions in different academic fields. In the hard sciences a text is considered a primary source when it contains original research. Almost exclusively such sources are research papers published in peer-reviewed journals, and follow the structure of an introduction, followed by a methods section, a results section, and finally a discussion section. On the contrary, secondary sources are texts that do not describe their own experiments, but rather use primary sources as their main source material. Examples of secondary sources in the hard sciences are review articles or meta-analyses. (See *Primary vs secondary literature in the biomedical sciences*, n.d.)

For many of the humanities disciplines, these definitions differ somewhat. In historical research, a primary source is generally a historic artifact. Secondary sources are the later interpretations of primary sources by historians (*Primary sources: a research guide*, n.d.; see also Scheuler, 2014 for an in-depth discussion of the distinction between primary and secondary sources in history writing). This means that scholarly articles are secondary sources, even when they present original research (as discussed earlier the humanities trade in viewpoints rather than facts). These definitions are given at length here to avoid any confusion. As this paper is concerned with the writing of history papers, the last discussed definitions hold true for any future mention of secondary sources.

1.2.1 Dialogism

Bakhtin and Holquist (1981) introduced the idea of dialogism in writing. More specifically, Tang (2009) illustrated that student academic writing is dialogic in nature in two separate ways. The first dialogue that the student-author engages in is the one with a ‘specific tutor-reader’. The second is the dialogue with the ‘wider disciplinary community’ (Tang, 2009, p.170). The latter is of course true for all academic publications. Fryer et al (2013) argued that for researchers, one part of this dialogue is shaped by the anticipation of the response of their reader-audience (p.184). However, the academic dialogue primarily takes shape in the form of interaction with other publications.

The key work on academic dialogism done by Martin and White – built on Bakhtin and Holquist’s (1981) work on dialogism – established that expressions may either be ‘monoglossic’ or ‘heteroglossic’ in nature (2005, p.100). The category monoglossic being applied to those utterances that do not refer to the existence of other viewpoints, whereas statements may be termed heteroglossic when they ‘invoke, allow for, or in some way challenge other voices or viewpoints in the discourse’ (Fryer et al, 2013, p.186). Which of the two categories a statement belongs to is thus determined by whether or not it recognises dialogistic alternatives (Martin & White, 2005).

It is not exactly the case that monoglossic writing shuts the door to all further discussion. These types of expressions may further academic dialogism as ‘the disposition of the text may be such that the categorical, monoglossically asserted proposition is presented as very much in the spotlight – as very much a focal point for discussion and argumentation’ (Martin & White, 2005, p.101). However, such interaction must then always be the result of a subject being taken up by a later reader. It is not part of a foreseen dialogue invited by the original author. Hence the focus in the rest of this chapter will lie with heteroglossic statements.

These heteroglossic interactions with other works are most likely to be found in specific sections of an academic paper, most notably in the introduction, which necessarily relies heavily on the work of other authors as providing a basis and a reason for the current work, and in the discussion section, which invites future research to contribute to the topic. The methods and results sections, on the contrary, are made up primarily of monoglossic statements (Fryer et al, 2013). These findings moreover correlate with the frequency of markers of modality, which several studies have shown are found more often in the introduction and discussion sections of a paper (see Conrad, 1996).

The interaction with other academic texts can take on multiple forms. For example, authors can make the decision to acknowledge or even embrace the views on their topic that are presented in another academic work. They can also choose to dismiss the opinions expressed by their colleagues. Even to ignore certain ideas can imply a conscious decision by the author. All of these options function in one of two ways, according to Martin and White's 'engagement system', depending on whether they open or close the space for alternative positions (2005, pp.103-104). They can be dialogically expansive, meaning they open the discussion, or dialogically contractive, meaning they show that other options exist, but consider them less valuable (see Fryer et al, 2013; Martin & White, 2005; and Tang, 2009). Figure 1 below, taken from Martin and White (2005, p.104), gives a visual illustration of the basis of this engagement system, which Fryer et al described as 'a subsystem of appraisal dealing with writer/speaker resources for intersubjective positioning … [its] resources include what are generally dealt with under the headings of modality, hedging, and attribution, among others' (2013, p.183).

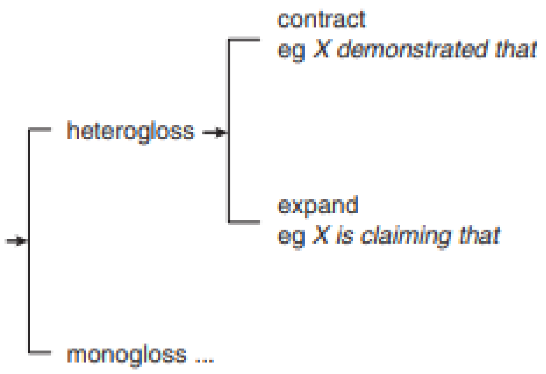

Figure 1: The basic division of the engagement system, taken from Martin and White (2003, p.104).

Fryer et al (2013) found that in a corpus of medical science writing, of all the heteroglossic statements that were discovered 66.51% belonged to the 'expand' category. Martin and White (2005) also described several subcategories within the 'contract' and 'expand' groups. The later study by Tang built on these structures, and established the following dialogically expansive categories with examples:

postulate	(it is possible that…)
evidentialize	(it appears that…)
hearsay	(some say that…)
acknowledge	(X says that…)
distance	(X claims that…) (Tang, 2009, p.173)

And the following categories for dialogically contractive possibilities:

pronounce	(I believe that…)
concurrence	(of course, …)
endorse	(as X argued, …) (Tang, 2009, p.174)

The work done by Tang served as the basis of the framework for the analyses that will be discussed in detail in the next chapter.

A final feature of the different forms of heteroglossic engagement that remains important to mention is their ‘gradability’ (Martin & White, 2005, p.136). This refers to the different forms that utterances within a single category may take. One such an example, taken from the category ‘pronounce’, would be the difference between the structures ‘I think’ and ‘I insist’. Whereas both of these statements tell the reader that the author holds a certain view to be true, there exists a marked difference in the intensity of these expressions. Similar structures could be put forward for each of the separate categories, where statements may convey a similar message with different notes of intensity.

1.2.2 Plagiarism

When students are taught to how handle secondary sources, the first and foremost rule they are introduced to is the zero tolerance policy on plagiarism. Of course, some students will still decide to copy others' works with the intent of passing it off as their own. Pecorari (2008) suggested that plagiarism be considered an 'act of language use' that in and of itself forms an important linguistic phenomenon (p.1). Randall's (2016) explored some of the motives students gave for their decision to copy text from other works. Most of them gave reasons – such as time management issues – that could be described as pragmatic violations of an ethical rule (Randall, 2016). However, other recent research has pointed out cases where plagiarism is not a clearly malevolent act. Pecorari (2008) revealed that the view of plagiarism as a conscious ethical violation is specific to Western society. In other cultures, imitation is traditionally encouraged, especially for beginning writers or for those not yet very familiar with the language.

Pecorari's work distinguished prototypical plagiarism – plagiarism committed with the intent to deceive – from patchwriting (2008, p.5). The term patchwriting is taken from Rebecca Howard (1999) and denotes the copying of certain text as a means to combat a personal inefficiency. This inefficiency may be due to a lack of proficiency in the language or familiarity with the text type. Importantly, patchwriting is done without any malicious intent. According to Pecorari, patchwriting is 'virtually inevitable as writers learn to produce texts within a new discourse community' (2008, p.5). This type of plagiarism is thus likely to be committed by students who are non-native users of English (Pecorari, 2008).

In her 2008 study, Pecorari used samples from student theses (i.e. wordcount approx. 3000 words per text). These were taken from the start of the documents as this is where citations occurred most frequently (Pecorari, 2008). The results revealed that many students had copied passages from other academic works, often in their entirety or with only very minor alterations, without adding quotation marks. Where it was possible to obtain access to the original source, it was found that on average 41% of the language used was not original to the student (Pecorari, 2008, p.63). Pecorari thus showed that language learning is intricately tied to plagiarism (i.e. patchwriting), that students are likely to use repeated language in their works, and that students lack transparency in terms of their source use (Pecorari, 2008).

1.2.3 Summary and research aim

Academic papers are always involved in a dialogue with both published and future researchers. Different parts of a research paper include different types of statements. The introduction generally features the most heteroglossic expressions. It is the authors' choice whether these expressions are dialogically expansive or contractive. Plagiarism occurs when information taken from other academic sources is not handled properly. Research has shown that L2 students form a high risk group for committing plagiarism, although this is not necessarily malignant. It can also be the result of cultural influences, or stem from an inability to successfully structure their own phrases (i.e. a lack of discourse competence).

Although it has been well established that academic writing should always be positioned within a broader discourse (Fryer et al, 2013; Tang, 2009), the dialogic aspect of academic writing is not currently placed at the forefront of academic writing courses. This is illustrated by the fact that aside from the section on correct referencing (i.e. the use of a specified style sheet) the dialogic aspect of academic writing is not represented in the TALPS. Section 1.2.1 illustrated the great variety in the ways an author can position their writing in relation to other academic works. The current study will question to which extent students employ the different dialogic stances in relation to the secondary academic sources used in their writing.

1.3: Differences in student and expert academic writing

This section will deal with previous research on the differences in academic language use between students and professional researchers. Due to the scope of this topic, most of this research has necessarily focused on only a single element of academic writing. In this section two such strands of research will be discussed in depth, these are the use of verbs and the use of linking adverbials. Verbs are an important as a carrier of modality, and research has moreover found that lexical verbs – specifically reporting verbs – take on an important role in academic writing (Granger & Paquot, 2009). When used correctly, linking adverbials are key markers of text relations which moreover attribute to the overall cohesion of the text. In addition they help reveal whether a student has the necessary vocabulary and understanding of the appropriate register. For these reasons, the use of verbs and of linking adverbials are solid indicators of academic writing proficiency.

1.3.1 Verbs

That academic language has a specific vocabulary has of course been an established fact for a long time. Several attempts have been made to construct wordlists that capture this vocabulary. One important example is the Academic Word List (AWL) that was produced by Coxhead (2000). However, academic writing is not limited to specialised vocabulary, and several common verbs are also used with a high frequency in scholarly texts. Some good examples of everyday verbs that feature often in academic writing are ‘to find’ and ‘to show’. This type of verb was not represented in the list created by Coxhead (2000). For this reason, Paquot (2007) compiled the English for Academic Purposes (EAP), a new wordlist that honoured frequently used ‘normal’ words in addition to more specialised academic terminology. The EAP includes words such as ‘aim’, ‘argue’, ‘suggest’ and ‘cause’ (Granger and Paquot, 2009, p.194).

Granger and Paquot (2009) explored differences in verb use by students and experts in an academic writing context. They argued that it can be challenging for non-native users of English to master the meaning of the large range of verbs that is used in academic writing, as well as to memorise the preferred lexico-grammatical category for each of them (e.g. ‘to suggest’ is generally combined with a subject, ‘to provide’ is used with an object, etc.). The importance of correct verb use should be stressed, as lexical verbs ‘enable writers to modulate their ideas and position their work in relation to other members of the discipline’ (Granger and Paquot, 2009, p.193). Verb use thus aids academic writers in taking a dialogic stance with their paper.

Using two corpora – one consisting of student essays and one of academic articles – Granger and Paquot (2009) created a wordlist for each corpus, and selected the 100 most frequently used verbs from each list for comparison. They found that certain verbs were significantly underused by students as compared to the experts. The majority of these verbs were academic verbs according to the EAP (Granger & Paquot, 2009). The top 20 most underused verbs are given in Table 2. Alternatively, some verbs were overused by the students in comparison to the experts. The most overused verb was ‘think’, and the rest of the list consisted mainly of similar verbs (i.e. those found in everyday speech). Examples included the verbs ‘say’ and ‘feel’ (Granger and Paquot, 2009, p.203)

1. describe	5. require	9. involve	13. include	17. appear
2. occur	6. contain	10. assume	14. record	18. attempt
3. note	7. obtain	11. derive	15. determine	19. demonstrate
4. suggest	8. identify	12. follow	16. remain	20. measure

Table 2: The 20 most underused verbs (given as lemmas) in academic writing by students in comparison to experts, taken from Granger and Paquot (2009, p.201).

The majority of the verbs in Table 2 fall into one of three categories. These are communication verbs (e.g. *suggest*), cognition verbs (e.g. *assume*) and relational verbs (e.g. *require*) (see Granger & Paquot, 2009, p.202 for a discussion of the role of these verb types in academic writing). The table gives the lemma for each verb, but the study by Granger and Paquot also looked at which specific verb forms were used. For example, this more in depth analysis showed that the forms ‘allowed’ and ‘conclude’ were overused, whereas ‘allowing’ and ‘concluded’ were underused (Granger & Paquot, 2009, p.204).

The corpora used in the study by Granger and Paquot consisted of academic writing samples collected from a database that includes great number of different academic disciplines. As mentioned in Section 1.1.3, academic writing is not a uniform genre. It would be interesting to examine whether the differences between the vocabulary and writing styles of these disciplines would influence the findings in a side by side comparison. The current study will therefore use the findings from Granger and Paquot (2009) in order to establish whether a student corpus consisting solely of history papers will show the same pattern of underused verbs in comparison to their professional counterparts.

1.3.2 Linking adverbials

Shaw (2009) presented the findings of his research relating to the difference in language use of students and experts, with a specific focus on their respective use of linking adverbials. He found that linking adverbials in an academic context most often function either as a means to place two or more text elements in apposition, or to denote result-inference (Shaw, 2009). The most frequently used linking adverbials – and their preferred function – in academic texts are: ‘however’ (contrast/concession), ‘thus’ (result-inference), ‘therefore’ (result-inference), ‘for example’ (apposition), and ‘then’ (result-inference) (Shaw, 2009, pp.215-216). Shaw’s (2009) research not only considered the preferred academic linking adverbials and their functions, but also looked at adverbial density (i.e. how often linking adverbials were used in a text, and how much they contributed to the total word count) and lastly also at the location they were most likely to occupy in a clause (i.e. whether they were found in initial, medial, or final position).

Previous research had found that learners of English typically used fewer linking adverbials, and placed those they used in the initial position more often than experts would do (Shaw, 2009; also Aarts & Granger, 1998). According to Shaw the use of linking adverbials is a writing skill that needs to be actively taught to L2 learners as an effective means of improving their writing. Longitudinal studies have indeed demonstrated the feasibility of achieving such a positive effect (Shaw, 2009). In his own paper, Shaw compared the use of linking adverbials by first-year students and professional researchers in literary studies. The results showed that students in fact used more linking adverbials than their professional counterparts did (Shaw, 2009). Both groups illustrated a preference for the five high frequency adverbials mentioned earlier, as well as for ‘yet’ and ‘indeed’. Academics were more likely to use sentence initial ‘and’ and ‘but’ than the students (Shaw, 2009, pp.226-227). This is most reasonably explained by university prescriptivism (i.e. students are taught to refrain from using coordinating conjunctions as adverbials) (Shaw, 2009). It is still interesting to note that the academics chose to take some liberties with this general rule. Lastly, some linking adverbials common to spoken English – ‘again’, though’, and ‘therefore’ – were used considerably more frequently by the students (Shaw, 2009, pp.226-227).

None of these differences, however, were particularly profound. In his final conclusion, Shaw argued that the reason for the higher density of linking adverbials found in the student texts was the fact that the propositions they linked were simpler and shorter than in the experts’ writing. The students were able to employ the same adverbials in much the same way, but they had less developed arguments and thus the adverbials made up a larger proportion of the total number of words (Shaw, 2009).

As mentioned above, the corpora used by Shaw (2009) consisted of student papers and professional research articles in literary studies. Like history, this is a humanities subject which relies on argumentation rather than experimentation. Because it may be assumed that the use of linking adverbials will be very similar in papers written by students of literature and students of history, the current paper will not include its own corpus based analysis of this aspect of academic writing. Rather, it will posit that the findings of Shaw’s study may be considered representational for humanities students in general. It would be useful to research whether these findings can be reproduced in relation to a corpus of hard science texts, however this lies outside the scope of the current study. For now it remains important to note that Shaw’s findings will be used to form the general hypothesis that at the lexical level, there are no real differences between the writing of university history students and of professional historians.

2. Methods

2.1 Corpora

In order to compare student academic writing with expert academic writing, two corpora were compiled, consisting of theses and articles by university students and professional researchers respectively. The following sections provide the selection and exclusion criteria per corpus.

2.1.1 Student corpus

For the student corpus, ten history theses at bachelor level were selected from Dutch universities. The decision to study this discipline in particular was based on the following arguments. Firstly, previous research has established that the soft sciences of which history forms a part are linguistically interesting, as their merit is primarily determined by the argumentation and narrative structure employed by the author. A further recommendation for the suitability of this discipline as a study subject comes from the fact that the BA is taught in Dutch, therefore the students whose texts were used were very unlikely to be L1 speakers of English. As has been discussed in the previous chapter, due to the universality of academic English, the majority of current (student) researchers is required to present their findings in a language other than their L1. It is thus the study of non-native users of English that will provide insight into this majority group. As the history BA's official language is Dutch, it follows naturally that the students could have written their final works in Dutch had they preferred to do so. The vast majority of BA theses in history in the Netherlands are in fact written in Dutch. However, the students whose works make up this corpus have each chosen to write in English, implying that they perceived their mastery of this L2 to be of a sufficiently high level not to be detrimental to their final grade. This was beneficial to the current study as it allowed the comparison of a range of academic writing skills at a high level, without the impairment of insufficient English grammar knowledge.

All the texts in the first corpus were taken from BA students. The BA thesis marks the end of a three-year university degree, and as such students who successfully complete this work are expected to have acquired decent academic writing skills as well as adequate knowledge of professional literature in their chosen field (this as opposed to first year university students, for example). Although MA students are arguably even closer in skill level to professional academics, the relatively short length of the BA thesis compared to the MA thesis allowed the inclusion of a greater number of participants while still ensuring the feasibility of the qualitative aspects of the methodology (which will be explained in more detail in later sections of this chapter).

The theses were obtained from the repositories of two Dutch universities: the Radboud University and Utrecht University. The theses were taken from two different universities to minimize the chance that the outcomes of this research proved unique to one specific institution. Following this argument, more universities were not included only because their repositories did not contain any theses that met the inclusion criteria (i.e. theses that were both written in English and not under embargo).

The theses were selected at random in the following manner: in the repository of the university website, the first English thesis on every other page was selected (if the page did not contain any thesis in English, or if the title in question was under embargo, then the process was continued on the next page). This process was repeated until ten theses were collected. All selected BA theses were written between 2015 and 2020.

2.1.2 Expert corpus

The second corpus contained 10 research articles from prominent historical journals. The main criterion for the selection of these journals was that they were influential journals with contributions by researchers working at the top of their field. As has been explained in the introduction, in the absence of a clearly demarcated list of characteristics for the ideal academic paper, the aim of this corpus is to represent the very highest level of academic history writing.

Four papers were selected from the *American historical review* and the same number of texts was selected from the *Journal of global history*. These are both examples of highly influential journals which cover numerous historical subjects. Their publications are representative for the high quality research that the students would ideally have been exposed to during their degree, and which they would have been aiming to emulate while constructing their own research. In addition, one paper was taken from the *Journal of Hellenic* studies and the *Journal of the History of Sexuality* each. As their names imply, these are two examples of more specialized journals. As mentioned in chapter 2.1.3, there are many different areas of study within the discipline of history. Similarly, historical journals are often concerned only with a specific time period, a certain topic, or a preferred methodology. The inclusion of the two articles from the abovementioned journals serves the primary purpose of including range of topics also present in the students' corpus, in order to represent the particular linguistic features of each of them.

The papers were randomized as follows: the first article in the first issue from a year between 2015-2020 was used. Again, the papers that were selected to make up the expert corpus were not necessarily written by native speakers. As mentioned in the introduction, this study's primary interest is how the academic language of students compares to that of professional historians. It therefore does not rely on a comparison between L1 and L2 speakers.

The final exclusion criteria was the absence of a (primarily) argumentative methodology. In practice, this meant that one paper that met the randomization criteria (i.e. it was the first article in the first issue of the year) was excluded from the corpus, as it relied almost exclusively on numerical data, and did not contain an argumentative interpretation of textual or visual primary sources. This methodology was not considered comparable to the BA theses, and thus this paper was not included in the final corpus.

2.2 Editing

The selected texts were made to undergo some basic editing before being added to the corpora. In order to ensure comparable text formats, the included theses and articles were edited to have a range from introduction to conclusion (i.e. the introduction and conclusion were both included in the final version, as well as all text between these two sections). In practice this entailed the deletion of such sections as abstracts, words of thanks, lists of abbreviations and 'about the author' sections. These were not included in the analyses for the dual reason that they usually do not contain the academic language of interest to this study, and the fact that they were not universal to all papers, and might thus have hindered fair comparisons. Bibliographies were not included in the corpora, although for the texts that included this section the number of secondary sources cited in each text was noted to be used in later analyses.

The final student corpus consisted of 85508 words (excluding footnotes). The final expert corpus contained 101021 words (excluding footnotes). The number of references and the number of sources that are mentioned in the analyses refer only to the secondary (scholarly) sources used (as defined in the introduction to chapter 2.2). Primary sources were not included in any of the analyses that make up this study. The dialogic interaction with secondary literature is a linguistic feature of academic texts that is comparable between multiple authors, but primary sources are specific to each topic, and the way they are used will primarily highlight the (historical) research skills of the individual author, rather than their academic writing prowess. This type of source would not be comparable across disciplines in any case.

It was a relatively simple task to filter out the primary (historical) sources, as in most cases their exclusion could reliably be based on their year of publication. Of course this approach could not hold true for modern history topics. Whenever there existed any uncertainty a close reading of text – where necessary combined with a search for the original source – revealed beyond any doubt whether the author meant to refer to a primary or a secondary source. For most works in the corpora, the topic under discussion was sufficiently removed in time from the scholarly publications it cited that the first approach sufficed for almost all references. In the case of one thesis in particular the second methodology had to be applied for the majority of its sources, as some of the primary sources under discussion were in fact scholarly publications (these were taken from 20th century medical journals). For this text, a closer look at the context of the references combined with the content of the medical articles was necessary to ensure that no primary sources were included in the final reference count that was used for the analyses that are explained in the next section.

2.3 Basic analyses

The first basic quantitative analyses were related to the interaction with secondary literature. In order to shed light on their use, the number of secondary sources (taken from the bibliographies where possible, and counted from the text itself in other cases) was determined, as well as the number of times a single source was referenced throughout the text. For each article the 'density' of references to secondary literature was determined by calculating the number of sources used per 1000 words. This was done in order to compare source use in texts of different length. In addition, the number of references per source was determined (i.e. the number of times each publication was referred to).

2.4 Interaction with secondary sources

2.4.1 The role of the historical debate

The study performed by Fryer et al (2013) showed that most heteroglossic statements were to be found in the introduction and discussion sections of academic papers. For history papers in particular, interaction with other academic sources primarily takes place at the level of the historical debate. This section of the paper forms part of the introduction and is a standard element in historical research. It contains past and current views on the topic by other historians and outlines the debate that the paper in question is attributing to. More or less explicitly this section thus forces the dialogue between the author(s) and their colleagues, because the standard practice implies that all research is defined in relation to other academic literature. Because the historical debate is a section that occurs in each historical discipline it is comparable across all texts.

As mentioned, the historical debate forms part of the introduction of the paper. For this reason concordance searches were used to explore the first ten references to secondary sources in each text, using an adaptation of the dialogic framework discussed in section 2.2.1 as well as a subsequent qualitative analyses of these findings. The fact that exactly ten sources were considered for each text moreover ensured that works with more frequent references were not overrepresented in the final analysis.

2.4.2 Dialogic categories

The first ten scholarly sources used in each text were thus analysed according to the abovementioned framework. Every in-text reference to a secondary source was considered and marked according to the dialogically contractive and dialogically expansive categories established by Tang (2009, for descriptions see section 2.2.1). Not all of the possible dialogical stances that were introduced in section 2.2.1 were present in the texts. The references could be divided into eight separate categories. Although all categories discussed below are based on the work by Tang and others as it was presented in the previous chapter, some of them have been adapted to more accurately portray the findings of the current study. Below follows a brief description of each category.

In those cases where there was no mention of the author or title of the secondary source in the text, but where the reference was simply part of a statement introducing new information to the reader, this was marked as the first category (I). These references could be viewed as a means provided to the reader to *fact-check* the information given in the text, or as an indication where to find more information on a given topic. Potentially these instances should also be considered as providing proof that the information is retrieved from a legitimate (i.e. professional and peer-reviewed) source.

References were marked as *acknowledgements* when the author referred to a secondary source in a neutral manner. Common phrases used in the texts included 'X stated that … ' and 'X argued that …'. The term acknowledgement is owed to Martin and White (2005) , however, as this description (i.e. a neutral reference) was still rather a broad category, I chose to make further distinctions between three separate subtypes of acknowledgement. The examples above would fall under the category singular acknowledgement (II), where there is a neutral reference in the text to a single secondary work (this work may be written by more than one author, the main point is that the text refers only to a single viewpoint). The next category thus becomes the plural acknowledgement (III). Common phrase structures found to meet this category included ‘some scholars consider that…’ and ‘most historians agree that…’. It is worth noting here that Tang (2009) also provided a category *hearsay* – identified by phrases such as ‘some say…’ - that could possibly be confused with the type of structure put forward here. However, although it contains the structure ‘some say’ the expressions found in the current corpora were not considered as hearsay because of the explicit referencing in the footnotes. The ‘some’ has as its purpose to serve the plurality of the sources, not to refrain from naming specifics. The last form of neutral dialogism that was counted was what I have coined negative acknowledgement (IV). Phrase structures that fell under this particular subcategory would often take the form of ‘no scholars have argued that…’, backed up with a footnote providing one or more references. For the import of this last category I will refer to the Results and discussion chapter.

References were placed in the fifth category, *endorsed* (V), if they contained positive, non-neutral references to secondary material. The category *pronounced* (VI) also refers to references that contained a positive standpoint, but that included a more personal stance. This occurred either in a first person form (i.e. I/we) or by making the thesis itself the subject of the sentence (e.g. ‘this thesis recognises the use of …’). Note that these categories are dialogically contractive, as opposed to the acknowledgements above which are dialogically expansive.

Another category that was recognised was *distancing* (VII), which involved a negative stance toward the secondary source, usually presented as a disagreement with its contents. A clear example of a phrase structure involving distancing would be ‘X claims that…’. I chose to also include sentence structures such as ‘X stated that..’ if these were immediately followed by a contradiction (e.g. ‘X stated this, however this article will demonstrate that actually … ’) .

The final category (VIII) was for those sources that did not fit in any of the other categories. In practice this happened with two references, both taken from the same thesis. The structure used in both cases was ‘one could argue that…’. Again, this structure did not fit Tang’s (2009) hearsay category since a reference was in fact given, nor did the context allow for an interpretation of distance. Because this structure was only found twice, and was moreover specific to the writing of a single author, it was not considered as relevant to the discussion of general characteristic of student academic writing, and will not feature further in the discussion sections.

Table 3 shows all the categories with their description and example phrase structures. It should be noted that the assignment of dialogic categories, although based on clearly defined descriptions, is a qualitative approach that necessarily relies on the researcher's interpretation of context. The example mentioned above in relation to the category *distanced* serves to illustrate this fact.

ti	Type	Description	Example
I.	Fact-check	Fact presented with no in-text mention of author	'The head of a diocese was a *vicarius* or the so-called vicar, who became the superior of the governor' (Thesis 2, p.3, see appendix)
II.	Acknowledged (singular)	Neutral reference to a single source	'According to Daniel Byman, there are also positive sides of manipulation, especially identity manipulation' (Thesis 7, p.4, see appendix)
III	Acknowledged (plural)	Neutral reference to two or more sources	'… several imperial scholars, such as Holden Furber and P. J. Marshall, stressed economic factors in the process of British expansion' (Article 2, p.90, see appendix)
IV.	Acknowledged (negative)	Neutral reference given to establish a niche in the research subject	'These two late-eighteenth-century events are not unknown to historians, but until now, no one has pieced together their interconnected stories to bring their significance to British imperialism into full view' (Article 2, p.87, see appendix)
V.	Endorsed	Dialogically contractive, positive, impersonal reference	'For a rich synthetic account of Fetu's history, see Yann Deffontaine' (Article 4, p.31, see appendix)
VI.	Pronounced	Dialogically contractive, positive, personal reference	'This thesis turns to the historian James Cortada, who regards the concept of information as an umbrella term for a collection of facts which describe a thing, place, person or events'(Thesis 10, p.4, see appendix)
VII.	Distanced	Dialogically expansive, negative reference	'Green's method however greatly differs from the proposed approach of this thesis, seeing as Green claims to use a historical approach, however he defers from this approach' (Thesis 4, p.4, see appendix)
VIII.	Other	Reference that did not clearly fit any of the dialogic categories	'One could argue that the most important aspect of the Nazi experiments -namely the ethics of using these experiments- has been extensively covered by several experts' (Thesis 4, p.5, see appendix)

Table 3: The eight dialogic categories with descriptions and examples, adapted from Tang (2009; based on the work by Martin & White, 2005)

2.5 Verb use

For the next part of the analyses, Wordsmith tools was used to compile wordlists for both corpora. The student-wordlist contained 7475 different words, and the expert-wordlist contained 11084 different words.

Using these wordlists, a search was conducted for the frequency (both absolute and relative) of the top 20 most underused verbs by students according to the study by Granger and Paquot (2009). This was done in order to ascertain whether the results from the Granger and Paquot study were representational for a corpus limited to a single discipline (i.e. history) or whether this influenced the results. The following verbs were used in this analysis: ‘describe’, ‘occur’, ‘note’, ‘suggest’, ‘require’, ‘contain’, ‘obtain’, ‘identify’, ‘involve’, ‘assume’, ‘derive’, ‘follow’, ‘include’, ‘record’, ‘determine’, ‘remain’, ‘appear’, ‘attempt’, ‘demonstrate’, ‘measure’ (taken from Granger & Paquot, 2009, p.201).

As Granger and Paquot (2009) had noted that there was a difference not only in frequency, but also in the verb forms used, this analysis not only considered lemmas, but also the specific forms of every verb. This analysis thus included a separate count for each of the 20 verbs mentioned above for the infinitive, past participle, third person present tense, and the gerund (e.g. ‘describe’, ‘described’, ‘describes’, and ‘describing’).

Since some of the lemmas were potentially also used as nouns (e.g. ‘the attempt’, ‘the record’), this part of the analysis also searched for the frequency of the related nouns of other words (e.g. ‘the description’, ‘the suggestion’) in order to establish whether these might reasonably be considered to interfere with the other searches. These nouns were of course not included in the lemma count.

Granger and Paquot (2009) also suggested that certain words were overused by student writers, especially everyday speech words. This study therefore also looked at the frequency of the words ‘say’, ‘think’, ‘feel’ and ‘believe’. The results, as well as a qualitative analysis of their potential insight in student academic writing, are given in the next chapter.

2.6 Analysis of research questions

The final part of this study was aimed at providing insight into the manner in which students and expert writers position their own work. As has been discussed in Section 1.2.3, although it is a generally accepted fact that all academic writing is dialogic in nature and should be positioned within a broader discourse, this is not always reflected in the way students are taught. The current study attempts to use a close reading technique in order to provide insights in the way student writers position their own work within the academic discourse of their topic, and how this may differ from the approach taken by professional writers.

The second aim of this method was to discover whether there were identifiable traces of an authoritative stance to be found in the student theses. Previous research has linked authority to successful academic writing (Tang, 2009), however, it is difficult to reduce this aspect of writing to a single linguistic feature (see Section 1.1.2).

In practice, the analysis involved a qualitative discussion of the way the main research question of each paper was expressed. The analysis specifically focused on the introduction of the research topic because previous research has illustrated that this section is the most likely to involve heteroglossic engagement of related academic discourse. Moreover, this section of the paper was considered to have a universal structure for every historical subdiscipline, and as such was comparable across all the works in the corpora.

The results of this analysis are given in the form of a description and a discussion per corpus, followed by a comparison between student and expert practices. Although this was a purely qualitative analysis of the texts, patterns were identified when it seemed appropriate to do so, and these will also be presented in the following chapter.

3. Results and discussion

In this chapter are given the results of the analyses described in the methods sections. Because this study involved multiple analyses, it was considered helpful provide the reader with context for each of its findings. For this reason, all the results that are presented in this chapter are directly followed by a short discussion of their connection to the theoretical framework, their relation to findings from previous literature and their probable implications. For sections that included both quantitative and qualitative methods, the quantitative results will be given first. In addition, there will be a second, separate discussion section at the end of this chapter which has as its primary goal to tie all these findings together in order to answer the main research question(s).

3.1 Use of secondary material

3.1.1 Use of secondary sources

Tables 4 and 5 show the quantitative data concerning source use by the students and the experts respectively. They provide the total number of secondary sources that were used in each text (i.e. how many different scholarly articles it cited). However, in order to compare texts of different length, they also provide the average number of sources per 1,000 words. In addition, the tables give the number of references, and the number of references per 1,000 words, for each text. Lastly, the tables show the average number of references per source (i.e. the number of times a secondary source was likely to be cited in a footnote).

	Wordcount	No. secondary sources	No. sources per 1000 words	No. in-text references	No. references per 1000 words	No. references per source
Thesis 1	5,861	20	3.4	14	2.4	0.7
Thesis 2	6,093	31	5.1	61	10.0	2.0
Thesis 3	9,933	24	2.4	144	14.5	6.0
Thesis 4	10,073	38	3.8	84	8.3	2.2
Thesis 5	7,631	102	13.4	59	7.7	0.6
Thesis 6	7,354	46	6.3	85	11.6	1.8
Thesis 7	6,546	17	2.6	49	7.5	2.9
Thesis 8	11,882	24	2.1	57	4.8	2.4
Thesis 9	9,466	27	2.9	95	10.0	3.5
Thesis 10	10,669	24	2.3	58	5.4	2.4
Total	85,508	353		706		
Average	8,551	35	4.4 (4)	71	8.2 (8)	2.5 (3)

Table 4: The results of the quantitative analyses of source use by students.

	Wordcount	No. Secondary sources	No. Sources per 1000 words	No. In-text references	No. references per 1000 words	No. References per source
Article 1	10,536	77	7.3	50	4.7	0.6
Article 2	8,868	66	7.4	65	7.3	1.0
Article 3	10,432	110	10.5	43	4.1	0.4
Article 4	11,634	97	8.3	62	5.3	0.6
Article 5	9,129	67	7.3	73	8.0	1.1
Article 6	10,572	44	4.2	110	10.4	2.5
Article 7	9,303	56	6.0	78	8.4	1.4
Article 8	10,360	58	5.6	59	5.7	1.0
Article 9	9,661	44	4.6	43	4.5	1.0
Article 10	10,526	44	4.2	29	2.8	0.7
Total	101,021	663		612		
Average	10,102	66	6.5 (7)	61	6.1 (6)	1.0 (1)

Table 5: The results of the quantitative analyses of source use by experts.

In order compare the averages of the two groups the three key measurements are also given side by side in Figure 2.

Figure 2: The average number of sources per 1,000 words, the average number of references per 1,000 words, and the average number of times each source was referred to in the text given for the two corpora (i.e. student and expert).

The comparison of these averages indicated that the experts tended to use more references to secondary sources in their texts. On average they referenced seven other articles for every 1,000 words of their own texts, whereas the students referenced only four secondary sources for every 1,000 words they wrote. The second inference that can be drawn from this data sample is that experts are unlikely to refer to a given secondary source more than once per article. Not only is the average number of in-text references per secondary source 1 (see Figure 2), but when looking at the separate instances it can be seen that there was in fact only one article that referenced it sources in more than one location in the text (i.e. Article 6, see Table 5). The student corpus, on the other hand, revealed that the majority of thesis writers re-used a single secondary article as reference on more than one occasion in their texts. This holds true for all but two of the cases considered in this corpus. On average the students referred to each source three times per article.

This correlation (i.e. less sources per 1,000 words and more uses per source) is not surprising. Students' works had a reference on average every 125 words (1,000/8) and experts every 166 words (1,000/6). Of course, in practice source use was more concentrated in certain sections of the text. As indicated by previous research, the introduction (specifically the historical debate outline) featured the most frequent references to scholarly sources.

The quantitative results indicated a difference in source use between the two corpora. Probable explanations are reserved until the next section, which gives the findings of analysis of the dialogic categories used with the references. A final note to this part of the chapter has to include the possibility that (unwitting) plagiarism may have influenced the results. Previous studies have shown that students who are non-native speakers of English are a high risk group for patchwriting in particular. Because there was no effective way to test whether plagiarism had indeed been used in the student (or expert) corpus, the current study had to rely on the universities' efforts to combat this practice. It is assumed that these efforts were successful.

3.1.2 Interaction with scholarly sources

The first ten references to secondary sources in each text were marked according to their dialogic category. The results of this effort are presented below in Table 6.

Type	Theses	Articles
Fact presented with no in-text mention of author	58	20
Acknowledged (singular)	25	38
Acknowledged (plural)	8	17
Acknowledged (negative)	0	5
Endorsed	3	12
Pronounced	2	6
Distanced	2	2
Other	2	0

Table 6: The dialogic categories of the first ten references to secondary sources found in each text.

Students proved more likely to use sources for facts, such as descriptions of events. Such examples may be said to be of a monoglossic nature in spite of their reference to another text. It would be easy to conclude from this that the experts did not deem it necessary to include a lot of (background) factual information in their articles. However, a reading of the professional articles revealed that they do in fact contain similar narrative information. One such example is given here:

> ‘During the second half of the eighteenth century, a global network of private British investors provided extensive credit to Indian rulers and Chinese merchants.
> When the borrowers defaulted on their loans, their British creditors sought to recoup their investments through the use of force. They overthrew the East India Company’s government in Madras with the help of the Company’s own army, and they tried to bully the Chinese government in Canton through the might of the Royal Navy.’
> (Article 2, see appendix)

For the information in this paragraph no source was given in the text. The seeming paradox here is that should the author presume that the reader is familiar with these facts, they would not be included, yet if it is new information a source would have to be provided.

The large difference in the amount of phrase structures that match the first dialogic category can at least be partly explained by the second finding, namely that experts are more likely to present information as the opinion of a certain author or authors. All forms of acknowledgement were found more often in the articles than in the student writing. In addition it was found that when references to secondary sources were introduced with the statement ‘it has been argued that’ or a similar neutral structure, the theses would generally provide only a single source in the footnote. The articles, on the other hand, would name multiple sources, on average between 4-8.

The corpora contained multiple instances of references for statements such as ‘most historians agree that’ or ‘some scholars argue that’. Interestingly, several of the articles also gave numerous sources as proof that ‘no scholars had argued this yet’. Linguistically this is an incredibly interesting practice. It points to the fact that interaction with other sources forms such a primary element of the historical debate that even when there has been no precedent for researching a certain topic, authors will feel the need to provide references as proof of this fact. This is related to the academic practice of using secondary sources to back up statements made in the introduction of any scholarly work, and also to the necessity of defending a chosen research topic. In five cases of the expert corpus, the solution to the lack of citable sources was found in the referencing of renowned works on related topics for the information that they did *not* contain. Of course this practice requires in depth knowledge of both current and previous literature on a subject, which may explain why it was not attempted by any of the student writers.

The dialogic analysis revealed a further difference in the stance taken by the student and professional authors. The experts proved more likely to include a positive value judgement when introducing a secondary source (i.e. they had more statements that fit into the *pronounce* and *endorse* categories). This was true for both impersonal (endorse) and personal (pronounce) expressions. It is of course possible that students did not feel confident to pass judgement – positive or otherwise – on the quality of academic sources. Perhaps the more likely explanation for these findings is that for students a statements such as ‘X states that…’ are already conferring credibility on the statement simply because it was made by a published historian. This hypothesis is related to the finding that students are more likely to present information from scholarly articles as facts without an in-text mention of its author. The professionals on the other hand would be more likely to distinguish the merits of a specific view, especially since they proved to include a greater number of references in general (see figure 2).

3.1.3 Summary of source use

The findings discussed at the start of this chapter revealed that students on average used less secondary sources than the experts did. In addition, they referenced the ones they did use more often than the professional historians. The second analysis showed that students primarily referred to secondary sources when presenting information in the text as factual. They could be said to use the articles they referenced as a ‘textbook’, to provide information on the topic they are researching. This corresponds with the finding that even explicit in-text mentions of other works were almost exclusively *acknowledged* and almost in all cases only a single source was given. The students exhibited a strong tendency to present everything found in secondary sources as true.

The experts referenced a greater range of secondary source material, and tended to only address them in a single footnote per source. Sources were provided to show the historical debate, to point out alternative viewpoints to the one presented in the text, to refer the reader to a source of more in-depth study of a related – but not at present discussed – topic, or even to show that other sources failed to provide information on a certain subject. The experts were also more likely to give a value judgment of the sources they referenced.

Several reasons are likely to be the cause of these different approaches to secondary literature. Experts are better aware of the historiography of their topic. Students who have only just chosen a topic for this particular thesis cannot be expected to be aware of all past and current views, and most of them would not have had sufficient knowledge of what has already been written to be able to unhesitatingly make claims to the extent of ‘this has never been written before’. Experts make more assumptions about the knowledge of their readers. This is part of the interaction with their audience, as they know their papers will be read by other experts in their field (i.e. a select audience with detailed knowledge of the subject) making many explanations superfluous. One of the probable reasons students gave more citations for narrative (background) information is that these ideas would have been unknown to themselves not so very long ago.

Although students are taught to interact with current literature, it should be remembered that most of their exposure to secondary sources has been in the form of articles to be studied and in many cases probably even crammed for exams. They are not as yet likely to consider themselves on an equal level with the professional authors they are citing. The professional writers are naturally more willing to nuance or counteract claims made in other articles. The final conclusion of these analyses is that the students showed a tendency to approach scholarly sources as containing facts they can build their own argument on, whereas the experts treated them as merely another narrative that can and should be challenged.

3.2 Verb use

The next analysis concerned verb use by students and expert writers. The lemmas of the 20 most underused academic verbs by students were considered by means of the wordlists compiled in Wordsmith tools. The results are given below in Figure 3, which shows all the use of all 20 verbs in terms of their contribution to the total word count of the two corpora. In addition, the use of the everyday verbs ‘say’, ‘think’, ‘feel’ and ‘believe’ were compared across the two corpora. These results can be found in Figure 4. As has already been mentioned in the methods section, the possibility that the noun forms of certain verbs could interfere with the results was entertained, and this was accounted for by considering whether the frequency of the use of nouns related to the other verbs differed between the two groups. More specifically, because there was no practical way to filter out the instances where such words as ‘attempt’ or ‘record’ were cases of noun use without having to use a concord search for each individual word, the frequency of nouns such as ‘description’ and ‘suggestion’ were compared between the two groups in order to ascertain the likelihood that nouns counted as verbs influenced the results of this analysis. As the frequency of these latter nouns was found to be similar in both groups, it was assumed that any nouns that counted toward the final lemma findings for the verbs ‘attempt’, ‘note’ and ‘record’ would also have exhibited comparable frequencies in both corpora and therefore did not skew the results.

Neither the over- nor the underuse of the verbs tested for reached the same levels as the previous study by Granger and Paquot (2009), suggesting that the verb use of third year history students more closely resembles that of professional academic writers. Some lemmas, such as ‘demonstrate’, ‘identify’, and ‘record’ were indeed used markedly more by the experts (see Figure 3). For example, the lemma ‘demonstrate’ made up 0.05% of the expert corpus but only 0.02% of the student corpus. The lemma ‘include’ formed 0.09% of the expert corpus, compared to 0.06% of the student corpus.

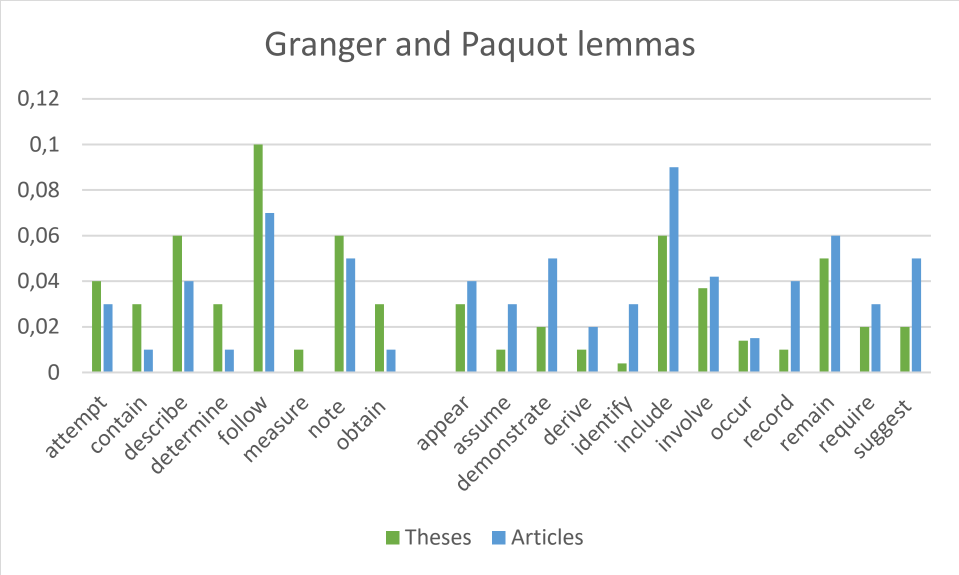

Figure 3: The lemma count for the 20 academic verbs analysed in the student and expert works, based on the study by Granger and Paquot (2009), given as percentage of the total corpus word count.

However, several of the lemmas were in fact used more often by the students than they were by the experts. These are easily recognizable in Figure 3, where they have been separated on the left side of the graph. Moreover, although the informal verbs were used more frequently by the students, they also appeared in the experts' texts (see Figure 4). A possible explanation for the difference between these findings and the outcome of the Granger and Paquot study involves the specific narrative structure of academic history writing. This hypothesis is certainly strengthened by the finding that some verb lemmas that more naturally belong to the hard sciences, most notably 'measure', are rarely used in either corpus. It would perhaps be logical to assume that since successful history writing is directly linked to argument structure, students of this discipline need a more profound knowledge of the English language than do those studying sciences based on experimental research. Further research involving concordance searches and a directly comparable corpus of hard science papers would be necessary to be able to confirm this idea.

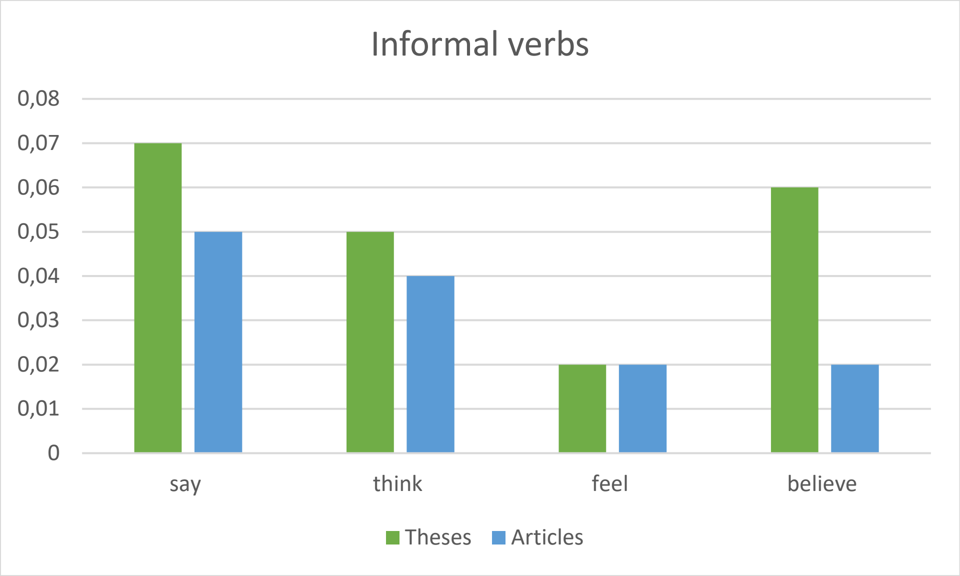

Figure 4: The lemma count for the informal verbs ‘say’, ‘think’, ‘feel’ and ‘believe’ in the student and expert works, given as percentage of the total word count for each corpus.

Overall it can be concluded that final year history students seem capable of employing the same range of academic verbs as experts. This is important among other reasons for the discussion on dialogism. The gradeability of dialogistic expressions is made possible largely through the use of different verbs. In addition, successful use of lexical verbs allows academic writers to position their text in relation to other academic works (Granger and Paquot, 2009). Therefore the knowledge of a broad vocabulary is a prerequisite for the participation in the academic dialogue.

The findings discussed in this chapter are helpful because they strengthen the hypothesis that the differences between student and professional academic writing are not found at the word level. This is in line with the conclusion drawn by Shaw (2009) in relation to linking adverbials. Shaw similarly illustrated that university students have to a large extent begun to master the vocabulary necessary for successful academic writing. The differences that still exist between their writing and that of professional authors are not found at a lexical level, but result from those elements of discourse competence that relate to the structuring of the argumentation of the text as a whole.

3.3 Self-positioning

This section describes some of the differences that came to light during the qualitative analysis of the introduction of the texts.

3.3.1 Student corpus

As discussed in the introduction to Section 2.2, academic texts are always embedded in a larger discourse. Hoe students approach this reality is determined for a large part by the way they choose to position their paper in relation to other academic material, as has been discussed at the start of this chapter from a dialogic framework. In addition, a qualitative analyses of the theses was conducted in order to reveal how the students referred to their own work in its introduction. This analysis concerned a more explicit form of self-positioning, and was directly specifically at the language students used to present their chosen research question, as well as the manner in which they attempted to legitimise their research. As far as possible these introductions were coded for similarities. The same approach was used for the expert corpus. The rest of this chapter gives a description of the findings as well as a comparison between the corpora.

The following excerpt (A) was taken from one of the theses that made up the student corpus:

> A: '**This study aims** to critically compare the current theses, analysing their individual strengths and weaknesses, in order to create a clear overview which will facilitate further research on this subject. In addition, **this essay attempts** to bridge the gap between the existing theses. The central question of this essay is: <u>what were the reasons for integrating the province of Mauretania Tingitana in the newly created diocese of Hispaniae</u>?' (Thesis 2, p.4, emphasis added, see appendix)

There are three elements in this particular excerpt that beg attention. The first – emphasised in bold – is the fact that linguistic agency is conferred to the thesis itself through the use the verbs 'aim' and 'attempt'. The second important aspect to notice is that this section refers to previous research, and does this in a way to imply that the current work will build on the findings of these earlier studies. The goal of this thesis is at least in part to 'bridge the gap' between existing hypotheses. Lastly, it is important to point out that the research subject of the thesis is framed as an explicit question (underlined).

With these initial remarks in mind it is easy to recognise similar points in the following excerpt (B), taken from another thesis:

> B: **'This thesis hopes** to shed more light on a mystery that already grabbed the attention of Kenneth Stow in 1989. Also, this has the potential to reveal Philip's involvement in the case. To be able to solve this mystery, **this thesis will seek to answer** the following questions: <u>what underlying rationale might have pushed pope Boniface VIII to make an unexpected decision, considering the protection given to Jews by other popes, on the matter of the Parisian host desecration case in 1295? And: what underlying rationale might have been at the root of Philip IV's decisions concerning the host desecration case?'</u> (Thesis 3, p.9, emphasis added, see appendix)

Here, again, the author has chosen to give agency to the thesis, illustrated by the expressions given in bold. In this case, the thesis is assigned a 'hope', which of course is very similar to the 'aim' of the thesis in the previous example. The second comparable aspect is the reference to previous literature on the subject. In this particular example the author informed the reader that they would build on the work done by Kenneth Stow (a professional historian). Just as in the first quotation, this paragraph also introduced the research topic in the form of a question.

Below is given one additional example to strengthen the idea that these form recurrent structures in student writing. Once again the text included a thesis with linguistic agency and an explicit research question.

> C: **'This thesis** therefore **wants** to examine how the British parliament arrived at the decision to intervene based on the news and reports available to them. The central question is therefore formulated as follows: <u>How has the British decision to intervene in the Taiping Civil War been influenced by the information in newspapers and official reports available to the British Parliament between 1853 and 1864?'</u> (Thesis 10, p.3, emphasis added, see appendix)

Not every single thesis included each of these elements. This is only logical as these are not prerequisites of university theses. However, it was discovered that every single student made at least one explicit reference to their own work in the introduction of their topic. Out of the ten works in that made up the student corpus, four chose to confer agency to their thesis (i.e. used an active verb such as ‘aim’, ‘hope’ or ‘want’). An additional two used the structure ‘this thesis will answer …’ as a definitive guarantee to the reader (in comparison to example A where it is merely presented as a goal). It is possible that students employed such structures in order to convey authority in their writing. As mentioned in Section 2.1.2, Hyland (2002) established that students are less likely than professional researchers to write in the first person. Conferring agency to their work may have been a means to emulate the first person singular viewpoint.

The explicit research question was present in the majority of cases: it was used in as many as seven out of the ten papers under consideration. This element might have been used to benefit the dialogue with the specific tutor-reader (see Tang, 2009). This hypothesis is based in part on the comparison with the expert corpus, which will be discussed in the next section. The last recurrent aspect was that students gladly ‘stood on the shoulders of giants’. This refers to the positive references to previous studies that were included in the introduction to the research topic (seen most clearly in excerpt B). It was found that four of the thesis explicitly admitted to build their paper on the research done by earlier historians. This also matches earlier findings (see Section 3.1.3) that students were more likely to assume what was written in other scholarly sources to be true and to use them as a basis for their own research.

3.3.2 Expert corpus

In order to illustrate that the points mentioned above are specific to student writing it is necessary to compare the theses to the articles written by professional historians. Besides the usual challenges of qualitative research, these two corpora proved almost incomparable in the sense that no similar structures were to be found in the articles. That is to say that the articles did not contain any explicit references to themselves, nor did any of them feature an explicitly phrased research question. Below are given excerpts from four of the articles in order to illustrate through which means these introduced their topic instead.

D: ‘A number of historians of U.S. missions have noted that renaming was a wide-spread form of fundraising, but until now the phenomenon has not been explored in depth.’ (Article 1, p.816, see appendix)

E: ‘These two late-eighteenth-century events are not unknown to historians, but until now, no one has pieced together their interconnected stories to bring their significance to British imperialism into full view …’ (Article 2, p.87, see appendix)

F: ‘The migration of the Field Code was a central event in Anglo-American legal history, but no historian has traced the extensive borrowings of the Field text in detail or recognized the political furor that greeted the code outside New York’ (Article 3, p.134 see appendix)

G: ‘By considering both the Native American and Russian material, bound together by the more loquacious western European middlemen, a more complete picture of the history of early modern sassafras is visible than when research is limited to the western European sources alone.’ (Article 5, p.5, see appendix)

Although they are not made explicit, it is fairly easy to construct the research questions that must underlie the articles to which the above given excerpts belong. For example, for excerpt D the main goal of the study was to discover the role that renaming (i.e. choosing a name to baptise a child) played in the success of fundraising activities for the spread of Christianity in the United States. This topic was introduced through the statement that ‘until now’ research has not yet addressed it. The author of excerpt E used the same expression (i.e. ‘until now’), and both the other examples may be said to follow this pattern.

Essentially, all the articles in the expert corpus followed a similar pattern. The historians all started their work with a short introduction to its topic, more often than not this consisted of a small anecdote or quotation in combination with some broad historical knowledge on the subject. This was then followed by the topic’s relevance for study in a general sense. Subsequently, they defended their own addition to the discourse by noting that there existed a certain aspect that had not been studied before. This is related to the negative reference discussed previously. In short, they identified the niche they had found and filled. There was a notable absence of any explicit research questions. The professional researchers did not give their ‘aims’ and ‘hopes’ for their study, nor did they makes guarantees as to its outcomes and the way these would benefit the field. They merely remarked what aspect has thus far remained unnoticed and explained its relevance to the discourse.

3.3.3 Summary and conclusions

Clearly the students and the experts made different (linguistic) choices with regard to the way they introduced their research topic. The students built their papers on earlier research, and generally gave some overview of what had been written before. However, they failed to identify a truly new aspect worthy of being studied. In this sense their work does not add anything to their topics discourse, at the very least they themselves do not appear to expect to make this contribution. The majority of the student theses framed their research topic as an explicit question (i.e. a full sentence ending in question mark). They also showed a tendency to explicitly refer to their own work ('this thesis', 'this study') in the first person, and to express their intentions for their work by giving agency to the text itself. This allowed the thesis to take on an explicit goal, which was most commonly phrased as a 'hope' or 'aim'. Some students also chose to foreshadow the outcome of the research, seemingly in an attempt to use this as legitimization for the research question.

Although the two corpora differed markedly from each other, the texts within a single corpus contained comparable linguistic structures. Section 1.2.1 discussed how all academic texts are placed in a broader discourse. In addition, Section 1.1.2 introduced the idea that (authoritative) voice in writing is based on socially available repertoires. Textual features can thus be 'an expression not only of the individual but also the surrounding social relations' (Matsuda & Tardy, 2007, p.237). This may form part of the reason why the structures employed by the two groups of authors were so similar within each group.

Another tentative hypothesis explaining the choices made by the student-authors is that in a university setting the importance of having a good research question is frequently emphasised. Potentially, students may interpret this as meaning that their research question must be made explicit (and thus easily recognisable) in their work. A similar argument was offered by Stapleton (2002) in relation to voice in academic writing. Stapleton argued that the prevalence of voice as a topic of academic study led to teachers placing a lot of stress on the role of voice in writing. In turn, students would then be led to believe voice to be 'critically important' to their writing, and would become more concerned with expressing voice than with creating good content (p.187). Stapleton perhaps took this concern to an extreme, but his views do offer a potential explanation for the findings in Section 3.3.1. It may be the case that students are imbued with a sense of the importance of a well formulated research topic, but are unaware of the correct means to implement this knowledge in their writing. The explicit research question may have also been initially introduced to accommodate the dialogue with the specific tutor-reader (see Tang, 2009). Future research on this topic is recommended. One aspect that would be of particular interest for future study is whether there exists any correlation between explicitly phrased research questions and (perceived) university success.

4. Conclusion

University students need to attain a very high level of discourse competence in English in order to successfully finish their degree (Bruce, 2008; Pot & Weideman, 2015). Several discourse elements, such as correct grammar and register, are applicable to any academic work. However, there are also several aspects of academic writing that differ according to an author's chosen academic discipline. A key example would be the vocabulary that is specific to a given field of study. In addition, there are some factors that take on a greater role in certain academic disciplines. Although some form of argumentation is found in every academic article, the argument structure of a history paper will contribute more to the overall value of the work than the argument structure of a biomedical paper. This is due to the fact that history is a humanities subject which relies solely on an argumentative and narrative structure to present its conclusions, whereas a hard science paper generally relies on experimental research methods.

It is part of the challenge for university students to identify and master the conventions of their chosen academic discipline. Previous corpus-based research has generally approached student academic writing as a universal genre. The current study challenged this approach, and instead focused on a single academic discipline, in order to do justice to the specific text type of a single academic subject. The study found that at the lexical level, the academic writing skills of final year history BA students proved very similar to that of expert academic writers. Specifically, it was discovered that in comparison to an earlier study relying on a corpus of several academic disciplines, the history students did not underuse academic verbs or overuse everyday speech verbs to the same extent. In order to state with certainty that this is due to the specifics of academic history writing as a genre, further comparative research between corpora of different disciplines is needed.

The student and expert corpus closely resembled each other at the lexical level. Still, it was found that there exist some remarkable differences between the student theses and the professional articles. Through the use of a dialogic framework, it was discovered that students approached their secondary sources as texts containing factual information. Professional researchers, on the other hand, considered other academic works as texts relaying a specific viewpoint, one which they did not necessarily agree with themselves. On average students cited fewer secondary sources. The sources they used were usually referred to multiple times throughout the text. Experts, however, would generally refer to a source in only a single footnote. These differences suggest that there is much to be gained by incorporating good practices for source use in university courses. Currently, the standard university curriculum focuses mainly on avoiding plagiarism (i.e. incorrect source use), but does not always address the variety of ways writers can interact with other academic works. Additional research, both in terms of a larger corpus and a possible longitudinal study design, are necessary to establish whether introducing students to the concepts of dialogism in writing is helpful in improving student academic writing.

This study also found distinct ways in which students and professional historians introduced their research topic to their readers. The majority of the theses phrased an explicit research question. In addition, several of them gave agency to their work (i.e. assigned a ‘hope’ or ‘aim’ to their thesis). None of the articles included an explicit research question. Instead, the articles clearly introduced a niche (i.e. a specific element that had thus far not been studied). In order to define this niche, several historian employed a *negative reference*. They cited one or more sources in relation to information that was not in fact present anywhere. For students who want to improve their academic writing, emulating these two practices – establishing a niche and interacting with other academic literature – would likely be beneficial. As the current study is limited both in terms of corpus size and focus (i.e. it only considered history students at Dutch universities), it is impossible based on these findings to state whether the patterns identified here are a general attribute of student writing, or whether they are specific to this study group. Based on the results of this study, it is recommended that future research relating to the typicality of student writing focus on the way in which the authors position themselves and their work, rather than on elements at the lexical level.

Bibliography

Aarts, J. & Granger, S. (1998). Tag sequences in learner corpora: a key to interlanguage grammar and discourse. In S. Granger (Ed.) *Learner English on computer* (pp. 132-141) London, England: Longman.

Bakhtin, M. & Holquist, M. (1982). *Dialogic imagination: four essays.* Austin, U.S. : university of Texas press.

Biber, D & Gray, B. (2016). *Grammatical complexity in academic English: linguistic change in writing*. Cambridge, England: Cambridge University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow, England: Longman.

Bruce, I. (2008). *Academic writing and genre: a systematic analysis.* London, England: Continuum.

Butler, G. (2010). The design of a postgraduate test of academic literacy: Accommodating student and supervisor perceptions. *Southern African linguistics and applied language studies*, 27 (3), 291-300. DOI: 10.2989/SALALS.2009.27.3.7.941

Conrad, S. (1996). *Academic discourse in two disciplines: professional writing and student development in biology and history*. ProQuest Dissertations Publishing.

Council of Europe. (2001). *Common European Framework for Reference of Languages*. Accessed from https://rm.coe.int/16802fc1bf.

Coxhead, A. (2000). A new academic word list. *TESOL quarterly*, 34, 213-238. DOI: 10.2307/3587951

Fryer, D., Bech, K & Andersen, G. (2013). Exploring the dialogism of academic discourse: heteroglossic engagement in medical research articles. *English corpus linguistics: variation in time, space and genre,* 77, 183-207. DOI: 10.1163/9789401209403_011

Granger, S & Paquot, M. (2009). Lexical verbs in academic discourse: a corpus-driven study of learner use. In M. Charles, D. Pecorari & S. Hunston (Eds.). *Academic writing at the interface of corpus and discourse* (pp. 193-214). London, England: Continuum.

Hyland, K. (2002). Authority and invisibility: authorial identity in academic writing. *Journal of pragmatics*, 34, 1091-1112. DOI: 10.1016/S0378-2166(02)00035-8

Martin, J.R and White, P.R.R. (2005). *The language of evaluation: appraisal in English.* Basingstoke, England: Palgrave Macmillan.

Matsuda, P.K. (2001). Voice in Japanese written discourse: implications for second language writing. *Journal of second language writing,* 10(1-2), 35-53. DOI: 10.1016/S1060-3743(00)00036-9

Matsuda, P.K. & Tardy, C.M. (2007). Voice in academic writing: the rhetorical construction of author identity in blind manuscript review. *English for specific purposes*, 26(2), 235-249. DOI: 10.1016/j.esp.2006.10.001

Patterson, R. & Weideman, A. (2013). The typicality of academic discourse and its relevance for constructs of academic literacy. *Journal for language teaching*, 47(1), 107–123.

Pecorari, D. (2010). *Academic writing and plagiarism: a linguistic analysis.* London, England: Continuum.

Pot, A & Weideman, A. (2015). Diagnosing academic language ability: an analysis of the Test of Academic Literacy for Postgraduate Students. *Language matters: studies in the languages of Africa*, 46(1), 22-43. DOI: 10.1080/10228195.2014.986665

Primary sources: a research guide. Healey library, University of Massachusetts Boston. Accessed from https://umb.libguides.com/PrimarySources/secondary.

Primary vs secondary literature in the biomedical sciences. Drexel university libraries. Accessed from https://libguides.library.drexel.edu/biomed-literature-types.

Prior, P. (2001). Voices in text, mind, and society: sociohistoric accounts of discourse acquisition and use. *Journal of second language writing,* 10(1-2), 55-81. DOI: 10.1016/S1060-3743(00)00037-0

Randall, M. (2016). *Pragmatic plagiarism: Authorship, profit and power.* Toronto, Canada: University of Toronto Press.

Robinson, N. J. (2017, June 27). Academic language and the problem of meaninglessness. *Current Affairs*, accessed from https://www.currentaffairs.org/2017/07/academic-language-and-the-problem-of-meaninglessness.

Salazar, D. (2014). *Lexical bundles in native and non-native scientific writing: applying a corpus-based study to language teaching.* Amsterdam, Netherlands: John Benjamins Publishing Company.

Scheuler, S. (2014). Primary and Secondary Sources in History: A Primer for Undergraduates, Challenges for Librarians. *The Reference librarian*, 55 (2), 163-167. DOI: 10.1080/02763877.2014.881274

Shaw, P. (2009). Linking adverbials in student and professional writing in literary studies: what makes writing mature. In M. Charles, D. Pecorari & S. Hunston (Eds.). *Academic writing at the interface of corpus and discourse* (pp. 215-235). London, England: Continuum.

Stapleton, P. (2002). Critiquing voice as a viable pedagogical tool in L2 writing: returning the spotlight to ideas. *Journal of second language writing*, 11(3), 177-190. DOI: 10.1016/S1060-3743(02)00070-X

Staudigl, M. (2013). Towards a relational phenomenology of violence. *Human studies*, 36, 43-66. DOI: 10.1007/s10746-013-9269-x

Tang, R. (2009). A dialogic account of authority in academic writing. In M. Charles, D. Pecorari & S. Hunston (Eds.). *Academic writing at the interface of corpus and discourse* (pp. 170-190). London, England: Continuum.

Weideman, A. (2003). Assessing and developing academic literacy. *Per linguam*, 19(1), 55–65.

[Pages 58-60 under Embargo]