12-17-2020

A constructional analysis of written academic English as a Lingua Franca: The case of unedited and edited research writing

Selahattin Yilmaz

A CONSTRUCTIONAL ANALYSIS OF WRITTEN ACADEMIC ENGLISH AS A LINGUA FRANCA: THE CASE OF UNEDITED AND EDITED RESEARCH WRITING

by

SELAHATTIN YILMAZ

Under the Direction of Ute Römer, PhD

ABSTRACT

Recent English as a Lingua Franca (ELF) research has started to document the general characteristics of academic writing by international scholars from different linguistic (similectal) and disciplinary backgrounds, as well as the complex patterns of variation that shape these characteristics. However, not only is this line of research in its infancy, studies are also generally small-scale, resulting in a limited understanding of the complexities of ELF academic writing. Adopting a Construction Grammar (CxG) approach, this study aims to comprehensively examine the distinctive constructions, here defined as multi-word sequences with discourse-functional properties in three corpora of academic writing from 50 disciplines in the social and natural sciences. These corpora are: (1) an unedited ELF corpus of 140 texts by non-native scholars from nine different similects (L1s) in the social and natural sciences; (2) an edited ELF corpus that matches the similects and disciplines in the unedited ELF corpus; and (3) an edited English as a Native Language (ENL) corpus that matches the disciplines in the ELF corpora. A range of corpus analytic methods were used to identify distinctive constructions around key function words. The target constructions were first analyzed in terms of forms and functions across corpora. Then, similectal

and disciplinary variation in the use of these constructions were investigated via robust multivariate statistical tests.

The findings support previous research in that ELF writers use conventional features of academic writing such as nominals, and passives more often than ENL writers. Their use of constructions including low-frequency prepositions with nominal complements, predicative adjectives, and determiners, however, show remarkable disciplinary and similectal variation. As a result, conventionality and simplification are argued to be two important "universals" of ELF, that is, general features across different disciplines and similects, regarding high-frequency constructions. However, the high degree of similectal and disciplinary variation in the use of the low-frequency constructions points to the complexity inherent in ELF as the only "universal" characterizing their usage. Implications of the study for future research on the diversity of written scientific communication and pedagogy are discussed in light of the findings.

INDEX WORDS: English as a Lingua Franca, English for Academic Purposes, Academic writing, Second language writing, Construction Grammar, Corpus Linguistics

A CONSTRUCTIONAL ANALYSIS OF WRITTEN ACADEMIC ENGLISH AS A LINGUA FRANCA: THE CASE OF UNEDITED AND EDITED RESEARCH WRITING

by

SELAHATTIN YILMAZ

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2020

A CONSTRUCTIONAL ANALYSIS OF WRITTEN ACADEMIC ENGLISH AS A LINGUA FRANCA: THE CASE OF UNEDITED AND EDITED RESEARCH WRITING

by

SELAHATTIN YILMAZ

Committee Chair: Ute Römer

Committee: Diane Belcher

Stephanie Lindemann

Anna Mauranen

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

December 2020

DEDICATION

This dissertation is dedicated to all fellow additional language users of English around the world. Without the tremendous influence of your unique voices on global scientific communication, this study would not be possible.

ACKNOWLEDGEMENTS

It has been over a decade since I started my undergraduate program in Teaching English as a Foreign Language in the fall of 2009. I have been continuously involved in the field since then. So, this dissertation, by marking the start of my second decade as an applied linguist, represents a crucial milestone in my professional career. Not only am I extremely happy to have completed it, but I also owe so much to so many inspirational mentors and peers I have met along the way. They have helped me grow in ways that would otherwise be impossible. Some of the early influences on my professional identity are my undergraduate professors in the English Language Teaching Department at Marmara University, Drs. Selma Karabınar, Zeynep Cumhur Çamlıbel Acar, Özlem Kaşlıoğlu, and Kamile Hamiloğlu, and especially Drs. Pınar Ersin and Gökçe Kurt Tiftik, whose hard work and success as doctoral candidates and lecturers were inspirational and instrumental in my decision to pursue my studies in the Foreign Language Education department at Boğaziçi University. During my MA studies, I had the chance to work as a research assistant in the English Language Teaching Department at Yildiz Technical University with amazing academics Drs. Suzan Kavanoz, Gülru Yüksel, Ahmet Başal, Celile Ökten, and with colleagues who also became good friends, Merve Celen, Ferda İlerten, Dr. Emrah Özcan, and Dr. Gülümser Efeoğlu. I consider myself lucky to have worked in a supportive work environment that boosted my confidence in my academic skills.

As I close a chapter in my life, I, of course, gratefully remember those who helped me open it in the first place, my MA thesis advisors Drs. Yasemin Bayyurt and Leyla Marti. Along with teaching the courses that laid the foundations of my current work, they also enabled me to work on my first major research project and encouraged me to do the exchange program in the Department of Applied Linguistics and ESL at Georgia State University. Taking their advice, I started my studies at GSU, and later had the chance to meet my dear doctoral advisor Dr. Ute Römer. She has been a true inspiration and a role model since the day we met, and hence, an excellent professor, advisor, and collaborator. Her patient guidance at every step of the way has helped me mature my academic work tremendously and bring this dissertation to completion. I also thank my dissertation committee members, Drs. Diane Belcher, Stephanie

Lindemann, and Anna Mauranen, whose scholarship in general and feedback on my work in particular, facilitated the perfecting of this dissertation greatly. I am also thankful for the Graduate School at GSU that supported me with the provost's dissertation fellowship during spring 2020. Having the time to focus on my dissertation project significantly improved the quality of my work.

During my time at GSU, I have learned from and worked with very supportive professors and exemplary experts in the field, such as Drs. Eric Friginal, YouJin Kim, Scott Crossley, John Murphy, and Viviana Cortes. My colleagues and good friends from the program, who have also been the best partners in crime one could ask for, also deserve mention for keeping me sane during this four-year-long hard work and intensive learning experience. So, Analynn Bustamante, Jillian Siehlmann, Sally Ren, Julie Carver, Laura Tan, Eunice Nam, Rurik Tywoniw, Jessica Lian, Katia Monteiro, Sanghee Kang, thank you for being there!

Aside from work, I am indebted to my dear parents Nilgün and Asaf, my aunt Belkıs, and of course, beloved friends Ebru Kıran Baylan and Andaç Özgür for their unconditional support and love for as long as I have known them. I am lucky to have you all by my side despite whatever there is for me in the cards.

I am also amazed by my own ambition and diligence in doing this despite myself. Once a colleague said that the struggles of a doctoral student could not be that bad as they could be considered as growing pains, and now I agree with them. The push for bettering myself every day has not always been pleasant, but it has definitely made a huge impact on me developing professionally and personally. This is very the reason why I am glad to have held on to it and reached the final steps of this process.

Lastly, I am happy that more and more people are vocal about the challenges of being in graduate school, including researchers, as well as institutions that continue to explore how the graduate school experience can be improved for everyone. In an era of constant and fast-paced change, we cannot afford to following the academic traditions that do not serve us well. Just like how many people helped me stay on track, I also hope to help improve others' experiences in my work.

So, thank you!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CH	Chinese
CxG	Construction Grammar
Cx	Construction
CW	Content word
CZ	Czech
DP_{norm}	Normed deviation of proportions
EAP	English for Academic Purposes
EC	Expanding Circle
ELF	English as a Lingua Franca
EN	English
ENL	English as a Native Language
ERPP	English for Research Publication Purposes
FI	Finnish
FR	French
FW	Function word
H_{norm}	Normalized entropy
IC	Inner Circle
IT	Italian
KFW	Key function word
KW	Keyword
L1	First language
L2	Second language
LE	Learner Englishes
NS	Native speaker

NNS	Non-native speaker
OC	Outer Circle
PO	Portuguese
SCI	Sciences
SP	Spanish
SSH	Social sciences and humanities
SW	Swedish
UB	Usage-based
WE	World Englishes
WrELFA	Written English as a Lingua Franca in Academic Settings

1 INTRODUCTION

The unprecedented sociopolitical and economic power of the English language has made it the indisputable lingua franca of international communication primarily used by non-native speakers (NNSs) from a wide variety of different linguistic and cultural backgrounds. This has led to an increasing acceptance of the highly dynamic, variable and pluralistic view of the language in recent years (Bardi & Muresan, 2014; Björkman, 2013; Crystal, 2003; Jenkins, 2017b; Mauranen, 2017; O'Neil, 2018; Phillipson, 1992; Seidlhofer, 2004). In light of these changes, English as a Lingua Franca (ELF) research has aimed to examine the complex nature and the prevalent global role of English. Ever since its preliminary stages in the early 2000s, ELF research has explored the use of English mainly by its non-native speakers from different angles, including phonology (Jenkins, 2000, 2002), lexicogrammar (Pitzl, 2012, 2018; Pitzl, Breiteneder, & Klimpfinger, 2008; Prodromou, 2008; Ranta, 2009; Seidlhofer, 2004, 2009a), and pragmatics (Cogo, 2009; Cogo & Dewey, 2012; Firth, 1990, 1996; House, 2013). Most of the early research employed general spoken corpora such as the Vienna-Oxford International Corpus of English (VOICE, 2009), as well as other smaller corpora that the researchers compiled in local ELF contexts (e.g., Cogo & Dewey, 2006, 2012; Prodromou, 2008).

Meanwhile, as the increasing internationalization of universities has also reinforced the role of English as the academic lingua franca (Mauranen & Ranta, 2008), a new line of ELF research was launched. For example, Mauranen's (2005, 2006b, 2007, 2009, 2012) work on the first academic ELF corpus, the English as a Lingua Franca in Academic Settings, henceforth ELFA (Mauranen & Ranta, 2008) corpus, has started to uncover the workings of spoken ELF in academic contexts. Björkman's (2008a, 2008b, 2009) studies on academic ELF interactions documented another series of comprehensive analyses of spoken academic ELF. As is the case with non-academic spoken ELF research, these studies generally pointed to the distinct and variable nature of ELF talk, what Firth (2009) called the "lingua franca factor," the similarities of ELF to English as a native language (ENL), and emphasized the communicative effectiveness of ELF speakers in most contexts regardless of the unconventionality observed in the structure and uses of their language.

Until recently, however, most ELF research focused primarily on the spoken mode; thus, the early definitions of the phenomenon relied on spoken communication as the core of ELF (O'Neil, 2018). Several researchers attributed the initial focus on spoken ELF to the higher degree of variation, naturalness of production (little to no editing), and lower levels of conformity to the primarily codified, that is, primarily written, standard language ideologies (e.g., Hynninen & Solin, 2017; Mauranen, 2003, 2012; Seidlhofer, 2001, 2004, 2017; Seidlhofer, Breiteneder, & Pitzl, 2006). However, these very differences between spoken and written language, and the distinct research traditions between the two modalities highlight the need for more research on written ELF in order to gain a deeper and more comprehensive understanding of the workings of ELF (Flowerdew, 2019; Horner, 2017; Marlina & Xu, 2018; Ranta, 2017). This need becomes even more critical when it comes to academia where the written text is the dominant medium (Ingvarsdóttir & Arnbjörnsdóttir, 2013; Mauranen, 2011, 2012; Mauranen, Pérez-Llantada, & Swales, 2010b). The research article, in particular, is considered the major written academic genre for sharing research findings with the international academic community, as well as for academic hiring and promotion decisions (Berkenkotter & Huckin, 1995; Mauranen, 2012). Moreover, as in all other ELF domains, the global expansion of academia stems primarily from the fast increase in the number of non-native writers and readers of English (Jenkins, 2017b; O'Neil, 2018). Hence, it is safe to argue that Mauranen's (2012) conceptualization of ELF in the title of her book as "[a]cademic English shaped by non-native speakers," despite initially pertaining to speaking, holds true for academic writing, as well.

Although there has recently been an increase of ELF perspectives in academic writing research, certain essential issues related to the characteristics and role of written academic ELF are yet to be adequately addressed. For instance, on the one hand, research shows that ELF writers face many challenges in the publishing world because they are encouraged to follow native speaker norms that many journal guidelines still favor today (Flowerdew, 2000, 2008; Hanauer, Sheridan, & Englander, 2019; McKinley & Rose, 2018; Mur-Dueñas, 2013). On the other hand, the non-nativelike language use of these scholars has been documented to be increasingly visible in unedited (Bondi & Morelli, 2018; Carey,

2013; Lafuente-Millán, 2018; Mauranen, 2016; Mur-Dueñas, 2018a; Murillo, 2018; Rowley-Jolivet, 2017; Vetchinnikova, 2015, 2017), published (Bardi & Muresan, 2014; Farley, 2018; Lorés-Sanz, 2016; Martinez, 2018; Mur-Dueñas, 2015, 2016, 2018b; Pérez-Llantada, 2013; Rozycki & Johnson, 2013; Tribble, 2017, 2019), and self-published research writing such as academic blogs (Luzón, 2018; Vetchinnikova, 2017). Moreover, there are seemingly opposing views on written academic ELF as an area of study. For instance, Rowley-Jolivet (2017) points to the potential challenges in the acceptance of non-nativelike features of ELF in academic writing unlike speaking, and Tribble (2017, 2019) argues against Jenkins' (2014) call for academic ELF as a new research paradigm in English for Academic Purposes (EAP) research. Several methodological limitations of these studies, such as using small corpora, investigating only a limited number of linguistic features, and the tendency to adopt a prescriptive approach based on standard language ideologies and researcher intuition, further problematize the generalizability of these findings.

One important strength of this recent research on written ELF, however, is the growing interest in usage-based (UB) arguments for the study of ELF that were first made in theoretical discussions and spoken ELF research (Alptekin, 2011, 2013; Baird, Baker, & Kitazawa, 2014; Hall, 2012, 2017; Mauranen, 2012). Different from spoken ELF research where there were critiques that usage data from corpora and quantitative analyses can offer only incomplete accounts of ELF (e.g., Seidlhofer, 2009c, 2011), such approaches have been mainstream in written academic ELF research. Several studies (Carey, 2013; Vetchinnikova, 2015, 2017) have also explained the use of written academic ELF directly through UB approaches. However, in line with the methodological constraints stated above, a comprehensive UB analysis of written academic ELF is still lacking, which is evident in recent research that factors in the complex nature of ELF through rich data and analytic methods (e.g., Deshors, 2019; Laitinen, 2018; Laitinen, Levin, & Lakaw, 2018). Therefore, in order to gain a better understanding of the linguistic characteristics of written academic ELF, this dissertation study aims to investigate the use of dominant constructions, defined by Goldberg (1995, 2003, 2006) as pairings of form and function at any level of complexity from morphemes to multi-word phraseological patterns, across unedited and edited

(published) ELF research writing by scholars from different first language (L1) backgrounds, in comparison to edited (published) ENL academic writing through an inductive, corpus-based analysis.

The following chapters of this dissertation are organized in a way that provides connections between previous research, and the design and findings of the current study from multiple perspectives. The second, third, and fourth chapters summarize the relevant literature by drawing on research on ELF, Construction Grammar (CxG), EAP, and second language acquisition (SLA). The sixth chapter provides the conceptual foundation, aim, and research questions of the dissertation, followed by the seventh chapter that describes its methodology, including the data and analysis employed. Following are three chapters (Chapters 7-9) that summarize the findings of the study and address the three main research questions that the study is driven by. Lastly, the tenth chapter concludes the dissertation with an overview of the findings and their implications for our understanding of ELF research and L2 academic writing pedagogy.

2 WRITTEN ACADEMIC ELF AS AN INTERDISCIPLINARY PHENOMENON

The complexity inherent in ELF is commonly agreed-upon today thanks to researchers who have long examined it by drawing from many different areas of linguistics and adopting a wide variety of theoretical and methodological perspectives, which makes ELF a fundamentally interdisciplinary phenomenon (e.g., Mauranen, 2012; Mauranen, 2018b). However, except for a few recent accounts (Horner, 2017; Mauranen, 2018a), such interdisciplinary connections have not been carefully examined, particularly for written academic ELF. Thus, with the aim of situating the present study, this chapter provides a theoretical overview of the current state of ELF research, by first defining ELF, and then discussing the contribution of different disciplines to its study with particular attention to written academic ELF, including standard language ideologies and native speaker (NS) norms, World Englishes (WE), second language acquisition (SLA), variation across modes and discourses, English for Academic Purposes (EAP), and a usage-based (UB) approach to language. Although not exhaustive, the list of areas covered here can be said to include a considerable number of research traditions that continue to have close links with ELF research.

2.1 Defining English as a Lingua Franca (ELF)

Early conceptualizations of lingua francas were predominantly about the functionality of any contact language used as a means of communication between people who do not share the same first language (Gnutzmann, 2000; Jordan & Fuller, 1975; Samarin, 1987). Following this research tradition, early research by Firth (1990, 1996) and House (1999) can be considered some of the earliest attempts of studying English as the contact language used among its non-native speakers from different first language backgrounds. However, it is the emphasis on English as the powerful international medium used for a great variety of sophisticated uses in speaking as well as writing that later led researchers to shift from a general lingua franca English perspective to conceptualizing it as only an immediate and simplified contact language including a mix of the first languages of its speakers (Björkman, 2013; Gnutzmann, 2000). As ELF research became more widespread and documented more of the unique nature of ELF

uses, it also became more inclusive by accounting for the presence of native speakers in interactions with non-native speakers (Seidlhofer, 2004, 2011).

While early conversations on ELF included debates as to whether it should be seen as a stable English variety (e.g., Mollin, 2006), the accumulating evidence from ELF research led Firth (2009) to suggest the existence of a "lingua franca factor," that is, ELF is a linguistic phenomenon that can be primarily characterized by its variability and complexity recurrently observed in the analyzed interactions. Around the same time as Firth formulated his argument, Seidlhofer (2009c) also emphasized the importance of an increasing trend in ELF research where an analysis of ELF forms was not adequate without also examining the accompanying functions of these forms. Similarly, Mortensen's (2013) definition of ELF as "the use of English in a lingua franca language scenario" (p. 36) called for research to focus more on the functions and contexts of ELF use to capture its complexity in more detail.

Another relevant development was Mauranen's (2012) introduction of two concepts describing the complexity of ELF at different levels: similects and second order language contact. Similects refer to the similarities in the way ELF is used by those of the same first language background even though they do not have a dialectal community where they use English among themselves (Mauranen, 2012, p. 28), while second-order language contact highlights the fact that ELF is used by a rather heterogeneous group of speakers from different similectal backgrounds (Mauranen, 2012, p. 29). The latter, together with Firth's (2009) lingua franca factor, is a serious attempt to explain the complexity inherent in ELF interactions primarily due to the encounters of ELF users with bi/multilingual linguistic repertoires that are very distinct from one another.

A recent theoretical discussion echoing these arguments on the complexity of ELF as an outcome of the linguacultural diversity of the speakers was initiated by Jenkins (2015), who critiqued the scope of ELF definitions and research by being almost always about the English language use of ELF speakers. She argued that this is largely influenced by the monolingual bias in SLA research. Therefore, she

suggested a new conceptualization, an organic follow-up to the current ELF research acknowledging its hybridity: English as a Multilingua Franca (henceforth, EMF). Jenkins' (2017b) paper also provides an up-to-date description of ELF which reads as follows:

> […] ELF by definition, involves the use of English among speakers who have different first languages, most of whom are themselves multilingual in that only a small minority of people who use English in intercultural communication are native English speakers, and a still smaller subsection of this minority are monolingual. Thus, ELF is by definition a multilingual phenomenon, and would not exist at all if it were not. (p. 3)

She continues to argue that considering multilingualism as the essence of ELF could help scholars make connections with research in multilingualism, critical approaches to SLA, and intercultural communication to better understand ELF. Thus, the complexity inherent in ELF due to multilingualism can be more robustly explained by, for example, factoring in the first language background of ELF users, which is a common research design in L2 research (Flowerdew, 2019). Accordingly, in addition to studies such as Mauranen (2012) and Martinez (2018) where similects were only controlling variables to ensure that the findings apply to a somewhat homogenous group of ELF speakers, analysis of individual ELF similects has also recently been taken up by ELF researchers (Bardi & Muresan, 2014; Farley, 2018; Luzón, 2018; Mur-Dueñas, 2018b; Zapletalová, 2018).

Following these developments in ELF research, this dissertation also approaches ELF as a complex and hybrid multilingual phenomenon influenced by the substantial linguistic and sociocultural diversity of its speakers' resources, which is reinforced and co-constructed through interactions of any sort. While ELF is acknowledged here to be the use of English around the world regardless of the nativeness status of its speakers, this study compares the use of written academic English by non-native and native speakersin order to examine the potentially distinct constructional profiles of the texts produced by both groups. Hence, for ease of reference, ELF in this study refers to the academic writing of NNS scholars, while non-ELF refers to that of NS scholars from Inner Circle countries. In terms of the approach to analyzing this diversity, this study first follows Mauranen (2012) to arrive at generalizable

findings on ELF in the form of form-function pairings occurring across similectal groups. Next, the potential influence of similectal and disciplinary influences on the usage of certain patterns are discussed across different author backgrounds.

2.2 Standard Language Ideologies, Native Speaker (NS) Norms, and ELF

Ever since its beginning phases, ELF research has problematized the common practice of judging the language use of ELF speakers against that of ENL speakers who are generally considered as the "norm providers" of English (Kachru, 1987, p. 254). This view can be said to be fueled largely by standard language ideologies, that is, a set of beliefs and attitudes in favor of an ideal, uniform and institutionalized variety, through which we perceive and produce languages in our contexts (Giles & Marlow, 2011; Lippi-Green, 2012). Reinforced through educational institutions and prestigious reference sources such as grammar books and dictionaries, standard language ideologies influence the views of majorities on how languages should be used (Milroy, 2006; Milroy & Milroy, 2012). For this type of influence, an essential factor is the codification of the standard in writing, the norms of which also impact speaking (Lillis & McKinney, 2013; Melchers, Shaw, & Sundkvist, 2019; Milroy & Milroy, 2012). As for the English language, the long tradition of standardization in British and American English places these two varieties as the primary sources of these normative views (Melchers et al., 2019, p. 32).

While standard language ideologies are quite influential in creating and reinforcing an idealized standard, this uniform perception of a prestigious common language is at odds with the reality of language variation (Crowley, 2003). Furthermore, even varieties closer to this ideal are not as uniform as they are perceived to be. As shown by Biber, Johansson, Leech, Conrad, and Finegan (1999) in their comprehensive reference book of English grammar, the so-called Standard English, that is, the "dialectal variety that has been codified in dictionaries, grammars, and usage handbooks" (p. 19), includes a high degree of variation across different contexts of use. Hence, by taking into account the complexity and great degree of variation in ELF across the globe, researchers critique the dominance of ENL standards for judging the efficiency of ELF because they are not as relevant for ELF as they are for ENL contexts

(Jenkins, 2007, 2014; Mauranen, 2012; Seidlhofer, 2004, 2011). Besides, the heterogeneity of ELF is likely to lead to even greater diversity and a relatively fast pace of change in the way ELF is used (Mauranen, 2012, 2015, 2018b), thus further contradicting the standard language ideologies based on the uniformity of standardized varieties in monolingual speech communities (Milroy & Milroy, 2012).

Language ideology research often prioritizes speech, and there is a tendency to approach writing as a more neutral and relatively stable mode (Canagarajah, 2012; Lillis, 2013; Lippi-Green, 2012; Turner, 2018). While arguments about the role of NS norms in ELF research initially focused on speaking, more recent research has expanded these discussions to the practices of EAP writing and scientific publishing (e.g., Hynninen & Kuteeva, 2017; Hynninen & Solin, 2017; Jenkins, 2014; Solin & Hynninen, 2018). Added to that, despite the face of academic writing changing due to the ever-increasing diversification of scientific communities all around the world (e.g., Canagarajah, 2012; Turner, 2018) and features not typical of NS norms becoming more visible in published writing (Rozycki & Johnson, 2013), NS norms, mainly of American and British English, remain to be the rather unquestioned standards of research writing (Casanave, 2017; McKinley & Rose, 2018; O'Neil, 2018). However, conforming to such standard written norms is not congruent with today's world where most authors publishing academically in English are non-native speakers, and there is a wide variety of social and individual factors further diversifying academic writing (Horner, 2017; Mauranen, 2018a). Hence, this dissertation study, by exploring the distinctive constructions of written academic ELF in unedited and edited research articles, aims to uncover the extent to which NS ideologies and norms play a role in shaping academic English today.

2.3 World Englishes, Second Language Acquisition, and ELF Connections

Thanks to the proliferation of research over the past few decades, ELF has established itself as a field of inquiry of its own; however, its connections with WE and SLA research have always been of high relevance to scholarly discussions (Jenkins, 2015; Mauranen, 2018b; Schneider, 2012). Hence, understanding the similarities and differences between these three areas of research can enable a more thorough conceptualization of ELF.

Starting from its early days, ELF research has always problematized the traditional ENL-normative L2 research and teaching because this view was not in line with the global reality of English and led to the deficit view of L2 use in the field (Jenkins, 2000, 2002; Mauranen, 2003; Seidlhofer, 2001, 2004). Thus, the mutual emphasis on "the need for a pluricentric rather than monocentric approach to the teaching and use of English" (Jenkins, 2006, p. 175) enabled important common ground for both ELF and WE research. Early ELF research, in particular, was greatly influenced by the WE paradigm as it was evident from the focus on formal descriptions of ELF and its varietal status (Jenkins, 2015, 2017a). However, due to its primary focus on local varieties of English in formerly colonized territories (Pakir, 2019; Seidlhofer, 2002, 2009b), and an approach to English spoken as a foreign language as learner language (Jenkins, 2006), WE research also differs remarkably from ELF research that evolved to focus more on the complexity and variability of ELF interactions in more linguaculturally heterogeneous environments (Jenkins, 2014; Jenkins & Mauranen, 2019; Seidlhofer, 2011).

Despite important differences, ELF and WE complement one another in many important ways (Jenkins, 2017a). For instance, research on spoken ELF has also documented remarkable similarities to postcolonial varieties, and patterns of change also observed in ENL varieties (Björkman, 2013; Breiteneder, 2005, 2009; Mauranen, 2012; Ranta, 2006, 2009). Some studies on academic writing from different native and nonnative varieties also reported similar findings (Edwards, 2017; Edwards & Laporte, 2015), and others argued that ELF academic writing has a unique status on the WE continuum that is unlike the Englishes in any of the Kachruvian circles (Laitinen, 2018; Laitinen et al., 2018). These findings thus provided support for the argument by some ELF and WE scholars that second language users are not just learners working to achieve nativelike competence (Edwards, 2017; Jenkins, 2014; Mauranen, 2018b; Seidlhofer, 2011).

Although both ELF and WE have moved beyond equating second language users and learners, it is essential to understand that both focus on the use of English as an additional language that is influenced by L2 speakers' former language learning experiences, multilingual repertoires and environments of English use (Mauranen, 2003, 2018b; Schneider, 2012). Besides, both areas of research are very much in

line with the SLA approaches that critique the monolingual NS ideologies (e.g., Cook, 1999; Firth & Wagner, 1997), and the recent calls for the need to take into account the ever-growing multilingualism in L2 research (Douglas Fir Group, 2016; Ortega, 2018, 2019). Hence, it is plausible to suggest that ELF, despite having its differences, still has fundamental connections with WE and SLA research.

2.4 Variation across Modes and Discourses: Implications for ELF

The focus of early ELF research on spoken rather than written modes of communication has often been justified by the argument that speaking is less norm-oriented and planned (edited); thus it is perceived to be more varied and open to change (Mauranen, 2012; Mauranen, Hynninen, & Ranta, 2010a; Ranta, 2017; Seidlhofer, 2004; Seidlhofer et al., 2006). As much as this trend started to change slowly in the early 2010s with the publication of written ELF research (e.g., Boyle, 2011; Carey, 2013; Rozycki & Johnson, 2013), the speech-heavy ELF research exploring the dynamic co-construction of utterance meaning observed in ELF interactions (Firth, 2009) can be said to have led to a theoretical gap. Conceptualizations of ELF today stem primarily from the interactional features of oral communication that are not directly relevant to writing, partly due to the distinctive demands of speech processing and production (Horner, 2017; Ranta, 2017). Thus, the current state of ELF theorization requires a thorough understanding of speaking-writing differences to better accommodate written ELF research.

Decades of research on the nature of ‘spokenness’ and ‘writtenness’ have repeatedly shown that there is a high degree of variation within and across the two modes (Biber, 1985, 1986, 1988; Biber et al., 1999; Chafe, 1982; Halliday, 1989; Redeker, 1984). While written registers tend to have greater linguistic diversity and complexity as opposed to the relatively more restricted range of variation across spoken registers, academic writing shows an even higher degree of variation (Biber, 2006b, 2009; Biber & Conrad, 2019; Biber, Conrad, Reppen, Byrd, & Helt, 2002). Such diversity may be partly due to the wide variety of disciplinary genre conventions (Berkenkotter & Huckin, 1995; Mauranen et al., 2010b; Swales, 1990), as well as the great linguistic and cultural diversity of academic writers who also take part in intricate processes of negotiation as they write (Horner, 2017). Hence, while certain parallels were

observed between spoken and written academic ELF in terms of the nature and frequency of non-nativelike language usage (e.g., Carey, 2013; Rozycki & Johnson, 2013), this evidence is rather scarce in comparison to the wealth of knowledge on speaking-writing differences in other areas of language research, which necessitates further research on written academic ELF.

2.5 English for Academic Purposes (EAP) and ELF

The interdisciplinary nature of written academic ELF research is partly due to its interconnectedness with academic writing research (Horner, 2017), which makes EAP research a crucial element of this dissertation as well. It is also a well-documented fact that the English language has been the primary medium of international academic publishing (Crystal, 2003, p. 93), and a primary focus of attempts to maximize the global outreach of universities around the world (Jenkins, 2014). Hence, as an organic extension of the shifting demographics of academia, EAP researchers have started to address issues pertaining to its NNS members (e.g., Belcher, 2007; Flowerdew, 2000, 2008). In addition, ELF research, starting from its early days, also focused on academic English, users of which, as in other domains, are mainly comprised of NNSs from all around the world (Mauranen, 2003, 2012).

As a well-established sub-field of English for Specific Purposes (ESP), EAP is primarily concerned with the description, instruction, and assessment of the linguistic and rhetoric characteristics of, generally written, academic discourse for non-native speakers of English in university settings (Charles, 2013; Hyland & Shaw, 2016). Besides, the field can be further divided into two strands: (1) English for General Academic Purposes (EGAP), that is, the commonalities in academic language found across different communities, and (2) English for Specific Academic Purpose (ESAP) that deals with the distinctive characteristics of academic language in each community (Hyland, 2016b). While the former tends to be related to common pedagogic applications of EAP research, the latter in particular pertains closely to the role of English in scholarly publishing, which is also studied under the names English for Professional Academic Purposes (Hyland, 2009) and English for Research Publication Practices (Flowerdew, 2013; J. Flowerdew, 2015). In addition, explorations of scientific texts in the genre research tradition popularized by Swales (1990), and analysis of linguistic and cultural differences in L2 writing

via Contrastive Rhetoric (Kaplan, 1966), later known as Intercultural Rhetoric (Connor, 2004) added immensely to our understanding of L2 academic writing. The wealth of literature in EAP paved the way for discussions that are more accepting of hybridity and variation in ELF (e.g., L. Flowerdew, 2015; Mauranen et al., 2010b; McIntosh, Connor, & Gokpinar-Shelton, 2017), and also informed ELF research (Björkman, 2008a, 2013; Mauranen, 2006a; Wang, 2017, 2018).

As much as the ENL-normativity in EAP research has also been problematized and the nativist ideologies in the field are documented by some researchers (Jenkins, 2014; Mauranen, Hynninen, & Ranta, 2016; McKinley & Rose, 2018), others highlight that nativeness can explain only part of the complexities of academic writing including social, contextual, disciplinary and generic dynamics (e.g., Curry & Lillis, 2019; Hultgren, 2019). The latter argument is especially relevant to written academic ELF research as it also acknowledges, in a way, the need for a renewed understanding of academic literacies as multifaceted, dynamic processes (Horner, 2017), which has also been adopted in recent ELF research that takes into account different factors such as mode of communication, genres, and backgrounds of ELF speakers when analyzing usage data (Deshors, 2019; Laitinen, 2018; Laitinen et al., 2018). Since such rigorous studies can help uncover the characteristics of "the lingua franca factor"(Firth, 2009), this dissertation approaches the analysis of written academic ELF as a complex phenomenon and draws from EAP research by controlling the data for genre and exploring the disciplinary characteristics of ELF texts.

2.6 A Usage-based (UB) Approach to the Analysis of ELF

Research on ELF has taken different shapes and forms over the years. The variety of ways in which ELF scholars conceptualize the linguistic analysis of ELF can be summatively divided into two strands: the traditional (competence-performance) approach and usage-based approaches (Vetchinnikova, 2015). The traditional approach, first seen in Seidlhofer (2001), follows the concept of 'virtual language'(Widdowson, 1997) to argue that different types of language use, whether ELF or ENL, are alternative actualizations of the same abstract system of rules deeming these and other uses of English possible. Building on this rather generativist (Chomsky, 1965) approach to language, Seidlhofer (2011)

puts forward that the differences between ELF and ENL in terms of the speakers' linguistic resources and the purposes of use lead them to "exploit the possibilities of virtual language" differently from one another (p.120). This line of research is common in mostly qualitative explorations of spoken ELF interactions where low-frequency lexicogrammatical items are investigated from a conversational pragmatic perspective, and findings point to the creative exploitation of the virtual language by ELF speakers based on contextual and interpersonal parameters (Pitzl, 2012, 2018; Seidlhofer, 2009a, 2011)

Usage-based approaches, on the other hand, focus on exploring ELF through the lens of the cognitive dimensions of second language use such as memory and processing, and the effect of input frequency on second language users' output (e.g., Alptekin, 2011, 2013; Hall, 2012, 2017; Mauranen, 2012; Vetchinnikova, 2015, 2017). Unlike the Chomskyan idea of a set of abstract general rules learned through innate language learning mechanisms, UB approaches suggest that learners grasp the patterned nature of languages through their experiences with them, that is, with the input they receive (Bybee, 2006, 2008; Bybee & Hopper, 2001; Tomasello, 2003). Learners engage with the salient input in their environment in a largely implicit manner through general human cognitive capacities, which, over time, leads to language learning (Ellis & Wulff, 2015b, pp. 75-76). In addition, these cognitive and environmental factors are considered to be "all inextricably intertwined in rich, complex, and dynamic ways in language, its use, and its learning" (Ellis & Larsen-Freeman, 2009, p. 91). Following this perspective, many researchers adopt a frequency-informed approach to the analysis of ELF corpora and explain the distinctive lexicogrammatical features of ELF via UB processes such as input frequency, entrenchment, and changes in the form-function associations of patterns (Carey, 2013; Hall, Joyce, & Robson, 2017; Mauranen, 2012; Vetchinnikova, 2015, 2017).

While these two theoretical approaches adopted in ELF research have both contributed tremendously to our understanding of ELF, the choice of linguistic theory still has significant implications for ELF research due to the recognizable differences in the conceptualization and investigation of the phenomenon (Vetchinnikova, 2015), Hence, following the definition of ELF as the dynamic, variable type of English language 'use' as a common medium of communication by mostly second language

speakers from linguaculturally diverse backgrounds (Firth, 2009; Jenkins, 2017b; Kecskes, 2019; Mauranen, 2012; Mortensen, 2013; Seidlhofer, 2011), this dissertation adopts a UB approach. Accordingly, it aims to investigate constructions, the building blocks of language from a UB perspective (Ellis & Wulff, 2015b), by employing a corpus-based approach and rigorous statistical analyses for the investigation of the complexities of written academic ELF by taking into account the writers' L2 backgrounds, disciplines, and the role of editing in publishing. With the aim of capturing the frequency-informed characteristics of its usage in detail, this study is hypothesized to further enhance our knowledge of written academic ELF.

2.7 Summary

A thorough conceptualization of written academic ELF relies on a variety of research areas that approach the study of English as an additional language in ways that differ from but also complement one another. The connections made between research in ELF, standard language ideologies, WE, SLA, EAP and UB are of fundamental value to this study as they informed the methodological decisions and the interpretation of findings. The next chapter elaborates on the theory of the chosen unit of analysis for this study: Goldbergian Construction Grammar (CxG).

3 BRIDGING LEXIS AND GRAMMAR: CONSTRUCTION GRAMMAR (CxG) FOR ELF

The study of multi-word units (MWUs) ranging from collocations to idioms and grammatical constructions has been crucial in language research, and a broad array of research perspectives employing terms such as phraseology, idiomaticity, formulaicity, and lexicogrammar have been employed to explore MWUs (Römer, 2009b; Siyanova-Chanturia & Pellicer-Sánchez, 2018). As a result of extensive research efforts on the topic, we now have ample evidence that languages are highly patterned, a view that blurs the traditional distinction between lexis and grammar (Biber & Conrad, 1999; Biber et al., 1999; Erman & Warren, 2000; Francis, Hunston, & Manning, 1996, 1998; Goldberg, 1995; Halliday, 1994; Jackendoff, 1995; Pawley & Syder, 1983; Sinclair, 1991). Similarly, SLA research has highlighted that phraseological features of languages are critical for the mastery of an additional language, and learners may have challenges acquiring them (e.g., Ellis, 1996; Pawley & Syder, 1983). ELF research has been motivated by this line of SLA research (Vetchinnikova, 2015), and the study of MWUs has been a central focus of research in ELF ever since its initial stages (e.g., Seidlhofer, 2001, 2002).

Among the vast array of approaches to the study of MWUs, UB approaches are especially important because of their conceptualization of successful SLA as the learning of MWUs, in the form of constructions (e.g., Ellis, 2002, 2003). The concept of constructions in this line of research is taken from the cognitive linguistic approach to grammar and refers to a network of form-meaning pairings of any capacity from morphemes to multi-word phraseological patterns (Goldberg, 1995, 2003, 2006). As a theoretically sound and broad usage-based (UB) approach that takes into account the intricate patternings in language at different levels, CxG has recently been influential in applied linguistic research (Gilquin & De Knop, 2016). Its applications in ELF (Hall, 2012, 2017; Hall et al., 2017; MacKenzie, 2014; Pirc, 2013, 2014) as well as the few studies on constructions of written academic ELF in particular (Vetchinnikova, 2017; Yilmaz & Römer, 2020) point to its great potential. Thus, CxG is selected as the theoretical framework for this dissertation. After a brief survey of relevant phraseology research, this

chapter defines CxG and provides a review of CxG-oriented applied linguistic research that informs the linguistic approach adopted in the current study for the analysis of written academic ELF.

3.1 Earlier Phraseological Approaches to the Study of Second Language Use: The Case of EAP Writing

Some of the early accounts of phraseology research that have influenced applied linguistics research to this day include Hallidayan (Systemic Functional Linguistics) lexicogrammar (Halliday, 1994); the Idiom Principle by Sinclair (1991) and the related comprehensive work on Pattern Grammar (Francis et al., 1996, 1998; Hunston & Francis, 2000); lexical bundles (Biber & Conrad, 1999; Biber et al., 1999); and lexical priming (Hoey, 2005). Studies on formulaicity (Wray, 2002), and idiomaticity (Pawley & Syder, 1983) have also been influential. As much as they are different theoretical frameworks, these approaches revolutionized how language use is viewed theoretically, methodologically, and pedagogically. Today, not only are we aware that phraseology is a crucial aspect of language research, but also that there are differences between first and second language acquisition because of the dynamics involved in the latter that deem mastering the phraseological aspect of language challenging (Vetchinnikova, 2019; Wray, 2018; Wulff, 2018).

Phraseological patterning of language is especially important when the focus of the inquiry is academic writing where we find highly complex and sophisticated language use shaped by powerful ideologies and conventions. Summarizing the importance of phraseological research for EAP from a 'formulaic language' perspective, Durrant (2018) explains:

> Formulas act as conventional frames in which academics set the novel content of their texts, that these frames carry important characteristic meanings which academics use to create knowledge in approved ways, and that they vary in interesting ways across academic communities and contexts (p. 223).

In the case of written scholarly communication in specific discourse communities such as English for Research Publication Purposes (ERPP), phraseological complexity becomes even more crucial as highly sophisticated language use is considered the norm for this kind of writing (Biber & Conrad, 2019, p.

302).By generally adopting a corpus-based approach, research has documented that this type of academic writing differs remarkably from less formal and informational types of writing and speaking due to its complex phraseological profile with a high degree of variability (e.g., Biber, 2006b, 2009; Biber & Conrad, 1999; Biber, Conrad, & Cortes, 2004; Biber et al., 2002; Biber et al., 1999; Gray & Biber, 2013; Paquot, 2019). The centrality of the 'phraseological dimension' (Paquot, 2019) for EAP writing is also evident in previous analyses of different academic genres such as research article introductions (Cortes, 2013; Lu, Yoon, & Kisselev, 2018) and book reviews (Römer, 2010, 2011b), as well as disciplinary variation (Durrant, 2015; Hyland, 2008b), and author expertise (Cortes, 2004; Hyland, 2008a).

Besides, phraseological approaches to L2 academic writing have led to the exploration of a wide variety of contiguous and non-contiguous word sequences across different disciplines, texts, and genres, and author demographics. Today, thanks to this research, we know that there is remarkable overlap in phraseological patterning between L1 and L2 academic writing, which can be better understood when both nativeness status, as well as expertise level of writers, are factored into the analysis (Lu, Kisselev, Yoon, & Amory, 2018; O'Donnell, Römer, & Ellis, 2013; Qin, 2014; Römer, 2009a, 2009b). Other studies, however, have pointed to certain structural and functional differences between L1 and L2 EAP writing in their findings. For instance, they reported a relatively limited number of lexical bundles (LBs) in L2 student writing (Ädel & Erman, 2012), which increased with expertise (Qin, 2014). Research on writing by students (Bychkovska & Lee, 2017; Chen & Baker, 2010) and experts (Pan, Reppen, & Biber, 2016) also pointed to the high frequency of clausal LBs as opposed to the more common use of nominal LBs in writing by ENL writers. Perez-Llantada (2014), however, showed that L2 English writers employed fewer clausal bundles than ENL writers did. In terms of functions, L2 student writers in Chen and Baker (2010) and Ädel and Erman (2012) were found to use fewer stance LBs, whereas the student writers in Bychkovska and Lee (2017) and expert writers in Pan et al. (2016) employed this function more frequently. Several grammatical and functional unconventionalites due to L1 effect in the case of L1 Chinese writers in Bychkovska and Lee (2017) and L1 Spanish writers in Pérez-Llantada (2014) were also identified. In sum, these findings, while uncovering important characteristics of L2 academic writing,

seem to also point to the complexity of L2 phraseology and the need to account for the multifaceted nature of the phenomenon in question.

3.2 Defining Construction Grammar (CxG)

CxG, while having similarities to the previously mentioned theories in terms of its main tenets, is a somewhat recent theoretical framework in phraseology research, which has developed from an established strand of cognitive linguistic descriptions of language. Even though the early (and mostly theoretical) work on the theory was carried out in the 1980s (Fillmore, 1985, 1988; Fillmore, Kay, & O'Connor, 1988; Lakoff, 1984), its impact on applied linguistics gained momentum especially after Goldberg's (1995) influential publication on argument structure constructions.

Unlike early CxG work that was rather theoretical (Römer, 2009b), its use in applied linguistics has been increasingly data-based, drawing from corpus linguistics, as seen in collostructional analysis (Stefanowitsch & Gries, 2003), extraction and interpretation of key-words and MWUs (Groom, 2018, 2019; Yilmaz & Römer, 2020), situating former phraseological research within CxG by identifying constructions from grammar patterns (Hunston, 2019; Hunston & Su, 2017), and measuring constructional knowledge of learners through psycholinguistic methods (Ellis, O'Donnell, & Römer, 2014; Ellis, Römer, & O'Donnell, 2016; Römer, O'Donnell, & Ellis, 2014). As De Knop and Gilquin (2016) argue, while still in its infancy, such applied CxG research is in fact, quite compatible with, if not identical to, other usage-based research paradigms (e.g., corpus linguistics, psycholinguistics), thus it is of great potential for second language research (p. 14).

Goldberg (1995) initially defined the term construction as "a form-meaning pair" that is non-compositional, meaning it is a complete, and generally phrasal pattern unlikely to be broken down into smaller semantic components (p.4). Later, Goldberg (2006) expanded the definition, stating that units of more compositional nature can also be considered constructions "as long as they occur with sufficient frequency" (p. 6). Acknowledging the limitations of her previous two definitions and by drawing on a vast amount of cognitive linguistic and psychological research, Goldberg (2019) proposed a recent and relatively more comprehensive definition of constructions as "emergent clusters of lossy memory traces

that are aligned within our high-(hyper!) dimensional conceptual space on the basis of shared form, function, and contextual dimensions" (p.7). As is evident from her mention of the role of usage frequency and the emergent nature of construction learning in the 2006 and 2019 definitions, Goldbergian CxG is a usage-based account of language, which is also supported by research on first (e.g., Tomasello, 2003; Tomasello, 2005, 2011) and second language acquisition (e.g., Ellis & Ferreira-Junior, 2009a, 2009b; Ellis et al., 2016; Eskildsen, 2012).

An essential strength of CxG is also that the pairing of form and meaning an be at any level such as "morphemes, words, idioms, partially lexically filled and fully general linguistic patterns" (Goldberg, 2003, p. 219), as well as discourse (Östman, 2005) and genre (Hoffmann & Bergs, 2018; Ruppenhofer & Michaelis, 2010). Moreover, in line with cognitive linguistics, CxG does not presume clear differences between syntactic and lexical forms, and its approach to meaning does not separate semantic, pragmatic and discoursal functions from one another (Goldberg, 1995, 2013). Hence, CxG has the potential for extending phraseological approaches to language description thanks to its strong theoretical basis and high explanatory adequacy (Groom, 2019).

The constructionist approach to language acquisition suggests that constructions lie at the heart of language learning as our constructional knowledge, built through usage and entrenchment, is comprised of partially abstracted and complex representations of our social experiences with language in the form of structural patterns matched with particular meanings (Ellis et al., 2016; Goldberg, 1995, 2003, 2006, 2019; Hilpert, 2014). While CxG research highlights remarkable resemblances in the processes of L1 and L2 acquisition, the commonly observed nonnative-like attainment by adult L2 learners, however, is explained through challenges of learners' generalizing from insufficient input and learning mechanisms that are highly attuned to L1 acquisition and usage over the years (Ellis & Wulff, 2015a, 2015b, 2019; Goldberg, 2019). In addition, the use of sophisticated methodologies employed in usage-based CxG research enables rich descriptions of both L1 and L2 use from different perspectives. Several studies have highlighted the importance of contextual, varietal and generic appropriateness of constructional use as determinants of successful acquisition (Ellis, 2012; Wulff & Gries, 2011). Some others employ data-

driven approaches first to identify the ENL usage norms to which L2 usage data are compared, as well (e.g., Gries & Adelman, 2014; Gries & Deshors, 2014; Wulff & Gries, 2019). Overall, CxG seems to have great potential for advancing research efforts in applied linguistics thanks to its broad scope and progressive agenda.

3.3 CxG-Oriented Academic Writing Research

In recent years, CxG has been influential in academic writing research in several ways. The International Corpus of Learner English (ICLE; Granger, Dagneaux, & Meunier, 2002; Granger, Dagneaux, Meunier, & Paquot, 2009) has been extensively used for constructional explorations of language produced by college-level, mostly advanced L2 writers. For example, research has shed light on the use of ditransitive and prepositional dative constructions by L1 Spanish (Valenzuela Manzanares & Rojo López, 2008) and L1German learners of English (Wulff & Gries, 2009); infinitival and gerundial construction use by L1 German (Wulff & Gries, 2009), and Spanish learners (Martinez-Garcia & Wulff, 2012); periphrastic causative construction across 16 L1s represented in ICLE (Gilquin, 2012), and phrasal verb constructions by French learners of English (Gilquin, 2015). These studies generally found that advanced L2 academic writers have a solid knowledge of the target constructions. However, several interesting findings about L2 writing were also reported: 1) influences of first language background (Martinez-Garcia & Wulff, 2012; Valenzuela Manzanares & Rojo López, 2008); 2) more frequent use of certain constructions by learners compared to L1 writers (Martinez-Garcia & Wulff, 2012), 3) high frequency of a limited number of different phrasal verbs(Gilquin, 2012, 2015; Wulff & Gries, 2009); and 34) unconventional lexical choices within constructions (verbs in causatives that are not in line with L1 use) (Gilquin, 2012).

A specific type of verb-argument construction (VAC), “V preposition n,” has also received considerable attention from researchers who looked into the use of constructions by German and Spanish (Römer, Roberson, O’Donnell, & Ellis, 2014; Römer, Skalicky, & Ellis, 2018), and by Turkish learners of English (Yilmaz & Römer, 2020). These studies similarly concluded that learners have constructional knowledge of especially the more frequent (thus more entrenched) verb-VAC associations and

highlighted the varying degrees of first language influence as well as of text and the task type in the selected corpora.

While most studies focused on opinion and argumentative essays produced by college-level writers, to the knowledge of the researcher, there has been only one CxG study on research writing by college-level students, which examined attended and unattended *this* constructions (Wulff, Römer, & Swales, 2012). The study did not report any differences between native and nonnative students, and findings were discussed by focusing on different expertise levels of the students. A number of studies also investigated constructions in different varieties of World Englishes including intransitive, monotransitive, and ditransitive constructions (Mukherjee & Gries, 2009), verb-preposition constructions (Wong, 2014), light verb constructions (Ronan & Schneider, 2015), and comparative constructions (Hoffmann, 2014). These studies pointed to differences across varieties and documented the nativization processes in new varieties of English (Schneider, 2007) towards the uniformity found in native varieties in relation to constructional usage. In sum, existing CxG studies have contributed to shedding light on a number of interesting patterns of second language usage across a wide variety of contexts and purposes of use..

3.4 A Usage-based Constructionist Approach to the Phraseology of Written Academic ELF

As stated earlier, phraseological explanations of language usage in ELF have been common since the initial stages of ELF research. To illustrate, a prominent figure in the field, Seidlhofer (2001, 2002, 2004) stated that 'unilateral idiomaticity,' that is, one speaker's conformity to ENL phraseological conventions more than the other(s) in an ELF interaction, could lead to miscommunication, based on which she suggested that, instead of following ENL standards, ELF phraseological competence is about accommodating to the dynamics of such interactions. Her famous (2004) study also listed the phraseological characteristics of ELF such as unconventional usage of *that*-clauses, adjectives in adjective-noun phrases, and prepositions in verb-preposition phrases (p. 220). Further analyses of phraseology in ELF conversations following Seidlhofer's work also reported seemingly unconventional and limited usage of MWUs due to different speaker backgrounds, and pragmatic dynamics such as accommodation strategies (Cogo & Dewey, 2006, 2012; Kecskes, 2007; Prodromou, 2008). Following

similar findings, Seidlhofer (2009a) made a substantial claim, later supported by Kecskes (2015) and Widdowson (2019), that the distinctive idiomatic competence of ELF speakers is related to their limited reliance on the idiom principle compared to native speakers. Along the lines of the theorization of ELF as an actualization of virtual language different from ENL (Seidlhofer, 2001; Widdowson, 1997) , this view of ELF idiomaticity, also associated with generativism, led to research on qualitative explorations of unconventional idiomatic language as instances of creativity in ELF (Hülmbauer, 2013; Pitzl, 2012, 2018; Seidlhofer, 2011; Seidlhofer & Widdowson, 2017; Widdowson, 2017). However, to the best knowledge of the researcher, this approach was adopted only in spoken ELF research.

The more usage-based ELF research, however, depicted a different picture of ELF phraseology. For instance, the corpus-based studies on phraseology in spoken academic ELF by Mauranen (2005, 2006b, 2007, 2009, 2011, 2012) repeatedly showed that there was significant overlap between ELF and ENL phraseology, with the former only showing a few instances of approximations that had partial differences from conventional ENL patterning (e.g., in/on/*from my point of view*, *as the matter of fact*). This finding was also supported through written academic ELF research (Carey, 2013; Vetchinnikova, 2015), which reported findings such as high ENL conformity with distinctive rhetoric, and, to a lesser extent, lexicogrammatical choices (e.g., Farley, 2018; Lorés-Sanz, 2016, 2018; Martinez, 2018; Mur-Dueñas, 2018a; Murillo, 2018; Pérez-Llantada, 2014; Tribble, 2017, 2019). Further, Wu, Mauranen and Lei's (2020) comprehensive analysis of syntactic complexity in written academic ELF demonstrated how common use of complex nominal structures in ELF contributed to communicative efficiency in texts.

Also within a usage-based tradition, research that examined constructions was able to capture more of the emergent and dynamic nature and variation of the phraseological patterning in ELF usage from a formal and functional perspective at the same time. Vetchinnikova (2017) and Hall et al. (2017) demonstrated how communal ELF was highly varied, while individuals relied on a restricted set of forms and functions of a construction. Yilmaz and Römer's (2020) analysis showed that ELF-dominant constructions in academic writing not only exhibited structural variability but also functional specificity such as referring to research activities.

In conclusion, there seem to be two major strands of phraseological research in ELF, offering remarkably distinct perspectives. While the more qualitative lens on ELF idiomatic creativity enables a closer look at individual patterns, the nature of the so-called lingua franca factor (Firth, 2009) as a variable, complex and dynamic phenomenon is yet to be adequately understood. Hence, to contribute to our understanding of ELF at a large scale, a comprehensive and usage-based CxG approach that employs both quantitative and qualitative methods to analyze the constructional profile of ELF academic writing is used in the current study.

3.5 Summary

This chapter aimed to explain the motivation behind the usage-based constructionist focus of this study. For this aim, selected phraseological research in applied linguistics, ELF, and EAP, as well as the theory of and research in CxG, were summarized. While the previous and current chapter provided an overview of the theoretical foundations of the study, the next chapter offers a closer look at research on written academic ELF.

4 RESEARCH ON WRITTEN ACADEMIC ELF

Written academic ELF research has become popular especially in the last decade as researchers have started to approach the analysis of academic writing by international scholars from a renewed perspective by connecting ELF and fields such as English for Specific Purposes (ESP), English for Academic Purposes (EAP), and Intercultural Rhetoric (IR). This initiative introduced a great diversity and robust application of theories and research designs to gain insights into the linguistic and rhetorical profile of written academic ELF from a new perspective that aimed to understand the distinctive uses of English in academia as it is being "shaped by its non-native users" (Mauranen, 2012). A primary focus of this research has been the corpus-based analysis of phraseological patterns in different academic genres. Since this dissertation also analyzes constructions as phraseological patterns with paired formal and functional properties, this chapter summarizes relevant research on the phraseology of written academic ELF to describe the current state of scholarship in the field, and identify the research gap the current study aims to fill.

The summary provided here is structured to match the design of the current study, so studies on edited and unedited ELF texts are presented separately. For each group, the data, methodologies, and findings reported in the selected studies are reviewed. It should, however, be noted that only studies which are contextualized explicitly within the ELF paradigm are included in this chapter.

4.1 Unedited Written Academic ELF

An essential line of written academic ELF research has focused on unedited writing in order to uncover the language of ELF texts that are not revised with input from 'literacy brokers' (Lillis & Curry, 2006), such as editors and reviewers, during the publishing process. One of the earliest relevant studies is Anderson (2010) that analyzed published yet unedited working papers by international scholars in a European university. This study was followed by Carey's (2013) analysis of papers from a pre-release version of the WrELFA (2015) corpus, the first written academic ELF corpus project that includes unedited (and mostly unpublished) ELF research writing in three genres (research articles, blogs, and doctoral examiner reports) by authors from different disciplines and similectal backgrounds. Several

WrELFA studies focused on exploring the complete corpus (Carey, 2013; Mauranen, 2016, 2019; Yilmaz & Römer, 2020), whereas most research examined only SciELF, the research articles sub-corpus of WrELFA (Mur-Dueñas, 2018a; Murillo-Ornat, 2019; Murillo, 2018; Rowley-Jolivet, 2017; Wu et al., 2020).

A number of other studies focused on a selection of texts from the SciELF sub-corpus based on disciplines such as economics (Bondi & Borelli, 2018; Vitali & Bondi, 2020) and social sciences (Lafuente-Millán, 2018; Shchemeleva, 2018). While Lafuente-Millán's (2018) analysis was limited to the introduction sections of the texts, Zapletalová (2018) included only ELF users whose first language was Czech. Some researchers examined the corpus-internal variation within SciELF by analyzing the differences between the L1s and disciplinary groupings represented in the corpus (Murillo-Ornat, 2019; Murillo, 2018; Rowley-Jolivet, 2017). In addition, Laitinen's (2018) comprehensive analysis of ELF only used the doctoral examiner reports from WrELFA (2015), while Mauranen (2013) examined blogs from the corpus.

Using different sources of ELF data, a few other studies investigated patterns in academic blogging (Luzón, 2018; Vetchinnikova, 2017), working papers (Anderson, 2010), student writing such as master's theses (Vetchinnikova, 2015, 2019) and undergraduate essays (Tai, 2019), as well as multiple genres of academic and non-academic writing (Laitinen et al., 2018). Luzón (2018) and Tai (2019) focused on Spanish and Chinese ELF respectively, and Laitinen et al. (2018) targeted ELF used in Sweden and Finland.

In this line of research, most studies used published NS academic writing as the reference data, except for a few that used a corpus of published research by both ELF and ENL writers (Anderson, 2010; Bondi & Borelli, 2018; Vitali & Bondi, 2020), or a corpus representing the linguistic exposure of the ELF users whose texts were analyzed (Vetchinnikova, 2015, 2017). Several other studies (Luzón, 2018; Mauranen, 2016, 2019; Murillo-Ornat, 2019; Tai, 2019) did not include any comparison data at all. Carey (2013) adopted a more comprehensive approach and compared ELF academic speaking and writing. Similarly, approaching written ELF from a World Englishes perspective, Laitinen (2018) and Laitinen et

al. (2018) further expanded the scope of research by examining several spoken and written, as well as academic and non-academic genres.

As for the phraseological items studied, we see that a number of studies examined specific patterns known for their important discourse functions such as textual organization, evaluation, and hedging by analyzing the forms and functions of the selected patterns such as '*according to', 'in addition', 'it is* ADJ *that'* and 'enabling verb + NP + *to*-infinitive' (Bondi & Borelli, 2018; Carey, 2013; Lafuente-Millán, 2018; Mur-Dueñas, 2018a; Murillo-Ornat, 2019; Murillo, 2018; Rowley-Jolivet, 2017; Shchemeleva, 2018; Vetchinnikova, 2017; Vitali & Bondi, 2020). However, several other research designs followed a broader scope. For instance, Mauranen (2016) and Luzón (2018) reported a list of distinctive written academic ELF features, including both single words and MWUs. Anderson (2010) coupled manual analysis of 'deviations' from ENL usage with a follow-up corpus analysis of contiguous and non-contiguous MWUs in which these unconventional ELF features occurred. Other researchers started from quantitative corpus analyses of contiguous MWUs (Zapletalová, 2018), non-contiguous MWUs (Vetchinnikova, 2015, 2019), MWUs showing variation in past tense marking (Tai, 2019), and constructions (Yilmaz & Römer, 2020). Several of the studies differ from the above-summarized ones in that their focus was predominantly on the structural patterns such as past tense marking (Tai, 2019), syntactic complexity (Wu et al., 2020), and several phrasal and clausal MWUs (Laitinen, 2018; Laitinen et al., 2018).

The research on unedited writing yielded important findings on the nature of written academic ELF. For instance, while a long list of distinctive ELF features such as instances of unconventional lexicogrammatical variation can be made, researchers also pointed to the infrequency of features (Carey, 2013; Mauranen, 2016; Vetchinnikova, 2015), and the tendency of ELF academic writers to rely on a limited number of conventional and frequent patterns (Carey, 2013; Mur-Dueñas, 2018a; Murillo, 2018; Wu et al., 2020; Yilmaz & Römer, 2020; Zapletalová, 2018). Several exceptions (Anderson, 2010; Luzón, 2018; Rowley-Jolivet, 2017) highlighted the L1 (similectal) influence, and particularly, its potential effect on the hybridization of rhetoric patterns with a mix of L1 and English conventions (Lafuente-Millán,

2018; Murillo-Ornat, 2019; Murillo, 2018). Other studies reported conformity to ENL norms as well as genre and disciplinary conventions, however, generally with less personal engagement and more hedging in unedited ELF academic writing in comparison to ENL (Mur-Dueñas, 2018a; Shchemeleva, 2018), as well as published texts from both ELF and ENL authors (Bondi & Borelli, 2018; Vitali & Bondi, 2020). Besides, as Luzón (2018) argues, in academic blogging, ELF writers seem to conform to NS norms to a lesser extent, possibly due to the more relaxed genre conventions allowing more hybrid language use. Lastly, the studies which compare speaking and writing (Carey, 2013; Laitinen, 2018; Laitinen et al., 2018) point to the remarkably higher degree of ENL-like usage in written ELF, particularly in academic writing.

4.2 Edited (Published) Written Academic ELF

Continuing the corpus-based EAP research tradition of text analysis, ELF research on published writing has recently been common as well. Almost all studies have examined research publications with few exceptions looking at the language of university webpages (Bernardini, Ferraresi, & Gaspari, 2010; Ferraresi & Bernardini, 2015; Palumbo, 2015). Similar to the research on unedited ELF academic writing, these studies targeted specific disciplines such as animal science (Farley, 2018), biomedicine (Tribble, 2017), business management (Mur-Dueñas, 2015, 2016, 2018b), food sciences, (Martinez, 2018), sociology (Lorés-Sanz, 2016, 2018), and engineering (Rozycki & Johnson, 2013), as well as selected contrasting disciplines in the hard and soft sciences (Palumbo, 2017; Tribble, 2019). While some of these studies (Farley, 2018; Ferraresi & Bernardini, 2015; Lorés-Sanz, 2018; Martinez, 2018; Palumbo, 2015, 2017; Rozycki & Johnson, 2013) ensured the representation of multiple different L1s, others included ELF corpora with a small proportion of ENL writing (Lorés-Sanz, 2018; Mur-Dueñas, 2015, 2016, 2018b; Palumbo, 2017). Farley (2018) compared international ELF texts with local ELF and L1 research articles published locally in Indonesia. Mur-Dueñas (2018b) and Lorés-Sanz (2018) similarly compared international ELF with Spanish research articles. Research on university webpages also had specific foci, including comparisons of pages produced by Anglo-American institutions with European (Ferraresi & Bernardini, 2015; Palumbo, 2015), and specifically Italian ones (Bernardini et al., 2010).

As in the case of ELF research on unedited academic writing, the majority of these studies also included a reference corpus of ENL texts with some exceptions that did not use any reference corpora (Lorés-Sanz, 2016; Rozycki & Johnson, 2013; Tribble, 2017). Other studies used ELF corpora that included texts predominantly from non-native writers, but a small number of ENL texts were also included (Mur-Dueñas, 2016, 2018a, 2018b; Tribble, 2017, 2019). In addition, Lorés-Sanz' (2016) and (2018) studies stand out not only because of the sole focus of the analysis being on abstracts but also because the latter included articles that were translated to English from Spanish. Providing yet another different research design, Martinez (2018) examined the occurrence of common ELF-like patterns across different similects between the years 2000 and 2015.

Several of these studies analyzed texts manually for moves and specific lexicogrammatical items which signify certain rhetorical features of ELF (Farley, 2018; Lorés-Sanz, 2016, 2018; Rozycki & Johnson, 2013; Tribble, 2017). Others examined patterns in corpora in a primarily quantitative fashion, by focusing on both form and function of the selected patterns with discourse functions such as evaluation and hedging (Mur-Dueñas, 2015, 2016, 2018b), or identifying distinctive ELF patterns in comparison to ENL data (Bernardini et al., 2010; Ferraresi & Bernardini, 2015; Martinez, 2018; Palumbo, 2015, 2017; Tribble, 2019).

Despite the variety of analytic foci and data used in these studies, findings seem to be quite consistent in that they generally identified a small number of lexicogrammatical and rhetoric features in published ELF texts which distinguish them from conventional, nativelike English academic writing (Rozycki & Johnson, 2013; Tribble, 2017, 2019). Focusing on structural patterns of complex noun phrases, Palumbo (2017), however, found that writers from certain similectal backgrounds used some phrasal structures (e.g., pre- versus post-modification of nouns) more than others. Of these studies, Martinez (2018), being the most comprehensive of all, reported a list of ELF-dominant patterns across different similects and periods, and showed an increased prevalence of these patterns in food science articles. Studies which investigated specific features such as evaluative and persuasive devices pointed to

their high frequency with largely conventional usage in internationally published ELF as opposed to both ENL and locally published ELF (Farley, 2018; Mur-Dueñas, 2015, 2016, 2018b).

In terms of rhetorical conventions, there seem to be two main opposing findings. Researchers looking into the writing of authors from specific similects (Farley, 2018; Lorés-Sanz, 2018; Mur-Dueñas, 2018b) highlighted a somewhat prevalent L1 influence in addition to conventional features, making ELF texts discursively hybrid (Mauranen et al., 2010b). However, those analyzing the rhetorical structure of published articles by authors from different similects found that there was a strong tendency to follow the traditional IMRD (introduction, methods, results, and conclusion) structure (Lorés-Sanz & Murillo-Ornat, 2018; Tribble, 2017, 2019). Lastly, similar to Luzón's (2018) results from her analysis of ELF blogging, the studies that examined different genres such as university webpages reported relatively common use of unconventional collocations, high freqeuncy of noun phrases, and distinctive use of evaluative devices in comparison to ENL (Bernardini et al., 2010; Ferraresi & Bernardini, 2015; Palumbo, 2015).

4.3 Summary

Written academic ELF research has recently expanded into new venues of research and started to employ many different theoretical and methodological perspectives. So far, there seems to be a consensus that, when usage frequencies are taken into account, written academic ELF is very much like ENL in terms of the use of high-frequency phraseological items. The conventions of 'academic tribes' (Becher, 2001), and especially in the case of edited (and mostly published) ELF, 'literacy brokering' practices (Lillis & Curry, 2006), can be said to lead to such convergence with ENL academic writing. While unconventional phraseological and rhetoric patterns were reported to have restricted usage, the degree of their visibility varies across modes, genres, publishing contexts, similectal backgrounds, and disciplines. It seems that research on internationally published ELF texts reports a high degree of conformity while the analyses of unedited texts from especially local contexts and in different less strictly regulated genres show a greater degree of variation. As also pointed out by Flowerdew (2019), a remarkable amount of evidence for ELF-specific usage also comes predominantly from texts in the hard sciences, thus signaling a potential discrepancy in terms of the acceptance of ELF features between hard and soft sciences.

In sum, the research findings so far can be said to confirm the existence of the 'lingua franca factor' (Firth, 2009) for written academic ELF to a certain extent, in that ELF texts are also inherently complex. However, the complexity in writing has a different nature than speaking in ELF due to the quite distinct characteristics of the two modes (Ranta, 2017), with the former being subject to strict language regulation (Hynninen & Solin, 2017). Further, the interaction between the intricacies of English academic writing and ELF lead to unique characteristics in written academic ELF that are yet to be fully understood. Not only is written academic ELF research still in its infancy, but the small number of studies based on large and multiple sources of usage data, narrow scope of analyses, and limited use of statistics restrict our understanding of the complex phenomenon written academic ELF is. These limitations clearly show that more research needs to be conducted before we can arrive at definitive conclusions on the nature of written academic ELF. In light of these limitations and their implications for future research, the next section summarizes the rationale and research design of the proposed study.

5 THE CURRENT STUDY

The previous chapters covered three main strands of research that informed the current study: conceptualizations of ELF as a complex phenomenon, the dynamics of academic writing, and the usage-based constructionist approach to the study of phraseology. This chapter presents a synthesis of this research with the aim of summarizing the conceptual framework and rationale of this study. It then introduces the research questions that this study addresses.

5.1 Proposing a Comprehensive Analytical Model for the Analysis of Written Academic ELF

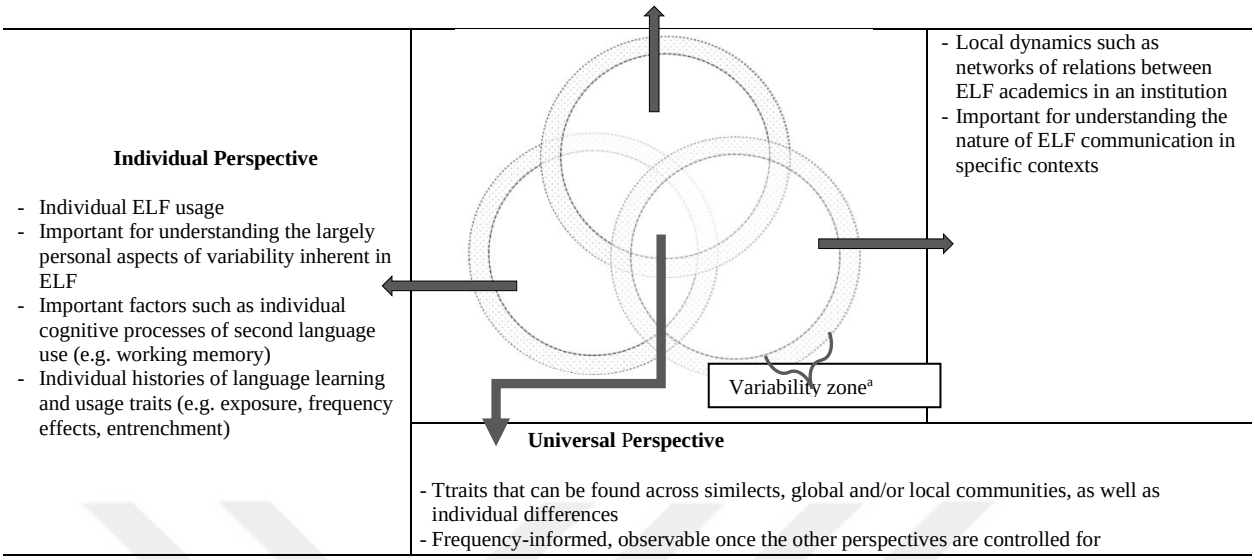
Following the increasingly agreed-upon argument that it is hardly possible to talk about ELF users as members of a homogenous and stable community with easily identifiable characteristics, Mauranen (2012, 2018b) proposed three perspectives of ELF, which are macrosocial, microsocial, and individual (cognitive) perspectives on its conceptualization and analysis. These perspectives are of major importance for the theoretical grounding of any ELF study as they can enable researchers to adopt different lenses, ranging from considering the multiple communities with which ELF speakers are affiliated to the multiple individual (cognitive) factors that can affect the use of ELF. Since this multilayered model is also in line with the argument for the lingua franca factor (Firth, 2009), as well as approaches to academic writing as a complex and dynamic phenomenon (Bhatia, 2002a, 2002b; Biber & Conrad, 2019; Horner, 2017; Turner, 2018) and the recent calls for the need to make 'transdisciplinary' connections to better understand multilingualism in second language research (Douglas Fir Group, 2016; Ortega, 2018, 2019), I argue that an adapted version of the model, as shown in Figure 5.1, is central to usage-based inquiries of written academic ELF. I hypothesize that this model can aid in making tangible claims about the academic writing by non-native scholars by taking into account the multiplicity of factors which influence their second language use.

According to Mauranen's (2012, 2018b) tripartite model, at the macrosocial level, for instance, we can look at the membership of loose-knit, distant, or sometimes imagined communities such as being part of the same similectal group or academic discourse community. The microsocial level, on the other hand, is more about actual, physical communities such as academics working in the same institution

where language is exchanged, and conventions are realized to facilitate personal interactions while working towards similar goals. These two social perspectives are shaped mainly by language regulation. As Solin and Hynninen (2018) state, despite being generally implicit, language regulation in academic writing can be easily seen in dominant perceptions and practices of "what kind of English language users deem appropriate, acceptable and functional in specific contexts" (p. 498). The individual (cognitive) level is about differences in the way language is accessed, stored, processed, and produced across language users. Not only is language use largely shaped by social factors, but it is also adapted through each user's individual experiences (Bybee, 2006). These levels are worth considering in ELF research because of both their distinct characteristics and the way in which they interact with each other. Besides, as Figure 5.1 depicts, the model takes into account the variability of the circles in terms of size and degree of convergence with one another, thus further accommodating the fluidity inherent in ELF.

One aspect of the model in particular, the intersection of the three perspectives, is crucial for making generalizable claims about ELF as a community of somewhat stable characteristics, which is what I call, "the universal perspective" in Figure 5.1. This fourth perspective that emerges from the convergence of the aforementioned three also represents the one crucial factor present at all levels: frequency effects. The impact of larger, virtual, or distant communities (macrosocial level), interactional dynamics (microsocial level), and ELF users' personal linguistic repertoires (individual-cognitive level) on language usage can all be explained via effects of frequency in the diverse, multilingual environments of ELF users (Mauranen, 2012, p. 12).

Macrosocial Perspective	**Microsocial Perspective**
- Communal properties of ELF in the broadest sense - Sources of influence on language use that are not necessarily present in one's context (e.g. virtual/imagined/long-distance/international communities) - Important factors such as similectal groups and academic discourse communities	- Interactional, or relational perspectives of ELF - Second language use in multilingual settings



[a]: This zone refers to the variability of the circles in terms of size and degree of convergence.

Figure 5.1 The Four Perspectives of ELF

While this theorization has great potential for ELF research, it should be remembered that it stems from work on spoken ELF, which is why the sizes and intersections of the three perspectives (represented by circles in Figure 5.1) need to be revisited for different domains and registers of language use such as written academic ELF. In comparison to speaking, the well-established expectations of 'good' academic writing homogenize usage across the three perspectives with the effect of implicit and explicit norms at the communal level such as standard language ideologies, and disciplinary and generic conventions of academic communities (Flowerdew, 2019; Seidlhofer, 2017; Solin & Hynninen, 2018). Thus, the regulating influence of academic discourse is also not a unidimensional phenomenon (Bhatia, 2002a, 2002b). For instance, while editing practices in academic publishing could limit the occurrence of unconventional ELF features (Bondi & Borelli, 2018; Vitali & Bondi, 2020), genres such as academic blogging offer a platform where less NS-normative language may become more visible (Luzón, 2018). As Flowerdew (2019) states, certain communities such as the hard sciences can be more accepting of these features than others (e.g., Martinez, 2018; Rowley-Jolivet, 2017; Rozycki & Johnson, 2013; Tribble, 2017, 2019). In addition, echoing the impact of the intersection of the three ELF perspectives on written academic ELF and adding the factor of individual style Fløttum, Dahl, and Kinn (2006) state:

> […] academic voices are developed through interchange within various kinds of communities which the authors belong to. However, there is also room in academic discourse for each author to develop his or her own individual voice, a voice which may be exploited in very personal ways to achieve the ultimate goal of scientific persuasion (p. 271)

In addition to acknowledging the undisputed reality of the ELF phenomenon shaped by a variety of contextual, ideological, individual and linguistic dynamics, it is important to mention that a remarkable amount of this compelling evidence comes from the analysis of ELF usage. As researchers have repeatedly stated, usage-based approaches are central to the linguistic descriptions of ELF (Alptekin, 2013; Mauranen, 2012, 2018b; Solin & Hynninen, 2018). However, the specific usage-based approach to be implemented in the analysis of ELF should also be theoretically and methodologically sound and flexible enough to accommodate the potential heterogeneity as well as uniformity of ELF. Hence, due to its broad scope and high explanatory power, Construction Grammar (CxG) (Goldberg, 1995, 2003, 2006, 2013, 2019) is the usage-based framework adopted in this study. While there have been partial applications of CxG in written academic (Vetchinnikova, 2015, 2017) and non-academic ELF research (Hall et al., 2017), to my knowledge, Yilmaz and Römer (2020) is the only attempt at describing written academic ELF comprehensively by adopting an inductive, corpus-based approach to the identification of a number of ELF constructions with distinctive usage characteristics. Hence, I claim that exploiting the great affordances of this approach is of major significance for a systematic description of written academic ELF as it is represented in usage data.

5.2 Research Questions

Based on the described research gap, that is, the need to more systematically and comprehensively describe the features of written academic ELF through usage, multiple sources of data, and robust analytical techniques, this dissertation will aim to address the following research questions:

1) What are the dominant constructions distinctive of unedited ELF, edited ELF, and edited ENL?
    a) How do distinctive constructions in ELF compare to those in ENL?

b) How do distinctive constructions in unedited ELF compare to those in edited ELF?

2) How do the identified sets of constructions distinctive of unedited ELF, edited ELF, and edited ENL compare statistically in terms of the background variables of editing status, similects, and disciplinary categories?

3) Based on the answers to research questions 1 and 2, is it possible to determine the constructional profile of "universal ELF," that is, a common and generalizable set of constructions in written academic ELF with little to no influence of the above-mentioned background variables?

The next chapter provides a detailed description of the methodology designed for the proposed study, including the data sets and methods used, and an outline of how each research question is aimed to be answered.

6 METHODOLOGY

This chapter first introduces the data used in the study, that is, the three corpora of edited and unedited research writing by ELF and ENL scholars. It then describes the data analysis procedure, including the extraction and statistical analyses of the constructions chosen to be the foci of the study.

6.1 Data: Corpora Used

Three different corpora were used for this study: (1) a corpus of unedited ELF research articles (henceforth, ELF_Unedited), (2) a corpus of edited and published ELF research articles (henceforth, ELF_Edited), and (3) a corpus of published research articles by native speakers of English (henceforth, ENL_Edited). The naming of the corpora prioritizes the texts' status of editing over publication mainly because the texts in the ELF_Unedited corpus were sampled from WrELFA (2015) which also includes published texts that were not edited for language. While I do not suggest that every published paper is edited for language, "literacy brokering" during publication processes is known to be a common practice that includes language revisions (Hartse & Kubota, 2014; Hyland, 2016a; Hynninen & Solin, 2017; Lillis & Curry, 2006).

6.1.1 The ELF_Unedited Corpus

The articles for the ELF_Unedited corpus come from the SciELF corpus (759,300 words)**,** one of the three sub-corpora of WrELFA (2015), compiled between the years 2011 and 2015. The corpus includes 150 unedited research papers by ELF writers from 10 different L1 backgrounds (Portuguese, Spanish, Chinese, Czech, Finnish, French, Italian, Russian, Romanian, Swedish), and a wide variety of disciplines categorized under two domains as sciences (N=78), and social sciences and humanities (N=72). While the majority of the texts are complete research articles by different writers, several texts were submitted by the same writers, as incomplete drafts, or different genres such as master's theses or book chapters. While 32 of the papers are reported to have been written between the years 2010 and 2013, no dates are available for the remainder of the papers.

To reduce the likelihood of the results being skewed towards certain textual and demographic variables, several texts in SciELF were not included in this study. These texts include the only four

social sciences articles by Romanian writers, three unfinished drafts, and the three texts that were not research articles (a thesis, a dissertation, and a book review) were excluded. The remaining texts form a more balanced corpus. Further, lengthy quotes, text in foreign languages, references, and appendices were also cleaned from the texts. As a result, the ELF_Unedited corpus includes 140 of the SciELF articles written by NNS scholars from nine different L1 backgrounds and makes up 668,110 words. L1 Finnish, Chinese, Czech, and French subsets are the largest in both the total and the domain-based summaries of the corpus (see Table 6.1).

Of the 140 articles, 75 are by authors in science (SCI) disciplines, whereas 65 are by authors in social sciences and humanities (SSH) disciplines. The number of texts is larger in SCI than SSH, but in terms of word counts, the former, with 328,888 words, is slightly smaller than the latter that has in 339,222 words.

As for the individual disciplines in the corpus, there is a remarkable degree of variation in both SCI and SSH. A total of four domains, two under SCI (natural sciences and medicine), and two under SSH (social sciences and humanities), which represent a total of 51 disciplines were defined in the corpus metadata (Please see Appendix A for more information on these disciplines and their distribution across the study corpora). As seen in Table 6.1, natural and social scientific domains have the highest degree of disciplinary diversity and similectal representation, both of which are less evenly distributed in the smaller domains of medicine, humanities, and behavioral sciences. However, since the binary grouping of SCI and SSH still shows distinctive lexicogrammatical qualities (Gray, 2015), the analysis focuses largely on the similarities and differences between the two. The high frequency and dispersion cut-offs used in this study, which are explained in the analysis section, also increased the control for disciplinary variation.

Table 6.1 Numerical Summary of ELF_Unedited

Similect	**Natural Sciences (SCI)**		**Social Sciences and Humanities (SSH)**	
	No. of Texts	**No. of Words**	**No. of Texts**	**No. of Words**
Chinese	11	41,843	10	42,321
Czech	10	45,587	9	39,389
Finnish	15	66,110	9	49,364
French	8	44,820	8	44,224
Italian	6	27,865	5	26,822

Portuguese	6	17,809	6	35,168
Russian	7	32,975	6	35,026
Spanish	6	28,489	6	33,592
Swedish	6	23,390	6	33,316
Total	**75**	**328,888**	**65**	**339,222**

Lastly, as observed by other corpus compilers, there is no one way of categorizing academic disciplines, and major corpora as well as institutions follow different conventions (Krishnamurthy & Kosem, 2007). Hence, the texts were also closely examined so that I could familiarize myself with the disciplines included in the corpus and better understand the disciplinary grouping. For instance, some texts categorized under educational sciences were found to be specifically on English as a Foreign Language (EFL) education. As the articles in ELF_Edited and ENL_Edited were sampled by closely matching the disciplines and topics in ELF_Unedited, I would argue that this stage of detailed examination led to a high level of comparability across corpora.

6.1.2 The ELF_Edited Corpus

The ELF_Edited corpus was compiled from published research articles matching the demographics of ELF_Unedited such as first language backgrounds, disciplines, and topics. Attention was also paid to select articles published after the year 2000. Keywords of the articles in the ELF_Unedited corpus were searched on Google Scholar to identify candidate texts for inclusion in this corpus. Then, the journal impact factors provided by Journal Citation Reports (JCR, 2019), CiteScore™ (CS, 2017) and The SCImago Journal & Country Rank (SCImago, 2019) were used to decide whether these texts were published in top-tier international journals (please see Appendix B for the complete list of articles in ELF_Edited).

There is no known foolproof method of determining the L1 status of the authors whose texts are included in corpora of published academic texts sampled online. However, as done in previous research (e.g., Hyland, 2016a; Marti, Yilmaz, & Bayyurt, 2019; Martinez, 2018; Rowley-Jolivet & Carter-Thomas, 2014; Wood, 2001), publicly available information including the authors' names, educational backgrounds, and institutional affiliations were closely examined to accurately determine the suitability of the texts for the corpus. Hence, the resulting ELF_Edited corpus, while being larger in terms of word

count (820,244 words), matches the ELF_Unedited corpus in terms of the total number of articles and first language backgrounds. As seen in Table 6.2 below, the composition of ELF_Edited is also similar to ELF_Unedited in that the word count of the SCI sub-corpus (404,041 words) is slightly larger than that of SSH (416,203 words) despite the opposite being true for the number of articles (SCI=75, and SSH=65).

Table 6.2 Numerical Summary of ELF_Edited

Similect	Natural Sciences (SCI)		Social Sciences and Humanities (SSH)	
	No. of Texts	No. of Words	No. of Texts	No. of Words
Chinese	11	53,904	10	60,218
Czech	10	67,590	9	53,345
Finnish	15	77,499	9	56,606
French	8	42,901	8	54,061
Italian	6	28,431	5	35,339
Portuguese	6	21,559	6	37,188
Russian	7	45,616	6	47,579
Spanish	6	32,669	6	31,034
Swedish	6	33,872	6	40,833
Total	**75**	**404,041**	**65**	**416,203**

6.1.3 The ENL_Edited Corpus

The third and last corpus of the study, ENL_Edited, contains published texts written by native speakers of English from Inner Circle countries. It was compiled to compare the two ELF corpora with ENL usage. Like ELF_Edited, ENL_Edited was designed to match the ELF_Unedited corpus in terms of the number and proportion of articles in different disciplines and topics. Similarly, measures were also taken to ensure the selection of articles from internationally renowned journals listed in Journal Citation Reports (JCR, 2019), CiteScore™ (CS, 2017), and The SCImago Journal & Country Rank (SCImago, 2019). Hence the three study corpora could be considered highly comparable samples (please see Appendix C for the complete list of articles in ENL_Edited). The resulting corpus is a 918,952-word corpus of 140 research articles written by ENL scholars. Unlike the ELF_Unedited and ELF_Edited corpora, however, the SCI sub-corpus of 75 articles, making up 469,895 words, is slightly larger than the 448,857-word SSH sub-corpus of 65 articles in terms of both word count and the number of texts (see Table 6.3).

Table 6.3 Numerical Summary of ENL_Edited

Matching Similects in ELF Corpora[a]	Natural Sciences (SCI)		Social Sciences and Humanities (SSH)	
	No. of Texts	No. of Words	No. of Texts	No. of Words
Chinese	11	90,785	10	66,067
Czech	10	70,876	9	65,478
Finnish	15	73,583	9	65,593
French	8	49,810	8	73,357
Italian	6	32,291	5	26,557

Portuguese	6	36,878	6	41,433
Russian	7	46,758	6	35,284
Spanish	6	32,287	6	35,743
Swedish	6	36,627	6	39,345
Total	**75**	**469,895**	**65**	**448,857**

[a]The matching of ELF and ENL texts mentioned here refers to that ELF texts were matched with ENL texts in terms of discipline, topic, and research methods used. Thus, every similectal subsets in ELF corpora have a matching subset in ENL_Edited. These matching subsets were primarily used to compare the two ELF corpora with ENL usage.

One important factor to consider in compiling an ENL reference corpus is the potential fallacy of seeing Inner Circle (IC) Englishes as a uniform phenomenon without any variation among them (Mauranen et al., 2010b). However, it is known that American English (AmE) and British English (BrE) are the dominant ENL varieties considered as the written norms of academic publishing (O'Neil, 2018, p. 158). Factors such as textual differences (e.g., register, genre, etc.) have also been reported to be much stronger predictors of lexicogrammatical variation than regional varieties (Biber et al., 1999). Moreover, as World Englishes research shows, especially IC varieties tend to show limited variation in academic writing, which researchers also attribute to other possibly more prominent factors affecting language usage such as textual and disciplinary conventions (Collins, 2008; Hundt, G., & Seoane, 2016; Seoane & Hundt, 2018).

Following the above-mentioned reasons and challenges, the ENL_Edited corpus was analyzed primarily as a uniform data set. Still, the impact of potential differences between IC varieties were considered when necessary via comparisons with either published research findings (Biber et al., 1999; Hundt et al., 2016; Liu, 2012; Seoane & Hundt, 2018) or corpus searches using the International Corpus of English (Greenbaum & Nelson, 1996). Previous ELF research (e.g., Bardi & Muresan, 2014; Ranta, 2009) has also drawn on and greatly benefitted from such comparisons.

Another limitation is that a second ENL corpus of unedited research writing was not included in this study. Unfortunately, there is no known public database from which such a corpus could be compiled. Besides, the fact that an ideal corpus of unedited research writing should also match the content of the other three corpora described above would double the challenge. Nevertheless, when available, such unpublished native speaker writing could be a useful addition to a one's research design,

enabling more fine-grained analyses of ELF-ENL differences, as demonstrated by Marti et al. (2019) and Römer (2009a).

Following is a detailed description of how these corpora were used to analyze the distinctive characteristics of constructions commonly used in written academic ELF.

6.2 Analysis

The data analysis for the study was a multi-step procedure that involved three main stages, each of which corresponds to one of the three main research questions of the study (see Table 6.4).

Table 6.4 Data Analysis Procedures per Research Question

Research question	How it was addressed
1. What are the dominant constructions distinctive of unedited ELF, edited ELF, and edited ENL?	Identifications of constructions via "key function words to constructions" (KFWs to Cxs) method
1a. How do distinctive constructions in ELF compare to those in ENL?	Comparisons of distribution, dispersion and keyness values of constructions in ELF_Unedited and ELF_Edited with ENL_Edited
1b. How do distinctive constructions in unedited ELF compare to those in edited ELF?	Comparisons of distribution, dispersion and keyness values of constructions between ELF_Unedited and ELF_Edited
2. How do the identified sets of constructions distinctive of unedited ELF, edited ELF, and edited ENL compare statistically in terms of the background variables of editing status, similects, and disciplinary categories?	Comparisons of construction frequencies via MANOVA tests, followed by univariate ANOVAs and posthoc pairwise tests between factor levels Reporting of distribution and dispersion of constructions and semantic domains proportions for their lexical fillers
3. Based on the answers to research questions 1 and 2, is it possible to determine the constructional profile of "universal ELF," that is, a common and generalizable set of constructions in written academic ELF with little to no influence of the above-mentioned background variables?	Reporting of the overall ELF-ENL differences that are not affected by similectal and disciplinary variation, as well as editing.

6.2.1 Addressing RQ.1: The "Key Function Words to Constructions" (KFWs to Cxs) Method

A multi-step, inductive method called the "key function words to constructions approach" (henceforth, "KFWs to Cxs"), adapted from Groom (2009, 2010, 2018), was used for the extraction of dominant constructions from the study corpora. The method could be summarized in three main steps: 1) identifying the function words (FWs) that were significantly more frequent in one corpus in comparison to the other two corpora sequentially; 2) extracting contiguous and non-contiguous patterns containing those FWs; and 3) using these patterns to identify dominant constructions in each study corpus through pairwise comparisons. The three pairwise comparisons included: (1) ELF-Unedited - ENL_Edited, (2) ELF_Edited - ENL_Edited, and (3) ELF_Unedited - ELF_Edited. The constructions identified at this stage also formed the basis of the following statistical comparisons of constructional usage across and

within the corpora. After presenting a review of research on key words (KWs) and function words to justify the "KFWs to Cxs" method in general, this section describes the steps of the analytic method in detail.

6.2.1.1 Research on Keyness and Function Words

Key words (KWs) are words that are statistically more (positive KWs) or less frequent (negative KWs) in a target corpus in comparison to a relevant reference corpus (Scott, 1997). Previous research (Baker, 2004a, 2004b, 2006) has mostly focused on key words that are content words, including nouns, verbs, adjectives, and adverbs. Key content words (KCWs) are generally used to explore the "aboutness" of a group of texts (Phillips, 1989; Scott, 1997; Scott & Tribble, 2006). Key function words (KFWs), on the other hand, are less commonly studied using KW analysis. Unlike content words, function words such as prepositions, articles, and conjunctions are generally those "which have little meaning on their own, but which show grammatical relationships in and between sentences (grammatical meaning)." (Richards & Schmidt, 2010, p. 126).

Function words, in general, have been instrumental in better understanding the phraseology of academic writing. For instance, lexical bundle research focused on noun phrases modified by prepositions to explain differences between writing and speaking (Biber & Conrad, 1999; Biber & Gray, 2010; Biber, Gray, & Poonpon, 2011; Biber et al., 1999), student and expert writing (Ansarifar, Shahriari, & Pishghadam, 2018; Benelhadj, 2019; Chen & Baker, 2010), and expert L1 and L2 writing (Esfandiari & Barbary, 2017; Pan et al., 2016). In addition to being highlighted in studies that focused on contiguous word sequences such as lexical bundles, function words have also been central to research on non-contiguous sequences like phrase- or lexical frames (Biber, 2009; Gray & Biber, 2013; Römer, 2010). An example frame would be *the # of the*, in which three function words frame a content word (here a noun such as *results*, *findings*, etc.). From an ELF perspective, research also documented a number of somewhat unconventional uses of FWs such as articles, prepositions, *that* in *that*-clauses, modal verbs, copular BE, relative pronouns, and negation markers (Björkman, 2008a, 2008b; Cogo & Dewey, 2006;

Mauranen, 2012, 2016; Ranta, 2009; Seidlhofer, 2004, 2009a; Vetchinnikova, 2015; Wang, 2018). Lastly, recent research on the analysis of changes in academic writing pointed to the frequent use of noun phrases modified by prepositional phrases, as well as coordinating and subordinating clauses as clear indicators of the increasing structural complexity of academic writing(Biber & Gray, 2016; Gray, 2015; Liardet, Black, & Bardetta, 2019).

Focusing on function words becomes even more critical when they are identified as key words in a corpus. the much smaller number of KFWs compared to KCWs might signal remarkable stylistic, generic and discoursal differences between the target and reference corpora of a study (Archer, 2009; Groom, 2010; Scott, 1997; Scott & Tribble, 2006; Xiao & McEnery, 2006). KFWs have been used in research on academic discourse as the starting point of the analysis of lexicogrammatical patterns (Bruce, 2015; Gledhill, 2000a, 2000b, 2011; Groom, 2009, 2010; Hunston, 2010; Whiteside & Wharton, 2019).

KFWs are especially important in Construction Grammar (CxG) research. As Fillmore (1987, as cited in Goldberg, 2995, p. ix) argued, the "meaning of an expression is arrived at by the superimposition of the meanings of open-class words with the meaning of grammatical elements." Thus, Groom's (2018) use of a method to identify constructions from corpora in an inductive way by starting from KFWs appears particularly relevant for this study. A similar approach was adopted by Hall et al. (2017) in an analysis of non-academic ELF writing centered on the KFW *can.* More recently, in Yilmaz and Römer (2020), the use of this method was extended to the identification of constructions typical of written ELF by using a range of corpus methods including the extraction of KFWs, their collocations, and the contiguous and non-contiguous word sequences in which they occur. Thus, this study builds on previous research with further improvements explained in the next section.

6.2.1.2 Identification of Constructions via the "KFWs to Cxs" Approach

LancsBox 4.5 (Brezina, Timperley, & McEnery, 2018) was used for most of the corpus analyses in the study. Prior to the analysis, the corpora were lemmatized and part-of-speech (PoS)-tagged automatically in LancsBox, which uses TreeTagger developed by Schmid (1994). As seen in Table 6.5, PoS tags were used to refer to parts of speech individually and in sequences that were also instrumental in extracting

grammatical patterns such as *noun-preposition-noun* ('N PREP [NP]'). In addition, raw frequencies as well as relative (normalized) frequencies per 10,000 words were automatically retrieved from LancsBox.

Table 6.5 Part-of-Speech (PoS) Tags Used in the Study

Part-of-Speech (PoS) Tag	**What it refers to**	**Example**
ADJ	All adjectives (central, comparative, superlative)	*good, better, best*, etc.
ADJ-SUPERL	Superlative adjective	*best*, etc.
DET	Determiner	*the, a, this, each*, etc.
N	All nouns (singular, plural, proper)	*book, books, Boston*, etc.
N-s	Plural noun	*books, people*, etc.
NN	Singular noun	*book, person*, etc.
(NP)[a]	Noun group: determiner, pronoun, adjective, and participle used in place of nouns	*role of* (NP) NP*: subject/this/exercising*, etc.
V-*be*	Verb *be* (all forms)	*be, was, were, am, is, are, been, being*
V-*ed*	Past participle	*forgotten*, studies, etc.
PREP	Preposition	*in, of, to, under*, etc.

Note. Tags are primarily based on Schmid (1994), partially adapted following Francis et al. (1996, 1998) and Goldberg (2006)
[a] Parentheses refer to the optionality of the PoS at the end of a search term.

In the first step of the analysis, lists of KFWs were extracted for each corpus in comparison to the other two. For KFW analyses, only function words included in the comprehensive, corpus-based list by Dang and Webb (2016) were selected as the focus of the analysis. Log-likelihood (LL) with a cut-off probability value of 0.001 (G^2=10.83) was used for significance testing, and differences between normalized frequencies (%DIFF, Gabrielatos & Marchi, 2012) were calculated to report effect sizes of the comparisons. High cut-off values and the reporting of effect sizes were employed mainly to control for the multiple pairwise comparisons were made across the three datasets. However, it is known that these statistical methods do not take into account the dispersion of the items being tested and are heavily influenced by the size of corpora, which might lead to inflated results (Bestgen, 2018; Brezina & Meyerhoff, 2014; Lijffijt, Nevalainen, Säily, Papapetrou, Puolamäki, & Mannila, 2016; Paquot & Bestgen, 2009). Hence, the keyness analysis was complemented by the probability values of Welch's independent-sample t-test, which Brezina (2018) argues to be appropriate for most corpus data as it controls for unequal variance. LancsBox automatically calculates this test based on frequencies of the same search word or pattern across two groups. The frequencies are set to be normalized on the basis of the number of words per text, which is the setting also used for this stage of the study. Lastly, because running multiple tests of comparisons is conducive to making false assumptions of significance, the Benjamini and Hochberg (1995) method was used to adjust the default alpha value of .05. The ranked p-

values from pairwise comparisons were adjusted on the bases of the number of total tests conducted, and a false discovery rate (FDR) of .05.

Once the KFW lists were extracted, MWUs of 2 to 6 n-grams, and PoS-grams (Stubbs, 2007) including the KFWs were extracted and keyness scores were calculated for these patterns as well. Since the lexicogrammar of academic writing is highly variable and contiguous MWUs such as n-grams are not as common in it as in speech (Biber, 2009), PoS-grams were instrumental to the identification of constructions from the candidate MWUs in this study. Recent research also used similar methods to facilitate the investigation of a large number of patterns in a summative fashion (e.g., Brett & Pinna, 2015; Gilquin, 2018; Pinna & Brett, 2018). In addition, to reduce the number of the tags for similar word classes that are represented separately in PoS tags, BFSU PowerConC 1.5 (Jiajin, Liang, & Jia, 2012) was used. PowerConC was helpful in identifying general PoS-grams and keyness scores of differences across corpora as it enabled the simplification of the TreeTagger tag set. For instance, thanks to the reduced tagging, singular and plural nouns, as well as verbs with different inflectional markings for person and tense could be merged when necessary.

Collocates of the KFWs were also extracted to potentially identify additional construction types or variants of already identified contiguous units. For the collocation analyses, following Danielsson (2007), the span was set to between three and nine words, that is a minimum one and a maximum of four words to the right and left of the KFWs. Collocates of the KFWs were identified based on the frequencies of co-occurrence, and ΔP (delta P) scores. ΔP was helpful in showing the directional strength of associations between constructions and their lexical fillers as it includes two scores; that is, one (ΔP1|2) for how well the search word attracts a collocate, and another (ΔP2|1) for the opposite attraction going from collocate to search word. ΔP has been argued to be a robust measure of association in human learning in general, and language acquisition in particular (Ellis, 2006; Ellis & Ferreira-Junior, 2009b; Gries, 2013). Further, unlike measures such as log-likelihood, it is not affected by the size of the corpus in which a collocation occurs (Gries, 2020, p. 109). A ΔP value close to 1.0 points to a highly powerful positive association, while -1.0 means the opposite, that is, negative association between the node and

collocate. To my knowledge, there is no commonly used minimum threshold for ΔP values as in the case of the mutual information score of 3.0. While Schneider (2018) found that strong collocations in her study had a minimum value of, a threshold also used by Brezina, McEnery, and Wattam (2015), Wahl (2015) indicates that his data did not point to such a threshold. In addition, both authors argued that high-frequency items such as function words tend to have very low ΔP values. Hence, in this study, after examining the 10 most frequent collocates of KFWs, any collocate with a minimum ΔP1|2 value of 0.001 was included in the functional analysis of especially high-frequency and polysemous patterns. Using the lowest possible attraction score enabled a comprehensive analysis that, at the same time, ensured at least a certain level of associational strength between KFWs and the surrounding words. Further, ΔP1|2 was chosen over the 2|1 value in line with the primary purpose of identifying constructions based on KFWs in this study.

As also done in previous research (e.g., Hyland, 2008a; Hyland, 2008b; Liu, 2012), the final set of constructions was required to meet a frequency cut-off of 20 per million words, and a range cut-off of 10%. The cut-offs meant that the constructions had to occur in at least 14 of the 140 texts in each corpus with minimum frequencies of 12, 16, and 18 respectively for ELF_Unedited, ELF_Edited, and ENL_Edited. Finally, concordances of the patterns were analyzed to examine their constructional status, that is, whether the grammatical patterns were systematically linked to semantic or functional features. In order to report the dispersional characteristics of the constructions, normed deviation of proportions (DP_{norm}; see Gries [2008)] for corpora with unequally sized subcorpora was calculated. This is a normalized measure between 0 and 1, where 0 indicates perfect dispersion, and 1 means a completely uneven dispersion. The calculation takes into account the differences between observed and expected proportions of each subcorpus. Gries (2010) also provided evidence for the validity of this measure of dispersion. Gries's (2020) categorization includes a DP_{norm} value of 0.08 - 0.24 as "minimal", 0.4-0.8 as moderate and 0.94-0.1 as "maximal" values (p.117).

Once the frequency-based identification of the constructions in each of the three corpora was completed, several measures of summarizing the frequency and distribution of the slot fillers in

constructions were employed to examine the productivity and predictability of the lexicogrammatical realizations of the constructions. These included ΔP (delta P) scores for documenting the strength of association from frequent fillers to constructions and vice versa, as well as normalized entropy (H_{norm}) scores for better understanding the nature of the overall distribution of the lexical fillers in a construction (Gries & Ellis, 2015). As Ellis and Gries (2015) state, entropy is a measure between 0 to 1 and the closer the score gets to 0, the more predictable and Zipfian a distribution gets (p. 235). Zipf's (1935) law applies to all natural language patterns where an inverse relation between the frequency and rank of a distribution is observed. Existence of such a relationship has been shown to have high explanatory power about the prototypicality and stability of certain patterns in learner language (Ellis, 2012), as well as in ELF (Vetchinnikova, 2017).

6.2.2 Addressing RQ2: Comparisons of Construction Usage across Corpora and Background Variables

In order to address the second research question on whether there are significant differences between the study corpora that can be explained by the factors of ELF status, editing status, broad disciplinary category and first language background, a multiple factorial analysis of variance test across group means was conducted. The normalized token frequencies of the individual constructions; that is, the dependent variables, were statistically compared across the independent variables including the three corpora, 10 L1 backgrounds and the two large disciplinary categories.

Normalization of raw frequencies is a straightforward and common procedure for comparing corpora of different sizes. As Brezina (2018) argues, normalization can be done by factoring in the sizes of corpora, subcorpora, or individual texts. However, Gries (2020) suggests that for an accurate representation of dispersion, sizes of the smallest meaningful units in corpora should be used (p. 109). Hence, while normalization was done using the sizes of individual texts at the exploratory (whole-corpus) stage (RQ2), the sizes of smallest similectal and disciplinary subsets (e.g., hard sciences texts by Chinese ELF writers, soft sciences texts by French ELF writers) were used for the more fine-grained comparisons

that included the factors of similectal and disciplinary grouping. This was considered a necessary adjustment due to the great discrepancies across the subsets, ranging from five texts in the Italian L1 text in social sciences and humanities, to 75 ENL texts in hard sciences.

Since the aim is to compare the groups of writers in terms of their construction usage, Analysis of Variance (ANOVA) was used for the study. ANOVA is suggested as a robust statistical technique that, unlike aggregate methods of significance testing (e.g., Log-likelihood), takes individual variation across data points into account (Brezina, 2018; Brezina & Meyerhoff, 2014). However, because the aim was to test the relation between multiple categorical independent variables (N=3, see Table 6.6) and multiple continuous dependent variables (construction frequencies), multivariate and multifactorial ANOVAs (MANOVAs) were used to measure the differences in construction frequencies across factor levels. Such multifactorial statistical testing is of fundamental importance since, as Baayen (2013) explains, "many phenomena can only be understood properly when a great many explananda are considered jointly" (p.1).

Table 6.6 Description of Independent Variables as Factors

Factor	**Level**	
	Number	**Label**
Corpus	3	ELF_Unedited, ELF_Edited, ENL_Edited
Discipline	2	SCI, SSH
Similect (L1)	10	Chinese, Czech, English, Finnish, French, Italian, Portuguese, Russian, Spanish, Swedish

The software environment R (R Core Development Team, 2019) was used for the MANOVAs. Of the R packages including MANOVA, "rankMANOVA" (Friedrich, Dobler, & Pauly, 2019) was chosen as the only known statistical package accommodating the multifactorial and multivariate analyses of non-normal and heterogenous data, as opposed to traditional MANOVAs where conservative parametric assumptions need to be met. Further, in order to increase the robustness of the analysis, the rank-based, Mann-Whitney-type MANOVA offered in this package, also includes bootstrapping; that is, reporting a mean probability value as a result of multiple reiterations of the same test by creating different samples from the data for each. As LaFlair, Egbert, and Plonsky (2015) argue, bootstrapping ANOVAs results in outcomes nearly as powerful as their parametric equivalents for small sample sizes and non-

parametric data. Since some levels of the factors in this study such as similectal groups are quite small and disproportionate across levels, and the data were found to violate the assumptions of heterogeneity of variance and multivariate normality, this test was considered a good fit for this study. The "rankMANOVA" function of the package was used for one-way and factorial MANOVAs, as well as the follow-up univariate comparisons. The bootstrapping feature was set to 10,000 iterations using the wild bootstrap (WildBS) approach. First, a one-way MANOVA was run to examine the differences between the three study corpora. Then two- and three-way ANOVAs were run to investigate the role of similectal and disciplinary groups in explaining the usage differences within and across the corpora. Following the developers of the package (D. Dobler, S. Friedrich, & M. Pauly, 2019; Friedrich et al., 2019) who state that alpha correction for the multiple follow-up tests is not necessary due to the nature of the test, the univariate ANOVAs were not corrected.

As for the post-hoc pairwise statistical comparisons, Kruskal-Wallis-based, bootstrapped posthoc tests provided in the "postHoc" R package (Labouriau, 2020) were run. As in the case of MANOVAs, 10,000 cycles of bootstrapping were used. Instead of the Benjamini and Hochberg (1995) procedure used in order to correct p-values for multiple comparisons across corpora, the Benjamini and Yekutieli (2001) procedure was used for this phase of the analysis because it is more sensitive to potential dependencies between the groups. As the large number of similectal and disciplinary subsets are likely to be moderately correlated with one another, this procedure was considered more accurate for minimizing FDR at this phase of the analysis, where 69 subsets based on the interactions between factor levels were taken into account.

For descriptive purposes, due to the unequal sizes of the corpus subsets, normalized median frequencies were reported for each subset instead of the sum of normalized frequencies reported in cross-corpus comparisons (for RQ1). The median frequencies were also provided as bootstrapped values by the posthoc tests. Normalized deviation of proportion (DP_{norm}) were also reported to take a closer look at the dispersion of constructions across subsets.

All contrasts were included in the analysis to produce visualizations of clustering plots based on the confidence intervals reported in these tests. The visualizations are comprised of clusters of groups that have similar distributions, and the positions of clusters in relation to one another visualize the degree of similarity, as well as the difference between them. The example plot provided by Labouriau (2020), the "postHoc" package developer, provides a plot for a treatment factor with five levels, A to E (see Figure 6.1). There are two clusters based on six different treatment types in Figure 6.1. One cluster includes treatments C, D, G, and the other has B, E, and D. The least amount of difference between the clusters is in treatment D, which is placed at the border of both clusters. However, maximal differences are observed between the treatments B and E in one cluster that are furthest from C and G in the other (Labouriau, 2020). With the large number of subsets involved in pairwise comparisons in this study, this plotting enabled a clearer interpretation of differences across groups.

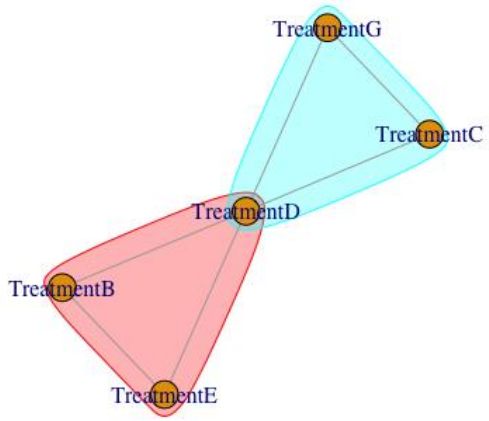


Figure 6.1 An Example of Clustering Based on the Plotting of Pairwise Comparisons in "postHoc" R package. From "Post-hoc Analysis Using the Package postHoc" by R. Labouriau (https://cran.r-project.org/web/packages/postHoc/vignettes/Post-hoc-analysis.html)

While all pairwise comparisons were calculated for visualization, not all significant differences were interpreted exhaustively. It is common to focus on certain contrasts over others in pairwise comparisons based on their relevance to the objectives of the study (Field, Miles, & Field, 2012; Levshina, 2015). Similarly, this portion of the study, while visualizing the grouping of subsets based on all contrasts, focused on two main types of contrasts to explain similectal and disciplinary variation in written academic ELF: (1) differences within corpora to capture corpus-internal variation; and (2) differences

across corpora to capture cross-corpus variation across relevant subsets. The excluded contrasts include those that report significant differences between subsets from different corpora and disciplines (e.g., SCI texts in ELF_Edited versus SSH texts in ENL_Edited), between different corpora and similects (e.g., unedited French ELF versus edited ENL texts matching Finnish ELF), and simultaneously between different disciplines and similects (e.g., SCI texts by Spanish ELF versus SSH texts by Chinese ELF). The decision to exclude these contrasts stems mainly from the fact that these subsets did not include comparable texts. Hence, their results would not be theoretically relevant to the study.

As for the analysis of constructional meanings, the proportions of semantic domains reported for each corpus in the previous chapter were analyzed in detail across the subsets within and across corpora. Since subdomains sometimes included very low and under-dispersed frequencies, only main domains were included in the analysis. To illustrate, the two main domains identified for 'N *of* (NP)' were abstract and concrete head nouns. While the abstract domain was divided into five subdomains including general, process, framing, cognitive and research nouns, the concrete domain included quantity nouns as a subdomain, and other less common nouns referring to humans, animals, and entities as another. With several of the subdomains being rather infrequent, only the domains of abstract and concrete nouns were used for the analysis of 'N *of* (NP)' for this chapter.

The examination of domain differences has two critical advantages. First, it allows us to gain deeper insights into constructional usage by factoring in the role of polysemy (Goldberg, 1995, p. 75). Second as Hoffmann (2004, 2005) argues, frequencies of broader concepts could complement those of individual forms, especially in the case of low-frequency items. Lastly, for easier comparisons across groups of unequal sizes, percentages of domain frequencies, as opposed to the total number of collocates annotated for meanings were reported. For instance, if semantic domains *a* and *b* were identified for a construction *X*, the total frequency of domains, that is $a + b$, was taken as 100% to examine the relative proportions of each domain at the same scale (e.g., $a = 54\%$, $b = 46\%$).

6.2.3 Addressing RQ3: Identifying ELF "Universals"

With the aim of answering the third research question on whether certain core ELF features could be identified regardless of different background variables, ELF vs. ENL differences that were found not to be significant across disciplinary and similectal backgrounds of the writers in the previous analyses (on RQs 1 and 2) were examined in further detail. The core features of ELF academic writing previously reported in research such as increased pervasiveness of distinctive lexicogrammatical choices regardless of similectal backgrounds (Martinez, 2018), a high degree of explicitness and varying degrees of structural complexification as well as simplification (Mauranen, 2019; Wu et al., 2020; Yilmaz & Römer, 2020), as well as the tendency to rely on a relatively limited number of high-frequency patterns (Carey, 2013) were revisited by taking a closer look at the more generally used constructions found in this study in terms of constructional frequencies and functions.

6.3 Summary

This chapter described in detail the three study corpora and the multi-step procedure followed to achieve the primary goal of this study; that is, adding to our understanding of the linguistic characteristics of written academic ELF from a usage-based, and empirically robust constructionist perspective. First, a summary of the characteristics of the study corpora ELF_Unedited, ELF_Edited and ENL_Edited were provided. Next, an overview of the data analysis, followed by further details as to how it was utilized to address the research questions of the study was provided.

The next three chapters report the findings of the study by focusing on each research question separately. The next chapter (Chapter 7) reports the results of the "KFWs to Cxs" method used to identify the target constructions in this study on the basis of statistically significant differences in the use of function words, and the patterns in which they commonly occur.

7 DOMINANT CONSTRUCTIONS IN UNEDITED AND PUBLISHED WRITTEN ACADEMIC ELF

7.1 Identified Key Function Words (KFWs): Prepositions across Corpora

As explained in the previous chapter, the "key function words to constructions" (KFW to Cxs) approach was used to identify the distinctive constructions in the study corpora. Of the function words identified in the three study corpora, prepositions were found to be the most common, also showing remarkable usage differences. 26 prepositions (*across, after, at, below, beyond, by, except, for, in, inside, near, of, off, versus, on, out, over, per, through, throughout, to, under, upon, versus, via, within*) were found to have significantly different frequencies in 36 pairwise comparisons across the three corpora. Six prepositions were found to be significantly more frequent in ELF_Unedited than in ENL_Edited, whereas 13 prepositions were identified as KFWs in ENL_Edited in comparison to ELF_Unedited (see Table 7.1). 15 key prepositions were identified when ELF_Edited and ENL_Edited were compared, six of which were found to have higher frequencies in the former, with the other nine being more frequent in the latter (see Table 7.2).

Table 7.1 Key Prepositions in ELF_Unedited and. ENL_Edited Sorted by Log-Likelihood (G^2) Values

	KFW	**ELF_Unedited**			**ENL_Edited**			G^2	**%DIFF**	p^a
		Raw	**Normed (per 10k)**	**DP_{norm}**	**Raw**	**Normed (per 10k)**	**DP_{norm}**			
ELF_Unedited vs. ENL_Edited	*of*	28,501	426.59	0.08	35,940	391.18	0.08	118.92	9.05	.013
	by	4,325	64.73	0.15	5,189	56.48	0.13	43.69	14.62	.013
	out	486	7.27	0.34	447	4.87	0.38	37.63	49.51	.029
	after	582	8.71	0.38	606	6.60	0.41	22.85	32.07	.010
	except	80	1.20	0.65	54	0.59	0.72	16.74	103.73	.042
	under	350	5.24	0.47	364	3.96	0.47	13.85	32.23	.013
	KFW	**ENL_Edited**			**ELF_Unedited**			G^2	**%DIFF**	p^a
		Raw	**Normed (per 10k)**	**DP_{norm}**	**Raw**	**Normed (per 10k)**	**DP_{norm}**			
ENL_Edited vs. ELF_Unedited	*across*	542	5.90	0.49	121	1.81	0.72	171.67	225.73	.006
	versus	151	1.64	0.73	19	0.28	0.91	78.85	477.93	.006
	at	3,486	37.94	0.25	2,035	30.46	0.27	63.11	24.57	.041
	upon	212	2.31	0.57	57	0.85	0.81	52.48	170.46	.006
	within	941	10.24	0.37	460	6.89	0.45	50.71	48.76	.011
	for	9,828	106.97	0.14	6,408	95.91	0.18	46.52	11.53	.039
	near	165	1.80	0.66	42	0.63	0.78	44.19	185.68	.019
	per	378	4.11	0.63	151	2.26	0.62	41.69	82.04	.041
	via	212	2.31	0.54	72	1.08	0.69	34.70	114.12	.035
	over	764	8.32	0.32	391	5.85	0.50	33.00	42.09	.011
	beyond	150	1.63	0.54	53	0.79	0.74	22.53	105.81	.013
	below	155	1.69	0.55	59	0.88	0.72	19.47	91.04	.013
	throughout	148	1.61	0.57	77	1.15	0.69	11.37	39.77	.041

[a]P-value for Welch's two-tailed t-test

Table 7.2 Key Prepositions in ELF_Edited and. ENL_Edited Sorted by Log-Likelihood (G^2) Values

	KFW	ELF_Edited			ENL_Edited			G^2	%DIFF	p^a
		Raw	Normed (per 10k)	DP_{norm}	Raw	Normed (per 10k)	DP_{norm}			
ELF_Edited vs. ENL_Edited	*of*	35,100	427.92	0.09	35,940	391.18	0.08	142.99	9.39	.019
	in	22,300	271.87	0.10	23,167	252.16	0.09	64.33	7.82	.041
	on	5,268	64.22	0.14	5,186	56.45	0.16	43.54	13.78	.013
	after	711	8.67	0.39	606	6.60	0.41	24.52	31.42	.017
	by	5,073	61.85	0.15	5,189	56.48	0.13	21.14	9.51	.041
	inside	26	0.73	0.76	60	0.65	0.89	17.95	12.01	.013
	KFW	**ENL_Edited**			**ELF_Edited**			G^2	%DIFF	p^a
		Raw	Normed (per 10k)	DP_{norm}	Raw	Normed (per 10k)	DP_{norm}			
ENL_Edited vs. ELF_Edited	*across*	542	5.90	0.49	150	1.83	0.62	193.55	222.59	.006
	for	9,828	106.97	0.14	7,815	95.28	0.15	58.58	12.27	.006
	near	165	1.80	0.66	57	0.69	0.75	43.30	158.44	.027
	over	764	8.32	0.32	491	5.99	0.41	32.93	38.92	.029
	through	842	9.16	0.32	559	6.82	0.37	29.97	34.48	.010
	upon	212	2.31	0.57	105	1.28	0.66	25.72	80.26	.013
	beyond	150	1.63	0.54	60	0.73	0.68	25.59	123.20	.034
	to	8,907	96.65	0.11	7,361	89.74	0.10	24.09	7.43	.006
	within	941	10.24	0.37	673	8.20	0.36	19.50	24.83	.041

[a]P-value for Welch's two-tailed t-test

'Of', 'by', and '*after'* are shared key prepositions with higher frequencies in comparisons of ELF_Unedited and ELF_Edited with ENL_Edited. Regarding the significantly more frequent prepositions in ENL_Edited compared to both ELF corpora, *'across', 'for', 'near', 'over',* and '*beyond'* are shared across both pairwise comparisons. As it is also clear from the comparisons of ELF data with ENL_Edited, there are fewer key prepositions in ELF as opposed to ENL_Edited, but these KFWs are much more frequent in general. While the G^2 values point to highly significant differences for these frequent prepositions, the %DIFF values point to relatively smaller-scale differences as opposed to the lower-frequency items in the lists.

Lastly, as Table 7.3 shows, '*off*' and '*versus'* were found to be the only two key prepositions in the pairwise comparisons between ELF_Unedited and ELF_Edited. Of these two KFWs, '*versus'* is especially interesting as the ELF_Unedited corpus was reported to have significantly fewer instances than both ELF_Edited and ENL_Edited. However, both prepositions are rather infrequent and poorly dispersed across the texts in the corpora.

Table 7.3 Key Prepositions in ELF_Unedited vs. ELF_Edited Sorted by Log-Likelihood (G^2) Values

ELF_Unedited vs. ELF_Edited	**KFW**	**ELF_Unedited**			**ELF_Edited**			G^2	%DIFF	p^a
		Raw	Normed (per 10k)	DP_{norm}	Raw	Normed (per 10k)	DP_{norm}			
	off	35	0.52	0.80	13	0.16	0.93	15.49	230.54	.041
ELF_Edited vs. ELF_Unedited	**KFW**	**ELF_Edited**			**ELF_Unedited**			G^2	%DIFF	p^a
		Raw	Normed (per 10k)	DP_{norm}	Raw	Normed (per 10k)	DP_{norm}			
	versus	93	1.13	0.80	19	0.28	0.91	39.27	298.69	.014

[a]P-value for Welch's two-tailed t-test

Building on these keyness results, the remainder of the chapter provides a detailed look into the findings of the "KFWs to Cxs" method, including the form and functions of 19 identified key constructions (Cxs), and how the findings relate to previous research in the field.

7.2 Identified Construction Candidates

The 19 distinctive constructions identified through the "KFWs to Cxs" method, are presented in five structural categories: (1) nominal ('N PREP [NP]'), (2) prepositional ('PREP N'), (3) adjectival ('ADJ PREP [NP]'), (4) passive ('[NP] V-*be* V-*ed* PREP [NP]') and (5) determiner-based ('DET *of* [NP]') constructions. Due to the cut-off thresholds of frequency, dispersion, as well as the significance requirement in multiple statistical tests, several of the low-frequency KFWs listed in the previous section (e.g. *'beyond,' 'except,' 'throughout,'* etc.) did not lead to the identification of any constructions. However, several of the highly frequent KFWs such as *'of'* and '*in'* were instrumental in the identification of more than one construction.

7.2.1 Constructions with Key Prepositions as Post-modifiers of Nouns: 'N PREP (NP)'

Some of the most remarkable differences across the corpora were identified in the use of nominal constructions with prepositional post-modifiers; that is, 'N PREP (NP)'. These five constructions, 'N *of* (NP)' (e.g., '*number/part of* [NP]'), 'N *of* NP *in* (NP)' (e.g., *role/number of* NP *in* [NP]), 'N *of* NP *of* (NP)' (e.g., '*analysis/part of* NP *of* [NP]'), 'N *for* (NP)' (e.g., '*value/data for* [NP]'), 'N *to* (NP)' (e.g., *'response/approach to* [NP])'), are of great importance for two reasons. First, research shows that expert academic writing is remarkably more phrasal than clausal, and nominal phrases make up the predominant phrasal structure in texts (Biber et al., 2011). Second, complex nominals, in general, are quite commonly utilized in ELF academic writing as a means of efficient and densely structured delivery of information (Wu et al., 2020). The pattern starts with a noun (N), singular, plural, or proper, followed by a prepositional phrase (PP), and ends with a noun phrase (NP). The difference between N and NP is that the former was specified as a noun (singular, plural or proper) in corpus searches, whereas NP stands for a noun group, that is nouns or other parts of speech that could be used in place of nouns. Following similar

groupings of different word classes by Francis et al. (1996, 1998) and Halliday (1994), 'NP', however, stands for not only nouns but also any word that can be used in place of a noun such as a determiner, pronoun, adjective, and participle. Further, as this slot was not specified in searches it was given in parentheses as '(NP)'.

As seen in Table 7.4, the five 'N PREP (NP)' constructions with significant differences across corpora employ three of the key prepositions '*of', 'for',* and *'to.'* Except for 'N *of* NP *in* (NP)*'* that does have a significantly higher frequency in ELF_Unedited than in ENL_Edited, the three nominal *of*-patterns are significantly more frequent in both ELF corpora than in ENL_Edited. For 'N *for* (NP) and 'N *to* (NP)', however, ENL_Edited was found to have significantly higher frequencies than both ELF corpora. As highly common patterns, all five construction have DP_{norm} values close to 1.0, indicating they are similarly well-dispersed items in the three corpora.

As the H_{norm} scores indicate, the distributions of head nouns as the lexical fillers of these constructions are similarly quite unpredictable across the corpora despite the above-described differences across corpora in terms of the construction frequencies. However, especially the normalized entropy scores suggest slightly more predictable distributions in the three corpora for 'N *of* (NP)', and 'N *to* (NP)' than for the other nominal constructions. Further, ELF_Edited and ENL_Edited have generally similar scores that are slightly lower than ELF_Unedited. This implies that the distributions of the head nouns in these nominal constructions are slightly more predictable in edited texts by both ELF and ENL writers.

Table 7.4 Numerical Summary of N PREP (NP)

	Cx	**ELF_Unedited**				**ELF_Edited**				**ENL_Edited**			
		Raw	**Normed (per 10k)**	**DP_{norm}**	**H_{norm} (nouns)**	**Raw**	**Normed (per 10k)**	**DP_{norm}**	**H_{norm} (nouns)**	**Raw**	**Normed (per 10k)**	**DP_{norm}**	**H_{norm} (noun)**
Frequencies	N *for* (NP)	3,866	57.99	0.21	0.91	4,899	59.72	0.17	0.90	6,355	62.91	0.15	0.89
	N *of* (NP)	35,624	384.37	0.09	0.85	31,737	386.89	0.09	0.84	31,879	346.83	0.08	0.83
	N *of* NP *in* (NP)	1,174	17.57	0.25	0.93	1,640	19.99	0.22	0.90	1,392	15.15	0.22	0.91
	N *of* NP *of* (NP)	1,257	18.81	0.27	0.92	1,529	18.64	0.29	0.91	1,105	12.03	0.27	0.91
	N *to* (NP)	1,998	29.91	0.21	0.87	2,399	29.25	0.20	0.85	3,330	36.24	0.18	0.85
	Cx	**ELF_Unedited vs. ENL_Edited**				**ELF_Edited vs. ENL_Edited**				**ELF_Unedited vs. ELF_Edited**			
		G^2	**%DIFF**	**p^a**		**G^2**	**%DIFF**	**p^a**		**G^2**	**%DIFF**	**p^a**	
Statistics	N *for* (NP)	64.84	17.15	.003		48.03	14.05	.003		2.20	3.62	.695	
	N *of* (NP)	216.81	13.68	.002		197.29	12.30	.002		1.99	1.23	.899	
	N *of* NP *in* (NP)	13.92	15.98	.081		58.18	31.97	.003		33.72	19.82	.180	
	N *of* NP *of* (NP)	117.88	56.43	<.001		125.10	54.99	<.001		4.70	7.92	.785	
	N *to* (NP)	46.84	21.20	<.001		64.79	23.92	.003		0.54	2.25	.487	

[a]P-value for Welch's two-tailed t-test

7.2.1.1 Nominal of*-constructions: 'N* of *(NP)', 'N* of *NP* in *(NP)', & 'N* of *NP* of *(NP)'*

Of-phrases are well-known to be one of the common nominal features of academic writing mainly due to their prevalence as post-modifiers of noun phrases (Biber et al., 1999), which seems to hold true for the corpora used in this study as well. Further, in line with the common use of nominals, another common feature of academic writing, the majority of *of*-patterns found in the corpora point to the 'N + *of*' sequence, which, along with being the predominant *of*-pattern, makes up 55.16%, 43.11%, and 32.40 % of all the 'N PREP (NP)' constructions in ELF_Unedited, ELF_Edited, and ENL_Edited respectively. In addition, ELF_Unedited and ELF_Edited, while being quite similar to one another, have significantly higher frequencies of the sequence than ENL_Edited. However, it is also important to note that, as Biber and Gray (2016) report, the use of *of*-phrases in written academic English decreased slightly in recent years (pp. 142-143). Thus, this construction, when examined closely, could help us examine the role of ELF writers in language change in academic English.

The analysis of the collocates of '*of*' indicates, as is seen in Table 7.5, that the types of head nouns and the degree of how strongly '*of*' attracts them (ΔP 1|2) are quite similar across corpora. However, the token frequencies are much higher in the two ELF corpora, with particularly ELF_Edited having the highest frequencies and more substantial differences to ENL_Edited.

Table 7.5 10 Most Frequent Head Nouns in 'N of *(NP)'*

ELF_Unedited						ELF_Edited						ENL_Edited					
Noun	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**	**Noun**	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**	**Noun**	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**
number	501	7.52	2.07	0.02	0.86	*number*	769	9.37	2.62	0.02	0.81	*number*	765	8.32	2.61	0.02	0.81
part	400	6.00	1.65	0.01	0.74	*part*	509	6.20	1.73	0.01	0.76	*use*	457	4.97	1.56	0.01	0.65
type	315	4.73	1.30	0.01	0.63	*level*	403	4.91	1.37	0.01	0.41	*effect*	455	4.95	1.55	0.01	0.41
use	307	4.61	1.27	0.01	0.69	*effect*	385	4.69	1.31	0.01	0.41	*level*	442	4.81	1.51	0.01	0.42
analysis	254	3.81	1.05	0.01	0.28	*type*	352	4.29	1.20	0.01	0.60	*value*	327	3.56	1.12	0.01	0.32
value	247	3.71	1.02	0.01	0.27	*use*	339	4.13	1.15	0.01	0.66	*part*	324	3.52	1.11	0.01	0.63
effect	234	3.51	0.97	0.01	0.39	*value*	322	3.93	1.10	0.01	0.30	*range*	319	3.47	1.09	0.01	0.62
result	220	3.30	0.91	0.01	0.21	*development*	300	3.66	1.02	0.01	0.41	*result*	314	3.42	1.07	0.01	0.25
level	201	3.02	0.83	0.01	0.32	*result*	281	3.43	0.96	0.01	0.22	*analysis*	286	3.11	0.98	0.01	0.23
kind	192	2.88	0.79	0.01	0.91	*analysis*	275	3.35	0.94	0.01	0.28	*type*	275	2.99	0.94	0.01	0.34

'Of' primarily denotes a "part of a whole" relation between two entities (Strauss, Feiz, & Xiang, 2018, p. 356), and as Sinclair (1991) states, it "combines with preceding nouns to produce elaborations of the nominal group" (p. 83). However, as showcased by Francis et al. (1998) in their comprehensive analysis

of 'N *of* (NP)', it is a rather polysemous sequence with a wide range of meanings (pp.176-179). Since these meanings are somewhat related through the larger "part-whole" meaning of the preposition *'of'*, this can be considered an example of constructional polysemy (Goldberg, 1992, p. 51). Further, since the high-frequency head nouns did not point to any central meanings preferred by the construction, all nouns with a minimum ΔP 1|2 score of 0.001, amounting to more than half of the three distributions, were examined closely for semantic domains. A categorization scheme adapted from previous research on the functional analysis of nouns in academic writing (Biber, 2006b; Cutting, 2012; Gray, 2015), MWUs containing '*of*' (Biber et al., 2004; Hyland, 2008a; Simpson-Vlach & Ellis, 2010), and 'N *of* (NP)' in particular (Francis et al., 1998; Gledhill, 2000b; Groom, 2010; Sinclair, 1991; Yilmaz & Römer, 2020) was used to carefully examine the meanings of the remaining patterns primarily by the head nouns in the sequences.

As previous research also shows, the 'N *of* (NP)' construction covers a large number of units with various meanings in the English language. However, the constructions selected here point to certain important differences in the proportions of the different uses across the study corpora. As documented in Tables 7.6 and 7.7, this construction was used to generally denote abstract meanings with the majority of the nouns having rather general meanings (e.g., '*type/form/effect of* [NP]'). While a smaller proportion of concrete nouns was identified, the quantity nouns in this domain (e.g. '*number/level of* [NP]') make up the most frequent subdomain in ENL_Edited and ELF_Edited. In addition, although ELF_Edited and ENL_Edited have comparable frequencies of concrete nouns, ELF_Edited also resembles ELF_Unedited regarding the high frequencies of nouns in the abstract domain. As well as expressing general meanings, these abstract nouns with higher frequencies in ELF than in ENL, are also used as framing devices (e.g., '*case/term of* [NP]'), and to refer to different general processes (e.g., '*development/use of* [NP]'), research terms (e.g., '*result/analysis of* [NP]'), and cognition (e.g., '*concept/notion of* [NP]'). Hence, the choice of head nouns differs remarkably across ENL and ELF writing, e the former includes mostly nouns referring to quantities, animate objects and materials, and the latter covers abstract concepts commonly known as shell nouns (Schmid, 1998, 2000).

Table 7.6 Semantic Domains Identified in Head Nouns of 'N of *(NP)' (min ΔP 1|2 = 0.001)*

Domain	Sub- domain	ELF_Unedited			ELF_Edited			ENL_Edited		
		Types	Tokens (raw)	Tokens (normed per 10k)	Types	Tokens (raw)	Tokens (normed per 10k)	Types	Tokens (raw)	Tokens (normed per 10k)
Abstract	General	47	3,718	55.77	44	4,294	52.35	40	4,311	46.90
	Process	28	2,021	30.32	31	2,617	31.90	26	2,220	24.15
	Framing	13	1,302	19.53	12	1,366	16.65	12	1,295	14.09
	Cognitive	17	1,172	17.58	17	1,412	17.21	14	1,075	11.70
	Research	8	873	13.10	8	1,060	12.92	7	1,117	12.15
	Subtotal	**113**	**9,086**	**136.29**	**112**	**10,749**	**131.03**	**99**	**10,018**	**108.99**
Concrete	Quantity	37	3,471	52.07	37	4,766	58.10	43	5,688	61.88
	Other	13	885	13.28	16	1,289	15.71	13	1,074	11.68
	Subtotal	50	4,356	65.34	53	6,055	73.81	56	6,762	73.57
Grand total		**163**	**13,442**	**201.64**	**165**	**16,804**	**204.85**	**155**	**16,780**	**182.56**

Table 7.7 Examples for Identified Semantic Domains of Head Nouns in 'N of *(NP)' (min ΔP 1|2 = 0.001)*

Domain	Sub-domain	ELF_Unedited	ELF_Edited	ENL_Edited
Abstract	General	***type*** *of wine* *other* ***kind*** *of weapon* ***effects*** *of DDPH*	***types*** *of sellers* ***effects*** *of piracetam* *alternative* ***forms*** *of development*	***effects*** *of emotions* ***types*** *of changes* *the* ***form*** *of activity*
	Process	***uses*** *of the word* ***development*** *of knowledge* ***process*** *of the construction*	***use*** *of (healthcare) services* ***development*** *of tourism* ***processes*** *of internationalization*	***the*** *use of the word* ***development*** *of language skills* ***distribution*** *of resources*
	Framing	***case*** *of substantive disagreement (in)* ***terms*** *of ownership*	*(in)* ***terms*** *of relations* *new* ***case*** *of hematologic cancer*	*(in)* ***terms*** *of language* *(in the)* ***case*** *of initial samples*
	Cognitive	***concept*** *of culture* ***notion*** *of movement* *main* ***ideas*** *of the text*	***concepts*** *of innovations* ***perceptions*** *of increased difficulty* ***notion*** *of entity*	***understanding*** *of systems* ***sense*** *of control* ***knowledge*** *of the systems*
	Research	***results*** *of the study* ***analysis*** *of variance* ***studies*** *of the interaction*	*the* ***results*** *of this study* ***analysis*** *of variance* ***studies*** *of science*	***results*** *of the study* ***analysis*** *of variance* ***study*** *of language policy*
Concrete	Quantity	***number*** *of studies* ***part*** *of the regions* ***values*** *of particle-size index*	***number*** *of measurements* ***part*** *of the book* ***values*** *of the parameters*	***number*** *of students* ***level*** *of democracy* ***values*** *of total patents*
	Other	***beginning*** *of the century* ***group*** *of students* ***central*** *region of the flame*	***end*** *of the century* ***members*** *of the community* ***Gulf*** *of Finland*	***time*** *of arrival* ***members*** *of digital culture* ***eastern*** *region of the country*

These findings are in line with those of previous research in some respects, and different from them in others. To illustrate, Palumbo (2015, 2017), for instance, claims that ELF writers, especially those who speak a Romance language as mother tongue, prefer post-modification (e.g., 'N *of/in* [NP]') over pre-modification, possibly because pre-modification is not common in their first language. However, his results show only slightly higher normalized frequencies of the pattern in ELF as opposed to ENL data. Yilmaz and Römer (2020) report a statistically higher proportion of 'N *of* (NP)' in unedited ELF than in edited ENL, and, their semantic analysis of the most frequent fillers of the pattern suggest indicates the higher frequencies of both qualitative (abstract) and quantitative meanings in ELF than in ENL. While this study similarly found a statistically significant difference between ELF and ENL texts with the former having much higher frequencies of the construction, the in-depth analysis of the meanings

conveyed by the construction also revealed an important semantic difference that abstract meanings of the pattern are more common in ELF, particularly in unedited texts, than in ENL writing. Highlighting two important diachronic changes in academic English regarding such quantifying patterns, Hyland and Jiang (2019) found that the use of these patterns increased in applied linguistics and sociology to provide further details of research activities, while showing a decrease in biology and electrical engineering, both of which were found to employ more abstract and evaluative patterns indicating a stronger authorial commitment (pp. 119-123). Hence, a closer examination of the pattern by considering the potential interactions of nativeness and disciplinary writing, as done in the next chapter, could enable deeper insights into the complexities of academic discourse construction.

A closer look at the textual environment of 'N *of* (NP)' led to the identification of two additional constructions extending from the prepositional object (NP) following '*of*': 'N *of* NP *of* (NP)' and 'N *of* NP *in* (NP).' While collocates of the key prepositions signaled the usage of multiple prepositions modifying a noun phrase (NP) in general, the high degree of variation in the structure of these complex patterns did not enable the identification of differences easily. Hence, I selected three PoS-grams corresponding to some of the commonly found patterns with two prepositions; that is, one following and the other preceding an NP with zero to two pre-modifiers that are determiners (e.g., '*the use*'), adjectives (e.g., '*big group*'), or nouns (e.g., '*peer review*').

These two patterns are quite relevant to research on lexicogrammatical complexity in academic writing since prepositional phrases, especially when there are more than one in one construction, are a structural characteristic of "postmodifier complexes" frequently observed in academic writing (Biber et al., 1999, p. 642). Biber et al. (2011) also proposed a list of developmental phases of grammatical complexity in academic writing, where the use of multiple prepositional post-modifiers of nouns was placed in the last phase, thus providing evidence of a particularly high level of complexity. Following this work, Ansarifar et al. (2018) found that expert writers employed these patterns more often than the graduate students with less expertise did. In addition, Benelhadj (2019) reported a higher frequency of

occurrence for multiple modifications in sociology than in medicine, and more consistent and efficient use of these patterns in published disciplinary texts than in student theses.

Although both of these constructions derive from 'N *of* (NP)', both have partially different characteristics as they are extended into structurally and functionally more complex sequences with the addition of a second prepositional phrase. The difference is also evident from the list of common head nouns presented in Table 7.8. For instance, nouns not listed among those commonly found in 'N *of* (NP)' such as *'occurrence', 'percentage', 'evaluation',* and *'proportion'* also signify different uses of these constructions.

Table 7.8 10 Most Frequent Head Nouns in 'N of *NP* in *(NP)' & 'N* of *NP* of *(NP)' (min ΔP 1|2 = 0.001)*

Cx	ELF_Unedited						ELF_Edited						ENL_Edited					
	Noun	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**	**Noun**	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**	**Noun**	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**
'N *of* NP *in* (NP)'	*number*	38	0.57	3.24	0.013	0.063	*role*	49	0.60	2.99	0.013	0.109	*number*	64	0.70	4.60	0.018	0.067
	role	30	0.45	2.56	0.010	0.084	*number*	49	0.60	2.99	0.013	0.050	*role*	45	0.49	3.23	0.013	0.106
	amount	22	0.33	1.87	0.008	0.108	*level*	41	0.50	2.50	0.011	0.049	*use*	37	0.40	2.66	0.011	0.052
	presence	19	0.28	1.62	0.007	0.105	*amount*	26	0.32	1.59	0.007	0.109	*importance*	25	0.27	1.80	0.007	0.093
	use	15	0.22	1.28	0.005	0.032	*absorption*	24	0.29	1.46	0.007	0.109	*level*	24	0.26	1.72	0.006	0.024
	occurrence	14	0.21	1.19	0.005	0.148	*effect*	23	0.28	1.40	0.005	0.023	*effect*	20	0.22	1.44	0.005	0.016
	development	14	0.21	1.19	0.005	0.037	*use*	22	0.27	1.34	0.006	0.042	*amount*	18	0.20	1.29	0.005	0.068
	analysis	14	0.21	1.19	0.004	0.015	*occurrence*	21	0.26	1.28	0.006	0.136	*result*	15	0.16	1.08	0.003	0.010
	level	13	0.19	1.11	0.004	0.022	*percentage*	20	0.24	1.22	0.005	0.124	*proportion*	13	0.14	0.93	0.004	0.091
	part	12	0.18	1.02	0.004	0.020	*analysis*	20	0.24	1.22	0.005	0.022	*degree*	13	0.14	0.93	0.004	0.048
'N *of* NP *in* (NP)'	*part*	34	0.509	2.705	0.013	0.065	*analysis*	35	0.43	2.29	0.011	0.042	*term*	27	0.29	2.44	0.011	0.045
	analysis	30	0.449	2.387	0.011	0.037	*part*	33	0.40	2.16	0.010	0.050	*analysis*	27	0.29	2.44	0.010	0.025
	term	19	0.284	1.512	0.007	0.044	*number*	31	0.38	2.03	0.009	0.031	*understanding*	23	0.25	2.08	0.009	0.114
	process	19	0.284	1.512	0.007	0.023	*result*	30	0.37	1.96	0.009	0.023	*number*	23	0.25	2.08	0.009	0.023
	point	18	0.269	1.432	0.007	0.038	*heat*	25	0.30	1.64	0.008	0.110	*part*	22	0.24	1.99	0.009	0.045
	result	18	0.269	1.432	0.006	0.017	*value*	24	0.29	1.57	0.007	0.022	*use*	21	0.23	1.90	0.008	0.029
	view	15	0.225	1.193	0.006	0.069	*evaluation*	22	0.27	1.44	0.007	0.102	*result*	21	0.23	1.90	0.008	0.016
	heat	14	0.210	1.114	0.005	0.095	*effect*	21	0.26	1.37	0.006	0.022	*study*	18	0.20	1.63	0.006	0.008
	understanding	13	0.195	1.034	0.005	0.111	*term*	20	0.24	1.31	0.006	0.034	*degree*	15	0.16	1.36	0.006	0.057
	number	13	0.195	1.034	0.004	0.019	*level*	20	0.24	1.31	0.006	0.022	*effect*	15	0.16	1.36	0.005	0.012

Zooming in on all the head nouns with a minimum value of (min ΔP 1|2 = 0.001) enables deeper insights into the semantic differences and similarities in these patterns' use in comparison to one another, as well as 'N *of* (NP)' (see Table 7.9). First, we see that a larger number of noun types are used in these patterns, with much lower token frequencies. It is also clear that cognitive, research, and framing categories are used less frequently. However, both patterns have similarities to 'N *of* (NP)' such as an

overall tendency to start with an abstract noun, with ELF_Edited having the highest frequency, and both ELF_Edited and ENL_Edited having a higher frequency of quantity nouns than ELF_Unedited.

Table 7.9 Semantic Domains of Head Nouns in 'N of *NP* in *(NP)' & 'N* of *NP* of *(NP)' (min ΔP 1|2 = 0.001)*

Cx	Domain	Sub-domain	ELF_Unedited			ELF_Edited			ENL_Edited		
			Types	Tokens (raw)	Tokens (normed per 10k)	Types	Tokens (raw)	Tokens (normed per 10k)	Types	Tokens (raw)	Tokens (normed per 10k)
'N *of* NP *in* (NP)'	Abstract	General	50	228	3.41	43	279	3.40	44	221	2.41
		Process	58	223	3.34	53	265	3.23	42	185	2.01
		Cognitive	9	35	0.52	8	42	0.51	7	39	0.42
		Framing	8	65	0.97	11	113	1.38	11	89	0.97
		Research	5	30	0.45	4	37	0.45	5	46	0.50
		Subtotal	**130**	**581**	**8.70**	**119**	**736**	**8.97**	**109**	**580**	**6.31**
	Concrete	Quantity	35	198	2.96	35	318	3.88	37	299	3.25
		Other	13	47	0.70	27	87	1.06	16	52	0.57
		Subtotal	**48**	**245**	**3.67**	**62**	**405**	**4.94**	**53**	**351**	**3.82**
	Grand total		**178**	**826**	**12.36**	**181**	**1,141**	**13.91**	**162**	**931**	**10.13**
'N *of* NP *of* (NP)'	Abstract	General	50	207	3.10	63	289	3.52	37	167	1.82
		Process	45	182	2.72	61	263	3.21	43	170	1.85
		Cognitive	16	91	1.36	13	80	0.98	14	71	0.77
		Framing	10	92	1.38	11	83	1.01	11	82	0.89
		Research	8	74	1.11	8	100	1.22	5	78	0.85
		Subtotal	**129**	**646**	**9.67**	**156**	**815**	**9.94**	**110**	**568**	**6.18**
	Concrete	Quantity	31	171	2.56	30	218	2.66	35	193	2.10
		Other	24	73	1.09	29	121	1.48	21	69	0.75
		Subtotal	**55**	**244**	**3.65**	**60**	**339**	**4.13**	**56**	**262**	**2.85**
	Grand total		**184**	**890**	**13.32**	**216**	**1,154**	**14.07**	**166**	**830**	**9.03**

Besides, as seen in Table 7.10, there are certain differences in areas of use for the two patterns. With the use of two distinct prepositions, 'N *of* NP *in* (NP)' has a richer functionality than 'N *of* NP *of* (NP)' that includes the repeated use of '*of*'. The spatial and metaphorical senses of containment or enclosure of the preposition '*in'* (Dirven, 1993; Radden & Dirven, 2007), enable the further specification of the part-whole meaning in 'N *of* (NP)' via the nouns it precedes (e.g., *'in + society/system/context'*, etc.). In contrast, *'of'* is repeated in 'N *of* NP *of* (NP)' to connect two parallel noun phrases. These features of the two constructions can also be said to explain the significant differences across the corpora. While 'N *of* NP *in* (NP)' had a significantly higher frequency in both ELF corpora than in ENL_Edited, only ELF_Unedited differs significantly from ENL_Edited in the use of 'N *of* NP *of* (NP)', which suggests that ELF writers might also refrain from using the structure in edited texts.

Table 7.10 Examples for Identified Semantic Domains of Head Nouns in 'N of *NP* in *(NP)' & 'N* of *NP* of *(NP)' (min ΔP 1|2 = 0.001)*

Cx	Domain	Sub-domain	ELF_Unedited	ELF_Edited	ENL_Edited
'N *of* NP *in* (NP)'	Abstract	General	***presence*** *of women in church* ***effects*** *of variables in a multivariate setting*	***effects*** *of peer review in the writing process* ***solubility*** *of CO_2 in the systems*	***presence*** *of black particles in the solvent* ***effect*** *of a rise in wealth inequality*
		Process	***use*** *of English in teaching* ***development*** *of mind in human society*	***absorption*** *of CO_2 in a solution* ***development*** *of robot contexts in Japan*	*the* ***use*** *of paraphrase in summary writing* *development*
		Framing	***role*** *of meanings in social life*	***role*** *of the L2 in a particular context*	*the* ***role*** *of English in this process*
		Cognitive	***concept*** *of sustainability in the supply chains*	***concept*** *of innovation in an enterprise*	*a* ***theory*** *of creativity in popular music*
		Research	***linguistic*** *analysis of modality in scientific texts*	*their* ***analysis*** *of metadiscoursal in texts*	*the* ***analysis*** *of various relationships in these data*
	Concrete	Quantity	***number*** *of heliostats in the first row* ***amount*** *of claystones in bed sediments*	*the* ***number*** *of boxes in the network* ***level*** *of proficiency in English*	*the small* ***number*** *of customers in each country* *the* ***amount*** *of nitrogen in these systems*
		Other	*dominant* ***group*** *of predators in our experiments*	*the big* ***group*** *of foreign tourists in the area*	*large* ***groups*** *of nodes in the network*
'N *of* NP *of* (NP)'	Abstract	General	*linguistic* ***context*** *of validation of an occurrence* *the* ***effect*** *of the size of halo orbits*	*the* ***effect*** *of the wake of the vanes indicators of goodness of fit*	***effect*** *of omission of these variables* *a* ***function*** *of the direction of the sun*
		Process	*the* ***process*** *of the construction of such texts* *the* ***construction*** *of a macrostructure of the text*	***evaluation*** *of the use of innovation concepts* ***process*** *of internationalization in business schools*	***examination*** *of the quality of the evidence* *an* ***assessment*** *of the quality of the support*
		Framing	*in* ***terms*** *of the production of the genre*	*the* ***point*** *of view of individual*	*in* ***terms*** *of direction of change*
		Cognitive	*the* ***probability*** *of the application of global approach*	*Hong's* ***understanding*** *of the nature of writing*	*our* ***understanding*** *of the range of borrowing strategies*
		Research	*an* ***analysis*** *of the performance of industrial IT*	***analysis*** *of the families of periodic orbits*	***analysis*** *of presence of beat*
	Concrete	Quantity	*another* ***part*** *of the context of validation* *the* ***number*** *of heliostats of the first row*	***number*** *of measurements of bedding planes* ***part*** *of the processes of evaluation*	*the* ***number*** *of citations of that play* ***part*** *of the process of semantic extension*
		Other	***heat*** *of absorption of CO_2 solution*	*differential* ***heat*** *of absorption of CO_2*	*the* ***product*** *of the number of transcriber pairs*

In previous research, 'N *of* NP *in* (NP)' was identified in published texts in pharmaceutical sciences to be a common pattern denoting research findings on the basis of study data (Gledhill, 2000a, p. 131). In their analysis of WrELFA compared against the academic subset of the Corpus of Contemporary American English (COCA), Yilmaz and Römer (2020) also identified the construction as a distinctive feature of ELF that refers to "a property of an entity in a source," by also noting that it has a wide variety of uses (p. 75). As for 'N *of* NP *of* (NP)', using the pattern as an example of phrasal complexity, Benelhadj (2019) argues that in sociology, "multiple modifications are needed to make successful references" (p.8). Citing Martin and Halliday (1993), she also adds, based on an example of the same kind, that such patterns lead to a decrease in clarity of meaning by increasing the lexical and informational density of texts. This argument is also in line with that of Biber and Gray (2016) on potential challenges of novice readers in

understanding what is expressed via complex prepositional structures in scientific texts (p. 239). In addition, the more common use of abstract nouns by ELF writers, also a general feature of L2 academic writing (Crossley & McNamara, 2011), could increase the vagueness and inexplicitness of texts (Cutting, 2012). However, along with the fact that texts are much shorter in the ELF corpora than in ENL_Edited, there seems to be an ELF-specific preference for more complex structures that group simpler parallel forms together in order to maximize "communication efficiency" (Wu et al., 2020, p. 8)

7.2.1.2 Nominal for-*construction: 'N* for *(NP)'*

The use of '*for'* with a preceding noun (N) and a following noun group (NP) led to the identification of another construction: 'N *for* (NP)'. Biber et al. (2011) found that this construction had significantly higher frequency in academic prose than in conversation, and Biber and Gray (2016) reported a remarkable increase in the use of the pattern in academic texts over the years. Not only was this structure found to be similarly common in the study corpora, but it also pointed to interesting differences across the corpora. As is seen in Table 7.11, the 10 most frequent head nouns in 'N *for* (NP)', aside from a certain degree of overlap, signify differences across the three corpora, with ELF_Unedited and ENL_Edited being the more distinct pair.

Table 7.11 10 Most Frequent Head Nouns in 'N for *(NP)' across Corpora*

ELF_Unedited						ELF_Edited						ENL_Edited					
Noun	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**	**Noun**	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**	**Noun**	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**
reason	66	0.99	1.77	0.01	0.28	*value*	79	0.96	1.67	0.01	0.07	*value*	82	0.89	1.33	0.01	0.08
factor	65	0.98	1.74	0.01	0.11	*need*	59	0.72	1.25	0.01	0.27	*need*	80	0.87	1.30	0.01	0.33
value	58	0.87	1.55	0.01	0.06	*result*	63	0.77	1.33	0.01	0.05	*taste*	72	0.78	1.17	0.01	0.62
data	48	0.72	1.29	0.01	0.07	*model*	61	0.74	1.29	0.01	0.04	*evidence*	72	0.78	1.17	0.01	0.16
need	36	0.54	0.96	0.01	0.26	*reason*	44	0.54	0.93	0.01	0.22	*demand*	70	0.76	1.14	0.01	0.43
condition	38	0.57	1.02	0.01	0.09	*preference*	43	0.52	0.91	0.01	0.31	*support*	67	0.73	1.09	0.01	0.33
method	39	0.59	1.04	0.01	0.06	*support*	40	0.49	0.85	0.00	0.23	*method*	67	0.73	1.09	0.01	0.08
tool	33	0.50	0.88	0.00	0.21	*condition*	42	0.51	0.89	0.00	0.08	*model*	67	0.73	1.09	0.00	0.04
potential	24	0.36	0.64	0.00	0.20	*tool*	38	0.46	0.80	0.00	0.20	*opportunity*	63	0.69	1.02	0.01	0.26
model	32	0.48	0.86	0.00	0.02	*method*	42	0.51	0.89	0.00	0.06	*data*	62	0.67	1.01	0.00	0.05

The preposition '*for'* is often used to refer to an intended purpose (e.g., '*salad for lunch'*), reason (e.g., '*march for justice'*), benefit (e.g., '*donations for children'*), recipient ('*a gift for you'*), or duration (e.g., '*for five days'*) in relation to an action, person or entity (Sinclair, 1996; Strauss et al., 2018; Tyler, 2012; Tyler & Evans, 2003). Consistent with the multiple meanings of the preposition, Francis et al. (1998)

identified 30 different semantic categories primarily determined by the head noun (N) in the pattern (pp. 148-149). However, to my knowledge, not much academic writing research has focused specifically on the meanings of the pattern. Schmid (2000) examined only certain lexical realizations of the pattern, as part of a semantic domain used to express a cause-effect relationship with the shell noun *'reward' (e.g., 'reward for our persistence'[p.110])* Biber and Gray (2016), focused specifically on one sub-type of the pattern where the noun group (NP) following '*for'* is restricted to an *-ing* clause. The study identified two major meanings of the pattern: 1) purpose of an entity when the head is a concrete noun (e.g. *'operations for determining….'*), and 2) rationale of ideas when followed by an abstract cognitive noun (e.g., *'good reasons for thinking….*') (pp. 201-202). Morley's (2018) corpus-based, and pedagogically-oriented list of academic phrases also included several MWUs based on the pattern as canonical means of talking about research processes and making discourse connections in academic texts (e.g., *reason/need/strategy/evidence for* [NP]')

In line with previous research, when all head nouns in the pattern with a minimum ΔP 1|2 score of 0.001 were examined, the distinctive lexical choices in the corpora were revealed to be rather complex. Of the abstract head nouns, the largest proportion in all three distributions, those denoting general meanings (e.g., '*factor for* [NP]'), cognitive (e.g., '*reason for* [NP]') and research (e.g., '*result for* [NP]') terms have the highest frequencies in ENL_Edited, and the lowest in ELF_Unedited. General process nouns, however, are the most frequent in ELF_Edited, and the least frequent in ENL_Edited. In terms of the head nouns representing the more quantitative aspects of the texts (e.g., '*value/rate for* [NP]'), ENL_Edited has the highest frequency, and ELF Edited has the lowest (see Table 7.12).

Table 7.12 Semantic Domains Identified in Head Nouns of 'N for *(NP)' in (min ΔP 1|2 = 0.001)*

Domain	Sub-domain	ELF_Unedited			ELF_Edited			ENL_Edited		
		Types	Tokens (raw)	Tokens (normed per 10k)	Types	Tokens (raw)	Tokens (normed per 10k)	Types	Tokens (raw)	Tokens (normed per 10k)
Abstract	General	25	396	5.94	29	555	6.77	25	635	6.91
	Cognitive	17	312	4.68	17	394	4.80	13	484	5.27
	Research	7	157	2.36	7	230	2.80	8	382	4.16
	Process	11	133	2.00	11	180	2.19	8	164	1.78
	Subtotal	**60**	**998**	**14.97**	**64**	**1,359**	**16.57**	**54**	**1,665**	**18.11**
Concrete	Quantity	14	244	3.66	11	242	2.95	17	459	4.99
	Other	4	67	1.01	3	70	0.85	6	101	1.10
	Subtotal	**18**	**311**	**4.67**	**14**	**312**	**3.80**	**23**	**560**	**6.09**
Total		**78**	**1,309**	**19.64**	**78**	**1,671**	**20.37**	**77**	**2,225**	**24.21**

As the examples in Table 7.13 illustrate, several nouns (e.g., *reason, explanation*) denote causality. Still, the majority have a purposive element, in line with the "orientational" nature of *'for'* (Radden & Dirven, 2007, p. 330). Thus, especially in the case of the abstract terms, the attribute in the head noun is explicitly assigned to the prepositional object. Also in line with research on stance nouns such as those in the cognition category in this study (e.g., *'need', 'demand'*), this relatively more explicit expression of intention points to a somewhat more visible authorial voice (Biber, 2006a).

Table 7.13 Examples for Identified Semantic Domains of Head Nouns 'N for *(NP)' (min ΔP 1|2 = 0.001)*

Domain	Sub- domain	ELF_Unedited	ELF_Edited	ENL_Edited
Abstract	General	*risk **factors** for nightmares* *optimal **conditions** for study* ***strategy** for gender equality*	***conditions** for inclusion* *risk **factors** for death* ***support** for the hypothesis*	***support** for collective provision* ***basis** for a detailed analysis* ***reference** point for the end*
	Cognitive	***reason** for summoning the council* ***need** for her assistance* ***potential** for winter tourism*	***need** for more training* ***reasons** for their choice* ***preferences** for redistribution*	***need** for language support* ***demand** for external finance* ***taste** for social status*
	Research	***methods** for the extraction* ***analytical** model for thinning*	*simulation **result** for angles* ***models** for entity type*	***evidence** for human presence* ***historical** methods for forecasting*
	Process	***physical** activity for young people* *the **search** for truth*	*heat of **absorption** for the solvents* ***explanation** for the high diversity*	*higher **agreement** for RaP* ***explanation** for the omissions*
Concrete	Quantity	***values** for the best model* ***data** for the solubility of CO_2* ***estimates** for various scenarios*	*similar **values** for the RBC count* *the FBI **data** for homicides* *F **test** for the linear model*	***values** for the model-free volatility* ***data** for these membranes* *extinction **rate** for each species*
	Other	***tool** for assessment*	***tool** for baseline measurements*	***site** for social change*

7.2.1.3 Nominal to-*constructions: 'N* to *(NP)'*

'*To'* is one of the KFWs, which, when used as a preposition, was found to be significantly more frequent in ENL_Edited than in ELF_Edited. While no significant difference was found between ELF_Edited and ELF_Unedited, the former still has a remarkably higher normalized frequency of the preposition. The most evident difference in the patterns, including '*to'* is in the 'N *to* (NP)' pattern, which also accounts for the largest proportion of the usage of '*to'* as a preposition in the study corpora.

As a directional preposition, '*to'* denotes a relation between two entities where one is directed or oriented to the other, often signifying a connection between the two in the form of activity, transfer, exchange, extent, location, and in terms of physical or logical relatedness (Radden & Dirven, 2007; Sinclair, 1996; Tyler & Evans, 2003). While their discussion is not about nominal patterns, Radden and Matthis (2002) also show that the expression of similarity and closeness (e.g., *'similar to'*), and difference (e.g., *'different from'*) are motivated by the spatial sense of the prepositions, which leads to the SIMILARITY IS CLOSENESS metaphor. This finding could also be associated with the different types of

relatedness that '*to*' is used to denote. To illustrate, in academic writing, several high-frequency *to*-patterns such as '*the extent to which*,' and '*with/in relation/response/regard(s) to* (NP)*'* have generally been identified as devices used for referring to specific aspects or attributes of content, and discourse organization (Benelhadj, 2019; Biber et al., 2004; Hyland, 2008a, 2008b; Simpson-Vlach & Ellis, 2010). Also known as complex prepositions (Quirk, Greenbaum, & Leech, 1985, p. 669), these are relatively fixed sequences that tend to occur with a small number of lexical units. As seen in Table 7.14, the most frequent nouns in the construction point to the common occurrence of these sequences.

Table 7.14 10 Most Frequent Head Nouns in 'N to *(NP)'*

ELF_Unedited						ELF_Edited						ENL_Edited					
Noun	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**	**Noun**	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**	**Noun**	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**
relation	65	0.97	3.25	0.010	0.238	*relation*	123	1.50	5.13	0.016	0.262	*response*	148	1.61	4.44	0.016	0.289
respect	61	0.91	3.05	0.010	0.589	*respect*	95	1.16	3.96	0.013	0.616	*approach*	143	1.56	4.29	0.015	0.207
approach	51	0.76	2.55	0.008	0.156	*approach*	70	0.85	2.92	0.009	0.157	*addition*	95	1.03	2.85	0.011	0.470
reference	49	0.73	2.45	0.008	0.282	*regard*	62	0.76	2.58	0.008	0.688	*respect*	94	1.02	2.82	0.010	0.582
addition	46	0.69	2.30	0.007	0.409	*response*	62	0.76	2.58	0.008	0.276	*relation*	75	0.82	2.25	0.008	0.311
attention	45	0.67	2.25	0.007	0.372	*access*	61	0.74	2.54	0.008	0.647	*access*	72	0.78	2.16	0.008	0.606
access	44	0.66	2.20	0.007	0.638	*addition*	56	0.68	2.33	0.008	0.409	*reference*	62	0.67	1.86	0.007	0.194
time	41	0.61	2.05	0.004	0.040	*reference*	52	0.63	2.17	0.007	0.217	*regard*	53	0.58	1.59	0.006	0.770
response	37	0.55	1.85	0.006	0.239	*attention*	39	0.48	1.63	0.005	0.311	*transfer*	45	0.49	1.35	0.005	0.118
transfer	34	0.51	1.70	0.004	0.142	*exposure*	32	0.39	1.33	0.004	0.262	*response*	148	1.61	4.44	0.016	0.289

Upon closely examining the instances of the construction, I was able to identify several semantic categories primarily through the head nouns. The first and most frequent one is framing *(*e.g., *'in relation to* [NP]*'*), which includes the above-mentioned complex prepositions. The second most common category for ELF_Edited and ENL_Edited is process (e.g., '*access to* [NP]'), denoting activities with a directional research process. The third category is comprised of general abstract nouns referring to labels and concepts (e.g., '*approach/attention to* [NP]'), which is, while slightly less frequent in ELF_Unedited than in ELF_Edited, the second most common category for the latter. Lastly, the concrete group entails nouns used to refer to objects, locations, and living beings (see Table 7.15 and Table 7.16).

Table 7.15 Semantic Domains Identified in Head Nouns of 'N to *(NP)' (min ΔP 1|2 = 0.001)*

Domain	Sub-domain	ELF_Unedited			ELF_Edited			ENL_Edited		
		Types	**Tokens (raw)**	**Tokens (normed per 10k)**	**Types**	**Tokens (raw)**	**Tokens (normed per 10k)**	**Types**	**Tokens (raw)**	**Tokens (normed per 10k)**
Abstract	**Framing**	12	389	5.82	13	584	7.12	14	716	7.79
	Process	17	225	3.37	22	298	3.63	24	470	5.12
	General	30	277	4.15	19	247	3.01	25	413	4.50
	Subtotal	**59**	**891**	**13.34**	**54**	**1129**	**13.76**	**63**	**1599**	**17.41**
Concrete		**14**	**128**	**1.92**	**7**	**55**	**0.67**	**13**	**137**	**1.49**
Grand total		**73**	**1,019**	**15.25**	**61**	**1,184**	**14.43**	**76**	**1,736**	**18.90**

Table 7.16 Examples for Identified Semantic Domains via Head Nouns in 'N to *(NP)' (min ΔP 1|2 = 0.001)*

Domain	Sub-domain	ELF_Unedited	ELF_Edited	ENL_Edited
Abstract	Framing	*in **relation** to questions posed* *with **respect** to bed sediment lithology* *with **reference** to the implications*	*in **relation** to other cites* *with **respect** to transport capacities* *with **regard** to local issues*	*in **response** to CP treatment* *in **addition** to describing the types* *with **respect** to macula status*
	Process	***access** to data services* ***transfers** to LL1 Halo orbits* ***answers** to the questions*	***access** to information* ***exposure** to such pollution* ***empirical** contribution to this issue*	***access** to finance* ***transfers** to halo orbits,* ***contributions** to this debate*
	Abstract	*the formal **approach** to collocations* *particular **attention** to people* *legal **value** to synods*	*various **approaches** to management* ***attention** to pronunciation traits* ***obstacles** to local self-governance*	*a new **approach** to management* *primary **attention** to meaning* ***sensitivity** to climate*
Concrete		***time** to full remission* *distribution **point** to clumped pattern*	*from the transition **point** to plane bed* *protection of **normals** to bedding*	*target **point** to the refueling station* ***time** to expiry*

7.2.2 Constructions with Key Prepositions and Their Nominal Complements: 'PREP N'

The key prepositions '*across', 'after,' 'over,' 'through,'* '*within,'* and *'in'* led to the identification of cross-corpus differences in the use of nouns as their complements. For the constructions with '*across', 'after', 'over' 'through',* and '*within',* unlike the 'N PREP (NP)' constructions where the searches did not control for the phrase-final '(NP)', the constructions in this section focus specifically on the nouns ('N') identified within a span of four words to the right of the target prepositions. While '*after* N', '*over* N', '*through* N' and '*within* N' included a noun (N) in any form, 'across' was found to attract primarily plural nouns (N-s). Hence, the construction was labeled as '*across* N-s', and the small number of singular nouns with which '*across'* occurs were excluded from the analysis. In addition, *'in order to* V*'* and *'in NN with* (NP)*'* were included in this section as both constructions are also extensions of 'PREP N'. While '*in order to* V' includes a *to*-infinitive, the 'NN' in '*in* NN *with* (NP)' refers to the nouns in the slot being predominantly singular as in 'in *line/comparison/agreement with* (NP).'

Of the five identified 'PREP N' patterns, '*after* N' was found to be the only one with significantly higher frequencies in the two ELF corpora than in ENL_Edited. As for '*across* N-s' and 'over N,' the opposite is true as ENL_Edited has a much higher frequency than ELF. In addition, ENL_Edited was found to have a significantly higher frequency for '*through* N' than ELF_Edited (but not ELF_Unedited), and for '*within* N' than ELF_Unedited (but not ELF_Edited). Further, the one significant difference between ELF_Unedited and ELF_Edited concerns '*over* N' as it occurs more often in the latter than the former. As for the distributions of nouns following the key prepositions in these four patterns, except for '*after* N,' we see a similar trend in that ELF_Unedited has the least predictable and least Zipfian

distribution and is closely followed by ELF_Edited. ENL_Edited, however, has a relatively more predictable distribution, especially in the cases of '*across* N-s' and '*within* N.' Furthermore, '*over* N' is where the largest difference between ELF_Unedited and ELF_Edited is seen, thus signaling a more Zipfian distribution for the latter. Finally, the distribution of verbs in '*in order to* V' were found to have no differences in entropy scores but a slightly lower TTR in ELF_Edited, which also was reported to have the lowest entropy for the singular noun in '*in* NN *with* (NP).' Both these results also point to the higher predictability of these two distributions in ELF_Edited (see Table 7.17). The following subsections delve into the functional differences across corpora as evidenced in the use of the common nouns with strong associations for each of these five patterns.

Table 7.17 Numerical Summary of 'PREP N' Constructions

	Cx	ELF_Unedited				ELF_Edited				ENL_Edited			
		Raw	**Normed (per 10k)**	**DP_{norm}**	**H_{norm} (nouns)**	**Raw**	**Normed (per 10k)**	**DP_{norm}**	**H_{norm} (nouns)**	**Raw**	**Normed (per 10k)**	**DP_{norm}**	**H_{norm} (noun)**
Frequencies	*across* N-s	61	0.92	0.83	0.96	114	1.39	0.83	0.94	338	3.68	0.59	0.89
	after N	481	7.22	0.41	0.94	594	7.24	0.41	0.95	495	6.03	0.42	0.94
	over N	272	4.08	0.47	0.93	464	5.44	0.42	0.89	692	7.53	0.33	0.86
	through N	466	6.99	0.44	0.97	528	6.56	0.38	0.97	802	8.73	0.32	0.95
	within N	431	6.47	0.45	0.94	648	7.90	0.37	0.93	900	9.79	0.37	0.93
	in order to V	378	5.66	0.41	0.93	406	4.95	0.43	0.93	304	3.31	0.50	0.93
	in NN *with* (NP)	177	2.66	0.44	0.79	250	3.05	0.39	0.74	165	1.8	0.44	0.81
	Cx	**ELF_Unedited vs. ENL_Edited**				**ELF_Edited vs. ENL_Edited**				**ELF_Unedited vs. ELF_Edited**			
		G^2	**%DIFF**	**p^a**		**G^2**	**%DIFF**	**p^a**		**G^2**	**%DIFF**	**p^a**	
Statistics	*across* N-s	133.12	75.12	<.001		92.05	62.21	.002		7.19	34.16	.236	
	after N	20.77	33.98	.044		23.80	34.16	.024		0.00	14.21	.650	
	over N	79.12	45.81	<.001		23.07	24.87	.003		18.75	27.87	.034	
	through N	14.75	19.89	.065		29.98	26.23	.018		1.68	8.60	.644	
	within N	52.40	33.97	.006		18.27	19.70	.063		10.01	17.78	.236	
	in order to V	48.93	71.06	.006		28.60	49.64	.041		3.50	14.31	.336	
	in NN *with* (NP)	12.50	46.75	.035		26.61	67.05	.012		1.73	-12.16	.454	

[a]P-value for Welch's two-tailed t-test

7.2.2.1 'After *N*'

For the analysis of '*after'*, in order to be consistent with the prepositional focus of the study, all the '*after* N' instances were examined manually to exclude uses of the pattern where *'after'* was a conjunction followed by a finite clause (e.g., *'after participants finished their reading tasks'*, '*after students had online resources'*). Hence, only those noun phrases following the preposition that are not part of a finite clause were included.

'After' denotes a sequential relation between two connected, where one follows the other in terms of location (e.g., '*turn after the light'*), time (e.g., *'after two hours')*, or event (e.g., *'after the accident'*)

(Strauss et al., 2018; Tyler & Evans, 2003). While it is not a commonly studied feature of academic writing, Gledhill's (2000b) analysis of the *after*-sequences in pharmaceutical sciences research writing, also finds that its dominant functions in the discipline are the expressions of time and methodological processes (pp. 128- 129). As is seen in Table 7.18, the frequent nominal collocates of *'after'* it precedes also seem to denote time (e.g., *minute, year*), and process (e.g., *treatment, disclosure*).

*Table 7.18 5 Most Frequent Nouns in '*after *N'*

ELF_Unedited						ELF_Edited						ENL_Edited					
Noun	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**	**Noun**	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**	**Noun**	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**
treatment	24	200	0.36	3.24	0.043	*year*	24	0.29	0.008	0.030		*time*	18	1446	0.20	2.28	0.021
minute	21	112	0.32	2.84	0.038	*day*	22	0.27	0.007	0.072		*disclosure*	16	31	0.17	2.03	0.020
surgery	13	31	0.20	1.76	0.024	*ischemia*	21	0.26	0.007	0.182		*event*	13	321	0.14	1.65	0.016
war	11	78	0.17	1.49	0.020	*war*	13	0.16	0.004	0.064		*period*	12	506	0.13	1.52	0.015
year	10	541	0.15	1.35	0.017	*exposure*	11	0.13	0.004	0.083		*festival*	11	177	0.12	1.39	0.014

The coding of the noun collocates attracted by '*after'* further shows that the study corpora employed mostly process nouns denoting events, and then temporal nouns, both of which have higher frequencies in the two ELF corpora than in ENL_Edited (see Table 7.19). As the examples in Table 7.20 also indicate, these process and time nouns are used predominantly to either sequence events (e.g., *'treatment', 'exposure',* etc.) or refer to a certain period of time in relation to which another event takes place.

*Table 7.19 Semantic Domains Identified via Nouns in '*after *N' (min ΔP 1|2 = 0.001)*

Domain	ELF_Unedited			ELF_Edited			ENL_Edited		
	Types	**Tokens (raw)**	**Tokens (normed per 10k)**	**Types**	**Tokens (raw)**	**Tokens (normed per 10k)**	**Types**	**Tokens (raw)**	**Tokens (normed per 10k)**
Abstract - Process	32	171	2.57	35	193	2.35	34	170	1.85
Concrete - Time	10	73	1.10	11	94	1.15	10	83	0.90
Total	42	244	3.66	46	287	3.50	44	253	2.75

*Table 7.20 Examples for Identified Semantic Domains via Head Nouns in '*after *N' (min ΔP 1|2 = 0.001)*

Domain	ELF_Unedited	ELF_Edited	ENL_Edited
Abstract - Process	*after dietetic **treatments*** *two years after **surgery*** *soon after the **war***	*damage after cerebral **ischemia*** *after **exposure** to cornstarch* *after **treatment** with biological DMARDS*	*after full **disclosure*** *immediately after the **event*** *months after the **festival***
Concrete - Time	*after less than 300 **minutes*** *after **years** of development*	*after one **year** in incubation* *after 5 **days***	*after the passage of **time*** *after the hatch **period***

7.2.2.2 'Across *N-s*'

A distinctive pattern including *'across,'* a KFW in ENL_Edited in comparison to both ELF_Edited and ELF_Unedited, was found to be '*across* N-s,' where the preposition is followed by a plural noun.

'*Across'* is a path-oriented spatial preposition signifying the relation generally between two locations (Radden & Dirven, 2007; Rauh, 1993; Tyler & Evans, 2003). However, as Table 7.21 also shows, the tendency of especially the ENL writers to use plural nouns as the object of the preposition seems to suggest that the pattern expresses a relatively limited range of semantic meanings in the study data.

The finding that a majority of the plural nouns following '*across'* are used to denote entities and classes of entities (see Table 7.22) points to the senses of the preposition that Sinclair (1996) identified as referring to an event that concerns people from different groups or the observing a phenomenon or event in all the entities mentioned after '*across'* (p.4). However, as the largely noun-specific semantic domains exemplified in Table 7.23 indicate, the construction seldom refers to people (e.g. '*across + sellers/labelers*'). Instead, in a much broader sense, it denotes the different types of concrete entities (e.g. '*across + segments/sites'*) and abstract classifying terms (e.g. '*across + types/groups*') involved in the nominal complements of the preposition. In relation to the differences across the corpora, this study shows that this particular function is preferred by ENL writers more often than by ELF writers.

*Table 7.21 5 Most Frequent Plural Nouns in '*across *N-s'*

ELF_Unedited						ELF_Edited						ENL_Edited					
N-S	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**	**N-S**	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**	**N-S**	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**
layers	6	0.09	7.50	0.013	0.079	*areas*	9	0.11	6.25	0.014	0.024	*types*	22	0.24	5.60	0.010	0.053
ecosystems	5	0.08	6.25	0.011	0.106	*data*	8	0.10	5.56	0.012	0.007	*groups*	20	0.22	5.09	0.009	0.046
groups	5	0.08	6.25	0.010	0.013	*groups*	8	0.10	5.56	0.013	0.019	*categories*	17	0.18	4.33	0.008	0.098
organisms	4	0.06	5.00	0.008	0.105	*types*	7	0.09	4.86	0.011	0.023	*sites*	16	0.17	4.07	0.007	0.054
segments	4	0.06	5.00	0.008	0.067	*studies*	7	0.09	4.86	0.011	0.008	*states*	12	0.13	3.05	0.005	0.041

*Table 7.22 Semantic Domains Identified via Nouns in '*across *N-s' (min ΔP 1|2 = 0.001)*

Domain	ELF_Unedited			ELF_Edited			ENL_Edited		
	Types	**Tokens (raw)**	**Tokens (normed per 10k)**	**Types**	**Tokens (raw)**	**Tokens (normed per 10k)**	**Types**	**Tokens (raw)**	**Tokens (normed per 10k)**
Entity	16	33	0.50	29	57	0.69	13	82	0.89
Class	6	12	0.18	11	39	0.48	5	65	0.71
Location	6	9	0.14	9	17	0.21	6	48	0.52
Animate	8	15	0.23	10	17	0.21	4	37	0.40
Quantity	10	12	0.18	9	12	0.15	5	22	0.24
Total	**46**	**81**	**1.22**	**68**	**142**	**1.73**	**33**	**254**	**2.76**

*Table 7.23 Examples for Identified Semantic Domains via Head Nouns in '*across *N-s' (min ΔP 1|2 = 0.001)*

Domain	ELF_Unedited	ELF_Edited	ENL_Edited
Entity	*distribution across the FSP **layers*** *SS varies across **ecosystems***	*model accuracy across **segments*** *stability across random **samples***	*trials across **tasks*** *consistent across **studies***
Class	*support across the two **groups*** *arguments across **taxa***	*vary across knowledge **areas*** *variance across categorical **groups***	*differs across **types** of entrepreneurs* *behaviors across exercise **groups***
Location	*benefit transfer across **sites***	*used across a range of **contexts***	*data pooled across **sites***

	*arrangements may vary across **countries***	*odonate individuals across **habitats***	*comparisons across **countries***
Animate	*species sorting across **organisms***	*characteristics across main types of **sellers***	*agreement across **labelers***
Quantity	*related across the data **sets***	*value for the S/E across all data **sets***	*increase across the **series***

7.2.2.3 'Over *N*'

Another spatial preposition identified as key in ENL_Edited as opposed to ELF_Unedited and ELF_Edited is '*over*', which led to the identification of the '*over* N' pattern covered in this sub-section. Cognitive linguistic research considered *'over'* to be a spatial preposition used primarily to talk about place, time, and quantity, as well as some metaphorical senses such as cause, contemplation, and preference (Dirven, 1993; Tyler & Evans, 2003). Despite the low frequencies of the pattern in the study corpora, the most common five nouns in the pattern highlight the salience of the temporal sense in academic writing (see Table 7.24).

Table 7.24 5 Most Frequent Nouns in 'over *N'*

ELF_Unedited						ELF_Edited						ENL_Edited					
Noun	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**	**Noun**	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**	**Noun**	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**
year	25	0.38	5.88	0.019	0.045	*year*	49	0.60	7.54	0.025	0.042	*time*	118	1.28	11.40	0.039	0.080
time	25	0.38	5.88	0.019	0.024	*time*	45	0.55	6.92	0.023	0.058	*year*	53	0.58	5.12	0.017	0.077
period	11	0.17	2.59	0.008	0.024	*period*	34	0.41	5.23	0.018	0.074	*period*	49	0.53	4.73	0.016	0.096
meaning	7	0.11	1.65	0.005	0.024	*decade*	18	0.22	2.77	0.009	0.187	*observation*	21	0.23	2.03	0.007	0.059
decade	6	0.09	1.41	0.005	0.105	*day*	12	0.15	1.85	0.006	0.039	*range*	21	0.23	2.03	0.007	0.040

A more in-depth analysis of the nouns attracted by *'over'* also shows that the temporal meaning is the primary sense of the construction. While there is a large gap in frequencies for the temporal sense (e.g., '*over + year/time'*) as opposed to the other senses that are much less frequent, it is clear that ENL_Edited also favors the use of quantity nouns. These nouns express the extent or degree of a quantifiable entity when used with *'over,'* which is followed by process nouns with much lower frequencies (e.g., '*over + observation + order'*). However, apart from the slightly higher frequency of quantity nouns in ELF_Edited, both ELF corpora use location nouns and general abstract terms more often than ENL_Edited. ENL_Edited, on the other hand, has no instances of general, entity, and animate nouns. The distribution of the semantic categories also points to a greater degree of variation in the use of '*over* N' in especially ELF_Unedited; that is, the study corpus with the lowest pattern frequency (see Table 7.25).

*Table 7.25 Semantic Domains Identified via Nouns in '*over *N' (min ΔP 1\|2 = 0.001)*

	Sub	**ELF_Unedited**	**ELF_Edited**	**ENL_Edited**

Domain	domain	Types	Tokens (raw)	Tokens (normed per 10k)	Types	Tokens (raw)	Tokens (normed per 10k)	Types	Tokens (raw)	Tokens (normed per 10k)
Concrete	**Time**	10	87	1.31	12	187	2.28	13	294	3.20
	Quantity	5	17	0.26	9	33	0.40	11	90	0.98
	Location	15	45	0.68	11	32	0.39	1	4	0.04
	Entity	4	12	0.18	-	-	-	-	-	-
	Animate	4	11	0.17	2	7	0.09	-	-	-
	Subtotal	**38**	**172**	**2.58**	**34**	**259**	**3.16**	**25**	**388**	**4.22**
Abstract	**General**	7	21	0.32	7	33	0.40	-	-	-
	Process	5	12	0.18	5	20	0.24	2	25	0.27
	Subtotal	**12**	**33**	**0.50**	**12**	**53**	**0.65**	**2**	**25**	**0.27**
Grand Total		**50**	**205**	**3.08**	**46**	**312**	**3.80**	**29**	**422**	**4.59**

As the examples in Table 7.26 document, some of the abstract nouns, are used in the metaphorical senses of 'over' listed by Tyler and Evans (2003), including cause (e.g., *'battle/struggle over* N*')* and preference (*'social over political order'*) in ELF_Unedited, and control (e.g., *'control/authority over* N'*)* in ELF_Edited. These senses are noticeably more frequent in ELF than in ENL. However, these abstract senses stem from a longer construction that also includes the noun preceding '*over* N', which was not identified as a key construction in this study.

*Table 7.26 Examples for Identified Semantic Domains via Nouns in '*over *N' (min ΔP 1|2 = 0.001)*

Domain	Sub-Domain	ELF_Unedited	ELF_Edited	ENL_Edited
Concrete	**Time**	*over the last three **years*** *cumulated over **time***	*change over **time*** *over the last 10 **years***	*change over **time*** *over the last 30 **years***
	Quantity	*over a wide range of **scales*** *10% over the winter **level***	*over the entire temperature **range*** *over 1000 **occurrences** in total*	*over a range of **temperatures*** *constant over the depth **interval***
	Location	*map over the journal **surface*** *distributed over an **area***	*control over the **university*** *dry air over the **gap***	*sampled over the total **plane***
	Entity	*distributions over **words***	-	-
	Animate	*power struggles over the **people***	*control over the **students***	-
Abstract	**General**	*discursive battle over the **meaning*** *options over the current **situation***	*statistical control over initial **abilities*** *authority over monetary **policy***	-
	Process	*social over the political **order***	*estimation over a CO_2 **loading***	*measured over long **observations***

7.2.2.4 'Through *N*'

The key ENL preposition '*through*,' when followed by a nominal complement, was also found to be a distinctive construction with a significantly higher frequency in ELF_Edited than in ENL_Edited. The preposition *'through'* is known primarily as a path preposition, which Radden and Dirven (2007) liken to "motion through a conduit" (p. 325). This path can be in the more concrete form of a place or time (e.g. *'through + tunnel/gate/time/year'*), as well as the more abstract senses of transmission (e.g., *'spread through media'*), and causal and procedural means for reaching a certain goal (e.g., *'success through failure*') (Dirven, 1993; Hilferty, 1999; Tyler & Evans, 2003). As seen in Table 7.27, the most frequent nouns following '*through'* can also be associated with some of these senses, and ENL_Edited, in

particular, seems to have a clear tendency to employ nouns denoting processes (e.g., '*through + process/ use/ development*).

*Table 7.27 5 Most Frequent Head Nouns in '*through *N'*

ELF_Unedited						ELF_Edited						ENL_Edited					
Noun	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**	**Noun**	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**	**Noun**	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**
sound	15	0.23	2.00	0.007	0.536	*analysis*	17	0.21	1.96	0.006	0.016	*process*	27	0.29	2.12	0.007	0.023
analysis	10	0.15	1.34	0.004	0.010	*L2*	9	0.11	1.04	0.004	0.053	*use*	24	0.26	1.88	0.006	0.034
education	6	0.09	0.80	0.003	0.023	*practice*	9	0.11	1.04	0.003	0.017	*time*	13	0.14	1.02	0.002	0.007
process	6	0.09	0.80	0.002	0.005	*dialogue*	8	0.10	0.92	0.004	0.210	*development*	11	0.12	0.86	0.003	0.018
time	6	0.09	0.80	0.001	0.005	*system*	8	0.10	0.92	0.002	0.005	*teacher*	10	0.11	0.78	0.002	0.009

A closer look at all the nouns attracted by '*through'* as its complements also confirms the observation that process nouns are the most salient type in ENL_Edited, closely followed by ELF_Edited. The use of nouns referring to concrete entities (e.g., '*through + microscope/system'*), seems to be the primary semantic choice in ELF_Unedited. In contrast, concrete nouns are only the second most common semantic domain in ELF_Edited and ENL_Edited. In ENL_Edited, concrete nouns denoting location and quantity are also more common than in ELF (e.g., *'through + series/space'*) (see Table 7.28).

*Table 7.28 Semantic Domains Identified via Nouns in '*through *N' (min ΔP 1|2 = 0.001)*

Domain	**Sub-domain**	**ELF_Unedited**			**ELF_Edited**			**ENL_Edited**		
		Types	**Tokens (raw)**	**Tokens (normed per 10k)**	**Types**	**Tokens (raw)**	**Tokens (normed per 10k)**	**Types**	**Tokens (raw)**	**Tokens (normed per 10k)**
Abstract	**Process**	13	60	0.90	21	107	1.30	14	127	1.38
	General	3	10	0.15	8	30	0.37	1	6	0.07
	Subtotal	**16**	**70**	**1.05**	**29**	**137**	**1.67**	**15**	**133**	**1.45**
Concrete	**Entity**	20	82	1.23	18	82	1.00	13	80	0.87
	Quantity	1	6	0.09	-	-	-	4	31	0.34
	Location	2	6	0.09	-	-	-	3	20	0.22
	Animate	5	18	0.27	2	10	0.12	2	15	0.16
	Subtotal	**28**	**112**	**1.68**	**20**	**92**	**1.12**	**22**	**146**	**1.59**
Grand total		**44**	**182**	**2.73**	**49**	**229**	**2.79**	**37**	**279**	**3.04**

However, as the usage examples in Table 7.29 showcase that, regardless of the semantic domains of nouns, '*through* NP', is used primarily to mention a concrete or abstract means that enables or causes the argument preceding '*through.'* This function is more common in ENL_Edited than in ELF.

*Table 7.29 Examples for Identified Semantic Domains via Nouns in '*through *N' (min ΔP 1|2 = 0.001)*

Domain	**Sub-domain**	**ELF_Unedited**	**ELF_Edited**	**ENL_Edited**
Abstract	**Process**	*processing through an **analysis*** *share through higher **education*** *through an iterative **process***	*filter through the **analysis** of texts* *perfected through **practice*** *learning through **dialogue***	*ethnogenesis through a **process** of creolization* *through the **use** of multiple samples* *improve language through professional **development***
	General	*through bearing oil **thickness***	*through a hyponymy **relation***	*history through another **lens***
Concrete	**Entity**	*passages through the **sound***	*learning through **L2***	*water moving through the **system***

		*observations though the **microscope***	*through the educational **system***	*connected through social **networks***
	Quantity	*constant through **time***	-	*third through fifth **grade*** *through a **series** of round table discussions*
	Location	*presented through **iso-surfaces***	-	*forward motion **through** space*
	Animate	*investments through global **vendors***	*realized through **client** contacts*	*through their **teachers***

7.2.2.5 'Within *N*'

Another key construction distinctive of ENL academic writing is '*within* N'. As in the case of *'in'*, Radden and Dirven (2007) identify '*within'* as a preposition of location used to express containment (p. 310). Sinclair (1996) provides further details on its other senses, including time (e.g., *'within a matter of weeks'*), proximity (e.g., '*within reach of* N'), extent (e.g., '*within + system/limit'*), as well as group membership (e.g., *'within society'*) and feelings (e.g. *'anger within himself'*) (p. 83). Particularly relevant to this study, Groom's (2007) analysis of the preposition in literature and history texts points to the contextualizing function of the patterns in which it occurs, with "contextual framing" as the most prevalent sub-function (p. 196).

The findings of this study seem to be in line with previous research in terms of the uses of '*within* N'. As the most frequent nouns in the construction suggest, the sequence is used to express both containment in a concrete sense (e.g. *'landslide', 'time', 'group'*) and abstract types of contextualization (e.g. '*field', 'context', 'framework'*) (see Table 7.30).

*Table 7.30 5 Most Frequent Head Nouns in '*within *N'*

ELF_Unedited						ELF_Edited						ENL_Edited					
Noun	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**	**Noun**	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**	**Noun**	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**
field	13	0.20	2.91	0.006	0.029	*system*	33	0.40	5.02	0.011	0.026	*system*	37	0.40	4.11	0.009	0.036
landside	13	0.20	2.91	0.007	0.043	*context*	17	0.21	2.58	0.006	0.040	*context*	33	0.36	3.67	0.008	0.070
framework	12	0.18	2.69	0.006	0.104	*framework*	15	0.18	2.28	0.005	0.067	*range*	31	0.34	3.44	0.008	0.060
community	12	0.18	2.69	0.006	0.043	*area*	15	0.18	2.28	0.005	0.017	*link*	18	0.20	2.00	0.005	0.110
group	12	0.18	2.69	0.005	0.012	*time*	15	0.18	2.28	0.004	0.012	*group*	18	0.20	2.00	0.004	0.016

A closer look at the larger number of nouns that '*within'* attracts as its object, with a minimum score of ΔP 1|2 = 0.001, points to a variety of different noun types used in this relatively infrequent pattern (see Table 7.31 and Table 7.32). However, the concrete references via entities, locations, measures, and groups seem to be much more common than the abstract ones used to frame abstract concepts and thus build arguments (e.g., *'within + framework/concept'*). While the differences between the corpora do not seem as large as documented via the statistical comparisons of overall pattern

frequencies (see Table 7.17), a slightly higher frequency of the abstract, discursive use of the pattern by ELF writers in comparison to ENL writers is observable from the analysis. While ENL writers make slightly more frequent use of the pattern for concrete references, the cross-corpus differences are not as pronounced as identified in statistical comparisons of the overall construction frequencies.

*Table 7.31 Semantic Domains Identified via Nouns in '*within *N' (min ΔP 1|2 = 0.001)*

Domain	Sub-domain	ELF_Unedited			ELF_Edited			ENL_Edited		
		Types	Tokens (raw)	Tokens (normed per 10k)	Types	Tokens (raw)	Tokens (normed per 10k)	Types	Tokens (raw)	Tokens (normed per 10k)
Abstract	Framing	8	49	0.74	7	55	0.67	5	63	0.69
	Process	10	37	0.56	9	49	0.60	4	20	0.22
	General	4	20	0.15	2	9	0.11	2	26	0.28
	Subtotal	**23**	**96**	**1.44**	**18**	**113**	**1.38**	**11**	**109**	**1.19**
Concrete	Entity	15	70	1.05	11	87	1.06	15	127	1.38
	Location	14	53	0.80	12	84	1.02	11	87	0.95
	Measure	12	55	0.83	10	72	0.88	6	74	0.81
	Group	6	42	0.63	7	49	0.60	5	49	0.53
	Subtotal	**46**	**220**	**3.30**	**40**	**292**	**3.56**	**37**	**337**	**3.67**
Grand total		**69**	**316**	**4.76**	**58**	**405**	**4.94**	**48**	**446**	**4.85**

Table 7.32 Examples for Identified Semantic Domains via Nouns in 'within *N' (min ΔP 1|2 = 0.001)*

Domain	Sub-domain	ELF_Unedited	ELF_Edited	ENL_Edited
Abstract	Framing	*studies within the* ***field*** *within the* ***framework*** *of a project*	*within the* ***context*** *of tourism* *within the* ***framework*** *of this project*	*within the* ***context*** *of field research* *within a phylogenetic* ***framework***
	Process	*activities within reenactment* ***festivals*** *boulders within landslide* ***accumulations***	*deformation within the* ***DSGSDs*** *studies within the* ***reviews***	*within the anti-vaccination* ***movement*** *within local government* ***planning***
	General	*sediment fining within the longitudinal* ***profile***	*overlap within analytical* ***uncertainty***	*contained within unique* ***links***
Concrete	Entity	*within the drainage* ***systems*** *businesses within the agricultural* ***sector***	*within the solar* ***system*** *relaxed zones within the rock* ***mass***	*working within other* ***systems*** *sediment within these* ***reaches***
	Location	*within the measurement* ***planes*** *work within the city* ***boundaries***	*ponds within the* ***area*** *faults verified within the* ***cave***	*within relatively small geographic* ***areas*** *sites within each* ***region***
	Measure	*within the* ***limits*** *of the funds allocated* *within a short* ***time*** *interval*	*within a specific* ***time*** *period* *mid-price changed within 1* ***day***	*within the range of eye* ***sizes*** *within two* ***years***
	Group	*transmitted within a* ***community*** *production of meaning within ethnic* ***groups***	*debates going on within the* ***community*** *practices within the child* ***group***	*respondents within each* ***group*** *within the medical* ***community***

Lastly, as much as the overall dominance of concrete nouns identified in this study for the '*within* N' construction might seem contradictory to Groom's (2007) findings, it is important to mention that his analysis of the *within*-sequences was not limited to a specific grammatical sequence as in the case of this study. Further, Groom's analysis did not include any hard sciences disciplines. Hence, a closer look into the disciplinary differences in the construction usage is necessary. Nevertheless, findings of the semantic analysis of the construction is still in line with previous research in the sense that the broad semantic domains of contextualization and containment are present among the nouns following '*within'*.

7.2.2.6 **'In order to *V' and* 'in *NN* with *(NP)'***

The first of the two *in*-constructions that are extensions of 'PREP N', is '*in order to* V,' the only infinitival pattern among the key constructions in this study. As summarized earlier in Table 7.17, this construction has a significantly higher frequency in ELF_Unedited and ELF_Edited than in ENL_Edited.

Being a relatively longer and more lexicalized pattern which is frequent in academic discourse, '*in order to* V' is commonly known to be used to denote a causal meaning (Simpson-Vlach & Ellis, 2010, p. 507). Similarly, as showcased in Tables 7.33 and 7.34, the same causal meaning, specifically stating the purpose of methods and processes, seems to be the dominant function of this pattern in the study corpora. An interesting difference, however, is evident from the choice of certain verbs that frequently occur in the pattern with the same function. For instance, the verb '*obtain,'* while being the most frequent in ELF_Unedited and ELF_Edited, occurs only once in ENL_Edited. Another difference is the use of high-frequency verbs in ELF_Unedited, two of them being *'get'* and *'give,'* which are the fifth and the ninth most frequent verbs in the list. Neither verb is used more than a few times in ELF_Edited and ENL_Edited, and thus is not among even the most frequent 20 verbs in this construction in the corpora.

*Table 7.33 5 Most Frequent Verbs in '*in order to *V'*

ELF_Unedited						ELF_Edited						ENL_Edited					
Verb	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**	**Verb**	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**	**Verb**	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**
obtain	18	0.27	4.76	0.003	0.174	*obtain*	16	0.20	3.94	0.002	0.134	*determine*	15	0.16	4.93	0.001	0.067
assess	11	0.17	2.91	0.002	0.246	*avoid*	11	0.13	2.71	0.001	0.171	*reduce*	11	0.12	3.62	0.001	0.070
avoid	10	0.15	2.65	0.001	0.144	*assess*	11	0.13	2.71	0.001	0.134	*rest*	9	0.10	2.96	0.001	0.100
improve	10	0.15	2.65	0.001	0.090	*determine*	10	0.12	2.46	0.001	0.087	*ensure*	8	0.09	2.63	0.001	0.078
get	8	0.12	2.12	0.001	0.072	*understand*	9	0.11	2.22	0.001	0.077	*assess*	8	0.09	2.63	0.001	0.060

*Table 7.34 Examples of in '*in order to *V' with Most Frequent Verbs across Corpora*

ELF_Unedited	ELF_Edited	ENL_Edited
***obtain** a staggered configuration*	***obtain** products with lower GI values*	***determine** the magnitude of this effect*
***assess** the actual influence of the parameter*	***avoid** the possible influence of vindication*	***reduce** sampling error*
***avoid** the child pressing an undesired key*	***assess** the stability of the estimated conditions*	***test** the hypothesized relationships*
***improve** the accuracy of the assessment*	***determine** the relationship between ST and TT*	***ensure** the data were reliable*
***get** a unique prediction for this model*	***understand** the history of the plant*	***assess** the effects of different components*

The next identified *in*-construction, '*in* NN *with* (NP)', was found to be a key construction in both ELF corpora in comparison to ENL_Edited. Both ELF corpora have four singular nouns ('NN') in the construction with a minimum ΔP 1|2 value of 0.001, as opposed to three in ENL_Edited (see Table 7.35).

Table 7.35 5 Most Frequent Singular Nouns in 'in NN *with (NP)' across Corpora*

ELF_Unedited						ELF_Edited						ENL_Edited					
Noun	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**	**Noun**	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**	**Noun**	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**
accordance	34	0.51	19.21	0.002	0.846	*accordance*	59	0.72	23.60	0.003	0.956	*line*	32	0.35	18.82	0.001	0.156
line	24	0.36	13.56	0.001	0.232	*line*	40	0.49	16.00	0.002	0.295	*accordance*	17	0.18	10.00	0.001	0.919
comparison	23	0.35	12.99	0.001	0.126	*comparison*	31	0.38	12.40	0.001	0.109	*conjunction*	16	0.17	9.41	0.001	0.775
agreement	15	0.23	8.47	0.001	0.127	*conjunction*	13	0.16	5.20	0.001	0.379	*agreement*	13	0.14	7.65	0.000	0.016
contrast	11	0.17	6.21	0.000	0.047	*combination*	13	0.16	5.20	0.000	0.121	*combination*	13	0.14	7.65	0.000	0.085

'*In* NN *with* (NP)' is a partially lexicalized version of the PNP (preposition-noun-preposition)-construction, as called by Hoffmann (2005) who extensively analyzed to explore the grammaticalization of complex prepositions in English across varieties and over time. Studies on PNP-constructions have documented these sequences in their findings based on general (Biber et al., 1999; Klégr, 1997; Quirk et al., 1985), as well as academic language data (Benelhadj, 2019; Hyland, 2008a, 2008b; Simpson-Vlach & Ellis, 2010).

Similar to the 'PREP N *to* (NP)' construction discussed earlier, these constructions are also instrumental in discourse organization, which is also in line with the combination of the senses of circumstance and conceptual association for *'in'* and *'with'* respectively (Dirven, 1993, pp. 80-81). Klégr's (1997) analysis suggests a more refined semantic categorization of PNP-constructions, where *'in line/accordance/agreement with'* is associated with the expression of agreement, *'in comparison/contrast with'* with that of contrast, *'in conjunction with'* with cooperation, and '*in combination with'* with manner (pp. 64-69). As the examples in Table 7.36 show, this categorization can also explain the usage of the constructions in this study. Hence, it is plausible to suggest that ELF writers have a somewhat stronger tendency than ENL writers to use this pattern to express contrasts. This tendency is most evident in ELF_Unedited as '*in contrast with* (NP)*'* is also a frequent MWU, despite ELF_Edited having a higher overall frequency of the '*in* NN *with* (NP)' construction.

*Table 7.36 Examples of '*in *NN* with *(NP)' (for 'NNs' with min. ΔP 1|2 = 0.001)*

ELF_Unedited	ELF_Edited	ENL_Edited
in accordance with *the Institutional and National Research Council's guideline*	***in accordance with*** *gravitationally widened L and D1 joint sets*	***in line with*** *previous studies*
examples ***in line with*** *empirically estimated model*	***in line with several*** *other reports*	***in accordance with*** *the stated values of the community*
in comparison with *the previous definition*	***in comparison with*** *the sedentary controls*	*when used* ***in conjunction with*** *amines*
in agreement with *the results of the study*	*Article 14* ***in conjunction with*** *Article 9*	

7.2.3 Constructions with Key Prepositions as Complements of Predicative Adjectives: 'ADJ PREP (NP)'

In addition to the nominal *of*-constructions discussed earlier, three constructions where predicative adjectives followed by an *of*-phrase as their complement were also found to be used differently across corpora. The first two of the 'ADJ PREP (NP)' constructions, 'ADJ *of* (NP)' and 'ADJ *at* (NP),' start with a central adjective (e.g., *important*, *remarkable*, etc.). 'ADJ-SUPERL *of* (NP),' however, starts with predicative adjectives that are superlatives.

Comparisons across corpora showed that with higher frequencies, ELF_Unedited and ELF_Edited differ significantly from ENL_Edited in the use of 'ADJ *of* (NP).' At the same time, only ELF_Unedited has a significantly higher frequency of 'ADJ *at* (NP)' compared to ENL_Edited. In the case of 'ADJ-SUPERL *of* (NP),' the opposite is observed; that is, both ELF corpora have significantly higher frequencies of this construction. In line with the frequencies of the predicative adjectives in the corpora, 'ADJ *of* (NP)' has a more predictable distribution of adjectives in. In contrast, the ELF corpora have higher predictability in 'ADJ-SUPERL *of* (NP).' As for 'ADJ *at* (NP),' despite the lowest item frequency observed, ELF_Unedited shows a more predictable distribution of the construction-initial adjectives by having the lowest H_{norm} value (see Table 7.37).

Table 7.37 Numerical Summary of 'ADJ PREP (NP)'

	Cx	ELF_Unedited				ELF_Edited				ENL_Edited			
		Raw	Normed (per 10k)	DP_{norm}	H_{norm} (ADJ)	Raw	Normed (per 10k)	DP_{norm}	H_{norm} (ADJ)	Raw	Normed (per 10k)	DP_{norm}	H_{norm} (ADJ)
Frequencies	ADJ *of* (NP)	327	4.91	0.36	0.84	409	4.99	0.31	0.84	627	6.82	0.28	0.75
	ADJ *at* (NP)	68	1.02	0.65	0.65	106	1.29	0.59	0.89	158	1.72	0.45	0.88
	ADJ-SUPERL *of* (NP)	136	2.04	0.50	0.16	144	1.76	0.49	0.23	82	0.89	0.61	0.38
	Cx	**ELF_Unedited vs. ENL_Edited**				**ELF_Edited vs. ENL_Edited**				**ELF_Unedited vs. ELF_Edited**			
Statistics		G^2	%DIFF	p^a		G^2	%DIFF	p^a		G^2	%DIFF	p^a	
	ADJ *of* (NP)	24.10	28.09	<.001		24.77	26.91	.003		0.05	1.62	.525	
	ADJ *at* (NP)	15.29	47.45	.027		8.34	34.43	.249		15.29	47.45	.166	
	ADJ-SUPERL *of* (NP)	36.46	128.67	<.001		26.01	96.77	<.001		1.58	16.22	.568	

[a]P-value for Welch's two-tailed t-test

Regarding 'ADJ *of* (NP),' a small number of adjectives are attracted by *'of'*, with a minimum ΔP 1|2 value of 0.001 for only two fillers in ELF_Unedited (*'aware', 'capable'*), and four in ELF_Edited (*'many', 'capable', 'aware', 'typical'*). ENL_Edited, on the other hand, has seven adjectives that meet this association criterion ('*many,' 'capable,' 'much,' 'independent,' 'aware,' 'representative,' 'typical'*).

'ADJ *at* (NP),' while having recognizably lower frequency in all three corpora, has a larger number of adjectives that are at least somewhat attracted by '*of*.' Nevertheless, as in the case of 'ADJ of (NP),' ENL_Edited has the highest number of adjectives in this construction, with a minimum ΔP 1|2 of 0.001 (N=15), followed by ELF_Edited (N= 12), and then ELF_Unedited (N=11). 'ADJ-SUPERL *of* (NP)' is different from the other two adjectival constructions in that only one lexical filler, i.e., '*most*,' has a ΔP 1|2 over 0.001, and the frequencies for the word in the ELF corpora are much higher (see Table 7.38).

Table 7.38 10 Most Frequent Adjectives in 'ADJ PREP (NP)'

Cx	ELF_Unedited						ELF_Edited						ENL_Edited					
	ADJ	Raw	Normed (per 10k)	%	ΔP 1|2	ΔP 2|1	ADJ	Raw	Normed (per 10k)	%	ΔP 1|2	ΔP 2|1	ADJ	Raw	Normed (per 10k)	%	ΔP 1|2	ΔP 2|1
ADJ *of* (NP)	*aware*	45	0.68	13.76	0.002	0.682	*Many*	41	0.50	10.02	0.001	0.092	*many*	110	1.20	17.54	0.002	0.147
	many	31	0.47	9.48	0.000	0.072	*capable*	37	0.45	9.05	0.001	1.000	*capable*	57	0.62	9.09	0.002	0.934
	independent	17	0.26	5.20	0.000	0.148	*aware*	25	0.30	6.11	0.001	0.781	*much*	50	0.54	7.97	0.001	0.303
	capable	16	0.24	4.89	0.001	0.640	*typical*	24	0.29	5.87	0.001	0.216	*independent*	43	0.47	6.86	0.001	0.250
	downstream	10	0.15	3.06	0.000	0.233	*independent*	21	0.26	5.13	0.000	0.144	*aware*	31	0.34	4.94	0.001	0.775
ADJ *at* (NP)	*available*	6	0.09	8.82	0.003	0.035	*significant*	21	0.26	19.81	0.007	0.034	*significant*	17	0.18	10.76	0.004	0.024
	present	5	0.08	7.35	0.002	0.015	*present*	6	0.07	5.66	0.002	0.011	*available*	14	0.15	8.86	0.004	0.039
	significant	4	0.06	5.88	0.001	0.005	*available*	4	0.05	3.77	0.001	0.015	*present*	13	0.14	8.23	0.003	0.035
	possible	3	0.05	4.41	0.001	0.005	*good*	4	0.05	3.77	0.001	0.011	*different*	8	0.09	5.06	0.001	0.005
	similar	3	0.05	4.41	0.001	0.007	*online*	4	0.05	3.77	0.001	0.056	*constant*	6	0.07	3.80	0.002	0.028
ADJ-SUPER *of* (NP)	*most*	124	1.86	91.18	0.004	0.405	*most*	133	1.62	92.36	0.004	0.360	*most*	69	0.75	84.15	0.002	0.161
	best	9	0.14	6.62	0.000	0.022	*best*	9	0.11	6.25	0.000	0.029	*best*	6	0.07	7.32	0.000	0.014
	closest	1	0.74	0.73	0.000	0.034	*least*	1	0.01	0.69	0.000	-0.038	*worst*	5	0.05	6.10	0.000	0.255
	largest	1	0.74	0.74	0.000	-0.029	*simplest*	1	0.01	0.69	0.000	0.100	*crudest*	1	0.01	1.22	0.000	0.961
	oldest	1	0.74	0.74	0.000	0.057	-	-	-	-	-	-	*largest*	1	0.01	1.22	0.000	-0.024

While this construction has not been the focus of research, two similar constructions with predicative adjectives, '*it* V-*be* (ADV) ADJ *to*' and '*it* V-be (ADV) ADJ *that*,' received considerable attention in uncovering aspects of evaluative language use in ELF academic writing (e.g., Mur-Dueñas, 2015, 2018a; Vetchinnikova, 2017). However, following Langacker's (1991) cognitive linguistic approach to the analysis of this pattern as a conventionalized unit, Osmond (1997) focused on the semantic function of how emotions are expressed via the rather fixed 'ADJ PREP (NP)' construction. In the corpus-based reference book on collocations by Benson, Benson, and Ilson (2010), 'adjective + preposition' was also given particular attention as an important adjectival collocation in English. As Biber et al. (1999) stated, predicative adjectives, in addition to being used commonly followed by prepositional phrases as their complements, function as "frame[s] for [intellectual] claims" in academic writing (p. 518). Lastly, Francis et al. (1998) provided a detailed analysis of the different realizations of the pattern including the identification of 13 semantic groups denoting meanings of attitude, emotion quality, and

state for 'ADJ *of* (NP)' (pp. 451-457), and three primarily emotive semantic groups for 'ADJ *at* (NP)' (pp. 428-430). A relatively more recent study of the 'ADJ *of* (NP)' pattern in a sample of the academic subcorpus of the British National Corpus (BNC) also pointed to the common use of the pattern as a means of evaluation and describing research findings (Chen & Chung, 2018).

As for this study, despite the small type and token frequencies of fillers that *'of'* attracts in this pattern, certain meanings expressed by these patterns are seen in the examples listed in Table 7.39. In the case of 'ADJ *of* (NP),' most adjectives denote descriptive qualities of the following noun group (e.g., '*capable/typical of* (NP)') with the exceptions of '*much'* and '*many,'* both of which have higher frequencies and ΔP 1|2 values in ENL_Edited than in the ELF corpora.

'ADJ *at* (NP)' has a larger number of adjectival filler types with much lower token frequencies than in 'ADJ *of* (NP)'. The construction refers to certain descriptive qualities (e.g., '*significant/similar at* [NP]), generally followed by a concrete noun group denoting time, place or quantity. With 'ADJ-SUPERL *of* (NP)' the dominance of *'most'* among the fillers shows that quantification, more specifically a part-whole relation between two entities, is the major meaning of the pattern, which is used much more frequently in the ELF corpora than in ENL_Edited. This is an interesting finding in relation to the fact that the quantity related head nouns in 'N *of* (NP)' were found to occur much more commonly in ENL_Edited than in ELF corpora. Hence, taken together, the findings on these two *of*-constructions explain more of the differences between ELF and ENL writers in terms of how they express quantification via *of*-phrases.

Table 7.39 Examples for 'ADJ PREP (NP)'

Cx	ELF_Unedited	ELF_Edited	ENL_Edited
ADJ *of* (NP)	***aware** of her self* ***many** of these women* ***independent** of temperature* ***capable** of broadening* ***downstream** of the leading edge*	***many** of its advocates* ***capable** of teaching* ***aware** of that fact* ***typical** of the material* ***independent** of scale*	***many** of the comments* ***capable** of killing* ***independent** of infection status* ***aware** of these beliefs* ***much** of the territory*
ADJ *at* (NP)	***available** at the website* ***present** at different stages* ***significant** at the 5% level* ***similar** at the beginning* ***possible** at that time*	***significant** at the 0.01 level* ***present** at low levels* ***available** at* (webpage link) ***good** at a lower price* ***online** at* (webpage link)	***significant** at 1.0%* ***available** at the time* ***present** at the site* ***different** at α= 0.05* ***constant** at 3.3m*
ADJ-SUPERL of (NP)	***most** of the time* *(to the) **best** of our knowledge*	***most** of the respondents* *(to the) **best** of our knowledge*	***most** of the variations* *(to the) **best** of my ability* ***worst** of all*

Lastly, the categorizing of 'ADJ-SUPERL of (NP)' following the PoS-tagger which identified it as a superlative adjective, might seem misleading as, for example, the most common filler, '*most'* is a pronoun denoting large quantities (Biber et al., 1999, p. 277). Biber et al. also continue to add that superlative forms are rare in academic prose. As stated by Li and Wharton (2012), academics are likely to avoid the use of superlatives due to their boosting effect when making scientific claims. However, '*most of* (NP)' is also identified as a commonly used quantifying sequence by Hyland (2008b), which may indicate that '*most'* is an exception to the general infrequency of superlatives in academic writing. Further, although previous research argues that superlatives are primarily boosting devices, the common use of the highly fixed sequences '*to the best of our/my knowledge/ability'* in the study corpora signify a hedged claim.

*7.2.4 Passive Constructions Followed by Key Prepositions: '(NP) V-*be *V-*ed *PREP (NP)'*

Three prepositional passive constructions, i.e., '(NP) V-be V-ed PREP (NP)' were also identified as core ELF constructions in the study. As is reported in Table 7.40, three passive constructions with prepositions *'by'* (e.g., '*habitats are characterized by* ….'),'*in'* (e.g., '*methods were used in* ….') and *'on'* (e.g., '*analysis was based on* ….') have significantly higher frequencies in ELF_Unedited and ELF_Edited than in ENL_Edited. *By*-passives and *in*-passives have higher overall frequencies in the two ELF corpora, both of which have higher frequencies than ENL_Edited. Although the *on*-passive construction is also significantly more common in the two ELF corpora, it is even more so in ELF_Unedited. Besides, according to the H_{norm} values, while all three constructions have more predictable distributions in ELF than ENL, the difference is more nuanced in the case of the *on*-passive. Both ELF corpora have remarkably more predictable distributions of verbs in *on*-passives (see Table 7.40).

*Table 7.40 Numerical Summary of '(NP) V-*be *V-*ed *PREP (NP)'*

		ELF_Unedited				ELF_Edited				ENL_Edited			
Frequencies	**Cx**	**Raw**	**Normed (per 10k)**	**DP_{norm}**	**H_{norm} (verbs)**	**Raw**	**Normed (per 10k)**	**DP_{norm}**	**H_{norm} (verbs)**	**Raw**	**Normed (per 10k)**	**DP_{norm}**	**H_{norm} (verbs)**
	(NP) V-*be* V-*ed by* (NP)	1,255	18.78	0.22	0.88	1,529	18.64	0.21	0.87	1,372	14.93	0.20	0.90
	(NP) V-*be* V-*ed in* (NP)	1,344	20.12	0.21	0.85	1,669	20.35	0.22	0.84	1,595	17.36	0.22	0.85
	(NP) V-*be* V-*ed on* (NP)	359	5.37	0.34	0.64	377	4.60	0.35	0.64	303	3.30	0.34	0.73
Statistics	**Cx**	**ELF_Unedited vs. ENL_Edited**				**ELF_Edited vs. ENL_Edited**				**ELF_Unedited vs. ELF_Edited**			
		G^2	**%DIFF**	**p^a**		**G^2**	**%DIFF**	**p^a**		**G^2**	**%DIFF**	**p^a**	
	(NP) V-*be* V-*ed by* (NP)	34.31	25.79	<.001		35.63	24.83	.002		0.04	0.77	.890	

	(NP) V-*be* V-*ed in* (NP)	15.75	15.87	.009	20.56	17.21	<.001	0.10	1.14	.808
	(NP) V-*be* V-*ed on* (NP)	39.31	62.93	<.001	18.65	39.37	.018	18.65	39.37	.069

[a]P-value for Welch's two-tailed t-test

Also known as long passives, these constructions introduce new information at the end of a clause, generally through a prepositional agent after the main verb (Biber et al., 1999; Talmy, 2000). Similarly, from a CxG perspective, Goldberg (2003) identifies the discourse function of the passive construction as "to make undergoer topical and/or actor non-topical" (p. 220). Similarly, since the early days of multidimensional register analysis, both passives in general and passives with a *by*-agent were found to be a defining characteristic of academic prose as a form of "procedural discourse" that is "typically abstract and technical in content, and formal in style"(Biber, 1988, p. 112).

7.2.4.1 By-*passive: '(NP) V*-be *V*-ed by *(NP)'*

The first construction is the passive with *by*-agent; that is, the *by*-passive. '*By,'* as the key preposition with second highest keyness score after '*of*' in both ELF corpora in comparison to the ENL corpus, is also the most common one that accompanies a passive sentence in general. While the primary sense of the preposition is identified as expressing a connection between two entities, aside from a temporal and spatial connectedness, it can be used to also denote a means, cause, or specification of abstract and quantitative properties (Dirven, 1993; Radden & Dirven, 2007; Sinclair, 1996; Strauss et al., 2018). Further, a passive verb followed by a *by*-agent could be associated with "de-emphasizing the subject within a long passive," thus communicating a somewhat impersonal stance (Baratta, 2009, p. 1409).

As seen in Table 7.41, the verb types most commonly used in *by*-passives are similar across corpora. However, ranking and token frequencies are especially different in ENL_Edited as the verbs are generally less frequent and towards the end of the list, the corpus has distinct verbs such as *'generate', 'drive', 'follow',* and *'calculate'* that are not found among the most common for the ELF corpora.

Table 7.41 10 Most Frequent Verbs in '(NP) V-be *V*-ed by *(NP)'*

ELF_Unedited						ELF_Edited						ENL_Edited					
V-*ed*	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**	**V-*ed***	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**	**V-ed**	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**
characterize	52	0.78	4.14	0.427	0.012	*characterize*	66	0.80	4.32	0.013	0.535	*determine*	47	0.51	3.43	0.009	0.97
explain	47	0.70	3.75	0.574	0.011	*affect*	56	0.68	3.66	0.011	0.461	*characterize*	45	0.49	3.28	0.009	0.54
determine	42	0.63	3.35	0.272	0.010	*explain*	52	0.63	3.40	0.010	0.536	*explain*	28	0.30	2.04	0.005	0.29
affect	36	0.54	2.87	0.539	0.008	*influence*	50	0.61	3.27	0.010	0.679	*affect*	28	0.30	2.04	0.005	0.36
influence	34	0.51	2.71	0.590	0.008	*determine*	47	0.57	3.07	0.009	0.277	*influence*	24	0.26	1.75	0.005	0.32

represent	28	0.42	2.23	0.367	0.006	*cause*	33	0.40	2.16	0.006	0.361	*support*	23	0.25	1.68	0.004	0.32
obtain	25	0.37	1.99	0.080	0.005	*represent*	29	0.35	1.90	0.006	0.217	*generate*	21	0.23	1.53	0.004	0.64
cause	22	0.33	1.75	0.299	0.005	*measure*	28	0.34	1.83	0.005	0.104	*drive*	19	0.21	1.38	0.004	0.40
support	21	0.31	1.67	0.561	0.005	*support*	25	0.30	1.64	0.005	0.589	*follow*	19	0.21	1.38	0.004	0.53
accompany	19	0.28	1.51	0.649	0.004	*form*	24	0.29	1.57	0.005	0.264	*calculate*	18	0.20	1.31	0.003	1.11

The meanings of the constructions were categorized using a coding scheme adapted from several studies on verbs in academic writing (Hyland, 1999; Thomas & Hawes, 1994; Thompson & Yiyun, 1991). These studies also included a category for cognitive (mental) activities, which was merged with the discourse group denoting textual connections for this analysis because of the great degree of overlap between the two categories for the instances of *by*-passives. This adaptation was considered necessary as Thompson and Yiyun (1991), the researchers that came up with the scheme, suggested, "these categories are, of course, not watertight—nor do they need to be. There is 'bleeding' from one category to another." (p. 370). As Tables 7.42 and 7.43 show, the *by*-passive is used predominantly to refer to a means utilized in discourse construction (e.g., '*characterize', 'explain'*) and the reporting of research processes (e.g., *'obtain', 'measure'*) and findings (e.g., '*determine'*). Although these two categories have quite similar frequencies in ELF_Unedited and ELF_Edited, ENL_Edited was found to have a noticeably lower smaller frequency of verb tokens in the discourse category.

Table 7.42 Semantic Domains Identified via Verbs in '(NP) V-be *V*-ed by *(NP)' across Corpora (min ΔP 1|2 = 0.001)*

Domain	ELF_Unedited			ELF_Edited			ENL_Edited		
	Types	**Tokens (raw)**	**Tokens (normed per 10k)**	**Types**	**Tokens (raw)**	**Tokens (normed per 10k)**	**Types**	**Tokens (raw)**	**Tokens (normed per 10k)**
Discourse	32	415	6.21	40	534	6.51	35	395	4.30
Process	73	444	6.65	79	533	6.50	77	504	5.49
Finding	6	63	0.94	7	83	1.01	9	91	0.99
Total	111	922	13.80	126	1150	14.02	121	990	10.78

Table 7.43 Examples for Identified Semantic Domains via Verbs in '(NP) V-be *V*-ed by *(NP)' (min ΔP 1|2 = 0.001)*

Domain	ELF_Unedited	ELF_Edited	ENL_Edited
Discourse	*secondary habitats are **characterized** by frequent allogeneic succession.* *contrast can be **explained** by similar X-ray scatter lengths.* *Absorption is mostly **affected** by the precipitation*	*channels are **characterized** by the occurrence of bedrock* *this finding is **explained** by the fact that children were **affected** by an increase in cognitive load*	*amplitude was **characterized** by its great deviation* *river system was **affected** by the Asian Monsoon climate* *this is **explained** by the apparent stabilizing effect.*
Process	*index (calcHI) was **obtained** by dividing the area* *this was **measured** by adding a tolerance on the location* *membranes were **estimated** by the weight loss*	*variables were **measured** by a 5-point Likert scale* *it is **formed** by one crevice* *bulk soil was **obtained** by total digestion*	*it was not **generated** by inversion* *heat was **calculated** by a piecewise polynomial function* *the heat flux is **given** by Fouriers law*
Finding	*values were **determined** by optical microscopy*	*mechanical behavior is **determined** by the matrix*	*tax rate is **determined** by the taste for status*

7.2.4.2 In-*passive: '(NP) V*-be *V*-ed in *(NP)'*

The second prepositional passive construction distinctive of ELF takes the preposition *'in'*, which, as explained earlier in section 7.2.1.1, primarily denotes a spatial enclosure or containment (Dirven, 1993; Tyler & Evans, 2003). As in the case of 'N *of* NP *in* (NP)', the *in*-phrase enables the further specification of what is expressed earlier in the construction. Thus, the activities expressed in the passive verbs are followed by spatial elaboration of these activities (e.g., '[NP] *V*-be *V*-ed *found/used/seen/presented in table/figure/sample/study'*)

As is seen in Table 7.44, the 10 most common verbs in the *in*-passive construction overlap largely for the three corpora, with slightly higher frequencies and ΔP values for ELF_Unedited. A functional categorization similar to that of the verbs in *by*-passive was employed to take a closer look at the semantics of the patterns primarily by its relation to the verbs, subjects and prepositional complements they attract through '*in'*. As reported in Table 7.45, the uses of the constructions to denote research processes (e.g., '[NP] *V*-be *V*-ed *used/included in [NP]'*) and report findings activities (e.g., '[NP] *V*-be *V*-ed *found/shown in [NP]*) are remarkably more common in both ELF corpora than in ENL_Edited. The uses of the construction specifically denoting a location (e.g., '[NP] *V*-be *V*-ed *located in [NP]'*) are quite infrequent in all three corpora.

Table 7.44 10 most frequent Verbs in '(NP) V-be *V*-ed in *(NP)'*

ELF_Unedited						ELF_Edited						ENL_Edited					
V-*ed*	**Raw**	**Normed (per 10k)**	**%**	0.083	**ΔP 2|1**	**V-*ed***	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**	**V-*ed***	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**
use	86	1.29	6.40	0.270	0.083	*use*	93	1.13	5.57	0.003	0.067	*use*	107	1.16	6.71	0.003	0.067
find	69	1.03	5.13	0.527	0.270	*find*	81	0.99	4.85	0.003	0.189	*include*	70	0.76	4.39	0.003	0.189
include	58	0.87	4.32	0.155	0.527	*observe*	72	0.88	4.31	0.003	0.294	*find*	69	0.75	4.33	0.003	0.294
show	46	0.69	3.42	0.208	0.155	*include*	61	0.74	3.65	0.003	0.473	*show*	69	0.75	4.33	0.003	0.473
observe	44	0.66	3.27	0.156	0.208	*present*	47	0.57	2.82	0.002	0.182	*present*	47	0.51	2.95	0.002	0.182
present	36	0.54	2.68	0.201	0.156	*show*	45	0.55	2.70	0.002	0.089	*see*	41	0.45	2.57	0.002	0.089
report	35	0.52	2.60	0.145	0.201	*involve*	44	0.54	2.64	0.002	0.283	*observe*	36	0.39	2.26	0.002	0.283
see	30	0.45	2.23	0.289	0.145	*give*	39	0.48	2.34	0.001	0.050	*summarize*	29	0.32	1.82	0.001	0.050
involve	29	0.43	2.16	0.129	0.289	*see*	37	0.45	2.22	0.001	0.154	*locate*	25	0.27	1.57	0.001	0.154
conduct	20	0.30	1.49	0.001	0.129	*report*	35	0.43	2.10	0.001	0.203	*report*	26	0.28	1.63	0.001	0.203

Table 7.45 Semantic Domains Identified via Verbs in '(NP) V-be *V*-ed in *(NP)' (min ΔP 1|2 = 0.001)*

Domain	**ELF_Unedited**			**ELF_Edited**			**ENL_Edited**		
	Type	**Token (raw)**	**Token (normed per 10k)**	**Type**	**Token (raw)**	**Token (normed 10k)**	**Type**	**Token (raw)**	**Token (normed per 10k)**
Process	11	252	3.77	7	274	3.34	7	254	2.76
Finding	11	205	3.07	7	289	3.52	5	216	2.35
Discourse	9	150	2.25	8	207	2.52	7	185	2.01

Location	2	22	0.33	1	17	0.21	1	25	0.27
Total	**33**	**629**	**9.41**	**23**	**787**	**9.59**	**20**	**680**	**7.40**

As the examples of the semantic domains in Table 7.46 show, the slightly different senses of enclosure are evident in the three lists. Whether it is a process, finding, a discourse connection, or coordinates of a place, the 'NP' after *'in'* aims to specify a location physically (e.g., '[NP] V-*be located in regions'*), conceptually (e.g., "[NP] V-*be used in a similar manner'*) or textually (e.g., "[NP] V-*be presented in appendix'*).

*Table 7.46 Examples for Identified Semantic Domains via Verbs in '(NP) V-*be *V-*ed in *(NP)' (min ΔP 1|2 = 0.001)*

Domain	ELF_Unedited	ELF_Edtied	ENL_Edited
Process	*that might be **used** in a number of areas* *it should be **included** in their innovation strategy*	*Bacterial culture methods were **used** in this study…* *that must be **included** in the mission*	*measurements are **used** in a similar manner* *the unemployed are **included** in the sample.*
Finding	*events were **found** in three studies* *loan terms that are **observed** in our sample*	*descriptions…. can be **found** in these references.* *effect was **observed** in all experimental runs*	*similar concepts can also be **found** in SCM which has been **shown** in many cultures*
Discourse	*results are **presented** in figure* *as can be **seen** in table*	*indicator is **presented** in table* *accounts are **given** in interviews*	*method is **presented** in appendix* *as can be **seen** in figure*
Location	*orbits are also **located** in regions*	*sample was **placed** in an aluminum block*	*stock is **located** in European regions*

7.2.4.3 On-*passive: '(NP) V-be V-ed* on *(NP)'*

The third and last prepositional passive construction is '(NP) V-be V-ed *on* (NP)'; that is, the *on*-passive. The collocation of '*base'* and *'on'* commonly occurs in lists of frequent sequences used in academic prose to specify certain attributes or describe research processes or findings (Cortes, 2013; Simpson-Vlach & Ellis, 2010). As seen in Table 7.47, this verb is also the predominant lexical choice for the *on*-passives found in the three study corpora.

*Table 7.47 5 Most Frequent Verbs in '(NP) V-*be *V-*ed on *(NP)'*

ELF_Unedited						ELF_Edited						ENL_Edited					
V-ed	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**	**V-ed**	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**	**V-ed**	**Raw**	**Normed (per 10k)**	**%**	**ΔP 1|2**	**ΔP 2|1**
base	165	2.47	0.69	0.272	0.041	*base*	177	2.16	63.90	0.015	0.102	*base*	110	1.20	36.30	0.021	0.170
focus	18	0.27	0.08	0.251	0.004	*focus*	17	0.21	6.14	0.013	0.290	*perform*	11	0.12	3.63	0.002	0.091
perform	15	0.22	0.06	0.090	0.004	*place*	8	0.10	2.89	0.011	0.546	*focus*	9	0.10	2.97	0.002	0.078
place	10	0.15	0.04	0.176	0.002	*premise*	3	0.04	2.53	0.008	0.174	*place*	8	0.09	2.64	0.001	0.102
conduct	7	0.10	0.03	0.048	0.002	*predicate*	2	0.02	2.53	0.008	0.228	*conduct*	8	0.09	2.64	0.001	0.058

The semantic domains identified here also point to the discoursal importance of this construction, with a remarkable difference between the ELF corpora and ENL_Edited. It seems that by utilizing the discoursal meaning expressed by the *on*-passive (e.g., '[NP] V-*be based/focused' on* [NP]), ELF writers emphasize

the importance of providing the foundation of their findings and claims, as illustrated by Biber et al. (1999, p. 1021)

*Table 7.48 Semantic Domains Identified via Verbs in '(NP) V-*be *V-*ed on *(NP)' (min ΔP 1|2 = 0.001)*

Domain	ELF_Unedited			ELF_Edited			ENL_Edited		
	Type	Token (raw)	Token (normed per 10k)	Type	Token (raw)	Token (normed per 10k)	Type	Token (raw)	Token (normed per 10k)
Discourse	2	183	2.74	2	194	2.37	2	119	1.30
Process	14	65	0.97	17	48	0.58	14	69	0.75
Location	1	10	0.15	1	8	0.10	2	12	0.13
Total	**17**	**258**	**3.86**	**20**	**250**	**3.05**	**18**	**200**	**2.18**

Table 7.49 Examples for Identified Semantic Domains via Verbs in *'(NP) V-*be *V-*ed on *(NP)' (min ΔP 1|2 = 0.001)*

Domain	ELF_Unedited	ELF_Edtied	ENL_Edited
Discourse	*analysis was **based** on the self-reports* *study is **focused** on lived experiences*	*indices that are **based** on the species richness* *this paper is **focused** on research that*	*analysis is **based** on a field visit* *we are less **focused** on children*
Process	*experiments were **performed** on the beamline* *analysis was **conducted** on the basis of …. interviews*	*focus is **set** on those strategies* *experiments were **performed** on the AMP-PC system*	*analysis was **conducted** on the data* *experiments were **performed** on cultures*
Location	*mask was **placed** on the keyboard*	*advertisement was **placed** on the first page*	*samples were **placed** on the moving platform*

Based on the analysis of the *by*-passive ([NP] V-be V-ed *by* [NP]), *in*-passive ([NP] V-be V-ed in [NP]), and *on*-passive ([NP] V-be V-ed *by* [NP]), it is plausible to suggest that there are certain parallels between these findings and previous research. For instance, Chen and Baker (2010) reported differences in these patterns between students and expert writers in terms of frequency and function. While native experts used the constructions more frequently, Chinese L2 students used them less often, and both L1 and L2 students used them more often for discourse organization and elaboration than ENL writers. Pan et al. (2016) and Pérez-Llantada (2012, 2014), on the other hand, showed that Chinese and Spanish L2 expert writers respectively, used them more often than the ENL experts. Similar results were reported in Yilmaz and Römer's (2020) analysis of WrELFA in comparison to American academic writing, where *on*-passives were found to be distinctive of ELF. As Pan et al. (2016), the study reported a distinctive function of the patterns, that is, to describe research-related entities and activities. Besides, Pérez-Llantada (2014) added that passives have an effect on the writers' construction of stance, helping them distance themselves from the propositions made. However, diachronic research on the changes in academic Englishes also points to a decrease in the use of passives for a clearer, more interactive, and

personal authorial style, which is considered to be characteristic of scientific discourse today (Hyland & Jiang, 2019; Seoane, 2013). Besides, Seoane (2006) passive in academic writing today "have become mere conventional ornaments and are functionally empty" (p. 204). Hence, the findings on the three prepositional passive constructions identified in this study have complex implications for our understanding of written academic ELF, which are further illustrated in the following chapters.

7.2.5 Constructions with Key Prepositions as Postmodifiers of Determiners: 'DET **of** *(NP)'*

The last construction identified in the study is 'DET *of* (NP)', which includes a determiner, quantifier, or demonstrative pronoun followed by a prepositional phrase as its complement (e.g., '*those/that/some of* [NP]'). The normalized frequencies of the construction were found to differ significantly across corpora, with ENL_Edited having the highest value. However, interestingly, unlike most other identified patterns, all three pairwise comparisons were significant, also indicating a difference between not only the ELF corpora and the ENL corpus but also among the former two. Regarding the distributional characteristics of the lexical fillers within the pattern, the H_{norm} values indicate that the determiners filling the first slot in the construction, unlike the content words that were investigated in the other patterns as the primary element of their form and functions, signified a quite predictable distribution for all the three corpora. This is an expected result as there is a limited number of determiners available. However, in line with the remarkably higher construction frequencies in ENL_Edited, both ELF corpora have slightly higher H_{norm} values. This difference points to a more unpredictable distribution of the DET fillers in ELF than in ENL (see Table 7.50).

Table 7.50 Numerical Summary of 'DET of *(NP)'*

Frequencies for Cz	**ELF_Unedited**				**ELF_Edited**				**ENL_Edited**			
	Raw	**Normed (per 10k)**	**DP_{norm}**	**H_{norm} (DET)**	**Raw**	**Normed (per 10k)**	**DP_{norm}**	**H_{norm} (DET)**	**Raw**	**Normed (per 10k)**	**DP_{norm}**	**H_{norm} (DET)**
DET *of* (NP)	280	4.20	0.34	0.74	454	5.53	0.21	0.74	708	7.70	0.26	0.71
Statistics for Cx	**ELF_Unedited vs. ENL_Edited**				**ELF_Edited vs. ENL_Edited**				**ELF_Unedited vs. ELF_Edited**			
	G^2	**%DIFF**	**p^a**		**G^2**	**%DIFF**	**p^a**		**G^2**	**%DIFF**	**p^a**	
DET *of* (NP)	79.59	45.47	<.001		30.83	28.15	.004		13.43	24.11	<.001	

[a]P-value for Welch's two-tailed t-test

In addition, a closer look at these fillers and the strength of '*of*' co-selecting the determiners, that is, DET fillers with a minimum ΔP 1|2 of 0.001, also showcases the cross-corpus differences. In ELF_Unedited

only two DET fillers (*'some,' 'that'*) pass the ΔP 1|2 threshold. In contrast, five DET fillers ('*some,' 'that,' 'each,' 'all,' 'those'*) meet the ΔP 1|2 criterion in ELF_Edited and ENL_Edited, where the latter has generally higher values (see Table 7.51).

Table 7.51 5 Most Frequent Determiners in 'DET of *(NP)'*

ELF_Unedited						ELF_Edited						ENL_Edited					
DET	**Raw**	**Normed**	**%**	**ΔP 1|2**	**ΔP 2|1**	**DET**	**Raw**	**Normed**	**%**	**ΔP 1|2**	**ΔP 2|1**	**DET**	**Raw**	**Normed**	**%**	**ΔP 1|2**	**ΔP 2|1**
some	105	1.58	37.50	0.002	0.074	*some*	127	1.55	27.97	0.003	0.090	*each*	223	2.43	31.50	0.005	0.112
all	39	0.59	13.93	0.000	0.017	*that*	81	0.99	17.84	0.002	0.100	*some*	144	1.57	20.34	0.003	0.109
each	37	0.56	13.21	0.000	0.008	*each*	80	0.98	17.62	0.001	0.047	*all*	121	1.32	17.09	0.002	0.059
that	34	0.51	12.14	0.001	0.029	*all*	66	0.80	14.54	0.001	0.039	*those*	79	0.86	11.16	0.001	0.066
those	27	0.41	9.64	0.000	0.012	*those*	63	0.77	13.88	0.001	0.063	*that*	73	0.79	10.31	0.001	0.055

The semantics of these phrases also seem to point to interesting similarities and differences across the study corpora. For instance, '*some'*, '*all'*, and '*each'* are identified as quantifiers by Biber et al. (1999), which indicates that this particular quantifying meaning is used less frequently in ELF_Unedited than in ELF_Edited and ENL_Edited. Biber and colleagues also highlight that academic texts often include the '*that/those of* (NP)' construction to refer to nearby nouns (p. 350). Gledhill (2000a) also found that the use of '*each'* in 'DET N' sequences (e.g. '*each dose level'*) in pharmaceutical sciences research articles denotes 'deictic refocusing'; that is, referring to "the distribution and repetition of a series of clinical processes" (p. 214). Similar to the quantifying meaning, ELF_Unedited has the lowest frequency for the referential use of the construction (see Table 7.52).

While the low frequencies of both the quantifying and referential uses of 'DET *of* (NP)' in ELF_Unedited are in line with the overall infrequency of the construction in the corpus, this could also have other implications. Throughout the results, ELF_Unedited was consistently found to have limited uses of constructions for quantification, which is unlike the increasing use of *of*-bundles as quantification devices in academic discourse (Hyland & Jiang, 2019). Further, the limited use of the referential meaning of the constructions might be related to the preference of ELF users to be as explicit as possible (Mauranen, 2012; Wu et al., 2020). Thus, they might avoid using referential devices, and make use of nominal sequences repeatedly instead to enhance textual clarity.

Table 7.52 Examples for Identified Semantic Domains via Determiners in 'DET of (NP)' (min ΔP 1|2 = 0.001)[a]

Domain	ELF_Unedited	ELF_Edited	ENL_Edited
Quantity	***some*** *of the most typical comments* ***all*** *of these methods* ***each*** *of the possible responses*	***some*** *of the related concepts* ***each*** *of the individual theories* ***all*** *of the factors from the questionnaire*	results of **each** of the three surveys **some** of the key results in these simulations **all** of the different simulation that were run
Referential	*equal to* ***that*** *of the same power wood plant* *same extent as* ***those*** *of men*	*together with* ***that*** *of same-sex relations*	similar to **those** of the reference assembly **twice** that of the parent membrane

[a] '*All*', '*each'*, and '*those*' do not meet the ΔP 1|2 threshold in ELF_Unedited.

7.3 Summary

This chapter provided a comprehensive overview of the findings from the initial stage of this dissertation study, that is, the corpus-based identification of distinctive constructions in ELF_Unedited, ELF_Edited and ENL_Edited based on key function words (KFWs). Prepositions were found to be the group of function words with the largest number of ELF-ENL differences. 26 items were identified as key in 36 pairwise comparisons across the study corpora. Starting from these prepositions, a corpus-based analysis of contiguous and non-contiguous word sequences led to the identification of 19 constructions that employed 12 of the key prepositions: *of, for, to, across, after, over, within, through, in, at, by,* and *on.* The prepositions '*of*' and '*in'* led to the identification of six and three constructions respectively. Each of the remaining prepositions enabled the identification of one construction. In terms of structure, the high-frequency prepositions were instrumental in identifying the target nominal (N PREP [NP]) and passive ([NP] V-*be* V-*ed* PREP [NP]) constructions in ELF. The relatively lower-frequency prepositions, however, made it possible to uncover a variety of prepositional (PREP NP) and adjectival (ADJ PREP [NP]) patterns, as well as a determiner-based one (DET PREP [NP]) generally found to be key in ENL (see Table 7.53 for an overview).

On the one hand, statistical comparisons of normalized frequencies showed that both ELF corpora had remarkably higher frequencies than ENL_Edited for the three nominal *of*-constructions, all three prepositional passives, three of the prepositional constructions ('*after* N,' '*in* NN *with* [NP]*,'* and *'in order to* [VP]') and the adjectival construction 'ADJ-SUPERL *of* (NP).' ENL_Edited, on the other hand, was found to have much higher frequency counts than both ELF corpora for the nominal *for*- and *to*-constructions, prepositional constructions with *'across,' 'over,'* and *'through,'* adjectival 'ADJ *of* (NP),'

and the determiner-based 'DET *of* (NP)' construction. In addition, ENL had significantly higher counts for '*within* N' than only ELF_Edited, and for 'ADJ *at* (NP)' than only ELF_Unedited. As for the differences between the two ELF corpora, only '*over* N' and 'DET *of* (NP)' were found to be significantly more frequent in ELF_Edited than in ELF_Unedited.

Table 7.53 Overall Summary of Identified Constructions across Corpora

Cx forms	**Cx Meanings & Examples**	**Key in ELF_Unedited vs.:**		**Key in ELF_Edited vs.:**		**Key in ENL_Edited vs.:**	
		ENL_ Edited	**ELF_ Edited**	**ENL_ Edited**	**ELF_ Unedited**	**ELF_ Unedited**	**ELF_ Edited**
N *of* (NP)	Abstract (e.g., *type/use of* [NP]) or concrete (e.g., '*number/part of* [NP]*'*) part-whole relation between two entities	✓	-	✓	-	-	-
N *of* NP *in* (NP)	Abstract (e.g., '*type/use of* [NP] *in* [NP]') or concrete (e.g., '*type/use of* [NP] *in* [NP]') part-whole relation between two entities at a location	✓	-	✓	-	-	-
N *of* NP of (NP)	Abstract (e.g., *'term/analysis of* [NP] *of* [NP]') or concrete (e.g., '*part/number of* [NP] *of* [NP]') part-whole relation between three entities	✓	-	✓	-	-	-
N *for* (NP)	Generally abstract (e.g., '*reason/need for* [NP]) and sometimes concrete (e.g., '*value/data for* [NP]') intentional relation between two entities	-	-	-	-	✓	✓
N *to* (NP)	Mostly abstract (e.g., '*relation/approach to* [NP]') and rarely concrete (e.g., '*time/point to* [NP]') directional relation between two entities	-	-	-	-	✓	✓
across N-s	Common occurrence of an entity, phenomenon in more than one concrete entities (e.g., *'across layers/data'*) or abstract concepts (e.g., '*across types/groups*')	-	-	-	-	✓	✓
after N	Sequential relation between processes (e.g., *'after + exposure, treatment'*) or time periods (e.g., *'after + time/year'*)	✓	-	✓	-	-	-
over N	Position of one entity in relation to the other in a concrete (specifically, temporal) (e.g., '*over + time/year'*) or abstract (e.g., '*over + meaning/policy'*) sense	-	-	-	✓	✓	✓
within N	Concrete (spatial) containment (e.g., '*within + system/sector'*) and abstract contextualization (e.g., '*within + field/context'*)	-	-	-	-	-	✓
through N	Path (more specifically, means) as a concrete entity (e.g., '*through + system/microscope'*), or abstract processes (e.g., '*through + analysis/process'*)	-	-	-	-	✓	-
in order to V	Purpose statement (e.g., '*in order to obtain/assess'*)	✓	-	✓	-	-	-
in NN *with* (NP)	Discourse connections that denote agreement (e.g., '*in line with* [NP]), or comparison (e.g., '*in comparison with* [NP]')	✓	-	✓	-	-	-
ADJ *of* (NP)	Qualifying (e.g., '*aware/capable of* [NP]') or quantifying (e.g., *'many/much of* [NP]') part-whole sense	-	-	-	-	✓	✓
ADJ-SUPERL *of* (NP)	Quantity-related part-whole sense (e.g., '*most of* [NP]')	✓	-	✓	-	-	-
ADJ *at* (NP)	Attributive quality (e.g., '*significant/similar at* [NP']) and location (e.g., *present/available at* [NP]') of an entity	-	-	-	-	✓	-
(NP) V-*be* V-*ed by* (NP)	Prepositional passive used with verbs that denote: discourse construction (e.g. '[NP] V-*be characterized/explained by* [NP]'); research processes (e.g. '[NP] V-*be obtained/measured by* [NP]); or research findings (e.g. '[NP] V-*be determined by* [NP]')	✓	-	✓	-	-	-
(NP) V-*be* V-*ed in* (NP)	Prepositional passive used with verbs that denote: discourse construction (e.g. '[NP] V-*be presented/seen in* [NP]'); research processes (e.g. '[NP] V-*be used/included in* [NP]; research findings (e.g. '[NP] V-*be found/observed in* [NP]'); or location (e.g. '[NP] V-*be placed/located in* [NP]')	✓	-	✓	-	-	-

(NP) V-*be* V-*ed* *on* (NP)	Prepositional passive used with verbs that denote: discourse construction (e.g. '[NP] V-*be based/focused on* [NP]'); research processes (e.g. '[NP] V-*be performed on* [NP]'; or location (e.g. '[NP] V-*be placed on* [NP]')	✓	-	✓	-	-	-
DET *of* (NP)	Quantifying (e.g. *'some/each of* [NP]') or referential (e.g., *'that/those of* [NP]') part-whole sense	-	-	-	✓	✓	✓

The findings seem rather consistent across the two ELF corpora regardless of the different editing statuses of the texts in each. In addition, the reported overall higher frequency of conventional constructions in general, and structurally complex nominal constructions with relatively dense packaging of information by ELF writers is in line with previous research on written academic ELF (Carey, 2013; Mauranen, 2019; Yilmaz & Römer, 2020). Besides, the finding regarding the more common occurrence of passive constructions overlaps not only with research on ELF (Yilmaz & Römer, 2020), but also on L2 academic phraseology (Chen & Baker, 2010; Pan et al., 2016; Pérez-Llantada, 2014). The commonality of the passives in ELF writing, which are also the only clausal structures among the target constructions, however, differs from the finding by Wu et al. (2020) that ELF writers prefer phrasal structures over clausal ones, which is also the trend with increasing expertise (Biber et al., 2011; Biber, Gray, & Staples, 2016). Further, the ELF writers' adherence to conventionality appears to create a gap between their writing and recent trends in academic writing such as the increasing use of 'N PREP (NP)' constructions with prepositions other than *of (*e.g., *'for,' 'to'*) (Biber & Gray, 2016) and sequences denoting quantification (Hyland & Jiang, 2019), as well as decreasing rates of passive usage (Seoane, 2013).

While findings reported in this chapter yielded important differences between the three study corpora in terms of construction usage, many similarities were also reported. For instance, the high-frequency nominal and passive constructions point to a great degree of overlap in the choice of lexical fillers in these constructions. For instance, despite the higher frequency of concrete head nouns in ENL than in ELF, large proportions of the nouns denote abstract concepts in all three corpora. The primarily procedural and discursive meanings expressed through the verbs in prepositional passives is another commonality. The predictability of the distributions of lexical fillers and the dispersion of the constructions also do not result in any ELF-ENL differences.

This chapter reported the results of a whole-corpus approach to the exhaustive analysis of ELF-ENL differences. However, the complex phenomenon of ELF academic writing is shaped by a variety of factors including the similectal and disciplinary diversity in the writers' backgrounds, which also increases the heterogeneity of the data remarkably. In order to capture the characteristics of the complexity of ELF in further detail, examining similectal and disciplinary variation in a statistically robust manner to further control for the heterogeneity of the construction frequencies is essential. Hence, the next chapter delves into a detailed statistical analysis and interpretation of the cross-corpus differences in terms of the potential links between the disciplinary and similectal variation represented in the corpora, and frequency and dispersion of the 19 identified constructions.

8 A MULTIVARIATE ANALYTIC APPROACH TO THE ANALYSIS OF VARIATION IN WRITTEN ACADEMIC ELF

This chapter reports the findings from the analyses of differences in the use of the target constructions identified in the previous chapter across the two broad disciplinary domains of hard sciences (SCI) and social sciences and humanities (SSH), as well as the nine similectal (L1) backgrounds represented in ELF_Unedited.

For the comparisons of the frequencies of the target constructions, statistical tests including bootstrapped Multivariate Analysis of Variance (MANOVA), and follow-up tests for posthoc pairwise comparisons were run. Descriptive data on the distribution and dispersion of the constructions, as well as clustering plots for posthoc tests were provided to illustrate the differences across disciplinary and similectal subsets. These analyses were then complemented by examining the proportions of semantic domain frequencies and examples for distinctive instances of the constructions. Finally, connections of the study findings with relevant literature were made in terms of common disciplinary and similectal characteristics of written academic ELF.

8.1 Beyond a Corpus-Level Approach: Zooming in on Disciplines and Similects across and within Corpora

For the analyses of disciplinary and similectal variation within and across the corpora, the raw frequencies of the focal constructions were normalized per 10,000 words based on the smallest unit of comparison relevant to the analyses; that is, the 54 possible combinations of the levels of three factors: 1) corpus (N = 3), 2) discipline (N = 2), and 3) similect (N = 9). Since the ELF_Edited and ENL_Edited corpora were compiled by strictly following the similectal and disciplinary representation of each text in ELF_Unedited, each of the 18 subsets created in both ELF corpora (e.g., Sciences [SCI] texts by Chinese [CH] L1 writers) were matched with the exact number of texts corresponding to each in the ENL_Edited Corpus (see Table 8.1). Splitting data this way enabled comparisons of ELF data with parts of ENL_Edited that best match them in terms of content.

Table 8.1 Corpus Subsets Created for Analysis of Disciplinary and Similectal Variation

Discipline	Similect	Interaction	No. of texts in each subset (equal across the three corpora	No. of words in each subset across corpora		
				ELF_Unedited	ELF_Edited	ENL_Edited
Hard Sciences (SCI)	Chinese (CH)	CH_SCI	11	41,760	53,867	90,924
	Czech (CZ)	CZ_SCI	10	45,486	67,592	70,920
	Finnish (FI)	FI_SCI	15	66,049	77,511	73,639
	French (FR)	FR_SCI	8	44,708	42,935	49,832
	Italian (IT)	IT_SCI	6	27,763	28,452	32,341
	Portuguese (PO)	PO_SCI	6	17,761	21,546	36,862
	Russian (RU)	RU_SCI	7	32,862	45,642	46,768
	Spanish (SP)	SP_SCI	6	28,380	32,665	32,349
	Swedish (SW)	SW_SCI	6	23,286	33,901	36,664
Social Sciences & Humanities (SSH)	Chinese (CH)	CH_SSH	10	42,296	60,225	66,034
	Czech (CZ)	CZ_SSH	9	39,354	53,392	65,489
	Finnish (FI)	FI_SSH	9	49,283	56,609	65,599
	French (FR)	FR_SSH	8	44,053	54,096	73,348
	Italian (IT)	IT_SSH	5	26,752	35,346	26,558
	Portuguese (PO)	PO_SSH	6	35,115	37,133	41,440
	Russian (RU)	RU_SSH	6	34,981	47,573	35,283
	Spanish (SP)	SP_SSH	6	33,473	31,005	35,760
	Swedish (SW)	SW_SSH	6	33,285	40,828	39,346

Then the matched similectal and disciplinary subsets were analyzed separately instead of including all in one multifactorial and multivariate model. As also done by Staples, Egbert, Biber, and Gray (2016), this step of regrouping data was considered necessary due to the mostly uneven sizes of subsets (e.g., FR_SCI vs. IT_SSH), and lack of representation for some factor levels (e.g., no ENL_Unedited data). Hence, in order to convert the multifactorial tests into unifactorial ones for possible interaction effects, combinations of the factors were manually created, including "corpus:similect:discipline", "corpus:similect", "corpus:discipline", and "similect:discipline".

Before proceeding with the comparisons, multicollinearity of the dependent variables, that is, the usage frequencies of the identified constructions, was checked. Of the 19 patterns identified in the previous chapter, 18 were included in this phase of the analysis. 'N *of* NP *of* (NP)' was excluded due to its high multicollinearity with 'N *of* (NP)' in ELF_Unedited ($rs[140]=.75$, $p < .001$) ELF_Edited ($rs[140]=.068$, $p < .001$), and ENL_Edited ($rs[140]=.66$, $p < .001$). Then, each of the 18 target constructions was examined via one-way bootstrapped MANOVAs. As can be seen in Table 8.2, interactions of the three factors led to significant results for most of the constructions. These results indicate that the three factors of corpus, discipline, and similect are not independently related to the usage of the target constructions. Instead, there are complex interactions between the three of them.

The three-way interactions between the factors of corpus, discipline and similect ("corpus:discipline:similect") and the two-way interactions between the factors of corpus and similect ("corpus:similect"), as well as similect and discipline ("similect:discipline") were found to be significant for all constructions except for '*after* N'. As for the two-way interactions between the factors of corpus and discipline ("corpus:discipline"), analyses for most constructions resulted in significant differences across subsets, with the exception of '*after* N', 'N *for* NP', '*in* NN *with* (NP)', and *in*-passive. '*After* N', the only construction that did not vary significantly across any of the similectal and disciplinary subsets within and across the corpora, is also unique in that it only showed a significant difference across the corpora. As this result is in line with what was already reported in the seventh chapter, this construction is excluded from the following analyses. Following Levshina (2015, p. 116), I proceeded with the further analyses of the above-mentioned significant interaction effects instead of the main effects of individual factors (corpus, discipline and similect).

Table 8.2 Multivariate and Univariate ANOVA Results for Disciplinary and Similectal Differences

Factor		**Corpus**		**Discipline**		**Similect**		**Similect: Discipline**		**Corpus:Discipline**		**Corpus: Similect**		**Corpus:Discipline: Similect**	
Value		**Z**	***p***	**Z**	***p***	**Z**	***p***	**Z**	***p***	**Z**	***p***	**Z**	***p***	**Z**	***p***
MANOVA results		58.05	< .001	39.53	< .001	1809.33	< .001	5250.14	< .001	249.88	< .001	2927.43	< .001	8737. 53	< .001
Univariate ANOVA results	N *of* (NP)	0.66	.264	2.45	< .001	217.84	< .001	478.54	< .001	9.11	.002	478.54	< .001	741.16	< .001
	N *of* NP *in* (NP)	2.28	.010	1.91	.001	108.34	< .001	279.79	< .001	10.87	< .001	279.79	< .001	483.56	< .001
	N *for* (NP)	1.83	.023	0.23	.242	185.31	< .001	407.22	< .001	4.92	.070	407.22	< .001	654.26	< .001
	N *to* (NP)	2.30	.007	8.91	< .001	133.64	< .001	467.18	< .001	31.37	< .001	467.18	< .001	713.39	< .001
	Across N-s	7.28	< .001	0.73	.007	74.90	< .001	242.30	< .001	18.09	< .001	242.30	< .001	314.51	< .001
	After N	1.58	.041	0.22	.248	38.21	.209	150.29	.101	4.75	.083	150.29	.101	243.88	.338
	Over N	8.88	< .001	1.59	.001	117.86	< .001	330.67	< .001	24.77	< .001	330.67	< .001	520.77	< .001
	Within N	4.33	< .001	2.24	< .001	58.77	.007	204.41	.001	16.35	< .001	204.41	.001	422.06	< .001
	Through N	2.61	.004	8.38	< .001	64.14	.008	296.12	< .001	30.83	< .001	296.12	< .001	524.41	< .001
	In order to V	2.70	.003	2.46	< .001	112.90	< .001	354.50	< .001	13.07	< .001	354.50	< .001	563.23	< .001
	In NN *with* (NP)	1.57	.033	0.34	.148	38.34	.158	157.78	.037	4.83	.068	157.78	.037	303.23	.028
	ADJ *of* (NP)	3.48	< .001	6.20	< .001	71.25	.001	273.74	< .001	26.27	< .001	273.74	< .001	468.16	< .001
	ADJ-SUPERL *of* (NP)	2.53	.002	0.43	.077	36.86	.139	168.28	.004	9.34	< .001	168.28	.004	267.16	.026
	ADJ at (NP)	2.93	< .001	0.88	.012	68.73	< .001	183.26	< .001	9.35	< .001	183.26	< .001	354.17	< .001
	(NP) V-*be* V-ed *by* (NP)	2.04	.015	1.07	.012	123.96	< .001	334.74	< .001	8.05	.005	334.74	< .001	579.73	< .001
	(NP) V-*be* V-*ed in* (NP)	1.27	.078	0.47	.089	175.64	< .001	425.12	< .001	4.13	.129	425.12	< .001	668.28	< .001
	(NP) V-*be* V-*ed on* (NP)	3.62	.001	0.05	.583	82.38	< .001	227.73	< .001	7.81	.006	227.73	< .001	437.92	< .001
	DET *of* (NP)	6.17	< .001	0.98	.014	100.26	< .001	268.47	< .001	15.97	< .001	268.47	< .001	477.67	< .001

The following subsections discuss the findings from these pairwise comparisons for the subsets identified by the interaction of disciplinary and similectal backgrounds. As also explained in the methodology chapter, medians for normalized construction frequencies are reported throughout the chapter to account for the non-normality of the data as well as differences in the sizes of subsets. Unlike

the sum of all normalized frequencies in Chapter 7, medians represent the frequency for the middle value. Thus, among a set of values such as '4, 7, 9, 12, 13', 9 is the median. Among a set of values such as '0, 0, 0, 4, 9', the median is 0. Similarly, in the results tables, a median of 0 for the occurrence of a construction in a subset means that the construction occurs in only a small number of the texts in the subset. In such cases, the median value of 0 is generally accompanied by a high DP_{norm} value (close to 1), which also shows that the construction is under-dispersed in the subset. In the cases of constructions that do not occur in a subset at all, no values are given in the tables. The only instance of this non-occurrence is for '*across* N-s' for the "similect discipline" subsets, some of which (e.g., Chinese ELF in SCI) have "NA"s for both the median and DP_{norm} values.

8.1.1 Differences in Nominal (N PREP [NP]) Constructions

Despite the significant MANOVA results for the three-way "corpus:discipline:similect" interaction, the posthoc test showed that no matching subsets were found to differ significantly. However, several important differences were found across subsets based on the two-way interactions, including "corpus:similect," "corpus:discipline" and "similect:discipline."

Results for the "corpus:similect" interaction indicate that there is a certain degree of variation among subsets (see Table 8.3). Although 'N *of* (NP)' seems to have a somewhat even dispersion as the low DP_{norm} values for all subsets suggest, frequencies fluctuate across groups with the Russian ELF subset having the highest frequency, and Finnish having the lowest. To illustrate, 97 of the 501 instances of '*number of* (NP)' in ELF_Unedited come from the Russian ELF subset, which also has nouns such as '*probability'*, '*set'*, and '*case'* with much higher frequencies than the whole ELF_Unedited corpus. For the other three nominal constructions, both dispersion and frequency of items vary greatly. In 'N *of* NP *in* (NP)', with the exception of the edited Portuguese ELF subset, the variation across groups is not substantial. The Portuguese ELF subset, however, is distinctive not only because of the frequency difference but also because it has a higher type-token ratio (0.46) than the whole ELF_Edited corpus (0.16), which indicates a higher degree of lexical diversity. In 'N *for* (NP)', several ENL subsets have

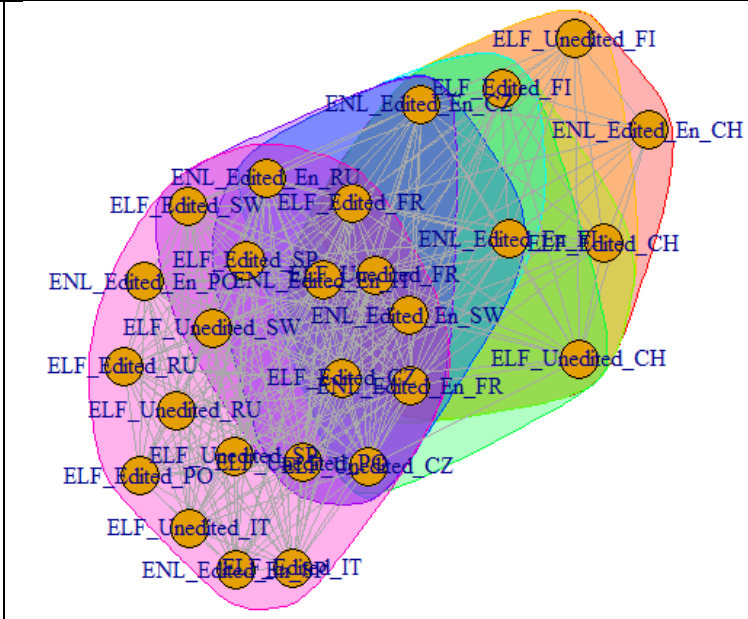
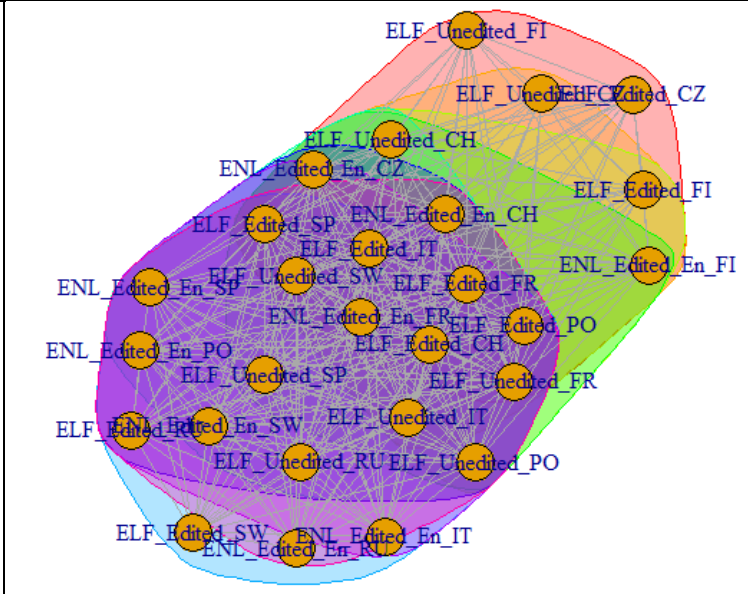
higher frequencies than ELF subsets, but the edited Swedish ELF texts have the highest frequency. As for 'N *to* (NP)', all ENL subsets except for those matching Swedish and Russian ELF texts, have higher frequencies than the ELF subsets.

Table 8.3 Bootstrapped Median and DP_{norm} Values for 'N PREP (NP)' Cxs across "Corpus:Similect" Subsets

Similect	Corpus	N *of* (NP)		N *of* NP *in* (NP)		N *for* (NP)		N *to* (NP)	
		Median (normed per 10k)	DP_{norm}	Median (normed per 10k)	DP_{norm}	Median (normed per 10k)	DP_{norm}	Median (normed per 10k)	DP_{norm}
CH	**Unedited**	28.02	0.07	1.19	0.24	4.79	0.19	2.36	0.25
	Edited	31.74	0.05	1.33	0.25	6.31	0.16	1.49	0.31
EN	**Edited**	28.38	0.07	1.21	0.25	5.61	0.09	3.97	0.14
CZ	**Unedited**	42.65	0.09	2.28	0.25	3.81	0.22	2.79	0.16
	Edited	44.01	0.07	2.37	0.21	4.14	0.21	2.44	0.16
EN	**Edited**	38.17	0.07	1.37	0.22	6.26	0.11	2.60	0.16
FI	**Unedited**	25.23	0.08	1.11	0.25	4.00	0.20	1.64	0.23
	Edited	22.21	0.11	1.27	0.27	4.15	0.16	1.73	0.18
EN	**Edited**	23.90	0.08	1.29	0.18	5.31	0.16	3.13	0.20
FR	**Unedited**	43.84	0.07	1.35	0.23	7.38	0.24	2.70	0.22
	Edited	41.90	0.08	2.07	0.18	6.43	0.20	2.56	0.18
EN	**Edited**	41.04	0.09	1.50	0.19	7.42	0.22	3.27	0.20
IT	**Unedited**	61.30	0.07	2.51	0.12	8.60	0.15	3.95	0.22
	Edited	65.37	0.07	2.81	0.21	9.14	0.21	5.66	0.23
EN	**Edited**	52.71	0.10	2.48	0.25	10.92	0.20	7.74	0.27
PO	**Unedited**	52.97	0.09	2.41	0.26	10.20	0.19	3.82	0.19
	Edited	65.51	0.07	3.38	0.16	7.58	0.20	4.28	0.21
EN	**Edited**	59.45	0.07	2.17	0.18	9.65	0.13	7.00	0.13
RU	**Unedited**	75.18	0.07	2.43	0.23	10.29	0.22	5.42	0.19
	Edited	72.52	0.08	2.73	0.16	10.30	0.13	5.67	0.15
EN	**Edited**	49.82	0.08	1.70	0.24	11.62	0.08	4.82	0.21
SP	**Unedited**	60.15	0.08	2.55	0.25	9.56	0.18	4.74	0.23
	Edited	49.00	0.10	2.36	0.20	7.23	0.22	4.68	0.16
EN	**Edited**	64.04	0.09	2.10	0.22	10.35	0.20	4.74	0.16
SW	**Unedited**	58.95	0.11	1.71	0.31	8.57	0.23	4.20	0.23
	Edited	59.18	0.09	2.82	0.29	12.44	0.11	5.08	0.15
EN	**Edited**	51.47	0.09	1.64	0.32	10.64	0.17	4.64	0.19

On the one hand, a number of "corpus:similect" subsets were found to differ significantly in the use of 'N *of* (NP)' and 'N *for* NP'. One the other hand, these differences were not significant between corresponding subsets (e.g., unedited Chinese ELF vs. edited Chinese ELF vs. corresponding edited ENL texts). Instead, certain subsets within corpora were found to be significantly different from one another. An interesting finding for 'N *of* (NP)' relevant to this chapter, however, is that many ELF subsets, including texts by writers with Indo-European L1s, as well as the matching ENL texts seem to cluster together. In the case of 'N *for* (NP)' there are fewer significant differences; however, Finnish and Czech subsets appear to stand out from the other subsets. Besides, results for both constructions point to variation within the ENL subsets in ways similar to the matching ELF subsets (see Table 8.4).

Table 8.4 Clusters and P-Values of Pairwise Differences for 'N PREP (NP)' across "Corpus:Similect" Subsets

Cx	Clusters	Pairwise Differences
***N of* (NP)**		**- Within ELF_Unedited:** FI - CZ (p =.027) RU - FI: (p =.007) IT - CH (p =.008) SP - CH (p =.008) IT - FI (p =.007) SP - FI (p =.007) PO - CH (p =.022) SW - CH (p =.044) PO - FI (p =.023) SW - FI (p =.026) RU - CH (p =.008) **- Within ELF_Edited:** IT - CH (p =.007) RU - FI (p = .009) IT - FI (p =.015) SP - CH (p =.009) PO CH (p =.007) SP - FI (p =.034) PO - FI (p =.010) SW - CH (p =.013) RU - CH (p = .009) SW – FI (p =.023) **- Within ENL_Edited:** IT - CH (p = .007) SP - CH (p = .007) PO - CH (p = .009) SP - CZ (p = .013) PO -FI (p = .007) SP - FI (p = .007) RU - CH (p = .007)
***N for* (NP)**		**- Within ELF_Unedited:** IT - FI (p = .039) RU - FI (p = .017) PO - CZ (p = .043) SP - CZ (p = .037) PO - FI (.038) SP - FI (p = .027) RU - CZ (p = .023) **- Within ELF_Edited:** RU - CZ (p = .017) SW - CZ (p = .017) RU - FI (p = .017) SW - FI (p = .017) **- Within ENL_Edited:** IT - CH (p = .048) RU- CZ (p = .028 IT - FI (p = .017) RU - FI (p = .017) PO - FI (p = .017) SP - FI (p = .024) PO - FI (p = .017) SW - FI (p = .018)

As shown in Table 8.5, the proportions of semantic domain frequencies for the 'N PREP (NP)' constructions point to the dominance of abstract over concrete nouns for all four constructions, and across all the "corpus:similect" subsets. However, the two *of*-constructions seem to have remarkably higher proportions of abstract than concrete nouns in ELF, with slightly higher proportions in the unedited

subsets, as opposed to ENL subsets. This observation is in line with the cross-corpus differences reported in the previous chapter in that abstract nouns are more common in the ELF corpora, while concrete nouns have higher frequencies in ENL texts. As for 'N *for* (NP)' and 'N *to* (NP)' this pattern persists even though these are key constructions in ENL_Edited. However, there is a certain degree of variation in the proportions of the noun domains found for these two constructions. While generally ELF_Edited has the highest proportions of abstract nouns, several ELF_Unedited subsets such as Swedish, Czech, and Russian ELF texts that match them have even lower proportions than observed in the matching ENL subsets.

Table 8.5 Percentages of Collocates with Abstract and Concrete Meanings in 'N PREP (NP)' across "Corpus:Similecs" Subsets

Cx		**N *of* (NP)**		**N *of* NP *in* (NP)**		**N *for* (NP)**		**N *to* (NP)**	
Domain		**Abstract**	**Concrete**	**Abstract**	**Concrete**	**Abstract**	**Concrete**	**Abstract**	**Concrete**
Examples		*type, use, analysis*	*number, part, value*	*use, role, effect*	*number, amount, level*	*reason, need, support*	*value, data, test*	*relation, approach, respect*	*time, point*
CH	**Unedited**	70.36	29.64	72.73	27.27	80.56	19.44	93.50	6.50
	Edited	67.09	32.91	70.00	30.00	89.84	10.16	98.62	1.38
EN	**Edited**	61.05	38.95	62.93	37.07	77.45	22.55	92.96	7.04
CZ	**Unedited**	60.01	39.99	62.18	37.82	68.75	31.25	84.00	16.00
	Edited	57.36	42.64	60.30	39.70	74.18	25.82	87.43	12.57
EN	**Edited**	58.95	41.05	59.31	40.69	71.92	28.08	90.27	9.73
FI	**Unedited**	63.78	36.22	67.11	32.89	78.51	21.49	75.00	25.00
	Edited	61.67	38.33	60.10	39.90	77.74	22.26	98.56	1.44
EN	**Edited**	59.05	40.95	62.09	37.91	73.87	26.13	95.22	4.78
FR	**Unedited**	72.24	27.76	78.38	21.62	74.47	25.53	97.87	2.13
	Edited	69.76	30.24	65.98	34.02	84.03	15.97	98.72	1.28
EN	**Edited**	62.42	37.58	64.60	35.40	69.78	30.22	92.04	7.96
IT	**Unedited**	72.65	27.35	81.13	18.87	86.87	13.13	95.93	4.07
	Edited	65.89	34.11	70.13	29.87	79.19	20.81	98.77	1.23
EN	**Edited**	59.66	40.34	68.92	31.08	77.86	22.14	88.51	11.49
PO	**Unedited**	72.07	27.93	86.96	13.04	82.95	17.05	94.94	5.06
	Edited	65.10	34.90	70.42	29.58	89.81	10.19	93.75	6.25
EN	**Edited**	56.42	43.58	70.79	29.21	76.24	23.76	88.27	11.73
RU	**Unedited**	66.95	33.05	66.33	33.67	70.33	29.67	83.84	16.16
	Edited	66.08	33.92	68.38	31.62	77.63	22.37	97.33	2.67
EN	**Edited**	57.14	42.86	55.13	44.87	72.56	27.44	95.76	4.24
SP	**Unedited**	71.47	28.53	63.10	36.90	82.18	17.82	91.86	8.14
	Edited	68.01	31.99	54.63	45.37	80.36	19.64	97.27	2.73
EN	**Edited**	63.48	36.52	63.51	36.49	82.84	17.16	92.86	7.14
SW	**Unedited**	62.94	37.06	75.00	25.00	65.81	34.19	81.00	19.00
	Edited	60.21	39.79	69.40	30.60	82.35	17.65	91.06	8.94
EN	**Edited**	57.26	42.74	51.67	48.33	70.51	29.49	87.91	12.09

Note. Sum of percentages for abstract and concrete meanings for each subset equal all Cx collocates included in semantic analysis.

As for the two-way "corpus:discipline" interaction, the overall prevalence of nominal constructions in SSH texts is evident from higher normed median frequencies and lower DP_{norm} values for all subgroups. Having twice as many occurrences of 'N *to* (NP)', SSH texts confirm this observation. Having the highest median frequencies, SSH subsets in ENL_Edited and ELF_Edited also include

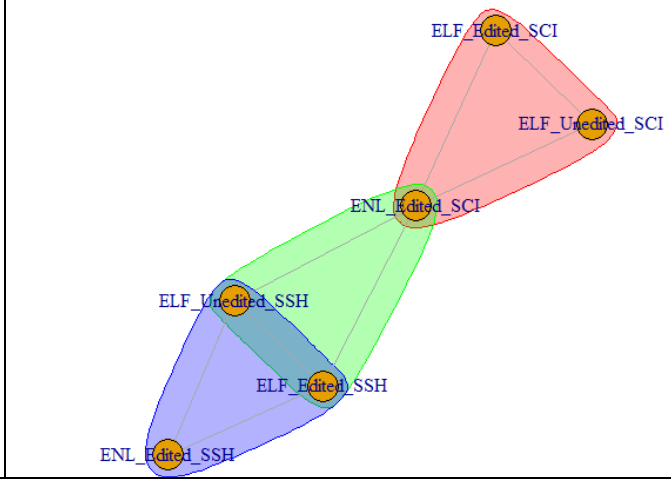
distinctive head noun choices, the most common one being '*approach.'* 126 of 157 instances in ENL_Edited and 62 out of the 70 instances of '*approach to (NP)*' come from the SSH texts *(e.g., 'an accommodationist approach to civil rights', education).* The only exception is the slightly lower frequency of 'N *for* (NP)' in SSH than in SCI within ELF_Unedited. In addition, the largest intradisciplinary variation seems to be between the SSH subsets of ELF_Unedited and ENL_Edited (see Table 8.6).

Table 8.6 Bootstrapped Median and DP_{norm} Values for 'N PREP (NP)' Cxs across "Corpus:Discipline" Subsets

Discipline	Corpus	N *of* (NP)		N *of* NP *in* (NP)		N *for* (NP)		N *to* (NP)	
		Median (normed per 10k)	DP_{norm}	Median (normed per 10k)	DP_{norm}	Median (normed per 10k)	DP_{norm}	Median (normed per 10k)	DP_{norm}
SCI	**ELF_Unedited**	40.54	0.08	1.68	0.27	6.94	0.24	2.41	0.27
	ELF_Edited	40.07	0.10	2.10	0.27	5.62	0.19	2.22	0.21
	ENL_Edited	38.78	0.08	1.32	0.24	7.11	0.15	2.96	0.19
SSH	**ELF_Unedited**	53.36	0.09	2.08	0.21	6.31	0.16	4.20	0.14
	ELF_Edited	48.06	0.08	2.31	0.17	7.31	0.15	4.77	0.17
	ENL_Edited	44.37	0.08	1.93	0.21	8.13	0.14	4.83	0.16

The significant differences found across subsets for the use of 'N *to* (NP)' also highlight a disciplinary divide across the SSH subsets with have higher frequencies than in the SCI subsets of the three study corpora individually. Although the maximal difference between the groups seems to be between both SCI subsets by ELF writers and the SSH subset in ENL_Edited, this difference is not statistically significant. (see Table 8.7).

Table 8.7 Clusters and P-Values of Pairwise Differences for 'N PREP (NP)' Cxs across "Corpus:Discipline" subsets

Cx	Clusters	Pairwise Differences
N *to* (NP)		**- Cross-disciplinary :** ELF_Edited: SSH - ELF_Edited: SCI (p <. 001) ELF_Unedited: SSH - ELF_Unedited: SCI (p <. 001) ENL_Edited: SSH - ENL: Edited: SCI (p = .001)

Semantic differences in terms of the proportion of abstract and concrete nouns across "corpus:discipline" subsets highlight the dominance of concrete nouns in these target nominal constructions, with an even more substantial discrepancy between the two domains observed in SSH subsets. As in the case of the "corpus:similect" interaction, the highest proportions of abstract nouns in the *of*-constructions are observed among subsets from ELF_Unedited, which are replaced by those from ELF_Edited in the case of 'N *for* (NP)' and 'N *to* (NP)'. the *of*-constructions, all SSH subsets have higher proportions of abstract nouns. Further, ENL_Edited has the second-highest ratio of abstract noun use for these constructions with values higher than those for ELF_Unedited subsets (see Table 8.8).

Table 8.8 Percentages of Semantic Domains in 'N PREP (NP)' across "Corpus:Discipline" Subsets

Cx		**N *of* (NP)**		**N *of* NP *in* (NP)**		**N *for* (NP)**		**N *to* (NP)**	
Domain		**Abstract**	**Concrete**	**Abstract**	**Concrete**	**Abstract**	**Concrete**	**Abstract**	**Concrete**
Examples		*type, use, analysis*	*number, part, value*	*use, role, effect*	*number, amount, level*	*reason, need, support*	*value, data, test*	*relation, approach, respect*	*time, point*
SCI	**ELF_Unedited**	58.53	41.47	64.94	35.06	68.18	31.82	82.40	17.60
	ELF_Edited	56.44	43.56	57.94	42.06	74.49	25.51	92.06	7.94
	ENL_Edited	52.51	47.49	55.82	44.18	66.78	33.22	89.67	10.33
SSH	**ELF_Unedited**	75.65	24.35	77.20	22.80	86.13	13.87	90.56	9.44
	ELF_Edited	71.10	28.90	71.58	28.42	87.72	12.28	97.49	2.51
	ENL_Edited	67.36	32.64	68.49	31.51	82.23	17.77	93.78	6.22

Note. Sum of percentages for abstract and concrete meanings for each subset equal all Cx collocates included in semantic analysis.

Lastly, taking the "corpus" factor out of the equation, the "similect:discipline" subgroups point to a different level of complexity due to the variation across the subsets. Besides, since comparisons between the ELF corpora do not result in significant differences except for the case of 'N *of* NP *in* (NP)', the examination of the findings of this interaction is of major importance. One crucial difference across "similect:discipline" groups is that the frequencies for nominal constructions are less uniformly distributed across the SCI subsets by ELF writers in particular, in comparison to SSH subsets that also have higher frequencies. Another difference is that the clear SSH-dominance identified across "corpus:discipline" subsets is not present among some subset differences. As the median values provided in Table 8.10 indicate, the median frequencies for the constructions are actually lower among the SSH than SCI subsets of many ELF and several ENL subsets. These are the French, Spanish, and Swedish ELF, and the ENL groups matching Portuguese, Spanish, and Swedish ELF for 'N *of* (NP)'; Czech ELF and the matching ENL subset, and Spanish ELF for N *of* NP *in* (NP), French, Portuguese, Russian,

Spanish and Swedish ELF for 'N *for* (NP)'; and lastly Spanish ELF for 'N *to* (NP)'. Hence, it is plausible to suggest that a higher degree of variation exists among ELF than ENL subsets, especially considering the distribution of the *of*-constructions found to be key in the ELF data. The variation across subsets is especially prominent in the case of 'N *for* (NP)' where several SCI subsets such as the Russian ELF subset and the two ENL subsets matching Italian and Portuguese ELF have much higher frequencies than most others (see Table 8.9).

Table 8.9 Bootstrapped Median and DP_{norm} Values for 'N PREP (NP)' Cxs across "Similect:Discipline" Subsets

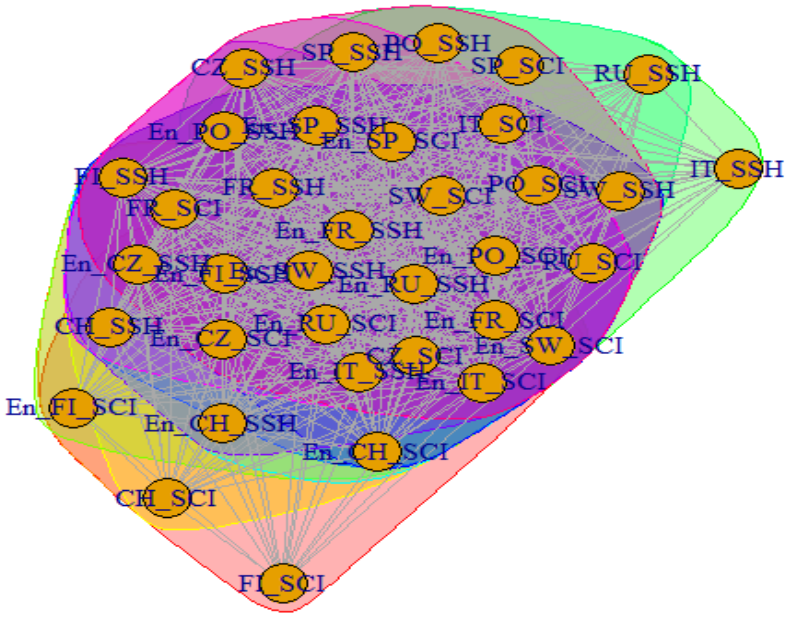
Discipline	Similect	N *of* (NP)		N *of* NP *in* (NP)		N *for* (NP)		N *to* (NP)	
		Median (normed per 10k)	DP_{norm}	Median (normed per 10k)	DP_{norm}	Median (normed per 10k)	DP_{norm}	Median (normed per 10k)	DP_{norm}
SCI	CH	28.21	0.05	1.15	0.30	4.79	0.21	1.30	0.39
	CZ	38.96	0.06	2.59	0.25	3.35	0.24	2.32	0.13
	FI	17.86	0.12	1.10	0.28	3.00	0.21	1.16	0.27
	FR	43.84	0.04	1.23	0.21	7.30	0.24	2.51	0.20
	IT	47.73	0.04	2.81	0.20	8.35	0.16	2.69	0.22
	PO	60.57	0.07	1.97	0.26	9.43	0.19	3.82	0.26
	RU	64.41	0.08	2.13	0.22	15.34	0.13	3.18	0.23
	SP	60.08	0.08	3.16	0.14	9.87	0.12	4.71	0.23
	SW	71.52	0.12	2.04	0.39	11.95	0.18	3.63	0.15
	EN_CH	22.99	0.05	0.88	0.28	4.07	0.10	2.75	0.15
	EN_CZ	34.55	0.07	1.41	0.22	6.35	0.10	1.84	0.19
	EN_FI	19.55	0.11	1.09	0.17	3.26	0.17	1.22	0.20
	EN_FR	36.42	0.12	1.01	0.22	7.53	0.15	2.31	0.27
	EN_IT	50.56	0.06	1.71	0.34	15.92	0.16	2.64	0.17
	EN_PO	64.71	0.05	2.04	0.10	15.06	0.08	6.24	0.17
	EN_RU	45.12	0.08	1.07	0.31	9.62	0.09	4.28	0.16
	EN_SP	70.02	0.09	1.71	0.29	9.43	0.27	4.50	0.19
	EN_SW	56.60	0.08	1.64	0.26	10.64	0.13	4.64	0.12
SSH	CH	32.93	0.07	1.83	0.20	6.81	0.13	2.55	0.17
	CZ	50.92	0.11	2.05	0.17	4.85	0.16	3.00	0.17
	FI	39.52	0.06	1.70	0.17	4.81	0.13	4.05	0.12
	FR	37.97	0.09	2.04	0.18	6.12	0.17	3.07	0.13
	IT	83.75	0.06	2.77	0.13	9.35	0.23	7.64	0.18
	PO	61.93	0.09	3.37	0.18	6.91	0.15	4.27	0.19
	RU	76.69	0.07	2.80	0.17	10.19	0.14	6.71	0.10
	SP	53.77	0.08	2.17	0.27	7.42	0.22	4.68	0.17
	SW	52.43	0.09	2.20	0.24	8.10	0.15	5.14	0.15
	EN_CH	32.41	0.09	1.36	0.22	6.44	0.10	4.54	0.12
	EN_CZ	38.17	0.07	1.37	0.25	6.26	0.12	3.36	0.12
	EN_FI	41.92	0.06	1.98	0.18	7.16	0.15	4.42	0.22
	EN_FR	43.77	0.07	1.64	0.20	6.00	0.27	3.48	0.16
	EN_IT	52.71	0.15	2.64	0.18	10.92	0.12	10.17	0.05
	EN_PO	55.51	0.09	2.29	0.26	9.17	0.13	7.00	0.11
	EN_RU	59.81	0.10	2.27	0.16	12.05	0.07	6.52	0.18
	EN_SP	62.08	0.09	2.52	0.18	10.49	0.14	5.32	0.16
	EN_SW	46.64	0.10	1.65	0.34	11.44	0.17	5.85	0.21

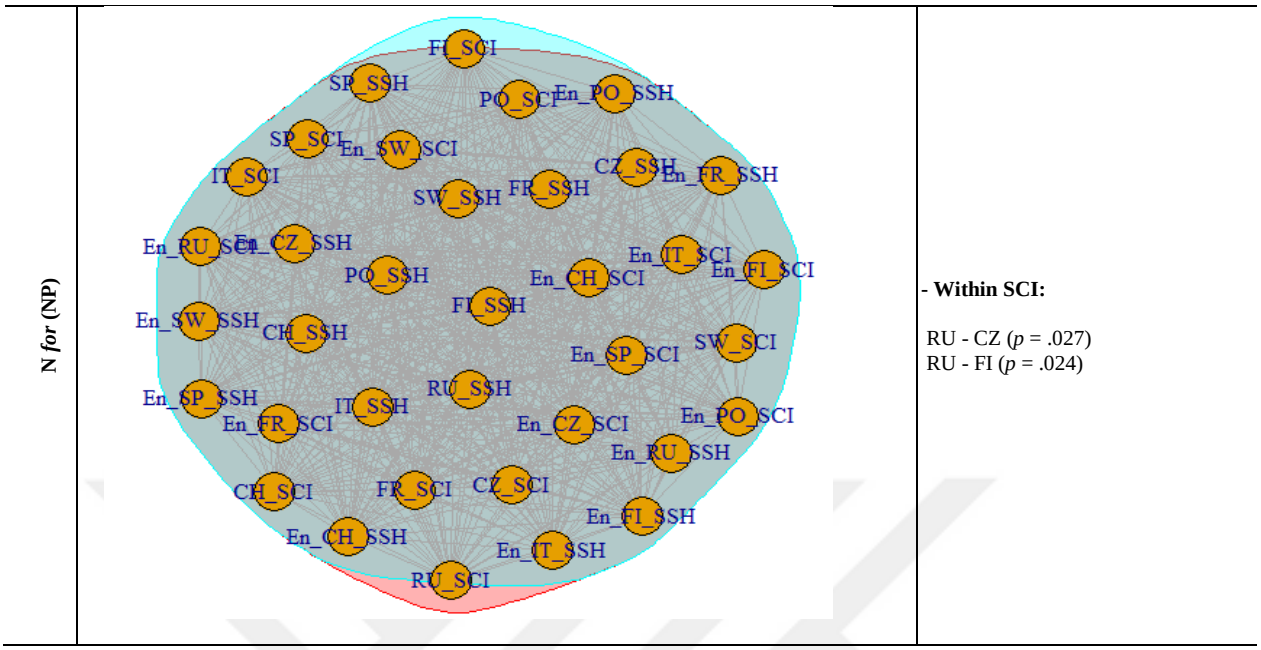
A concordance search shows that economics texts by Russian ELF writers use the construction with head nouns such as '*prediction'* and '*technique'* that generally less commonly occur in both ELF corpora (e.g., '*predictions for every subsample'* [ELF_Unedited_62], '*data mining techniques for movie forecasting'* [ELF_Edited_60]). The collocates from the ENL subset matching Italian ELF are dominated by the 72

instances of a term occurring in only one economics text (‘*taste for status,’* [ENL_Edited_49]). Lastly, the ENL subset matching Portuguese ELF generally aligns with the noun choices in the ENL_Edited corpus, with a slightly more frequent use of seemingly field-specific nouns such as ‘*intensity’*, and ‘*estimate’* (e.g., ‘*beam intensity for each arc’* [ENL_Edited_54, materials science], *‘estimates for large sample fraction’* [ENL_Edited_55, statistics]*).*

The clustering plots based on pairwise comparisons yield significant differences across “similect:discipline” subsets for ‘N *of* (NP)’ and ‘N *for* (NP). The plot for the ‘N *of* (NP)’ construction in particular highlights the variation within disciplines across several ELF subsets. The only significant cross-disciplinary comparison is present between the Finnish ELF writers’ SCI and SSH texts, where the latter have a much higher construction frequency. In addition, due to its lowest frequency among the subsets, the SCI subset by Finnish ELF writers is placed in a pink cluster separately from the rest (see Table 8.10). N *for* (NP), however, appears to be influenced by much less similectal and disciplinary variation, with the SCI subsets by Finnish and Russian ELF writers being the most distinct subsets.

Table 8.10 Clusters and P-Values of Pairwise Differences for ‘N PREP (NP)’ Cxs across “Similect:Discipline” Subsets

Cx	Clusters	Pairwise Differences
N *of* (NP)		- **Within SCI:** CH - IT (p = .006) CH - PO (p = .038) CH - RU (p = .034) CH - SP (p = .006) CZ - FI (p = .006) FI - FR (p = .017) FI - IT (p = .007) FI- PO (p = .011) FI - RU (p = .005) FI - SP (p = .005) FI - SW (p = .006) - **Within SSH:** CH - IT (p = .007) CH - PO (p = .016) CH - RU (p = .006) CH - SW (p = .049) CZ - IT (p = .049) FI - IT (p = .007) FI - RU (p = .023) - **Cross-disciplinary:** FI_SCI- FI_SSH (p = .011)



The proportions of semantic domain frequencies across the subsets for the "similect:discipline" interaction point to the only cases where higher proportions of concrete than abstract nouns are observed in the use of nominal constructions. All these exceptions are found in the use of *of*-constructions in SCI subsets, including: (1) Swedish ELF and ENL subsets matching Italian and Portuguese ELF in the use of 'N *of* (NP)', and (2) ENL matching Russian Spanish Swedish ELF subsets in the use of 'N *of* NP *in* (NP)'. As can be seen, in line with the generally higher occurrence of concrete nouns in ENL than in ELF, all these exceptions but one, are among the ENL subsets. In contrast, higher proportions of abstract nouns are observed for all subsets in the use of 'N *for* (NP)' and 'N *to* (NP)' It is also interesting to observe that several ELF subsets in both disciplinary groups have higher proportions of abstract nouns (e.g. Spanish and Portuguese ELF) than the rest of the subsets (see Table 8.11).

Table 8.11 Percentages of Semantic Domains for 'N PREP (NP)' Cxs across "Similect:Discipline" Subsets

Cx		**N *of* (NP)**		**N *of* NP *in* (NP)**		**N *for* (NP)**		**N *to* (NP)**	
Domain		**Abstract**	**Concrete**	**Abstract**	**Concrete**	**Abstract**	**Concrete**	**Abstract**	**Concrete**
Examples		*type, use, analysis*	*number, part, value*	*use, role, effect*	*number, amount, level*	*reason, need, support*	*value, data, test*	*relation, approach, respect*	*time, point*
SCI	**CH**	60.86	39.14	63.00	37.00	80.79	19.21	93.64	6.36
	CZ	52.14	47.86	55.30	44.70	68.33	31.67	79.14	20.86
	FI	54.68	45.32	58.04	41.96	72.47	27.53	80.53	19.47
	FR	60.37	39.63	56.25	43.75	69.05	30.95	100.00	0.00
	IT	61.06	38.94	70.83	29.17	77.08	22.92	90.38	9.62
	PO	57.16	42.84	67.50	32.50	75.53	24.47	94.55	5.45
	RU	61.46	38.54	66.99	33.01	65.20	34.80	91.45	8.55
	SP	67.25	32.75	54.29	45.71	80.00	20.00	91.14	8.86

	SW	45.92	54.08	73.47	26.53	62.43	37.57	88.89	11.11
	EN_CH	54.92	45.08	54.24	45.76	68.29	31.71	88.89	11.11
	EN_CZ	52.74	47.26	53.42	46.58	56.97	43.03	89.33	10.67
	EN_FI	52.70	47.30	58.33	41.67	67.05	32.95	95.80	4.20
	EN_FR	53.64	46.36	62.86	37.14	57.58	42.42	90.14	9.86
	EN_IT	42.70	57.30	72.41	27.59	76.53	23.47	90.00	10.00
	EN_PO	48.83	51.17	68.75	31.25	64.65	35.35	68.75	31.25
	EN_RU	53.73	46.27	46.30	53.70	67.83	32.17	97.18	2.82
	EN_SP	58.09	41.91	48.72	51.28	83.33	16.67	96.72	3.28
	EN_SW	51.74	48.26	42.31	57.69	57.53	42.47	80.77	19.23
SSH	**CH**	74.50	25.50	77.52	22.48	89.96	10.04	98.10	1.90
	CZ	65.35	34.65	70.29	29.71	74.71	25.29	91.53	8.47
	FI	73.77	26.23	75.49	24.51	88.27	11.73	93.69	6.31
	FR	78.70	21.30	80.37	19.63	88.06	11.94	97.66	2.34
	IT	75.26	24.74	79.31	20.69	87.50	12.50	99.34	0.66
	PO	74.15	25.85	83.00	17.00	93.01	6.99	94.17	5.83
	RU	72.91	27.09	67.86	32.14	89.68	10.32	92.42	7.58
	SP	72.21	27.79	63.22	36.78	83.13	16.87	97.44	2.56
	SW	74.18	25.82	67.50	32.50	87.69	12.31	85.43	14.57
	EN_CH	68.91	31.09	71.93	28.07	86.13	13.87	97.52	2.48
	EN_CZ	66.14	33.86	65.28	34.72	85.33	14.67	90.91	9.09
	EN_FI	66.97	33.03	67.57	32.43	81.53	18.47	94.87	5.13
	EN_FR	66.75	33.25	65.38	34.62	76.54	23.46	93.08	6.92
	EN_IT	81.78	18.22	66.67	33.33	80.95	19.05	87.96	12.04
	EN_PO	63.70	36.30	71.93	28.07	90.24	9.76	99.13	0.87
	EN_RU	64.95	35.05	75.00	25.00	83.67	16.33	94.68	5.32
	EN_SP	68.88	31.12	80.00	20.00	82.35	17.65	88.24	11.76
	EN_SW	63.12	36.88	58.82	41.18	77.08	22.92	90.77	9.23

Note. Sum of percentages for semantic domains for each subset equal all Cx collocates included in semantic analysis.

In conclusion, the results for the analysis of the relationship between the disciplinary and similectal background of the writers of ELF and ENL texts and the use of the target nominal constructions point to complexities that extend beyond the cross-corpus differences identified in the previous chapter. The findings suggest that the usage of 'N *of* (NP)' and, to a lesser extent, N *for* '(NP)' is at least partially related to the similectal background of the writers when combined with either editing status or disciplinary backgrounds. 'N *to* (NP)', however, seems to vary remarkably across disciplinary groups in the study corpora. Despite yielding a distinctive pattern of clustering across "corpus:discipline" subsets, 'N *of* NP *in* (NP), an analysis of which did not result in any significant findings in pairwise comparisons, does not seem to be noticeably related to the background of the writers.

The findings regarding the relationship between similectal background and the use of 'N *of* (NP)' are in line with previous research. For instance, Pan et al. (2016) found that research writing by L1 Chinese scholars had fewer instances of the construction than ENL writers did. Palumbo's broader (2015) comparison of university webpages in European Union (EU) countries with those in British and American contexts, pointed to higher frequencies of the construction in EU pages. Palumbo (2017) reported slightly higher frequencies of the construction in L1 Italian and Spanish scholars' academic writing in English.

Besides, a more general finding on the greater proportion of abstract nouns among ELF writers with Slavic and Romance L1s, as well as Swedish provides further evidence for previous research findings which indicate that L1 and L2 English scientific writers with these L1 backgrounds prefer to construct their discourse around abstractions, as well as conceptual and theoretical arguments (Bennett & Muresan, 2016; Pérez-Llantada, 2012; Tåqvist, 2018; Vassileva, 2000; Vassilieva, 1998). This preference, along with the finding reported in the previous chapter that ELF texts in general have higher frequencies for abstract nouns, could be interpreted from two contradicting perspectives. One is the vague and inexplicit nature of many of these abstract nouns (Cutting, 2012), which is considered a general characteristic of L2 writing (Crossley, Salsbury, & McNamara, 2010). The other is that, while quantification (e.g., use of bundles such as '*number/size/some/much of* [NP]') is also increasing, the documented diversification of abstract meanings of commonly used nominal patterns in academic discourse over the years is remarkable (Biber & Gray, 2011, 2016; Hyland & Jiang, 2019). Therefore, a diachronic approach to the analysis of written academic ELF could yield important findings on how ELF is related to this ongoing change in academic writing.

In the case of 'N *to* (NP)', the SSH-dominance observed across all three corpora indicates that the use of this construction is in line with disciplinary conventions even though there are differences between the datasets. The examples given above from the corpora (e.g. '*approach to* [NP]') also suggest that, similar to the 'noun + *to*-infinitive' patterns commonly studied in academic writing research, this construction is also used commonly to express authorial stance in relation to cognitive processes (Biber, 2006a). As found by Gray (2015), such use of the pattern with a *to*-infinitive is common in social sciences (Gray, 2015, p. 151).

Finally, a finding of crucial importance is that no subset was found to differ significantly from the other two matching subsets in the data, indicating the lack of remarkable links between the factors of corpus, similect and discipline and the usage of target constructions when texts are carefully matched. This might be due to two reasons. One is the use of prepositional phrases for nominal post-modification in academic writing, which is commonly known and encountered as an established feature of the register

(Biber & Gray, 2011). Another is the increased uniformity in the use of nominal structures between L1 and L2 writers of English when the latter group has high language proficiency and expertise (Biber et al., 2011; Parkinson & Musgrave, 2014).

8.1.2 *Differences in 'PREP N' Constructions*

With the exception of '*after* N' in all interactions and '*in order to* V' in "corpus:discipline", 'PREP N' constructions led to significant results in univariate comparisons. While no significant results were found in follow-up pairwise tests for the three-way "corpus:discipline:similect" interaction, all 'PREP N' constructions except for '*after* N' showed significant differences across the "corpus:similect", "corpus:discipline", and "similect:discipline" subgroups. Hence, the results of these two-way interactions are the focus of this section.

'PREP N' constructions have considerably lower frequencies than 'N PREP (NP) constructions. Hence, despite meeting the range threshold of 10 percent for each corpus, their usage frequencies point to relatively varied patterns of dispersion across the subgroups. As seen in Table 8.12 for the "corpus:similect" interaction, many subsets have medians of zero, and DP_{norm} values close to one, indicating rather uneven dispersions. '*Across* N-s', for instance, has a median value over zero for only four ENL subsets. For the rest of the 'PREP N' constructions (*'over N,' 'within N,' 'through N,' 'in NN with [NP],' 'in order to [VP]'*), a larger number of subsets have median values over zero, and lower DP_{norm} values than '*across* N-s'. However, the frequency differences seem larger across subsets for these constructions. Nevertheless, the median frequencies are mostly consistent with the findings reported in the previous section that '*across* N-s', *'over* N', '*within* N', and '*through* N' are key in ENL, while '*in order to* V' and '*in NN with* (NP)' are key in the ELF corpora.

Table 8.12 Bootstrapped Median and DP_{norm} Values for 'PREP N' Cxs across "Corpus:Similect" Subsets

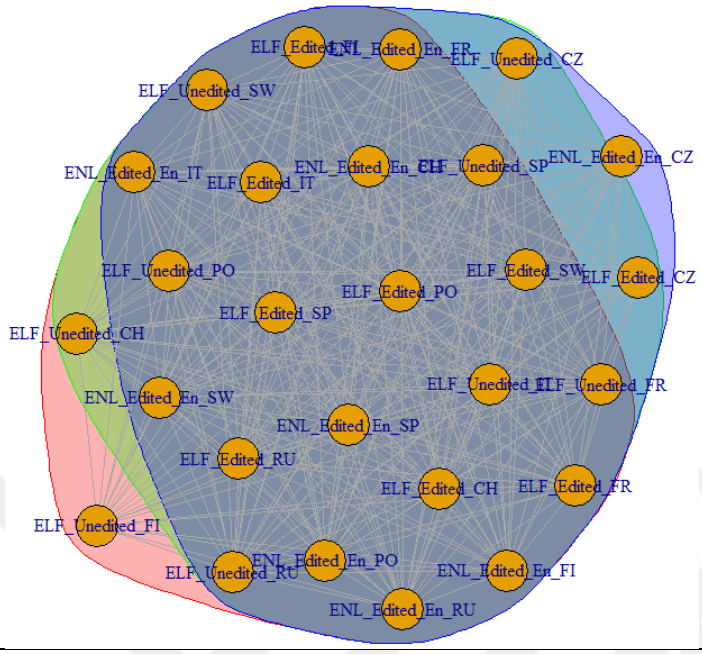
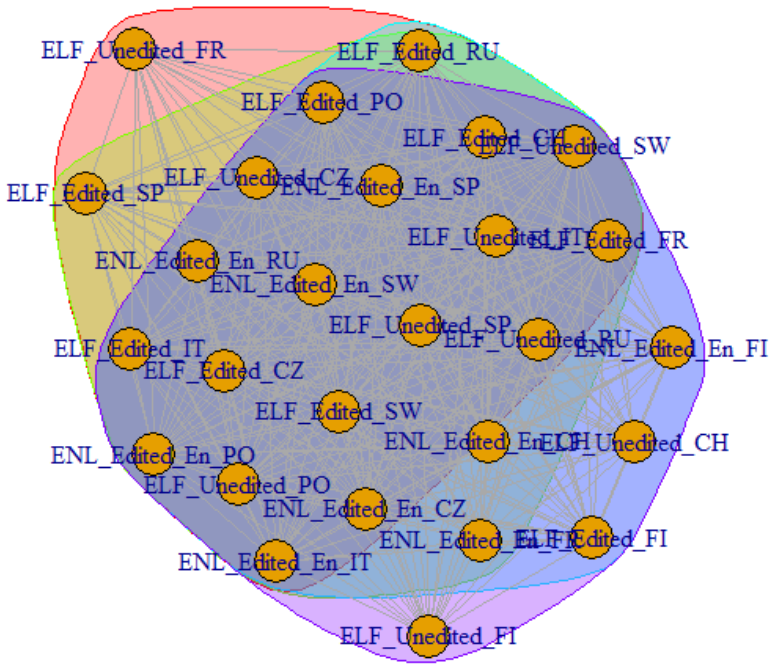
Discipline	Corpus	***across* N-s**		***over* N**		***within* N**		***through* N**		***in order to* V**		***in NN with* (NP)**	
		Median (normed per 10k)	**DP_{norm}**	**Median (normed per 10k)**	**DP_{norm}**	**Median (normed per 10k)**	**DP_{norm}**	**Median (normed per 10k)**	**DP_{norm}**	**Median (normed per 10k)**	**DP_{norm}**	**Median (normed per 10k)**	**DP_{norm}**
CH	**Unedited**	0.00	0.79	0.24	0.52	0.24	0.40	0.47	0.29	0.24	0.40	0.00	0.51
	Edited	0.00	0.86	0.19	0.50	0.33	0.33	0.56	0.30	0.19	0.45	0.17	0.53
EN	**Edited**	0.00	0.68	0.30	0.43	0.55	0.36	0.91	0.26	0.15	0.55	0.11	0.34
CZ	**Unedited**	0.00	0.88	0.00	0.56	0.76	0.31	0.25	0.45	0.25	0.45	0.44	0.24

	Edited	0.00	0.93	0.15	0.50	0.75	0.37	0.30	0.44	0.15	0.47	0.30	0.36
EN	**Edited**	0.15	0.52	0.92	0.24	1.27	0.25	0.76	0.35	0.46	0.44	0.15	0.45
FI	**Unedited**	0.00	0.86	0.25	0.32	0.15	0.44	0.00	0.55	0.00	0.59	0.15	0.50
	Edited	0.00	0.85	0.46	0.32	0.39	0.30	0.13	0.44	0.07	0.48	0.18	0.36
EN	**Edited**	0.07	0.61	0.41	0.35	0.61	0.28	0.61	0.30	0.15	0.53	0.14	0.38
FR	**Unedited**	0.00	0.90	0.23	0.68	0.56	0.43	0.45	0.42	1.02	0.27	0.22	0.45
	Edited	0.00	0.91	0.30	0.49	0.47	0.36	0.70	0.31	0.42	0.38	0.42	0.31
EN	**Edited**	0.20	0.50	0.60	0.31	0.82	0.44	0.71	0.35	0.00	0.55	0.14	0.46
IT	**Unedited**	0.00	0.71	0.37	0.48	0.72	0.46	0.72	0.43	1.12	0.42	0.00	0.64
	Edited	0.00	0.54	0.35	0.41	0.70	0.52	0.70	0.40	0.35	0.46	0.35	0.59
EN	**Edited**	0.00	0.72	1.24	0.32	0.62	0.36	0.38	0.46	0.00	0.62	0.31	0.58
PO	**Unedited**	0.00	0.91	0.00	0.64	0.14	0.52	0.71	0.44	1.14	0.30	0.57	0.36
	Edited	0.00	0.96	0.37	0.54	1.39	0.44	0.68	0.45	0.93	0.28	0.00	0.49
EN	**Edited**	0.41	0.48	0.51	0.42	0.97	0.49	1.36	0.31	0.63	0.37	0.24	0.48
RU	**Unedited**	0.00	0.82	0.57	0.44	0.30	0.64	0.57	0.42	0.30	0.41	0.29	0.52
	Edited	0.00	0.73	0.63	0.37	1.05	0.26	0.63	0.35	0.66	0.34	0.42	0.32
EN	**Edited**	0.00	0.57	1.13	0.30	0.57	0.45	1.13	0.32	0.21	0.52	0.00	0.54
SP	**Unedited**	0.00	0.95	0.30	0.36	0.33	0.38	1.16	0.35	1.41	0.28	0.15	0.51
	Edited	0.00	0.97	0.63	0.45	0.65	0.34	0.95	0.31	1.09	0.29	0.16	0.49
EN	**Edited**	0.00	0.68	0.44	0.44	0.73	0.45	0.84	0.50	0.00	0.58	0.00	0.55
SW	**Unedited**	0.00	0.93	0.65	0.28	0.52	0.51	1.65	0.42	0.90	0.33	0.43	0.45
	Edited	0.00	0.96	0.61	0.38	0.54	0.56	0.66	0.46	0.42	0.45	0.39	0.33
EN	**Edited**	0.00	0.70	1.23	0.21	0.79	0.44	1.50	0.28	0.76	0.40	0.26	0.45

Despite the fluctuation in the frequencies, only '*within* N' and '*in order to* V' were found to have significant differences across several "corpus:similect" subsets (see Table 8.13). Regarding '*within* N', many subsets appear to cluster together while Czech (green cluster) and Finnish data in ELF_Unedited (pink cluster) are the subsets with the only significant difference. With respect to '*in order to* V', Finnish and Spanish subsets in ELF_Edited differ significantly, and the unedited French ELF (pink cluster) is maximally different from the rest of the subsets. Despite '*within* N-s' being a key construction in ENL, and '*in order to* V' being a key construction in the ELF corpora, both have one of the lowest frequency counts for the Finnish L1 datasets. For the latter, one of the few ELF-ENL differences across the subsets is identified between the unedited French ELF and the matching ENL data that has a median value of zero.

Table 8.13 Clusters and P-Values of Pairwise Differences for 'PREP N' Cxs across "Corpus:Similect" Subsets

Cx	**Clusters**	**Pairwise Differences**

within N		**Within ELF_Unedited:** CZ- FI (p = .043)
in order to V		**- Within ELF_Edited:** FI - SP (p = .027) **- ELF vs. ENL:** ENL_Edited: FR - ELF_Unedited: FR (p = .018)

The semantic domain frequencies point to generally higher proportions of concrete than abstract nouns in *across*-, *over*- and *within*-constructions (see Table 8.14). Several exceptions are (1)‘*across* N-s’ in edited Spanish and unedited Portuguese ELF, as well as ENL matching Spanish ELF: (2) ‘*over* N’ in unedited Italian ELF, and (3) ‘*within* N’ in unedited Italian ELF and ENL matching Russian ELF. Other

constructions show more variation in the proportion of semantic domains. Regarding '*through* N', the majority of ENL and ELF_Unedited subsets also have higher proportions of concrete noun use, as opposed to ELF_Edited where most of the subsets have larger proportions of abstract nouns. The two subsets with the largest proportions of abstract nouns in the construction are the edited Italian ELF and the ENL subset that matches the Italian ELF texts. A concordance search for the abstract nouns used in these subsets however shows that these proportions of abstract nouns do not necessarily imply high frequencies of abstract nouns (29 in edited Italian ELF, and 32 in the matching ENL subset), and there are no dominant nouns among the collocates. For example, '*belief*' is the most frequent abstract noun in edited Italian ELF with only three instances in one text (e.g., *'truth approximation through belief revision'*, in ELF_Edited_39, SSH, philosophy). For the ENL subset matching the Italian ELF texts, the most frequent nominal collocates are the few nouns that occur twice, including '*use'*, '*work'*, and '*corpus'*, all of which come from the same text (e.g., *'through the use of computers as a tool for the humanities*', '*through work on corpora'* by ENL_Unedited_42, SSH, linguistics). As for '*in* NN *with* (NP)', only ELF subsets make use of the comparisons sense of the construction, however in small proportions.

Table 8.14 Percentages of Semantic Domains for 'PREP N' Cxs across "Corpus:Similect" Subsets

Cx		***across* N-s**		***over* N**		***within* N**		***through* N**		***in NN with* (NP)**	
Domain		**Abstract**	**Concrete**	**Abstract**	**Concrete**	**Abstract**	**Concrete**	**Abstract**	**Concrete**	**Agreement**	**Comparison**
Examples		*types, areas, groups*	*layers, tasks segments*	*order, ability, meaning*	*year, time, level*	*field, context, framework*	*limit, time, size*	*analysis, lens, process*	*time, grade, client*	*line, accordance*	*comparison*
CH	**Unedited**	0.00	100.00	7.69	92.31	31.82	68.18	32.26	67.74	92.31	7.69
	Edited	26.32	73.68	31.03	68.97	18.42	81.58	63.16	36.84	100.00	0.00
EN	**Edited**	7.14	92.86	20.83	79.17	27.03	72.97	45.45	54.55	100.00	0.00
CZ	**Unedited**	9.09	90.91	25.00	75.00	27.63	72.37	60.00	40.00	58.62	41.38
	Edited	9.52	90.48	0.00	100.00	21.49	78.51	58.62	41.38	84.00	16.00
EN	**Edited**	22.73	77.27	2.86	97.14	17.20	82.80	30.77	69.23	100.00	0.00
FI	**Unedited**	19.35	80.65	31.58	68.42	6.67	93.33	30.43	69.57	85.71	14.29
	Edited	21.43	78.57	9.30	90.70	25.00	75.00	69.57	30.43	86.49	13.51
EN	**Edited**	70.00	30.00	0.00	100.00	14.06	85.94	62.22	37.78	100.00	0.00
FR	**Unedited**	42.86	57.14	12.28	87.72	21.95	78.05	35.71	64.29	62.50	37.50
	Edited	83.33	16.67	2.78	97.22	36.84	63.16	55.56	44.44	54.55	45.45
EN	**Edited**	33.85	66.15	0.00	100.00	21.43	78.57	58.06	41.94	100.00	0.00
IT	**Unedited**	20.00	80.00	50.00	50.00	48.28	51.72	71.43	28.57	100.00	0.00
	Edited	44.44	55.56	44.44	55.56	56.25	43.75	90.00	10.00	72.73	27.27
EN	**Edited**	3.33	96.67	0.00	100.00	0.00	100.00	87.50	12.50	100.00	0.00
PO	**Unedited**	50.00	50.00	0.00	100.00	21.43	78.57	53.85	46.15	80.00	20.00
	Edited	20.00	80.00	9.09	90.91	27.27	72.73	44.44	55.56	50.00	50.00
EN	**Edited**	36.84	63.16	0.00	100.00	27.69	72.31	48.00	52.00	100.00	0.00
RU	**Unedited**	0.00	100.00	31.58	68.42	42.86	57.14	0.00	100.00	42.86	57.14
	Edited	22.22	77.78	11.76	88.24	20.45	79.55	52.38	47.62	27.27	72.73
EN	**Edited**	30.43	69.57	0.00	100.00	60.71	39.29	41.67	58.33	100.00	0.00
SP	**Unedited**	50.00	50.00	0.00	100.00	25.00	75.00	57.14	42.86	100.00	0.00

	Edited	70.00	30.00	20.00	80.00	48.00	52.00	50.00	50.00	100.00	0.00
EN	**Edited**	0.00	100.00	0.00	100.00	20.00	80.00	44.44	55.56	100.00	0.00
SW	**Unedited**	0.00	100.00	15.79	84.21	37.21	62.79	34.09	65.91	100.00	0.00
	Edited	0.00	100.00	25.81	74.19	41.18	58.82	60.42	39.58	88.89	11.11
EN	**Edited**	15.79	84.21	0.00	100.00	40.91	59.09	37.50	62.50	100.00	0.00

Note. Sum of percentages for semantic domains for each subset equals all Cx collocates included in semantic analysis.

Looking at the usage data of the 'PREP N' constructions across "corpus:discipline" subsets, it is apparent that the highest frequency and most even dispersion of '*across* N-s' is present in SSH texts by ENL writers. '*Over* N' is more frequent in both ENL subsets than in ELF subsets. However, the ELF vs. ENL difference is larger in SCI than in SSH writing. Similarly, the ELF vs. ENL difference in the use of '*within* N', and '*through* N' is limited among the SSH subsets in the three study corpora, as opposed to the larger-scale differences in SCI. However, frequencies of '*through* N' also indicate that the SSH texts by ENL writer have by far the highest construction frequencies. As for '*in order to* V', SSH texts by ELF writers show higher frequencies and more even dispersions, while SCI of ENL_Edited seems to have only few outlier values. Lastly, differences in the use of '*in NN with* (VP)' seem almost negligible, with ELF texts from both disciplinary subsets having somewhat higher frequencies and more even dispersions (see Table 8.15).

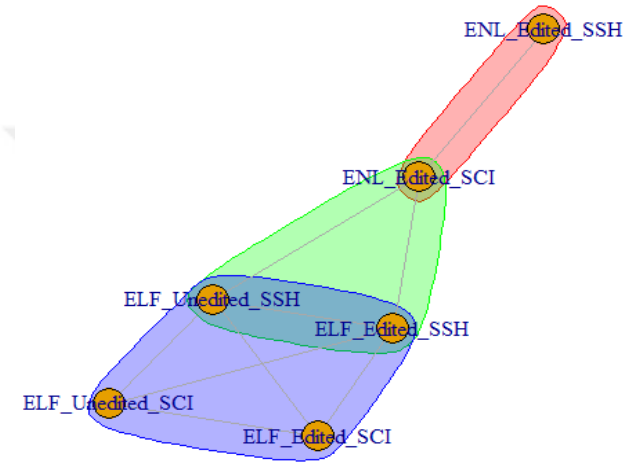
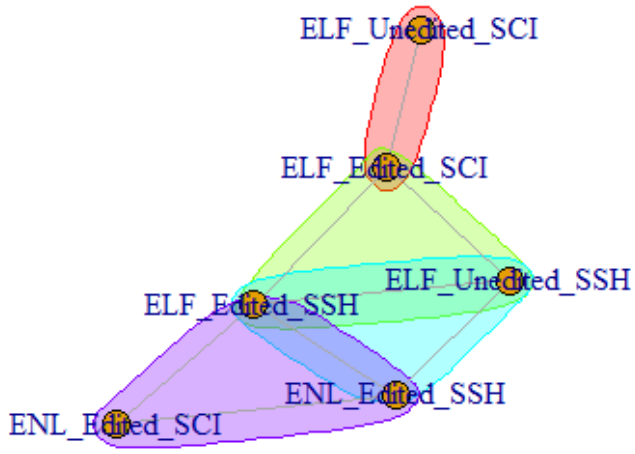
Table 8.15 Bootstrapped Median and DP_{norm} Values for 'PREP N' Cxs across "Corpus:Discipline" Subsets

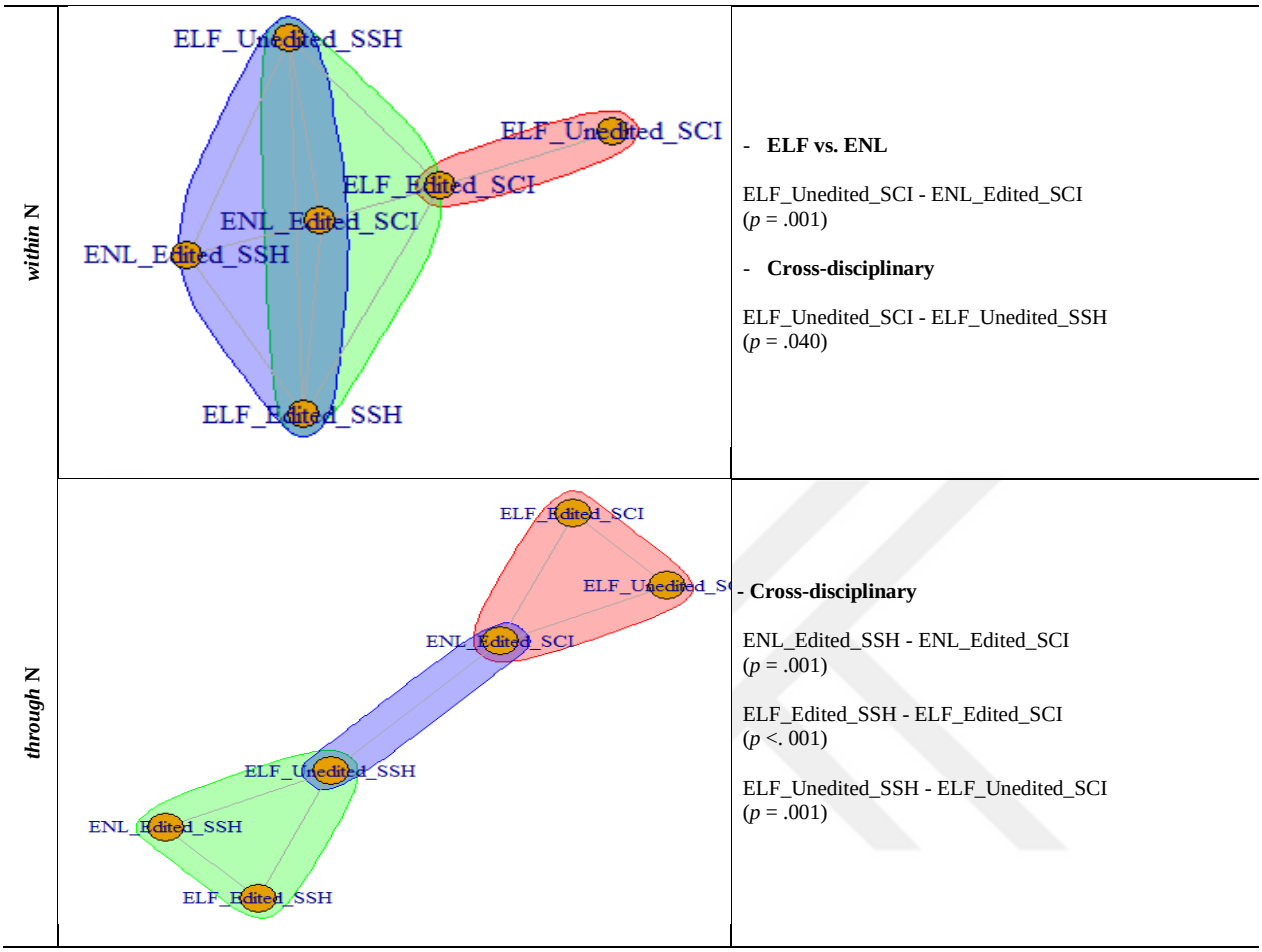
Discipline	**Corpus**	***across* N-s**		***over* N**		***within* N**		***through* N**		***in order to* V**		***in NN with* (NP)**	
		Median (normed per 10k)	**DP_{norm}**	**Median (normed per 10k)**	**DP_{norm}**	**Median (normed per 10k)**	**DP_{norm}**	**Median (normed per 10k)**	**DP_{norm}**	**Median (normed per 10k)**	**DP_{norm}**	**Median (normed per 10k)**	**DP_{norm}**
SCI	**ELF_Unedited**	0.00	0.85	0.15	0.53	0.30	0.47	0.24	0.50	0.24	0.48	0.22	0.44
	ELF_Edited	0.00	0.88	0.26	0.48	0.44	0.43	0.30	0.43	0.19	0.49	0.26	0.37
	ENL_Edited	0.00	0.69	0.68	0.32	0.68	0.35	0.55	0.32	0.00	0.57	0.11	0.49
SSH	**ELF_Unedited**	0.00	0.82	0.41	0.42	0.57	0.43	0.76	0.38	0.68	0.34	0.25	0.44
	ELF_Edited	0.00	0.79	0.56	0.37	0.75	0.31	0.88	0.32	0.50	0.37	0.21	0.41
	ENL_Edited	0.15	0.50	0.76	0.35	0.82	0.40	1.27	0.29	0.28	0.43	0.15	0.41

For '*across* N-s', clusters and significant pairwise differences point to a clear ELF-ENL distinction where, expectedly, the SSH subset of ENL_Edited is seen to be maximally different from the SCI subsets in the two ELF corpora. As the median frequencies also suggest, the use of '*over* N' differs significantly between ENL and ELF_Edited, as well as ELF_Unedited in SCI writing. Similarly, the frequency of '*within* N' in the SCI subset of ELF_Unedited is much lower than all the other subsets. '*Through* N' points to a relatively clearer disciplinary distinction where all SSH subsets have higher

frequencies of the construction than the SCI subsets. This disciplinary divide, while evident across the SCI and SSH subsets of ELF_Unedited, is shown to be more complex in the clustering plot where the latter is seen to neighbor SCI in ENL_Edited (see Table 8.16).

Table 8.16 Clusters and P-Values of Pairwise Differences for 'PREP N' Cxs across "Corpus:Discipline" Subsets

Cx	Clusters	Pairwise Differences
***across* N-s**		- **ELF vs. ENL:** ELF_Edited_SCI - ENL_Edited_SCI (p = .002) ELF_Unedited_SCI - ENL_Edited_SCI (p = .007) ELF_Unedited_SSH - ENL_Edited_SSH (p < .001) ELF_Edited_SSH - ENL_Edited_SSH (p < .001)
***over* N**		- **ELF vs. ENL** ELF_Unedited_SCI - ENL_Edited_SCI (p < .001) ELF_Edited_SCI - ENL_Edited_SCI (p = .001) - **Cross-disciplinary** ELF_Unedited_SCI - ELF_Unedited_SSH (p = .008)



Considering the distribution of semantic domains across the "corpus:discipline" subsets, it is apparent that '*across* N-s,' '*over* N,' and '*within* N' are commonly used with concrete nouns regardless of the disciplinary and similectal grouping of the texts. However, SSH texts in ELF_Edited in particular have higher ratios of abstract nouns than other subsets. The proportions of domains for '*through* N' point to relatively contrasting differences between ELF and ENL Both of the ELF_Unedited subsets have the highest proportions of concrete nouns, whereas the highest proportions of abstract nouns are present in both ELF_Edited subsets. Among the key ENL constructions, '*across* N-s', '*over* N', '*within* N', and '*through* N,' '*through* N' is the only one where a higher proportion of abstract than concrete nouns is observed in ENL (in its SSH subset). In the case of '*in* NN *with* (NP)', a key ELF construction, agreement is the dominant semantic domain, while the comparison domain is only used by ELF writers (see Table 8.17).

Table 8.17 Percentages of Semantic Domains for 'PREP N' Cxs across "Corpus:Discipline" Subsets

Cx		***across* N-s**		***over* N**		***within* N**		***through* N**		***in NN with* (NP)**	
Domain		**Abstract**	**Concrete**	**Abstract**	**Concrete**	**Abstract**	**Concrete**	**Abstract**	**Concrete**	**Agreement**	**Comparison**
Examples		*types, areas, groups*	*layers, tasks segments*	*order, ability, meaning*	*year, time, level*	*field, context, framework*	*limit, time, size*	*analysis, lens, process*	*time, grade, client*	*line, accordance*	*comparison*
SCI	**ELF_Unedited**	16.67	83.33	14.52	85.48	15.57	84.43	31.71	68.29	77.36	22.64
	ELF_Edited	20.97	79.03	9.52	90.48	18.55	81.45	56.00	44.00	69.51	30.49
	ENL_Edited	28.70	71.30	9.17	90.83	13.22	86.78	40.91	59.09	100.00	0.00
SSH	**ELF_Unedited**	20.00	80.00	23.42	76.58	39.27	60.73	44.68	55.32	74.42	25.58
	ELF_Edited	34.78	65.22	28.46	71.54	40.57	59.43	60.80	39.20	88.89	11.11
	ENL_Edited	27.89	72.11	0.00	100.00	36.28	63.72	51.32	48.68	100.00	0.00

Note. Sum of percentages for abstract and concrete meanings for each subset equals all Cx collocates included in semantic analysis.

The last comparisons for 'PREP N' are across the "similect:discipline" datasets. These comparisons show that some of the differences observed earlier across "corpus:similect" and "corpus:discipline" subsets are also present in the interaction of similect and discipline, without considering corpus-based differences. For instance, '*across* N-s' is rather under-dispersed in SCI data as it either has several outliers (e.g., Czech ELF, ENL matching Chinese ELF) and no instances in several subsets (e.g., Chinese, French, Spanish Swedish ELF subgroups, and ENL matching the Swedish ELF data; see Table 8.18).

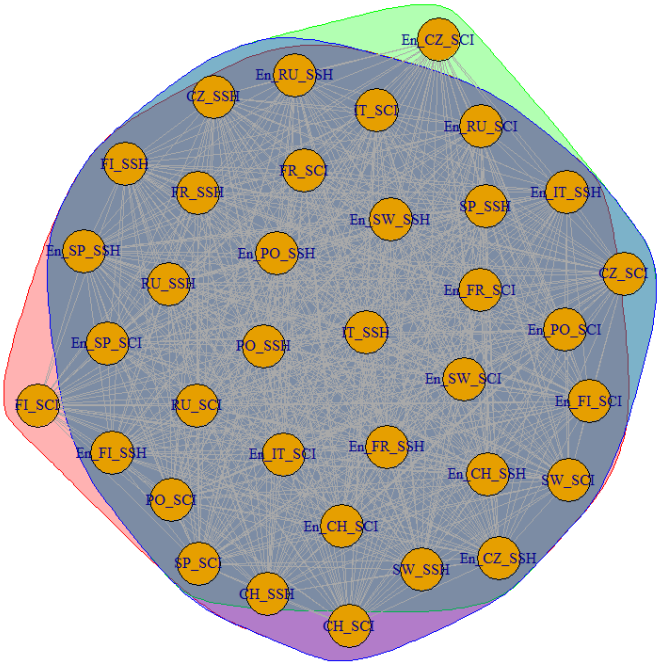
Table 8.18 Bootstrapped Median and DP_{norm} Values for 'PREP N' Cxs across "Similect:Discipline" Subsets

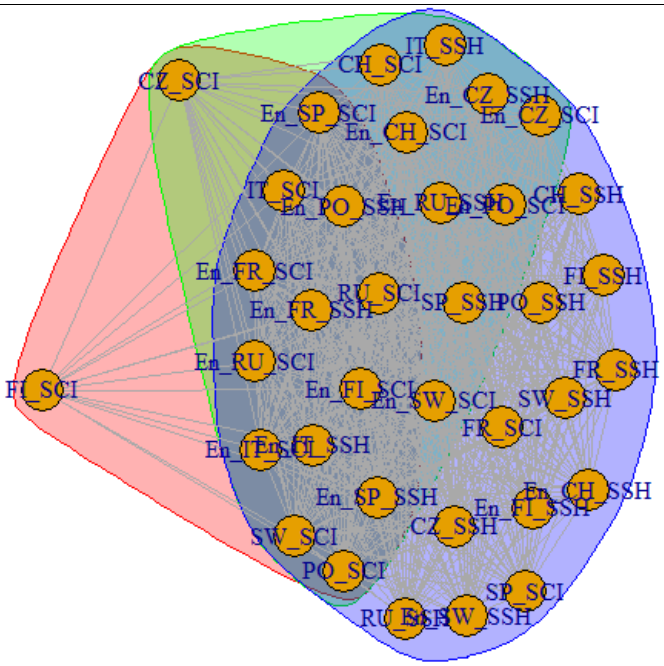
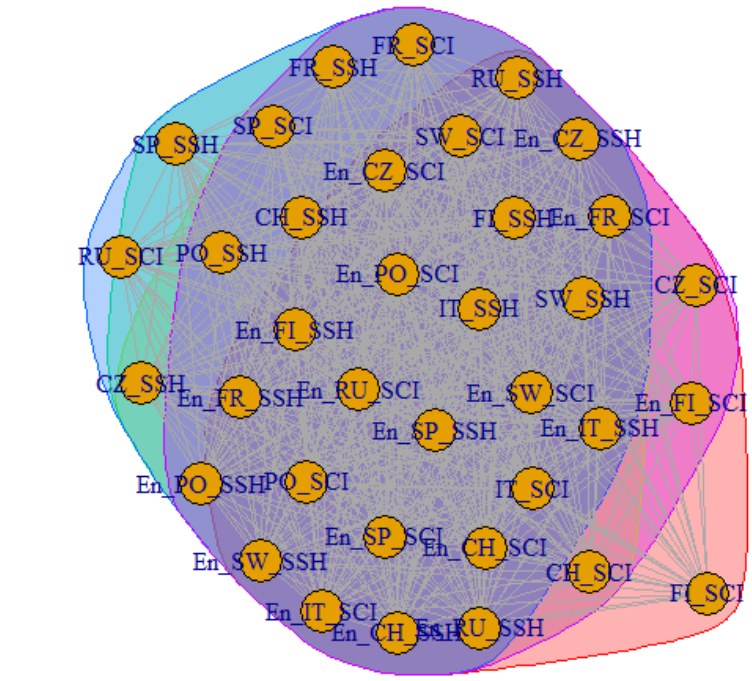
Discipline	**Similect**	***across* N-s**		***over* N**		***within* N**		***through* N**		***in order to* V**		***in NN with* (NP)**	
		Median (normed per 10k)	**DP_{norm}**	**Median (normed per 10k)**	**DP_{norm}**	**Median (normed per 10k)**	**DP_{norm}**	**Median (normed per 10k)**	**DP_{norm}**	**Median (normed per 10k)**	**DP_{norm}**	**Median (normed per 10k)**	**DP_{norm}**
SCI	**CH**	NA	NA	0.00	0.58	0.31	0.35	0.48	0.27	0.19	0.49	0.19	0.46
	CZ	0.00	0.98	0.00	0.65	0.74	0.39	0.19	0.37	0.00	0.50	0.26	0.28
	FI	0.00	0.85	0.21	0.37	0.15	0.39	0.00	0.56	0.00	0.62	0.15	0.37
	FR	NA	NA	0.34	0.50	0.35	0.51	0.59	0.33	0.82	0.29	0.23	0.37
	IT	0.00	0.54	0.35	0.40	0.36	0.54	0.18	0.60	0.36	0.46	0.18	0.63
	PO	0.00	0.78	0.46	0.53	0.28	0.55	0.23	0.57	0.28	0.42	0.00	0.61
	RU	0.00	0.67	0.37	0.49	0.44	0.39	0.37	0.40	0.90	0.31	0.11	0.46
	SP	NA	NA	0.48	0.41	0.48	0.43	1.59	0.32	1.47	0.23	0.35	0.35
	SW	NA	NA	0.36	0.51	0.43	0.32	0.59	0.55	0.51	0.43	0.29	0.37
	EN_CH	0.00	0.98	0.66	0.40	0.44	0.32	0.77	0.25	0.00	0.64	0.11	0.32
	EN_CZ	0.14	0.62	0.71	0.32	1.55	0.18	0.35	0.45	0.49	0.49	0.14	0.45
	EN_FI	0.00	0.67	0.41	0.27	0.27	0.26	0.14	0.29	0.00	0.79	0.00	0.52
	EN_FR	0.00	0.72	0.70	0.34	0.30	0.64	0.80	0.30	0.00	0.67	0.10	0.61
	EN_IT	0.00	0.53	0.78	0.33	0.62	0.28	0.16	0.54	0.00	0.80	0.16	0.70
	EN_PO	0.48	0.45	1.36	0.36	1.77	0.40	1.36	0.26	0.41	0.56	0.14	0.68
	EN_RU	0.21	0.63	0.86	0.29	0.64	0.46	0.43	0.47	0.86	0.32	0.00	0.79
	EN_SP	0.00	0.79	0.31	0.38	0.47	0.51	0.47	0.48	0.00	0.65	0.00	0.67
	EN_SW	NA	NA	1.50	0.19	0.96	0.26	1.50	0.31	1.50	0.35	0.69	0.35
SSH	**CH**	0.00	0.65	0.47	0.36	0.33	0.36	0.49	0.31	0.24	0.37	0.00	0.57
	CZ	0.00	0.84	0.44	0.40	0.76	0.18	0.66	0.34	0.56	0.26	0.37	0.34
	FI	0.00	0.84	0.53	0.25	0.76	0.28	0.95	0.30	0.19	0.42	0.18	0.50
	FR	0.00	0.79	0.21	0.66	0.71	0.29	0.62	0.38	0.62	0.37	0.28	0.37
	IT	0.00	0.67	1.13	0.41	1.87	0.42	0.80	0.29	1.41	0.40	0.00	0.59
	PO	NA	NA	0.00	0.62	0.43	0.49	0.83	0.42	1.25	0.25	0.54	0.32
	RU	0.00	0.93	0.74	0.26	0.96	0.44	1.00	0.30	0.29	0.39	0.50	0.32
	SP	0.00	0.89	0.30	0.47	0.65	0.30	0.90	0.32	0.78	0.34	0.00	0.68
	SW	0.00	0.89	0.82	0.20	1.19	0.48	1.49	0.39	0.90	0.36	0.55	0.36

EN_CH	0.15	0.49	0.23	0.33	0.91	0.45	1.21	0.24	0.30	0.50	0.15	0.40
EN_CZ	0.31	0.43	0.92	0.18	0.92	0.25	1.07	0.25	0.46	0.42	0.15	0.48
EN_FI	0.15	0.58	0.46	0.46	0.91	0.31	1.37	0.26	0.15	0.37	0.30	0.22
EN_FR	0.96	0.36	0.41	0.22	0.89	0.40	0.55	0.40	0.28	0.49	0.14	0.40
EN_IT	0.00	1.00	1.88	0.38	0.38	0.52	1.88	0.42	0.38	0.53	0.38	0.53
EN_PO	0.48	0.54	0.48	0.50	0.72	0.66	1.81	0.38	0.85	0.26	0.24	0.47
EN_RU	0.00	0.62	1.28	0.36	0.57	0.47	1.28	0.20	0.00	0.88	0.28	0.50
EN_SP	0.00	0.66	0.70	0.44	0.84	0.44	1.40	0.48	0.14	0.57	0.14	0.50
EN_SW	0.51	0.46	1.02	0.27	0.76	0.59	1.53	0.25	0.51	0.55	0.13	0.60

The pairwise statistical test for '*within* N,' reveals another one of the few ELF-ENL differences across subgroups, which is found between the SCI texts by Czech writers and the matching ENL texts. In the case of '*through* N', the low frequencies and uneven dispersions of the construction in SCI texts by Finnish and Czech ELF writers, separate them from the rest of the subsets that are clustered together in the large central circle (see Table 8.19). In fact, the significant higher frequency of '*through* N' in SSH compared to SCI texts by Finnish and Czech ELF writers also confirms the rarity of the construction in the SCI texts of these similectal groups. Lastly, regarding '*in order to* V,' the subset of SCI texts by Finnish ELF writers (pink cluster) is similarly separated from the majority of the groups. At the other end plot in the green cluster, however, the SCI texts by Russian ELF and SSH subsets by Spanish and Czech ELF writers also stand apart from the other subsets.

Table 8.19 Clusters and P-Values of Pairwise Differences for 'PREP N' Cxs across "Similect:Discipline" Subsets

***within* N**		**- Within SCI:** CZ - EN_CZ (p = .022)

***through* N**		**- Within SCI:** FI - CH (p = .001) FR - FI (p = .002) FR - SP (p = .003) FR - CZ (p = .047) **- Cross-disciplinary**: CZ_SCI – CZ_SSH (p = .029) FI_SCI – FI_SSH (p = .001)
***in order to* V**		**- Within SCI:** RU - CH (p = .047) RU - CZ (p = .012) RU - FI (p = .001) SP - FI (p = .015) FR - FI (p = .009) **- Cross-disciplinary:** CZ_SCI - CZ_SSH (p = .009)

The semantic domain proportions for the 'N PREP (NP)' constructions also show a remarkable degree of variation across "similect:discipline" subsets from all three study corpora. The domain frequencies for '*through* N' across subsets are the clearest indicator of this variation, as neither ELF nor ENL subsets show a clear preference of one domain over the other. The only consistent patterns are present in the use of '*over* N' among the SCI subsets that include higher proportions of concrete than

abstract nouns, and also the dominant use of '*in* NN *with* (NP)' for denoting agreement. Nevertheless, it is important to mention that the variation is somewhat more pronounced among ELF subsets compared to those in ENL_Edited, which is most evident from the fact that many more ENL than ELF subsets make use of only one semantic domain, especially in the cases of '*within N*' and '*in NN with* (NP)' (see Table 8.20).

Table 8.20 Percentages of Semantic Domains for 'PREP N' Cxs across "Similect:Discipline" Subsets

	Cx	***across* N-s**		***over* N**		***within* N**		***through* N**		***in NN with* (NP)**	
	Domain	**Abstract**	**Concrete**	**Abstract**	**Concrete**	**Abstract**	**Concrete**	**Abstract**	**Concrete**	**Agreement**	**Comparison**
	Examples	*types, areas, groups*	*layers, tasks segments*	*order, ability, meaning*	*year, time, level*	*field, context, framework*	*limit, time, size*	*analysis, lens, process*	*time, grade, client*	*line, accordance*	*comparison*
SCI	**CH**	0.00	0.00	0.00	100.00	6.25	93.75	32.26	67.74	94.44	5.56
	CZ	0.00	100.00	0.00	100.00	18.12	81.88	60.00	40.00	78.79	21.21
	FI	20.37	79.63	14.58	85.42	2.04	97.96	36.36	63.64	83.33	16.67
	FR	100.00	0.00	0.00	100.00	17.24	82.76	36.36	63.64	45.45	54.55
	IT	14.29	85.71	0.00	100.00	50.00	50.00	75.00	25.00	68.42	31.58
	PO	25.00	75.00	0.00	100.00	7.69	92.31	50.00	50.00	50.00	50.00
	RU	20.69	79.31	15.79	84.21	17.14	82.86	30.00	70.00	20.00	80.00
	SP	0.00	100.00	21.43	78.57	42.31	57.69	62.50	37.50	100.00	0.00
	SW	0.00	0.00	26.09	73.91	36.36	63.64	34.78	65.22	75.00	25.00
	EN_CH	0.00	0.00	27.40	72.60	8.82	91.18	37.50	62.50	100.00	0.00
	EN_CZ	14.29	85.71	4.65	95.35	15.00	85.00	22.22	77.78	100.00	0.00
	EN_FI	78.26	21.74	0.00	100.00	6.38	93.62	40.00	60.00	100.00	0.00
	EN_FR	25.00	75.00	0.00	100.00	3.70	96.30	83.33	16.67	100.00	0.00
	EN_IT	3.33	96.67	0.00	100.00	0.00	100.00	0.00	0.00	100.00	0.00
	EN_PO	0.00	100.00	0.00	100.00	0.00	100.00	75.00	25.00	100.00	0.00
	EN_RU	38.89	61.11	0.00	100.00	70.00	30.00	0.00	100.00	100.00	0.00
	EN_SP	0.00	0.00	0.00	100.00	0.00	100.00	100.00	0.00	100.00	0.00
	EN_SW	0.00	0.00	0.00	100.00	0.00	100.00	28.57	71.43	100.00	0.00
SSH	**CH**	20.83	79.17	32.76	67.24	42.86	57.14	63.16	36.84	100.00	0.00
	CZ	12.50	87.50	21.05	78.95	37.29	62.71	58.82	41.18	57.14	42.86
	FI	20.00	80.00	21.43	78.57	39.02	60.98	54.29	45.71	90.48	9.52
	FR	58.33	41.67	15.09	84.91	35.48	64.52	51.52	48.48	75.00	25.00
	IT	41.67	58.33	69.57	30.43	51.43	48.57	84.62	15.38	100.00	0.00
	PO	25.00	75.00	11.11	88.89	34.78	65.22	47.62	52.38	100.00	0.00
	RU	0.00	100.00	23.08	76.92	41.18	58.82	47.06	52.94	50.00	50.00
	SP	72.73	27.27	11.11	88.89	33.33	66.67	40.00	60.00	100.00	0.00
	SW	0.00	100.00	18.52	81.48	39.76	60.24	52.17	47.83	96.15	3.85
	EN_CH	7.14	92.86	0.00	100.00	42.50	57.50	50.00	50.00	100.00	0.00
	EN_CZ	37.50	62.50	0.00	100.00	21.21	78.79	38.10	61.90	100.00	0.00
	EN_FI	42.86	57.14	0.00	100.00	35.29	64.71	68.57	31.43	100.00	0.00
	EN_FR	34.43	65.57	0.00	100.00	32.56	67.44	52.00	48.00	100.00	0.00
	EN_IT	0.00	0.00	0.00	100.00	0.00	100.00	87.50	12.50	100.00	0.00
	EN_PO	50.00	50.00	0.00	100.00	52.94	47.06	35.29	64.71	100.00	0.00
	EN_RU	0.00	100.00	0.00	100.00	37.50	62.50	62.50	37.50	100.00	0.00
	EN_SP	0.00	100.00	0.00	100.00	23.53	76.47	34.78	65.22	100.00	0.00
	EN_SW	15.79	84.21	0.00	100.00	42.86	57.14	50.00	50.00	100.00	0.00

Note. Sum of percentages for abstract and concrete meanings for each subset equals all Cx collocates included in semantic analysis.

As opposed to the other five 'PREP N' constructions with two identified broad semantic domains, only one domain; that is, purpose, was identified for '*in order to* V', which is why the comparative analysis of the proportion of domain frequencies was not necessary for the construction. Thus, its analysis is kept separate from the 'PREP N' constructions, and is based on normed median frequencies across subsets

from all three interactions (see Table 8.21). As the semantic domains were identified only for the verbs that the construction strongly attracts, the median frequencies for these verbs care generally in line with the overall frequencies of the construction across subsets (as seen in Table 12 and Table 15). The SCI texts in the subsets for the "similect:discipline" interaction in particular, have almost identical overall and collocate frequencies. In line with the fact that most significant pairwise differences in the use of '*in order to* V' were also identified among the SCI texts by ELF writers, as shown in Table 19, the frequencies for collocating verbs also vary greatly from one subset to another. The two subsets with the highest frequencies, SCI texts by Spanish ELF writers and SCI texts in ENL_Edited that match the Swedish ELF texts have the same median values for both the frequency of the construction and the collocates. SCI texts by Spanish ELF writers include 41 different verbs that are used 57 times, and '*simplify'* and '*clarify'* are the highest-frequency verbs with three occurrences each. As for the SCI texts in ENL_Edited that match the Swedish ELF texts, 21 different verbs are used altogether 31 times. The two highest-frequency verbs in this subset are '*determine'* with four and '*reduce'* with three occurrences. While both subsets have a variety of verb types, the more frequent ones for the ENL subset align with the top collocates of the overall ENL corpus; those in the ELF subset in question are among the less frequent collocates of the ELF corpora (see Table 8.21).

*Table 8.21 Median Values for the '*in order to *V' Cxs Denoting Purpose across Subsets*

Domain: Purpose				**Examples:** *obtain, avoid, assess, test, understand*				
Corpus:Similect			**Corpus:Discipline**			**Similect:Discipline**		
	Unedited	**Edited**		**SCI**	**SSH**		**SCI**	**SSH**
Similect	**Median (normed per 10k)**	**Median (normed per 10k)**	**Corpus**	**Median (normed per 10k)**	**Median (normed per 10k)**	**Similect**	**Median (normed per 10k)**	**Median (normed per 10k)**
CH	0.24	0.19	**ELF_Unedited**	0.24	0.51	**CH**	0.19	0.24
CZ	0.25	0.15	**ELF_Edited**	0.19	0.42	**CZ**	0.00	0.54
FI	0.00	0.07	**ENL_Edited**	0.00	0.28	**FI**	0.00	0.19
FR	0.91	0.37				**FR**	0.82	0.50
IT	1.12	0.35				**IT**	0.36	1.31
PO	1.14	0.87				**PO**	0.28	1.14
RU	0.30	0.66				**RU**	0.90	0.29
SP	1.24	1.09				**SP**	1.47	0.78
SW	0.90	0.41				**SW**	0.51	0.90
EN_CH	NA	0.15				**EN_CH**	0.00	0.30
EN_CZ	NA	0.42				**EN_CZ**	0.00	0.54
EN_FI	NA	0.15				**EN_FI**	0.00	0.15
EN_FR	NA	0.00				**EN_FR**	0.00	0.21
EN_IT	NA	0.00				**EN_IT**	0.00	0.38
EN_PO	NA	0.63				**EN_PO**	0.41	0.72
EN_RU	NA	0.21				**EN_RU**	0.86	0.00
EN_SP	NA	0.00				**EN_SP**	0.00	0.14

EN_SW	NA	0.63			**EN_SW**	1.50	0.38

In summary, findings for the 'PREP N' constructions highlight a number of interesting dissimilarities between the corpora in terms of construction frequencies across similectal and disciplinary subsets. While usages of '*after* N' and '*in* NN with (NP)' were not found to vary significantly on the basis of disciplinary and similectal groups within and across the corpora, '*over* N' was found to show significant differences only across the "corpus:discipline" subsets. Analyses of '*across* N-s', "*through* N" and '*in order to* V' pointed to significant differences across the "corpus:similect" and "similect:discipline" subsets. '*Within* N', however, was found to be the only 'PREP N' construction, analysis of which led to the identification of significant differences across groups in all three two-way interactions.

Unlike the results for the nominal 'N PREP (NP)' constructions, several differences between ELF subsets and their matching ENL subsets were also identified in the use of 'PREP N' constructions. The differences involving similectal groups are observed between the matching ENL subsets and unedited French ELF for '*in order to* V,' as well as SCI texts by Czech ELF for '*within* N.' All other significant ELF vs. ENL differences were found among the "corpus:discipline" subsets, including both SCI subsets by ELF writers for '*over* N;' and the SCI subset for '*within* N.' Except for the case of '*in order to* V' where the unedited French ELF subset has a higher frequency of the construction, all the ELF vs. ENL differences stem from the significantly higher frequencies of the constructions in ENL than in ELF. The finding on '*across* N-s' that all "corpus:discipline" subsets in ENL differ from all of those in ELF indicates a global ELF-ENL difference. However, regarding '*over* N' and '*within* N,' only SCI texts by ELF writers are remarkably different from the matching ENL subset. Thus, it is clear that both the differences between ELF subsets, and ELF-ENL differences point to the finding that SCI writing in ELF shows distinctive characteristics regarding the use of these low-frequency 'PREP N' constructions.

These results exhibit interesting parallels to previous research on academic writing in general and written academic ELF in particular. Research on academic writing generally focuses on higher frequency prepositions such as '*of,' 'in,' 'on'* as they often appear in highly conventional phrases used in academic

writing, which is why '*in order to* V' often appear in lists of lexical bundles (e.g., Biber et al., 1999; Hyland, 2008b; Simpson-Vlach & Ellis, 2010). Research has shown that such conventional and high-frequency sequences are used more often by ELF than ENL writers (Carey, 2013; Mauranen, 2012; Yilmaz & Römer, 2020). This finding is partially corroborated by the findings for the 'PREP N' constructions distinctive of ELF, which are, '*after* N' and '*in NN with* (NP)', and '*in order to* V'. These constructions generally include higher-frequency items than '*across* N-s, '*through* N', '*over* N,' '*within* N'. Among the key ELF constructions in this section, '*after* N' and '*in* NN *with* (NP)' were found to differ only across the corpora, highlighting their common use as a general ELF feature only. However, in the case of *'in order to* V,' only one ELF vs. ENL difference, for the unedited French ELF subset, was found. The similectal variation for '*in order to* V', although limited, hints also to the potential heterogeneity across similectal backgrounds even in the use of such common academic writing sequences.

While this section focused on constructions including prepositions with limited frequencies in academic writing, recent research on language change points to the gradual yet steady increase in the frequency of prepositions that are traditionally infrequent in academic writing (Biber & Gray, 2016). In the previous chapter which focused on cross-corpus differences, I argued that this was not the case for written academic ELF on the basis of the lower frequencies of '*across* N-s,' '*within* N,' '*over* N,' and '*through* N' in ELF corpora than in the ENL corpus. The results of this chapter are consistent with this claim for '*across* N-s' and '*through* N-S' as interactions of factors including similectal and disciplinary groups did not yield remarkable differences between subsets. However, the findings for '*within* N' and '*over* N' are contradictory as only SCI texts by ELF and ENL writers show such differences. Since the SCI texts by ELF writers from across different similectal backgrounds also showed substantial variation, these findings support Flowerdew's (2019) claim based on previous ELF research that the differences between hard sciences texts by ELF and ENL writers are more noticeable than in soft sciences due to relatively more relaxed norms in the former. Along with a preference for higher-frequency conventional items, this might also be the reason that ELF writers in SCI disciplines differ from their ENL counterparts to a greater extent than in SSH disciplines.

8.1.3 Differences in 'ADJ PREP (NP)' Constructions

As the median and DP_{norm} values in Table 8.3 show, the univariate comparisons for all three adjectival constructions pointed to significant differences across the subsets in the interactions of corpora, disciplines and similects. As no relevant significant differences were found in pairwise comparisons for the three-way "corpus:discipline:similect" subgroups, this section focuses on the findings for comparisons across the "corpus:similect," "corpus:discipline," and "similect:discipline" subgroups.

Across the "corpus:similect" subgroups, a high degree of variation across subsets is clear from the normed median frequencies, and DP_{norm} values, which resemble those for the low-frequency 'PREP N' constructions analyzed in the previous sections. Being the adjectival construction with the highest frequency, 'ADJ *of* (NP),' however, differs from the other two with higher frequencies and DP_{norm} values across subsets. Still, while ENL texts, in general, have higher frequencies that are spread across subsets more equally than in ELF subsets, the ENL texts matching the Italian and Spanish ELF texts, and the edited Italian ELF subset have the highest frequencies among the subsets. Further, in line with the keyness analyses reported in the previous chapter, 'ADJ-SUPERL *of* (NP)' has slightly higher frequencies and dispersion among ELF subsets. However, the large number of median values of zero and high DP_{norm} values indicate that the construction frequencies are sparsely dispersed with occurrences in only few texts in both edited and unedited Chinese ELF, Italian ELF and Swedish ELF subsets. 'ADJ *at* (NP)' has higher frequencies and more even dispersion across ENL than ELF subsets, only a few of which have medians larger than zero (edited Portuguese and Russian, edited and unedited Spanish ELF subsets)(see Table 8.22).

Table 8.22 Bootstrapped Median and DP_{norm} Values for 'ADJ PREP (NP)' Cxs across "Corpus:Similect" Subsets

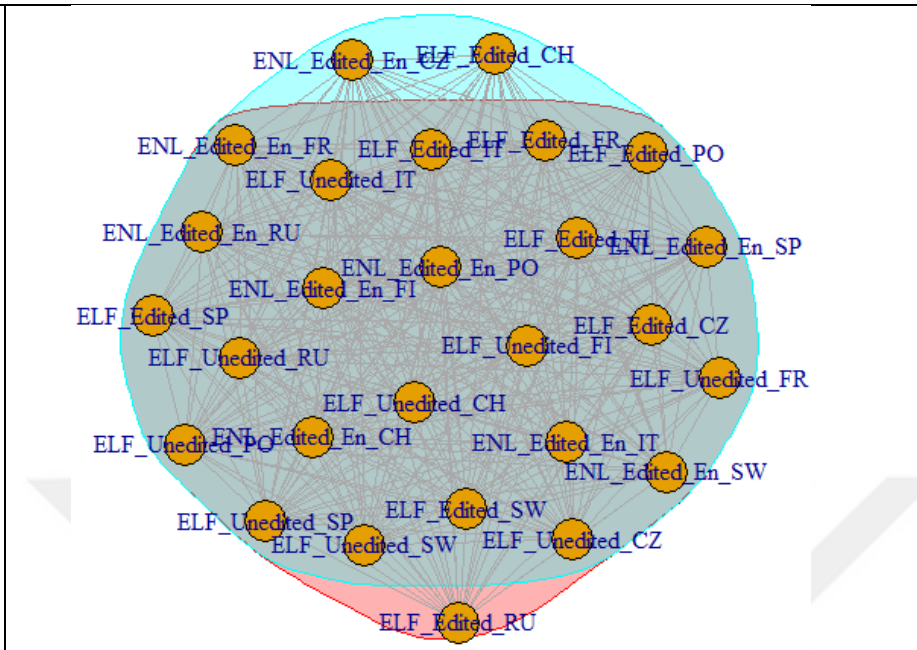
Similect	Corpus	ADJ *of* (NP)		ADJ-SUPERL *of* (NP)		ADJ *at* (NP)	
		Median (normed per 10k)	DP_{norm}	Median (normed per 10k)	DP_{norm}	Median (normed per 10k)	DP_{norm}
CH	**Unedited**	0.24	0.49	0.00	0.62	0.00	0.83
	Edited	0.33	0.35	0.00	0.64	0.00	0.77
EN	**Edited**	0.55	0.26	0.00	0.53	0.11	0.48
CZ	**Unedited**	0.22	0.45	0.22	0.42	0.00	0.58
	Edited	0.30	0.24	0.00	0.65	0.00	0.66
EN	**Edited**	0.46	0.30	0.00	0.69	0.14	0.49
FI	**Unedited**	0.38	0.31	0.15	0.51	0.00	0.79
	Edited	0.35	0.36	0.13	0.48	0.00	0.68

EN	**Edited**	0.41	0.29	0.00	0.60	0.14	0.45
FR	**Unedited**	0.56	0.41	0.22	0.45	0.00	0.59
	Edited	0.51	0.35	0.00	0.51	0.00	0.63
EN	**Edited**	0.75	0.29	0.00	0.61	0.20	0.42
IT	**Unedited**	0.72	0.28	0.00	0.81	0.00	0.69
	Edited	1.05	0.46	0.00	0.63	0.00	0.60
EN	**Edited**	1.51	0.20	0.00	0.54	0.31	0.52
PO	**Unedited**	0.56	0.37	0.42	0.26	0.00	0.66
	Edited	0.68	0.39	0.41	0.41	0.14	0.51
EN	**Edited**	0.73	0.35	0.00	0.78	0.24	0.48
RU	**Unedited**	0.30	0.38	0.30	0.46	0.00	0.67
	Edited	0.84	0.27	0.22	0.25	0.21	0.32
EN	**Edited**	0.64	0.33	0.00	0.53	0.21	0.45
SP	**Unedited**	0.80	0.21	0.30	0.45	0.15	0.54
	Edited	0.63	0.20	0.16	0.45	0.16	0.59
EN	**Edited**	1.04	0.21	0.00	0.82	0.14	0.53
SW	**Unedited**	0.86	0.39	0.00	0.81	0.00	0.60
	Edited	0.54	0.29	0.00	0.56	0.00	0.60
EN	**Edited**	0.79	0.30	0.00	0.76	0.26	0.38

The only relevant significant difference across pairs of subsets compared, however, is between edited Russian and Chinese ELF texts in the use of 'ADJ SUPERL *of* (NP).' As is also visualized in the clustering plot in Table 8.24, these two subsets form the most different pair among the subsets. Interestingly, the ENL subset matching the Czech ELF texts is the group closest to the edited Chinese ELF as both groups have only few extreme counts of occurrences for the construction. While the edited Russian ELF subgroup has a moderate median value of 0.22, it has the lowest DP_{norm} value across the subgroups with 0.25, meaning it has the most even dispersion of the construction across the datasets (see Table 8.23).

Table 8.23 Clusters and P-Values of Pairwise Differences for 'ADJ-SUPERL of (NP)' across "Corpus:Similect" Subsets

Cx	**Clusters**	**Pairwise difference**

ADJ-SUPERL*of* (NP)		**Within ELF_Edited:** RU - CH (p = .033)

The frequencies of semantic domains for 'ADJ *of* (NP)' illustrate that, with the exception of edited Russian and Italian ELF, all ELF subsets have higher frequencies for the abstract than the concrete domain. This difference between semantic domains is particularly large in the case of ELF_Unedited, mainly because the adjectives with concrete (quantifying) meanings in this corpus did not meet the collocational threshold to be included in the semantic analysis. There is greater variation in the proportions of domains for 'ADJ *at* (NP).' However, it is also largely evident that the proportion of the abstract domain is generally higher among the ELF_Edited and ENL_Edited subsets than in ELF_Unedited. With the exceptions of unedited Italian and Chinese ELF, all ELF_Unedited subsets show a tendency opposite to ENL subsets (see Table 8.24).

Table 8.24 Percentages of Semantic Domains for 'ADJ PREP (NP)' Cxs across "Corpus:Similect" Subsets

Cx		**ADJ *of* (NP)**		**ADJ *at* (NP)**	
Domain		**Abstract**	**Concrete**	**Abstract**	**Concrete**
Examples		*capable, independent, aware*	*many, much*	*significant, possible, similar*	*available, present, online*
CH	**Unedited**	100.00	0.00	100.00	0.00
	Edited	89.47	10.53	100.00	0.00
EN	**Edited**	38.33	61.67	77.78	22.22
CZ	**Unedited**	100.00	0.00	0.00	0.00
	Edited	81.82	18.18	100.00	0.00
EN	**Edited**	48.84	51.16	60.00	40.00
FI	**Unedited**	100.00	0.00	33.33	66.67
	Edited	68.42	31.58	80.00	20.00
EN	**Edited**	18.00	82.00	66.67	33.33
FR	**Unedited**	100.00	0.00	28.57	71.43
	Edited	72.73	27.27	80.00	20.00

EN	**Edited**	77.78	22.22	50.00	50.00
IT	**Unedited**	100.00	0.00	100.00	0.00
	Edited	28.57	71.43	100.00	0.00
EN	**Edited**	55.00	45.00	87.50	12.50
PO	**Unedited**	100.00	0.00	66.67	33.33
	Edited	70.00	30.00	66.67	33.33
EN	**Edited**	28.57	71.43	50.00	50.00
RU	**Unedited**	100.00	0.00	33.33	66.67
	Edited	33.33	66.67	60.00	40.00
EN	**Edited**	16.67	83.33	75.00	25.00
SP	**Unedited**	100.00	0.00	100.00	0.00
	Edited	77.78	22.22	25.00	75.00
EN	**Edited**	21.05	78.95	0.00	100.00
SW	**Unedited**	100.00	0.00	0.00	100.00
	Edited	77.78	22.22	60.00	40.00
EN	**Edited**	44.83	55.17	100.00	0.00

Note. Sum of percentages for semantic domains for each subset equals all Cx collocates included in semantic analysis.

Across the "corpus:discipline" subgroups, the overall tendency for 'ADJ *of* (NP)' and 'ADJ-SUPERL *of* (NP)' is that SSH texts in all corpora have higher frequencies of these three adjectival constructions than SCI texts. For 'ADJ *at* (NP),' it is the opposite as the SCI subset in ENL_Edited has a noticeably higher frequency and more even dispersion of the construction than other subsets. Further, these differences, despite variation across subsets, are in line with the cross-corpus differences, which indicates that the cross-corpus comparisons are affected by certain groups over others especially regarding the analysis of 'ADJ-SUPERL *of* (NP)' and 'ADJ *at* (NP)' (see Table 8.25).

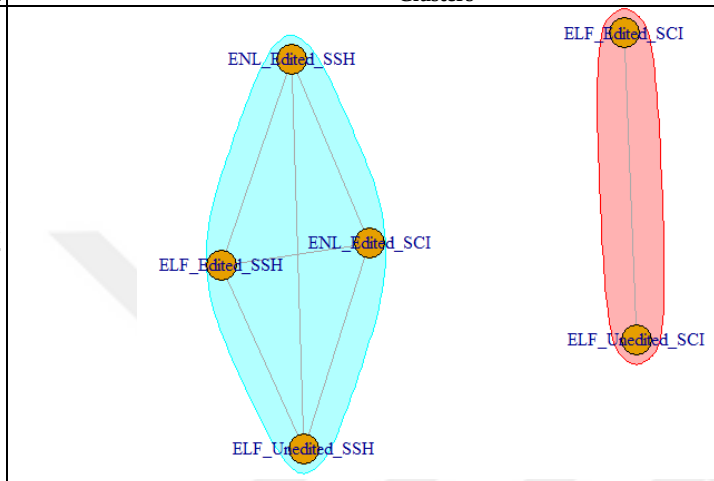
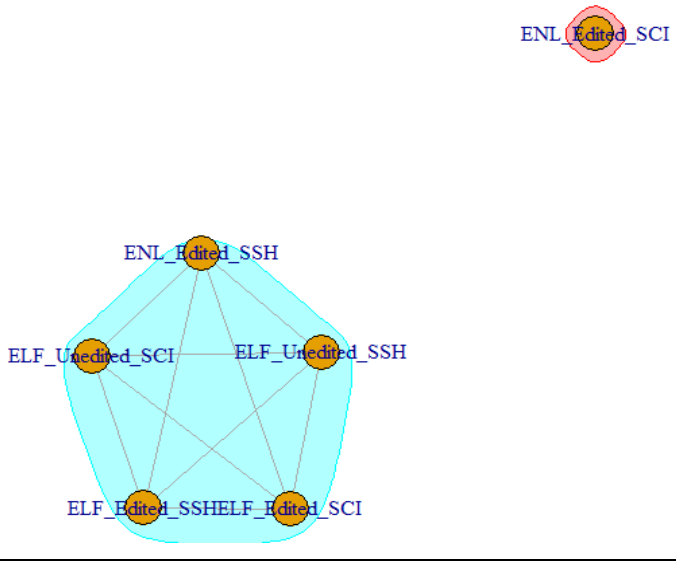
Table 8.25 Bootstrapped Median and DP_{norm} Values for 'ADJ PREP (NP)' Cxs across "Corpus:Discipline" Subsets

Discipline	**Corpus**	**ADJ *of* (NP)**		**ADJ-SUPERL *of* (NP)**		**ADJ *at* (NP)**	
		Median (normed per 10k)	**DP_{norm}**	**Median (normed per 10k)**	**DP_{norm}**	**Median (normed per 10k)**	**DP_{norm}**
SCI	**ELF_Unedited**	0.24	0.40	0.00	0.56	0.00	0.65
	ELF_Edited	0.30	0.34	0.00	0.51	0.00	0.61
	ENL_Edited	0.55	0.30	0.00	0.63	0.21	0.38
SSH	**ELF_Unedited**	0.71	0.32	0.24	0.47	0.00	0.66
	ELF_Edited	0.74	0.27	0.17	0.47	0.00	0.57
	ENL_Edited	0.84	0.26	0.00	0.59	0.00	0.54

The significant findings for pairwise comparisons and the clustering of "corpus:discipline" subsets also support the differences identified in distribution and dispersion. These findings further show that for 'ADJ *of* (NP)', the SCI texts in ENL_Edited are not as different from SSH texts as those by ELF writers. Regarding 'ADJ *at* (NP)', the significant pairwise differences illustrated in the clustering plot are also along the lines of the distributional and dispersional characteristics of the construction. The SCI subset in

ENL_Edited, with its higher and more evenly dispersed frequencies of occurrence, is separated from all other subsets (see Table 8.26).

Table 8.26 Clusters and P-Values of Pairwise Differences for 'ADJ PREP (NP)' across "Corpus:Discipline" Subsets

Cx	Clusters	Pairwise Differences
ADJ *of* (NP)		**- ELF vs. ENL** ELF_Unedited: SCI - ENL_Edited: SCI (p = .004) ELF_Edited: SCI - ENL_Edited: SCI (p = .048) **- Cross-disciplinary** ELF_Edited: SCI - ELF_Edited: SSH (p = .001) ELF_Unedited: SCI - ELF_Unedited: SSH (p < .001)
ADJ *at* (NP)		**- ELF vs. ENL** ENL_Edited: SCI - ELF_Edited: SCI (p = .017) ENL_Edited: SCI - ELF_Unedited: SCI (p = .003)

As for the frequencies of the semantic domains across the "corpus:similect" subsets, abstract meanings seem especially common among ELF texts regardless of disciplinary grouping, while ENL_Edited, especially its SSH subset, has a much higher proportion of the concrete domain used to denote vague quantification. A similar pattern is also present for 'ADJ *at* (NP)' with the SCI subset of ELF_Unedited having a remarkably lower proportion of adjectives from the abstract domain (see Table 8.27).

Table 8.27 Percentages of Semantic Domains for 'ADJ PREP (NP)' Cxs across "Corpus:Discipline" Subsets

Cx		ADJ *of* (NP)		ADJ *at* (NP)	
Domain		Abstract	Concrete	Abstract	Concrete
Examples		*capable, independent, aware*	*many, much*	*significant, possible, similar*	*available, present, online*
SCI	ELF_Unedited	100.00	0.00	52.63	47.37
	ELF_Edited	68.09	31.91	81.82	18.18
	ENL_Edited	52.29	47.71	73.85	26.15
SSH	ELF_Unedited	100.00	0.00	80.00	20.00
	ELF_Edited	67.50	32.50	79.31	20.69
	ENL_Edited	31.88	68.13	37.50	62.50

Note. Sum of percentages for semantic domains for each subset equals all Cx collocates included in semantic analysis.

Zooming in on the instances of the quantity-denoting constructions '*many of* (NP)' and '*much of* (NP)' shows that these are commonly used when making general statements that do not require precision (e.g. '*many of + previous studies/comments/accounts/examples*' in ENL_Edited). They signal a casual style (Biber et al., 1999, p. 276), and are used when the explicit mention of a numerical quantity is not deemed necessary (Cutting, 2012, p. 285). With 'ADJ *at* (NP),' the differences are more complex. SSH texts by ENL writers seem to employ the concrete domain more often, generally referring to the presence of an entity at a location (e.g., '*food-related activities available at the festival,'* [ENL_Edited_29, economics]). However, the higher proportions of the abstract domain in the SSH subsets of the ELF corpora denote not only qualifying meanings ('*adept at reducing the length'* [ENL_Unedited_10, Chinese, education], but also concrete ones (e.g. '*significant at the 0.05 level*' [ELF_Edited_16, Czech, economics].

Across the "similect:discipline" subsets, the two ENL subsets matching the Italian ELF data have the highest frequencies and most uniform distribution in 'ADJ *of* (NP).' In contrast, the SCI texts by Finnish and Chinese ELF writers and SSH texts by Czech ELF writers have some of the lowest frequency counts and most fluctuating distributions (see Table 8.28).

Table 8.28 Bootstrapped Median and DP_{norm} Values for 'ADJ PREP (NP)' Cxs across "Similect:Discipline" Subsets

Discipline	Similect	ADJ *of* (NP)		ADJ-SUPERL *of* (NP)		ADJ *at* (NP)	
		Median (normed per 10k)	DP_{norm}	Median (normed per 10k)	DP_{norm}	Median (normed per 10k)	DP_{norm}
SCI	CH	0.19	0.46	0.00	0.87	0.00	0.19
	CZ	0.30	0.23	0.00	0.52	0.08	0.30
	FI	0.15	0.37	0.13	0.47	0.00	0.15
	FR	0.34	0.34	0.11	0.48	0.00	0.34
	IT	0.36	0.44	0.00	0.59	0.18	0.36
	PO	0.23	0.51	0.00	0.51	0.00	0.23

	RU	0.30	0.37	0.26	0.37	0.22	0.30
	SP	0.35	0.33	0.33	0.40	0.00	0.35
	SW	0.73	0.34	0.00	0.92	0.00	0.73
	EN_CH	0.55	0.23	0.00	0.58	0.11	0.55
	EN_CZ	0.42	0.32	0.00	0.83	0.28	0.42
	EN_FI	0.27	0.37	0.00	0.65	0.14	0.27
	EN_FR	0.90	0.33	0.10	0.49	0.30	0.90
	EN_IT	1.55	0.15	0.00	0.56	0.31	1.55
	EN_PO	0.27	0.54	0.00	0.90	0.27	0.27
	EN_RU	0.64	0.37	0.00	0.67	0.21	0.64
	EN_SP	0.78	0.28	0.00	0.88	0.31	0.78
	EN_SW	0.82	0.29	0.00	0.74	0.27	0.82
SSH	**CH**	0.58	0.28	0.09	0.44	0.00	0.58
	CZ	0.22	0.45	0.00	0.57	0.00	0.22
	FI	0.71	0.26	0.09	0.53	0.00	0.71
	FR	0.83	0.33	0.23	0.47	0.00	0.83
	IT	1.12	0.38	0.00	0.82	0.00	1.12
	PO	1.11	0.31	0.56	0.31	0.00	1.11
	RU	0.84	0.28	0.36	0.43	0.11	0.84
	SP	0.94	0.09	0.30	0.47	0.30	0.94
	SW	0.61	0.35	0.12	0.48	0.12	0.61
	EN_CH	0.68	0.30	0.08	0.48	0.00	0.68
	EN_CZ	0.46	0.31	0.00	0.60	0.00	0.46
	EN_FI	0.91	0.23	0.00	0.59	0.15	0.91
	EN_FR	0.75	0.25	0.00	0.74	0.14	0.75
	EN_IT	1.51	0.30	0.00	0.63	0.38	1.51
	EN_PO	1.09	0.23	0.00	0.80	0.00	1.09
	EN_RU	1.14	0.32	0.14	0.45	0.00	1.14
	EN_SP	1.12	0.21	0.00	0.85	0.00	1.12
	EN_SW	0.77	0.34	0.00	0.85	0.13	0.77

Despite the differences observed in in terms of descriptive values, none of the three adjectival constructions were found to have significant pairwise differences across matching "similect:discipline" subsets.

Similar to the results for the "corpus:similect" interaction, both quality- and quantity-denoting collocates of 'ADJ *of* (NP)' are used somewhat interchangeably among ENL subsets whereas the quality orientation is clear from the proportions of collocational meanings among the ELF subsets. With the exceptions of texts by French and Portuguese ELF writers, SCI texts, where the majority of the occurrences are accumulated, seem to lean towards predominantly qualitative meanings regarding the use of "ADJ-SUPERL *of* (NP)" (see Table 8.29).

Table 8.29 Percentages of Semantic Domains for 'ADJ PREP (NP)' Cxs across "Similect:Discipline" Subsets

Cx		**ADJ *of* (NP)**		**ADJ *at* (NP)**	
Domain		**Abstract**	**Concrete**	**Abstract**	**Concrete**
Examples		*capable, independent, aware*	*many, much*	*significant, possible, similar*	*available, present, online*
SCI	**CH**	100.00	0.00	0.00	0.00
	CZ	78.57	21.43	100.00	0.00
	FI	80.00	20.00	71.43	28.57
	FR	50.00	50.00	37.50	62.50
	IT	0.00	100.00	100.00	0.00
	PO	100.00	0.00	50.00	50.00
	RU	54.55	45.45	50.00	50.00

	SP	66.67	33.33	75.00	25.00
	SW	100.00	0.00	75.00	25.00
	EN_CH	42.86	57.14	77.78	22.22
	EN_CZ	40.91	59.09	57.14	42.86
	EN_FI	25.00	75.00	91.67	8.33
	EN_FR	90.32	9.68	62.50	37.50
	EN_IT	83.33	16.67	100.00	0.00
	EN_PO	33.33	66.67	50.00	50.00
	EN_RU	7.69	92.31	75.00	25.00
	EN_SP	37.50	62.50	0.00	100.00
	EN_SW	83.33	16.67	100.00	0.00
SSH	**CH**	92.59	7.41	100.00	0.00
	CZ	91.67	8.33	100.00	0.00
	FI	83.33	16.67	0.00	100.00
	FR	83.33	16.67	75.00	25.00
	IT	62.50	37.50	100.00	0.00
	PO	82.35	17.65	75.00	25.00
	RU	30.77	69.23	50.00	50.00
	SP	100.00	0.00	50.00	50.00
	SW	85.71	14.29	0.00	100.00
	EN_CH	34.38	65.63	0.00	0.00
	EN_CZ	57.14	42.86	100.00	0.00
	EN_FI	11.54	88.46	16.67	83.33
	EN_FR	60.87	39.13	33.33	66.67
	EN_IT	12.50	87.50	50.00	50.00
	EN_PO	27.27	72.73	0.00	0.00
	EN_RU	27.27	72.73	0.00	0.00
	EN_SP	9.09	90.91	0.00	0.00
	EN_SW	17.65	82.35	100.00	0.00

Note. Sum of percentages for semantic domains for each subset equals all Cx collocates included in semantic analysis.

Similar to ‘*’in order to* ‘V’ where only one semantic domain was identified, ‘ADJ-SUPERL *of* (NP)’ is also separated from the other two adjectival constructions because only one domain was identified among the adjectives in the construction. Czech Swedish and Spanish ELF subsets seem to dominate the distribution of the domain frequencies in the “corpus:similect” interaction regardless of the editing status. Unlike the frequency counts for ‘ADJ-SUPERL *of* (NP)’ where the SSH texts in ELF_Edited have a slightly lower median than those in ELF_Unedited, the subset has a median of zero for the occurrence of the quantifying collocates strongly attracted to the constructions. This result can be attributed to the relatively higher fluctuation of the frequencies for the domain in the subset. The “similect:discipline” subsets also confirm the commonality of this domain in Spanish and Swedish regardless of the editing status of the texts (see Table 8.30).

Table 8.30 Distribution of 'ADJ-SUPERL of *(NP)' with Quantifying Meaning (‘*most of *[NP]’) across Subsets*

Corpus:Similect			**Corpus:Discipline**			**Similect:Discipline**		
	Unedited	**Edited**		**SCI**	**SSH**		**SCI**	**SSH**
Similect	**Median (normed per 10k)**	**Median (normed per 10k)**	**Corpus**	**Median (normed per 10k)**	**Median (normed per 10k)**	**Similect**	**Median (normed per 10k)**	**Median (normed per 10k)**
CH	0.00	0.00	**ELF_Unedited**	0.00	0.20	**CH**	0.00	0.00
CZ	0.00	0.00	**ELF_Edited**	0.00	0.17	**CZ**	0.00	0.15
FI	0.15	0.13	**ENL_Edited**	0.00	0.00	**FI**	0.13	0.09

FR	0.22	0.00		**FR**	0.00	0.23
IT	0.00	0.00		**IT**	0.00	0.00
PO	0.28	0.27		**PO**	0.00	0.56
RU	0.29	0.21		**RU**	0.22	0.25
SP	0.30	0.00		**SP**	0.16	0.30
SW	0.00	0.15		**SW**	0.00	0.12
EN_CH	NA	0.00		**EN_CH**	0.00	0.00
EN_CZ	NA	0.00		**EN_CZ**	0.00	0.00
EN_FI	NA	0.00		**EN_FI**	0.00	0.00
EN_FR	NA	0.00		**EN_FR**	0.10	0.00
EN_IT	NA	0.00		**EN_IT**	0.00	0.00
EN_PO	NA	0.00		**EN_PO**	0.00	0.00
EN_RU	NA	0.00		**EN_RU**	0.00	0.00
EN_SP	NA	0.00		**EN_SP**	0.00	0.00
EN_SW	NA	0.00		**EN_SW**	0.00	0.00

Overall, adjectival constructions analyzed in this section yielded limited differences across similectal and disciplinary groups. One significant difference was identified among two "corpus:similect" subsets, namely between Russian and Chinese edited ELF subsets for 'ADJ-SUPERL *of* (NP).' Besides, no significant differences were identified between "similect:discipline" subsets. Both of these findings are interesting as they point to a rather limited influence of the similectal background of ELF writers on the use of these three adjectival constructions.

A number of significant differences, however, were identified across "corpus:discipline" subsets for 'ADJ *of* (NP)' and 'ADJ *at* (NP),' for which the SCI frequencies in both edited and unedited ELF data differed significantly from that of ENL_Edited. Similar to '*over* N' and '*within* N,' SCI texts by both edited and unedited ELF corpora have much fewer occurrences of these two constructions than the SCI subset of ENL_Edited. ELF_Unedited in particular differs from ENL_Edited in that '*over* N' and '*within* N' are generally used to denote abstract meanings. Proportions of the semantic domain frequencies for the 'N PREP (NP)' construction with the prepositions '*of*' and '*for'* also similarly showed that even in SCI texts, ELF writers prefer nouns denoting abstract concepts and discursive connections more often than do the ENL writers. This might indicate that vagueness is a distinctive feature of hard sciences texts produced by ELF writers. Examining the use of abstract nouns in linguistics abstracts for two academic events, Cutting (2012) found that vagueness in the form of abstract nouns was a defining feature of abstracts which were based on incomplete research projects. Similarly, the fact that ELF_Unedited includes mostly unpublished texts is likely to imply that these texts did not go through the editing

processes that led to the higher precision in edited ELF and ENL writers' texts when they write about their research design and findings.

In addition, as explained earlier, a general preference for more conventional patterns in academic writing might be another factor related to the limited use of these two patterns. The much more conventional adjectival pattern also found commonly in academic writing, is 'adjective + noun' ('ADJ N'), which is much more frequent in ELF corpora (unedited: 723.34 per 10k, edited: 726.36 per 10k) than in ENL_Unedited (703.01 per 10k). The difference is even larger in SCI as ELF_Unedited and ELF_Edited have the relative (per 10k) frequencies of 367.34 and 268.59 respectively, as opposed to 244.47 in SCI of ENL_Edited for 'ADJ N.' Gray (2013) also similarly reported that the use of predicative adjectival structures is more common in more involved, soft sciences disciplines that focus on the scientific study of human behavior and cognition, as opposed to the more quantitative hard sciences texts that often make use of attributive adjectival structures with a heavy informational focus. Hence, similar to the findings for some of the 'PREP N' constructions, SCI writing points to greater ELF-ENL differences than SSH.

Lastly, as argued in the previous chapter when we interpreted the higher frequency of '*most of* (NP)' in ELF than in ENL, ELF writers' choice of '*most'* over '*much/many'* could be interpreted as having a boosting effect (Li & Wharton, 2012), highlighting a distinctive choice of stance construction by ELF writers as opposed to ENL writers.

*8.1.4 Differences in '(NP) V-***be** *V-***ed** *PREP (NP)' Constructions*

Regarding the disciplinary and similectal variation in the use of the three prepositional passive constructions, the three-way "corpus:similect:discipline," as well as the two-way interactions of corpus, similect and discipline were found to be significant in univariate comparisons, with the exception of the *in*-passive construction, '(NP) V-*be* V-*ed in* (NP),' across the "corpus:discipline" subsets. Since again no pairwise comparisons yielded significant differences across subsets for the three-way interactions, this

section concentrates on the results for the comparisons across subsets on the basis of the three two-way interactions.

For the comparisons of "corpus:similect" subsets, there seems to be a primarily frequency-related variation for *by*-passives and *in*-passives, while both frequencies and dispersions vary in the case of *on*-passives (see Table 8.31). The former has remarkably lower values compared to the other two passive constructions. Edited Italian and Swedish, as well as unedited Portuguese and Russian ELF subsets, have some of the highest frequencies of *by*-passives, with each having a normalized median frequency value of over three. In contrast, the ENL subset matching the Finnish ELF texts have the lowest as the only group with a value of below one. The highest frequencies for the *in*-passives come from the unedited Swedish and Italian ELF subsets. Both subsets have verb choices similar to those in ELF_Unedited in general. However, some verbs such as *'report, divide, find'* for the unedited Italian, and *'use, see, find'* for unedited Swedish make up a higher proportion of the verbs than in the whole corpus. The ELF-ENL difference is more complex here than in the *by*-passives as several ENL subsets (e.g., those matching Portuguese and Swedish ELF texts) have high frequency values comparable to those of their ELF counterparts. With respect to the *on*-passives, both Swedish ELF subsets and the unedited French ELF subset have the highest frequencies and lowest DP_{norm} values. Except for Swedish and Russian ELF, all ELF subsets have higher frequencies among the unedited texts than the edited ones.

*Table 8.31 Bootstrapped Median and DP_{norm} Values for '(NP) V-*be *V-*ed *PREP (NP)' Cxs across "Corpus:Similect" Subsets*

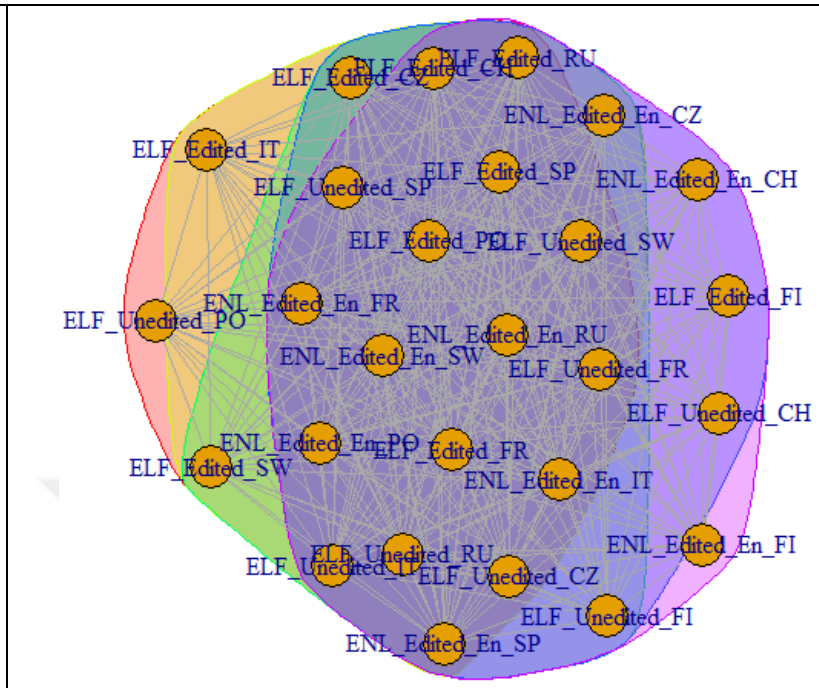
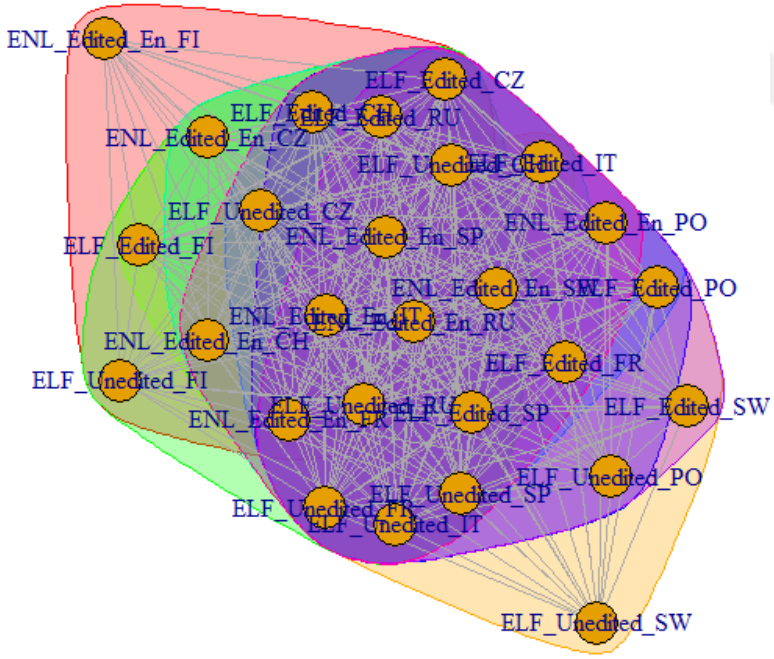
Similect	Corpus	(NP) V-*be* V-*ed by* (NP)		(NP) V-*be* V-ed *in* (NP)		(NP) V-*be* V-*ed on* (NP)	
		Median (normed per 10k)	DP_{norm}	Median (normed per 10k)	DP_{norm}	Median (normed per 10k)	DP_{norm}
CH	**Unedited**	1.18	0.27	1.67	0.18	0.47	0.38
	Edited	1.99	0.23	1.83	0.25	0.17	0.45
EN	**Edited**	1.10	0.24	1.97	0.19	0.30	0.30
CZ	**Unedited**	2.41	0.19	1.52	0.14	0.51	0.32
	Edited	2.62	0.20	1.50	0.20	0.37	0.34
EN	**Edited**	1.69	0.16	1.53	0.23	0.31	0.26
FI	**Unedited**	1.04	0.19	1.29	0.23	0.43	0.41
	Edited	1.16	0.19	1.23	0.24	0.31	0.32
EN	**Edited**	0.95	0.18	1.16	0.24	0.27	0.28
FR	**Unedited**	2.45	0.15	2.36	0.18	1.01	0.22
	Edited	2.67	0.19	3.15	0.19	0.46	0.35
EN	**Edited**	1.93	0.17	2.43	0.27	0.41	0.34
IT	**Unedited**	2.87	0.24	4.31	0.29	0.72	0.29
	Edited	3.40	0.21	3.17	0.26	0.57	0.27
EN	**Edited**	1.88	0.23	2.48	0.34	0.38	0.51
PO	**Unedited**	3.25	0.25	3.39	0.15	0.56	0.52

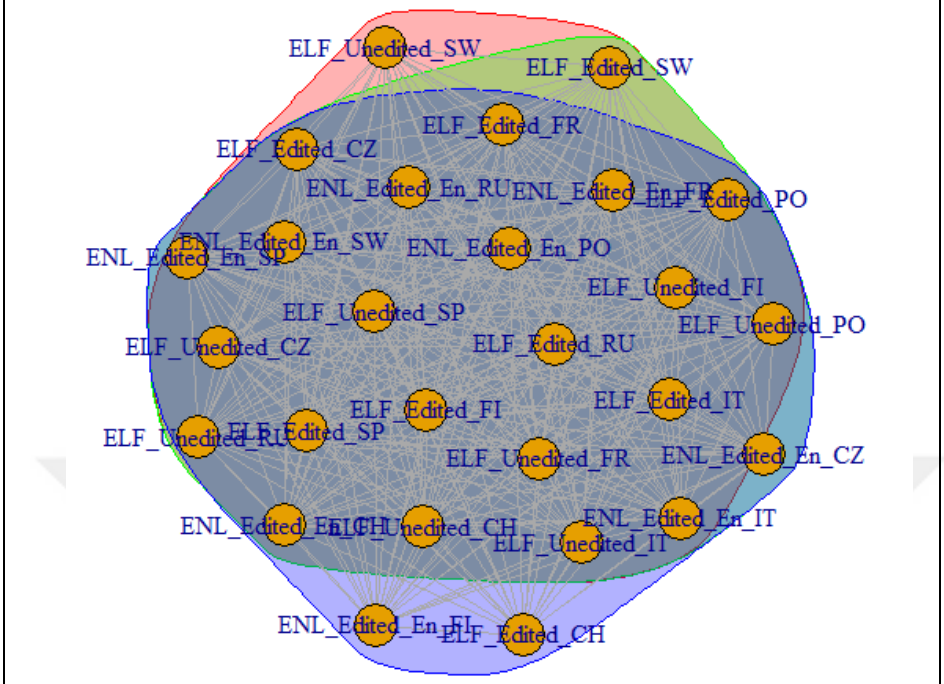
	Edited	1.87	0.21	3.71	0.20	0.37	0.27
EN	**Edited**	1.66	0.25	3.70	0.14	0.60	0.41
RU	**Unedited**	3.03	0.18	2.12	0.18	0.61	0.30
	Edited	1.97	0.15	1.97	0.19	0.63	0.39
EN	**Edited**	2.27	0.21	1.92	0.16	0.43	0.37
SP	**Unedited**	2.25	0.26	3.42	0.32	0.80	0.39
	Edited	1.58	0.24	3.55	0.23	0.32	0.37
EN	**Edited**	2.33	0.18	2.95	0.18	0.31	0.45
SW	**Unedited**	2.49	0.16	4.07	0.17	1.09	0.28
	Edited	3.07	0.16	3.49	0.14	1.20	0.19
EN	**Edited**	2.29	0.18	3.70	0.16	0.27	0.41

Similar to the majority of the results reported so far, pairwise statistical comparisons for these constructions point to differences across the subsets within the corpora instead of cross-corpus differences for the matching datasets. When clustered, the unedited Portuguese ELF subgroup (in the pink cluster) is separated from others in terms of the use of *by*-passives and is maximally different from the Chinese and Finnish texts in the same corpus (see Table 8.32). The frequencies of *in*-passives point to substantial variation within the three study corpora. It is interesting to observe that both edited and unedited Portuguese, Czech, and Finnish ELF subsets along with the matching ENL data for these similectal groups exhibit significant differences in all three corpora. This is also in line with the similarity of usage data among these matching groups in terms of frequency and dispersion. Lastly, the clustering of *on*-passives points to a more homogeneous distribution of the construction across subsets, with only the edited Chinese (in the blue cluster) and Swedish (in the green cluster) datasets being significantly different from one another.

Table 8.32 Clusters and P-Values of Pairwise Differences for '(NP) V-be V-ed PREP (NP)' Cxs across "Corpus:Similect" Subsets

Cx	**Clusters**	**Pairwise Differences**

(NP) V-*be* V-*ed by* (NP)		**- Within ELF_Unedited:** PO - CH (p = .037) PO - FI (p = .043)
(NP) V-*be* V-ed *in* (NP)		**- Within ELF_Unedited:** PO - CZ (p = .039) PO - FI (p = .011) SW - CH (p = .019) SW - CZ (p = .011) SW - FI (p = .003) **- Within ELF_Edited:** FR - FI (p = .047) PO - FI (p = .015) SW - CH (p = .044) SW - FI (p = .011) **- Within ENL_Edited:** PO - CZ (p = .047) PO - FI (p = .013) SW - FI (p = .043)



- Within ELF_Edited:

CH – SW (p = .027)

In terms of the proportions of semantic domains in the prepositional passive constructions across "corpus:similect" subsets, unedited Portuguese ELF and matching ENL texts, as well as both edited and unedited Russian ELF subsets have considerably higher proportions of discourse verbs, while all the other subsets have more process-oriented verbs among the instances of *by*-passives included in the semantic analysis (see Table 8.33). The proportions of verbs denoting research findings in the construction are generally lower than the other domains for all subsets. In the case of *in*-passives, there is an even greater degree of fluctuation across subsets in terms of the proportions for domain frequencies. However, the dominant semantic domains seem to be those including verbs that denote research processes and findings. Edited Chinese and unedited Portuguese ELF have higher proportions of process verbs, while edited Spanish ELF has the highest proportion of verbs denoting research findings. The ENL subset matching the Italian ELF texts has the largest proportion of discourse verbs. Further, ENL datasets matching Czech and Russian ELF are the only subsets with over nine percent of the verbs attracted by *in*-passives that refer to locations. Finally, for *on*-passives, in line with the overwhelming majority of instances which have '*base'* and '*focus'* as the main verb, discourse verbs make up the most substantial proportion for

most subsets. The ENL data matching the Swedish ELF texts, however, is unique in having a discourse verb in only 8.33% of the *on*-passives. This subset includes the majority of the construction with process verbs such as *'teachers were not rated on the ELL items'* (ENL_Edited_67, SSH, education)', and '*data for Ba2NF (1) were collected on the D1A instrument'* (ENL_Edited_75, SCI, chemical engineering)

Table 8.33 Percentages of Semantic Domains for '(NP) V-be V-ed PREP(NP)' Cxs across "Corpus:Similect" Subsets

Cx		**(NP) V-*be* V-*ed by* (NP)**			**(NP) V-*be* V-*ed in* (NP)**				**(NP) V-*be* V-*ed on* (NP)**		
Domain		**discourse**	**process**	**finding**	**discourse**	**finding**	**location**	**process**	**discourse**	**process**	**location**
Examples		*characterize, explain, affect*	*obtain, form measure*	*determine, report, show*	*present, give describe*	*find, show, observe*	*locate, place*	*use, apply, include*	*base focus*	*perform, conduct, test*	*place, locate*
CH	**Unedited**	21.84	65.52	12.64	30.14	21.92	5.48	42.47	40.00	60.00	0.00
	Edited	37.65	53.70	8.64	24.44	22.22	1.11	52.22	66.67	20.00	13.33
EN	**Edited**	41.58	47.37	11.05	31.13	27.36	0.94	40.57	68.57	28.57	2.86
CZ	**Unedited**	46.99	47.59	5.42	15.63	32.81	7.81	43.75	82.14	17.86	0.00
	Edited	51.03	41.38	7.59	19.49	48.31	0.00	32.20	93.33	6.67	0.00
EN	**Edited**	41.30	49.28	9.42	25.00	30.68	9.09	35.23	53.33	36.67	10.00
FI	**Unedited**	45.27	51.35	3.38	22.22	36.36	1.01	40.40	82.22	15.56	2.22
	Edited	48.33	48.33	3.33	27.01	26.28	2.19	44.53	87.50	12.50	0.00
EN	**Edited**	42.45	47.48	10.07	25.24	33.98	2.91	37.86	73.53	20.59	5.88
FR	**Unedited**	48.46	47.69	3.85	26.53	35.71	4.08	33.67	47.73	38.64	13.64
	Edited	42.96	50.70	6.34	32.00	36.00	7.00	25.00	65.38	30.77	3.85
EN	**Edited**	33.11	57.62	9.27	26.61	36.70	0.92	35.78	50.00	47.50	2.50
IT	**Unedited**	37.84	51.35	10.81	14.52	41.94	1.61	41.94	80.00	20.00	0.00
	Edited	47.06	47.06	5.88	27.14	44.29	0.00	28.57	92.86	7.14	0.00
EN	**Edited**	29.09	61.82	9.09	45.00	30.00	0.00	25.00	66.67	22.22	11.11
PO	**Unedited**	54.21	40.19	5.61	24.44	17.78	0.00	57.78	83.33	16.67	0.00
	Edited	41.07	48.21	10.71	24.56	36.84	3.51	35.09	100.00	0.00	0.00
EN	**Edited**	48.08	42.31	9.62	24.64	26.09	0.00	49.28	60.00	26.67	13.33
RU	**Unedited**	52.75	36.26	10.99	39.13	28.26	0.00	32.61	84.62	15.38	0.00
	Edited	54.37	39.81	5.83	39.13	23.91	0.00	36.96	77.27	18.18	4.55
EN	**Edited**	47.19	50.56	2.25	12.90	35.48	9.68	41.94	68.75	25.00	6.25
SP	**Unedited**	49.23	43.08	7.69	13.43	47.76	2.99	35.82	82.86	14.29	2.86
	Edited	49.06	39.62	11.32	26.09	50.72	0.00	23.19	57.14	42.86	0.00
EN	**Edited**	33.73	53.01	13.25	22.22	31.11	2.22	44.44	77.78	22.22	0.00
SW	**Unedited**	40.74	51.85	7.41	30.67	24.00	6.67	38.67	63.64	30.30	6.06
	Edited	39.22	52.94	7.84	24.00	42.00	4.00	30.00	65.12	25.58	9.30
EN	**Edited**	41.94	51.61	6.45	29.21	33.71	8.99	28.09	8.33	83.33	8.33

Note. Sum of percentages for semantic domains for each subset equals all Cx collocates included in semantic analysis.

As we can see in Table 8.34, the distributional and dispersional characteristics of "corpus:discipline" subsets do not appear to differ as drastically as in the case of "corpus:similect" datasets, with ENL subsets generally having lower frequencies. However, the SCI-SSH differences within the ELF corpora are minimal when SSH texts in ENL_Edited have visibly lower frequencies than their SCI counterparts, thus also leading to a slightly larger ELF-ENL difference in the use of *by*- and *in*-passives. However, no relevant significant differences across disciplines and corpora exist between pairs of subsets for these constructions.

*Table 8.34 Bootstrapped Median and DP_{norm} Values for '(NP) V-*be *V-*ed *PREP (NP)' Cxs across "Corpus:Discipline" Subsets*

Discipline	Corpus	(NP) V-*be* V-*ed by* (NP)		(NP) V-*be* V-*ed in* (NP)		(NP) V-*be* V-*ed on* (NP)	
		Median (normed per 10k)	DP_{norm}	Median (normed per 10k)	DP_{norm}	Median (normed per 10k)	DP_{norm}
SCI	ELF_Unedited	2.19	0.21	2.23	0.23	0.61	0.35
	ELF_Edited	2.07	0.20	2.22	0.22	0.31	0.37
	ENL_Edited	1.83	0.16	2.09	0.21	0.28	0.30
SSH	ELF_Unedited	2.00	0.21	2.13	0.16	0.60	0.33
	ELF_Edited	1.99	0.21	2.42	0.21	0.49	0.33
	ENL_Edited	1.51	0.19	1.68	0.23	0.30	0.37

Proportions of semantic domains for *by*-passives across the "corpus:similect" subsets show that process verbs make up the largest proportion of verbs used in *by*-passives among SCI subsets. For SSH subsets, however, both ELF groups have larger proportions of discourse verbs, such as *'characterize'* and *'explain,'* whereas the SSH texts in ENL_Edited still have a higher proportion of process verbs. In the case of *in*-passives, finding and process verbs generally have greater proportions, with the exception of SCI texts in ENL_Edited that has comparable frequencies of finding and discourse verbs. Lastly, the proportions of semantic domains for verbs used in *on*-passives point to a general tendency for the use of discourse verbs in this constructions. This tendency is even more pronounced in the case of the SSH texts in ELF_Unedited and both disciplinary subsets in ELF_Edited (see Table 8.35).

Table 8.35 Percentages of Semantic Domains for '(NP) V-be V-ed PREP(NP)' Cxs across "Corpus:Discipline" Subsets

Cx		(NP) V-*be* V-*ed by* (NP)			(NP) V-*be* V-*ed in* (NP)				(NP) V-*be* V-*ed on* (NP)		
Domain		Discourse	Process	Finding	Discourse	Finding	Location	Process	Discourse	Process	Location
Examples		*characterize, explain, affect*	*obtain, form measure*	*determine, report, show*	*present, give describe*	*find, show, observe*	*locate, place*	*use, apply, include*	*base focus*	*perform, conduct, test*	*place, locate*
SCI	ELF_Unedited	38.92	52.36	8.73	24.41	33.89	3.79	37.91	60.29	36.76	2.94
	ELF_Edited	42.31	49.73	7.96	27.54	40.04	2.54	29.87	77.50	20.00	2.50
	ENL_Edited	36.65	51.43	11.92	30.05	31.25	4.81	33.89	62.00	28.00	10.00
SSH	ELF_Unedited	55.01	41.26	3.72	22.71	29.95	2.90	44.44	82.79	12.30	4.92
	ELF_Edited	52.98	41.74	5.28	24.44	31.75	1.59	42.22	78.91	17.19	3.91
	ENL_Edited	46.48	49.85	3.67	22.73	32.58	1.89	42.80	57.00	41.00	2.00

Note. Sum of percentages for semantic domains for each subset equals all Cx collocates included in semantic analysis.

Lastly, examining the "similect:discipline" subsets enables more in-depth insights into the intersection of disciplinary and similectal factors that are not affected by editing status of ELF texts in the use of prepositional passive constructions. Except for Finnish, Italian, and Russian ELF and the corresponding ENL subsets, normed median frequencies are higher in SCI subsets regarding the use of *by*-passives. As is observed across "corpus:similect" subsets, the usage frequencies for *in*-passives vary greatly across both ELF and ENL subsets, especially among SCI texts. For instance, Spanish, Italian, and

Swedish ELF texts and the ENL texts corresponding to the Swedish ELF texts in SCI have median values that are more than twice as high as some subsets in both SCI (e.g., Finnish ELF) and SSH (e.g., ENL subset matching Finnish ELF texts). Similarly, the SCI subset with Swedish ELF texts has the highest median value for *on*-passives. At the same time, the ENL subset matching SCI texts in Italian ELF has a median of zero and also the highest DP_{norm} value, but of which indicates the infrequency of the construction in the subset (see Table 8.36).

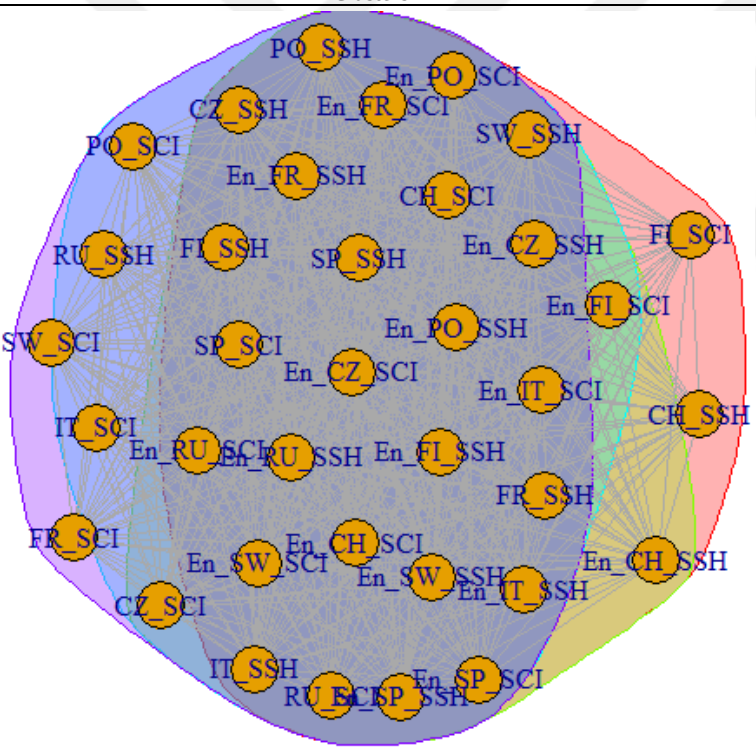
*Table 8.36 Bootstrapped Median and DP_{norm} Values for '(NP) V-*be *V-ed PREP (NP)' Cxs across "Similect:Discipline" Subsets*

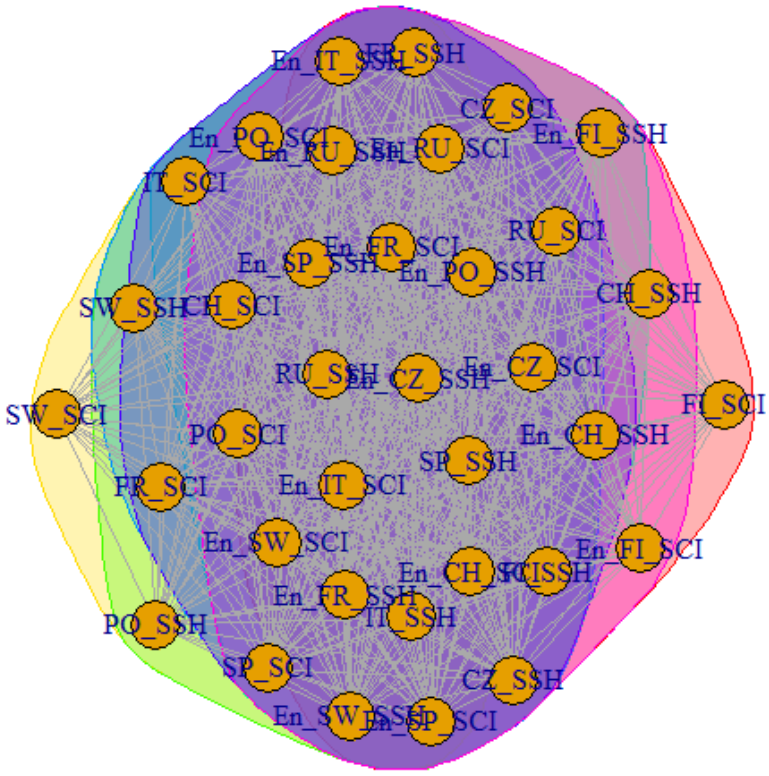
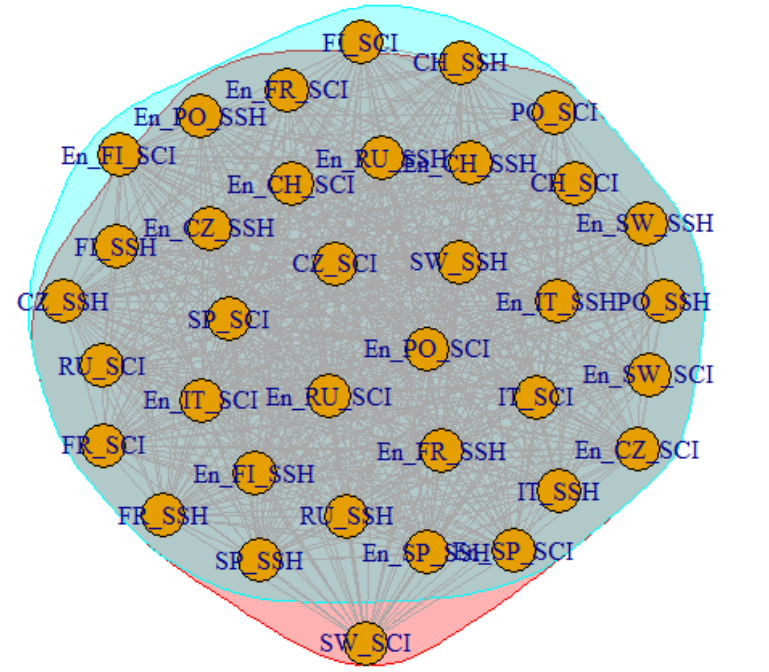
Discipline	Similect	(NP) V-*be* V-*ed by* (NP)		(NP) V-*be* V-*ed in* (NP)		(NP) V-*be* V-*ed on* (NP)	
		Median (normed per 10k)	DP_{norm}	Median (normed per 10k)	DP_{norm}	Median (normed per 10k)	DP_{norm}
SCI	CH	1.95	0.17	2.10	0.16	0.31	0.40
	CZ	2.85	0.18	1.54	0.20	0.30	0.29
	FI	0.99	0.18	1.21	0.25	0.21	0.44
	FR	2.92	0.16	3.31	0.13	0.89	0.33
	IT	2.87	0.20	4.85	0.25	0.70	0.27
	PO	2.79	0.26	3.71	0.20	0.51	0.43
	RU	2.05	0.20	2.05	0.16	0.66	0.34
	SP	2.13	0.24	4.91	0.24	0.53	0.37
	SW	2.95	0.13	4.21	0.21	1.68	0.21
	EN_CH	1.54	0.20	2.09	0.15	0.22	0.25
	EN_CZ	1.76	0.10	1.41	0.23	0.35	0.21
	EN_FI	0.95	0.13	1.22	0.25	0.27	0.27
	EN_FR	2.31	0.18	2.71	0.24	0.50	0.22
	EN_IT	1.86	0.21	2.95	0.33	0.00	0.66
	EN_PO	2.04	0.25	3.94	0.10	0.95	0.35
	EN_RU	1.92	0.22	1.50	0.20	0.43	0.34
	EN_SP	2.95	0.12	3.26	0.13	0.31	0.48
	EN_SW	3.14	0.08	4.78	0.17	0.41	0.38
SSH	CH	0.95	0.25	1.33	0.19	0.17	0.42
	CZ	2.49	0.16	1.52	0.14	0.51	0.30
	FI	1.53	0.21	1.42	0.22	0.61	0.24
	FR	1.94	0.16	2.27	0.17	0.71	0.26
	IT	2.99	0.23	2.54	0.22	0.66	0.29
	PO	2.21	0.26	3.46	0.16	0.28	0.39
	RU	2.55	0.16	2.10	0.23	0.43	0.35
	SP	1.93	0.27	3.06	0.28	0.46	0.37
	SW	2.55	0.21	3.15	0.11	0.75	0.23
	EN_CH	0.91	0.22	1.21	0.18	0.38	0.31
	EN_CZ	1.37	0.18	1.53	0.24	0.31	0.27
	EN_FI	1.22	0.17	1.07	0.20	0.30	0.27
	EN_FR	1.77	0.18	1.91	0.29	0.34	0.42
	EN_IT	1.88	0.19	1.13	0.30	0.38	0.27
	EN_PO	1.57	0.29	3.50	0.22	0.24	0.53
	EN_RU	2.27	0.20	2.69	0.24	0.29	0.46
	EN_SP	1.82	0.25	1.96	0.24	0.28	0.48
	EN_SW	1.91	0.11	2.54	0.15	0.13	0.50

As seen in the pink cluster in Table 8.37, Chinese SSH and Finnish SCI texts are separated from the others due to the infrequency of the *by*-passives in these subsets. The distinctly low frequency of *in*-passives in the SCI subset by Finnish ELF writers (in the pink cluster), along with the particularly high

frequency of the construction in SCI texts by Swedish ELF writers and SSH texts (in the yellow cluster), and by Portuguese ELF writers (in the green cluster) lead to these subsets being clustered separately from the others. Similarly, regarding *on*-passives, the highest frequency of the construction among SCI texts by Swedish ELF writers (in the pink cluster), and the lowest among SCI texts by Finnish ELF writers (in the cyan cluster) are clustered separately at two distant ends of the clustering plot. These complex pairwise differences seem to mirror the distribution of median frequencies that vary to a greater extent in SCI than in SSH subsets within ELF data.

*Table 8.37 Clusters and P-Values of Pairwise Differences for '(NP) V-*be *V-*ed *PREP (NP)' Cxs across "Similect:Discipline" Subsets*

Cx	Clusters	Pairwise Differences
(NP) V-*be* V-*ed by* (NP)		**- Within SCI:** CH - CZ (p = .020) FI - CZ (p = .020) FI - FR (p = .019) FI - IT (p = .020) FI - PO (p = .020) FI - SW (p = .022) **- Within SSH:** CH – RU (p = .033)

(NP) V-*be* V-*ed in* (NP)		**- Within SCI:** FI - CH (p = .048) FI - FR (p = .001) FI - IT (p = .016) FI - PO (p = .021) FI - SP (p = .036) FI - SW (p = .001) SW - CZ (p = .021) SW - RU (p = .028) **- Within SSH:** PO - CH (p = .013) PO - CZ (p = .021) SW - CH (p = .016)
(NP) V-*be* V-ed *on* (NP)		**- Within SCI:** SW - CH (p = .022) SW – CZ (p = .022) SW - FI (p = .021)

An analysis of semantic domains for *by*-passives across "similect:discipline" subsets points to the common use of verbs denoting discourse and process, proportions of which vary across the groups.

Despite this variation, the matching ENL subsets have a parallel ratio for these two domains with the exception of the Czech ELF subset having a higher proportion of discourse than process verbs, unlike the matching ENL texts. However, primarily due to the differences in the proportions among ELF subsets, the ELF-ENL discrepancy grows larger among SSH subsets, where only Portuguese and Spanish ELF texts have proportions of discourse and process verbs similar to those found for ENL texts. One of the most remarkable differences is that, while almost every subset has higher proportions of discourse verbs in SCI, the difference is particularly substantial for subsets such as those including texts by Portuguese and Chinese ELF writers. As for the *in*-passive, proportions of the domains seem to be relatively more diverse among the SCI texts for both ELF and ELF subsets. However, among ELF subsets, excluding Spanish ELF, process verbs make up the highest proportion of use in SSH subsets. Besides, SSH texts in ELF subsets seem to have higher proportions of discourse verb use in *on*-passives with a generally smaller proportion of verbs denoting location (see Table 8.38).

Table 8.38 Percentages of Semantic Domains for '(NP) V-be V-ed PREP(NP)' Cxs across "Similect:Discipline" Subsets

Cx		**(NP) V-be V-ed *by* (NP)**			**(NP) V-be V-ed *in* (NP)**				**(NP) V-be V-ed *on* (NP)**		
Domain		**Discourse**	**Process**	**Finding**	**Discourse**	**Finding**	**Location**	**Process**	**Discourse**	**Process**	**Location**
Examples		*characterize, explain, affect*	*obtain, form measure*	*determine, report, show*	*present, give describe*	*find, show, observe*	*locate, place*	*use, apply, include*	*base focus*	*perform, conduct, test*	*place, locate*
SCI	**CH**	23.76	65.19	11.05	30.28	21.10	3.67	44.95	28.57	4.76	66.67
	CZ	48.60	43.93	7.48	14.68	46.79	4.59	33.94	91.30	0.00	8.70
	FI	47.62	49.05	3.33	27.39	35.03	1.27	36.31	86.36	0.00	13.64
	FR	36.65	56.52	6.83	28.93	38.84	8.26	23.97	41.03	7.69	51.28
	IT	37.84	51.35	10.81	17.89	47.37	1.05	33.68	72.73	0.00	27.27
	PO	28.38	58.11	13.51	18.18	29.55	0.00	52.27	75.00	0.00	25.00
	RU	52.89	38.84	8.26	48.28	20.69	0.00	31.03	83.87	3.23	12.90
	SP	44.29	47.14	8.57	17.02	48.94	2.13	31.91	72.50	2.50	25.00
	SW	33.33	53.85	12.82	34.58	37.38	3.74	24.30	64.10	2.56	33.33
	EN_CH	40.52	47.06	12.42	37.10	17.74	0.00	45.16	53.33	6.67	40.00
	EN_CZ	38.30	48.94	12.77	27.87	26.23	13.11	32.79	61.90	14.29	23.81
	EN_FI	37.89	47.37	14.74	26.03	36.99	4.11	32.88	72.22	5.56	22.22
	EN_FR	23.68	64.47	11.84	20.83	41.67	2.08	35.42	75.00	8.33	16.67
	EN_IT	25.00	63.89	11.11	45.45	33.33	0.00	21.21	50.00	16.67	33.33
	EN_PO	43.75	43.75	12.50	21.05	34.21	0.00	44.74	66.67	11.11	22.22
	EN_RU	53.73	43.28	2.99	16.67	33.33	8.33	41.67	70.00	10.00	20.00
	EN_SP	18.18	61.36	20.45	31.82	22.73	4.55	40.91	100.00	0.00	0.00
	EN_SW	36.36	54.55	9.09	40.00	34.55	9.09	16.36	0.00	16.67	83.33
SSH	**CH**	54.41	38.24	7.35	20.37	24.07	1.85	53.70	85.71	7.14	7.14
	CZ	51.85	42.96	5.19	23.29	36.99	0.00	39.73	88.00	0.00	12.00
	FI	45.76	50.85	3.39	20.25	21.52	2.53	55.70	84.21	1.75	14.04
	FR	58.56	38.74	2.70	29.87	31.17	1.30	37.66	70.97	12.90	16.13
	IT	52.31	44.62	3.08	29.73	32.43	0.00	37.84	94.44	0.00	5.56
	PO	67.42	30.34	2.25	29.31	27.59	3.45	39.66	100.00	0.00	0.00
	RU	54.79	36.99	8.22	23.53	35.29	0.00	41.18	76.47	0.00	23.53
	SP	56.25	33.33	10.42	26.19	50.00	0.00	23.81	75.00	0.00	25.00
	SW	46.15	51.28	2.56	14.71	29.41	7.35	48.53	64.86	13.51	21.62
	EN_CH	45.95	48.65	5.41	22.73	40.91	2.27	34.09	80.00	0.00	20.00
	EN_CZ	47.73	50.00	2.27	18.52	40.74	0.00	40.74	33.33	0.00	66.67

EN_FI	52.27	47.73	0.00	23.33	26.67	0.00	50.00	75.00	6.25	18.75
EN_FR	42.67	50.67	6.67	31.15	32.79	0.00	36.07	39.29	0.00	60.71
EN_IT	36.84	57.89	5.26	42.86	14.29	0.00	42.86	100.00	0.00	0.00
EN_PO	55.00	40.00	5.00	29.03	16.13	0.00	54.84	50.00	16.67	33.33
EN_RU	27.27	72.73	0.00	0.00	42.86	14.29	42.86	66.67	0.00	33.33
EN_SP	51.28	43.59	5.13	13.04	39.13	0.00	47.83	66.67	0.00	33.33
EN_SW	55.56	44.44	0.00	11.76	32.35	8.82	47.06	16.67	0.00	83.33

Note. Sum of percentages for semantic domains for each subset equals all Cx collocates included in semantic analysis.

The comparisons of prepositional passives in terms of disciplinary and similectal groups within and across the corpora yielded important results highlighting considerable variation in the construction frequencies among "corpus:similect" and "similect:discipline" subsets. However, none of the significant pairwise test results were identified between ELF and ENL subsets. Instead, in a fashion similar to the findings for nominal constructions, the interaction of similectal background with the factors of corpora and disciplinary group uncovered complex patterns of variation among ELF subsets.

These passive constructions are also distinct from the others in the study because none were found to differ significantly across the "corpus:discipline" subsets. Variation in the use of these constructions across ELF subsets was captured primarily in the two-way interactions including the similectal background of the writers. The observed limited role of discipline without similect is partially in line with Hundt, Schneider, and Seoane's (2016) finding that passive use in New Englishes varies across varieties much more remarkably than it does across disciplines. However, a larger number of significant pairwise differences across "similect:discipline" subsets exist in SCI writing for all three constructions. On the one hand, these differences seem contradictory, in a way, to the earlier claim in support of Hundt et al. (2016) on the limited role of disciplines in passive usage, as hard sciences writing was found to be related to a higher degree of variation than in SSH writing across ELF subsets. On the other, Hyland (2008b) also reported that SCI writing has higher frequencies of passives than in SSH writing. Since a clear absence or presence of an SCI – SSH divide was not present in the current study, it is fair to conclude that the usage of prepositional passive constructions in written academic ELF is influenced by the lingua franca factor (Firth, 2009) "complexity" inherent in ELF.

The link between editing status of ELF texts and similectal backgrounds of their writers further complicates the findings but also points to important implications. First, the limited variation found only

among three unedited ELF subsets for *by*-passives (Portuguese, Chinese, and Finnish) could be attributed to the established usage of the construction in academic writing. The *on*-passive, usage of which differed significantly between only two edited subsets (Swedish, Czech), is also the least lexically diverse construction among the three. As was illustrated in Chapter 7, the majority of the verbs used in *on*-passives were identified as the two discourse verbs *'base,' and 'focus'.* Thus, both its high conventionality and fixedness are probable reasons for limited variation in its use. Besides, *in*-passives have the highest degree of variation in ELF, and the only significant differences for ENL subsets, most of which match the ELF subsets with varied usage of the construction. Along with being highly frequent, *'in'* is a highly polysemous preposition, which is also showcased in long lists of its senses (e.g., Sinclair, 1996; Tyler & Evans, 2003). Similarly, compared to the three domains identified for the other two prepositional passives, four domains were identified for *in*-passives, with three of these having comparable proportions to one another. As frequency and semantic diversity are shown to go hand in hand (Baayen, Milin, & Ramscar, 2016) and polysemy is shown to increase with second language proficiency (Crossley et al., 2010), the variation similarly found in both ELF and matching ENL subsets can be considered an expected outcome.

Lastly, the findings concerning the interactions involving different similectal backgrounds have interesting similarities to and differences from previous research. For instance, Chinese ELF, especially in SSH, has one of the lower-frequency subsets of passives, which is in line with Chen and Baker's (2010) findings on Chinese L1 student writing, but contradicts those of Pan et al. (2016) that reported higher frequencies in telecommunications texts by L1 Chinese compared to texts by ENL scholars. Besides, despite not differing from matching ENL subsets, Swedish ELF texts in the edited and SCI subsets have some of the higher construction frequencies. This finding is somewhat in contrast with Ädel and Erman (2012) who reported more common use of passives with more varied and complex structures by ENL academic writers than advanced L1 Swedish students writers. As the writers of the texts in the study corpora have higher academic writing expertise, this is a plausible difference.

Low frequencies in the Finnish subsets and high frequencies in the Portuguese subsets could also be related to the linguistic or cultural backgrounds of the writers. The limited use of prepositional passives in Finnish ELF texts might be linked to the different ways in which L1 Finnish users could express an impersonal stance, and more importantly, to the fact that "the Finnish passive allows virtually no expression of agent" (Mauranen, 1993, p. 13). The high frequencies of *by*- and *in*-passives, especially in SCI texts by Portuguese ELF writers are likely to be an epistemological choice to highlight the objective and empirical nature of the texts. This choice may have been motivated by the writers' perceived need to align more closely with a modernist perception of Anglo-American scientific discourse, as opposed to the L1 Portuguese discourse traditionally known to be more personally involved (Bennett, 2010, 2014).

8.1.5 Differences in the 'DET **of** *(NP)' Construction*

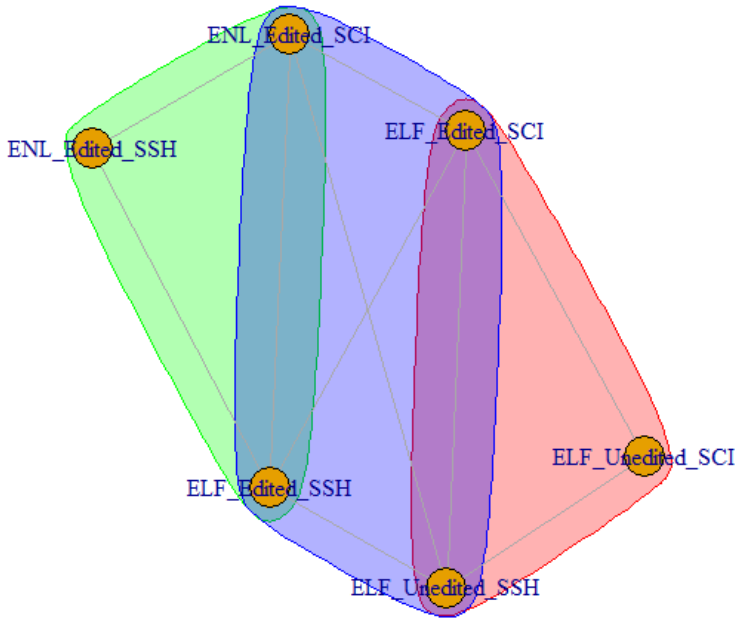
The univariate comparisons across similectal and disciplinary subsets within and across corpora for the use of 'DET *of* (NP)' led to significant results for all the interactions and main effects. However, relevant significant differences were only identified for the subsets as a result of the two-way "corpus:discipline" interaction. Nevertheless, in order to ensure comprehensive coverage of all the constructions with significant univariate results, all two-way interactions are covered in this section.

Among the "corpus:discipline" subsets, only few ENL subsets such as those matching Portuguese, Russian, and Swedish ELF texts have higher frequencies than the rest of the subsets (see Table 8.39). Most edited ELF subsets have slightly higher median frequencies and more even dispersions than their unedited counterparts. Further, all ELF and ENL "corpus:discipline" subsets point to higher usage frequencies in SSH than in SCI writing. The only two significant differences were identified for the "corpus:discipline" interactions of ELF_Unedited and ENL_Edited, where both disciplinary subsets have remarkably higher frequencies of the construction than the matching subsets of ELF_Unedited (see Table 8.40).

Table 8.39 Bootstrapped Median and DPnorm Values for 'DET of *(NP)' Cxs across "Corpus:Similect" Subsets*

Corpus:Similect					Corpus:Discipline					Similect:Discipline				
	Unedited		Edited			SCI		SSH			SCI		SSH	
Similect	Median (normed per 10k)	DP_{norm}	Median (normed per 10k)	DP_{norm}	Corpus	Median (normed per 10k)	DP_{norm}	Median (normed per 10k)	DP_{norm}	Similect	Median (normed per 10k)	DP_{norm}	Median (normed per 10k)	DP_{norm}
CH	0.24	0.45	0.33	0.34	**ELF_Unedited**	0.30	0.38	0.60	0.30	**CH**	0.37	0.33	0.21	0.36
CZ	0.25	0.41	0.44	0.33	**ELF_Edited**	0.47	0.31	0.56	0.31	**CZ**	0.33	0.37	0.31	0.33
FI	0.20	0.35	0.39	0.27	**ENL_Edited**	0.64	0.28	0.84	0.24	**FI**	0.30	0.33	0.47	0.30
FR	0.34	0.33	0.74	0.29						**FR**	0.57	0.36	0.55	0.29
IT	0.72	0.40	0.85	0.33						**IT**	0.00	0.69	0.85	0.26
PO	0.57	0.36	0.68	0.28						**PO**	0.46	0.35	0.83	0.31
RU	0.86	0.15	0.84	0.30						**RU**	0.91	0.25	0.63	0.14
SP	0.70	0.32	0.95	0.33						**SP**	0.70	0.25	0.78	0.37
SW	0.90	0.30	0.91	0.36						**SW**	0.73	0.47	0.90	0.26
EN_CH	-	-	0.45	0.29						**EN_CH**	0.44	0.30	0.69	0.21
EN_CZ	-	-	0.61	0.26						**EN_CZ**	0.71	0.32	0.61	0.23
EN_FI	-	-	0.48	0.28						**EN_FI**	0.41	0.29	0.61	0.17
EN_FR	-	-	1.00	0.18						**EN_FR**	1.00	0.22	1.16	0.17
EN_IT	-	-	0.62	0.28						**EN_IT**	0.78	0.26	0.38	0.35
EN_PO	-	-	1.69	0.25						**EN_PO**	1.22	0.44	1.69	0.13
EN_RU	-	-	1.50	0.22						**EN_RU**	0.86	0.26	1.99	0.18
EN_SP	-	-	0.73	0.30						**EN_SP**	0.93	0.37	0.70	0.27
EN_SW	-	-	1.19	0.25						**EN_SW**	1.09	0.33	1.27	0.19

Table 8.40 Clusters and P-Values of Pairwise Differences for 'DET of *(NP)' across "Corpus:Discipline" Subsets*

Clusters	Pairwise Differences
	**- ELF vs. ENL** ELF_Unedited: SCI - ENL_Edited: SCI (*p* = .008) ELF_Unedited: SSH – ENL_Edited: SSH (*p* = .016)

As discussed in detail in the previous chapter, only two lexical fillers (*'that,' 'some')*, met the association strength criterion in ELF_Unedited. In contrast, five fillers '*each, some, all, those, that*'. met the criterion

in both ELF_Edited and in ENL_Edited. Hence, the significant differences between the subsets of ELF_Unedited and the other corpora are in line with the limited construction frequency in the former.

As for the proportions of semantic domains, it is clear that the general tendency in most subsets is to use the quantifying 'DET *of* (NP)' constructions instead of the qualifying (referential) ones. Several exceptions with opposite tendencies include ELF subsets among the "corpus:similect" and "similect, discipline" groups, including unedited and edited Chinese ELF, SCI, and SSH texts in Chinese ELF, and SCI texts in French ELF. That the Chinese subsets consistently have higher proportions of the abstract domain is largely due to the limited use of the quantifiers '*each, some, all'* by these writers, both in this specific construction and in general (see Table 8.41).

Table 8.41 Percentages of Semantic Domains for 'DET of *(NP)' Cxs across Subsets*

Domain: Quality (referential)			**Examples:** *that, those*		**Domain: Quantity**					**Examples:** *some, all, each*				
Corpus:Similect					**Corpus:Discipline**					**Similect:Discipline**				
Similect	**Unedited**		**Edited**		**Corpus**	**SCI**		**SSH**		**Similect**	**SCI**		**SSH**	
	Abstract	**Concrete**	**Abstract**	**Concrete**		**Abstract**	**Concrete**	**Abstract**	**Concrete**		**Abstract**	**Concrete**	**Abstract**	**Concrete**
CH	70.59	29.41	70.73	29.27	**ELF_Unedited**	31.82	68.18	17.81	82.19	**CH**	75.68	24.32	61.90	38.10
CZ	4.35	95.65	21.28	78.72	**ELF_Edited**	40.09	59.91	28.50	71.50	**CZ**	22.73	77.27	3.85	96.15
FI	21.05	78.95	42.03	57.97	**ENL_Edited**	24.55	75.45	22.90	77.10	**FI**	39.62	60.38	34.29	65.71
FR	36.36	63.64	47.17	52.83						**FR**	64.71	35.29	23.33	76.67
IT	20.00	80.00	10.26	89.74						**IT**	25.00	75.00	8.33	91.67
PO	11.11	88.89	35.71	64.29						**PO**	33.33	66.67	22.58	77.42
RU	12.50	87.50	19.23	80.77						**RU**	15.00	85.00	25.00	75.00
SP	26.32	73.68	28.89	71.11						**SP**	12.00	88.00	38.46	61.54
SW	21.05	78.95	32.56	67.44						**SW**	40.74	59.26	20.00	80.00
EN_CH	-	-	13.48	86.52						**EN_CH**	18.18	81.82	8.89	91.11
EN_CZ	-	-	32.47	67.53						**EN_CZ**	26.67	73.33	40.63	59.38
EN_FI	-	-	20.00	80.00						**EN_FI**	18.31	81.69	23.08	76.92
EN_FR	-	-	32.00	68.00						**EN_FR**	41.86	58.14	24.56	75.44
EN_IT	-	-	28.95	71.05						**EN_IT**	43.48	56.52	6.67	93.33
EN_PO	-	-	14.29	85.71						**EN_PO**	12.50	87.50	16.22	83.78
EN_RU	-	-	28.57	71.43						**EN_RU**	15.38	84.62	37.84	62.16
EN_SP	-	-	17.14	82.86						**EN_SP**	7.14	92.86	23.81	76.19
EN_SW	-	-	29.41	70.59						**EN_SW**	41.67	58.33	18.52	81.48

Note. Sum of percentages for semantic domains for each subset equals all Cx collocates included in semantic analysis.

The analysis of the use of 'DET *of* (NP)' concerning the disciplinary and similectal grouping does not point to any further differences within and across corpora in terms of similectal backgrounds. However, the significant differences between both disciplinary subsets of ELF_Unedited and ENL_Edited highlight that the difference between these two corpora is more prominent than the one between ELF_Edited and ENL_Edited identified in the previous chapter. As the quantifying domain has a much larger frequency and proportion among the SSH texts in ENL_Edited in particular, the results are quite similar to those of 'ADJ *of* (NP)' where the same pattern was observed in the use of the quantifying

construction '*many/much of* (NP).' These quantification devices seem to be common in published social sciences and humanities writing by ENL writers, not only because precise numeric quantities are not needed (Cutting, 2012), but also because these vague quantifying devices have an important face-saving function (Ruzaitė, 2007). Hence, it makes sense that they are not used as commonly in unedited texts by ELF writers as in edited texts by ENL. As the edited texts by ELF writers have frequencies similar to those in ENL, these two constructions can be considered to be distinctive features of edited versus unedited academic writing.

8.2 Summary

This chapter aimed to investigate the relationship between the target constructions in the three study corpora (ELF_Unedited, ELF_Edited, and ENL_Edited) and the disciplinary and similectal backgrounds of the writers whose texts were included in each corpus in terms of form and meaning. 18 of the 19 target constructions identified in the previous chapter on the basis of comparisons of the three study corpora, were included in the statistical tests. The statistical test results were followed by an analysis of semantic domain proportions and interpretations of findings in light of previous research.

The interaction of the independent variables, corpus, similect and discipline, pointed to a remarkable degree of variation within and across the subsets of the corpora. However, except for a few differences between ELF and ENL subsets for 'PREP N,' 'ADJ PREP (NP)', and 'DET *of* (NP),' most of the differences were identified across ELF texts from the same corpus or disciplinary group. The lack of ELF vs. ENL differences among subsets for the nominal and passive constructions indicates that these highly conventionalized features of academic writing yield only significant global differences that are consistent across the corpora. The variation across the ELF subsets in terms of the "corpus:similect," "corpus:discipline," and "similect:discipline" interactions, can be said to confirm the existence of the 'lingua franca factor' (Firth, 2009), in the sense that there is "inherent diversity" across ELF subsets (p.162). This finding is also apparent from the discursive hybridity (Mauranen et al., 2010b) that was identified for some subsets in relation to the similectal background and local discourse conventions of the

writers. Another pattern of variation is that hard sciences (SCI) texts by ELF writers show a greater degree of variation among each other, compared to the relatively limited differences across SSH subsets, a finding which is in line with Flowerdew (2019), indicating that SCI writing is likely to be tolerant of more relaxed norms.

In contrast, the lack of ENL-ELF differences in the use of the canonical features of academic discourse including the 'N PREP (NP)' and '(NP) V-*be* V-*ed* PREP (NP)' constructions also provides strong evidence in support of previous ELF research (Carey, 2013; Mur-Dueñas, 2018a; Murillo, 2018; Wu et al., 2020; Yilmaz & Römer, 2020; Zapletalová, 2018) in that ELF writers prefer to make safe lexicogrammatical choices in favor of highly conventionalized and frequent sequences. These choices are especially evident in edited ELF writing where variation across subsets is slightly less substantial than in unedited ELF. This finding, however, has complex implications for research and pedagogy. On the one hand, considering the continued dominance of Anglo-American norms in academic writing (McKinley & Rose, 2018; O'Neil, 2018), conformity to these norms is a plausible strategy for non-native members of international scientific communities. On the other hand, diachronic research points to certain ongoing changes in academic discourse towards a more colloquial and personally involved style, including a decrease in the use of passives and high-frequency prepositions, as well as an increase in the expression of quantification, and an even more so, in the use of abstractions (Biber & Gray, 2011, 2016; Hundt et al., 2016; Hyland & Jiang, 2019; Seoane, 2006, 2013). Thus, the ENL-normative views that do not reflect this change in authentic academic discourse might soon lead to larger ELF-ENL discrepancies in written scientific communication, especially among scholars who already have distinctive discourse styles partially due to their similectal, cultural, and disciplinary backgrounds. Considering that many ELF scholars actually feel the need to learn to be objective, distanced, and impersonal, thus adopting the somewhat outdated conventions of Anglo-American academic writing (e.g., Bennett & Muresan, 2016; Dontcheva-Navratilova, 2014), an ELF perspective that is more accepting of hybridized discourses and the dynamic nature of academic language could significantly improve non-native scholars' writing practices.

This chapter showed that the cross-corpus differences in the usage of the target constructions as identified in the previous chapter could not straightforwardly be considered the uniform characteristics of written academic ELF for writers of all different similectal and disciplinary backgrounds. Controlling for variation in terms of these background variables, as also done by Martinez (2018), is essential for making conclusive claims on 'universal' ELF features. Thus, the next chapter focuses on the feasibility of arguing for 'universals' of written academic ELF by closely examining the general constructional characteristics which were homogeneously observed across different similectal and disciplinary backgrounds in the current study.

9 "UNIVERSALS" OF WRITTEN ACADEMIC ELF

As ELF perspectives became more widespread in L2 research, several studies in recent years have looked into the general ("universal") characteristics of academic writing by second language users from a variety of similectal and disciplinary backgrounds (Carey, 2013; Mauranen, 2016, 2019; Wu et al., 2020; Yilmaz & Römer, 2020). Although these studies have contributed immensely to our understanding of written academic ELF, we also know that ELF discourses are largely hybridized owing to the diverse linguistic and cultural dynamics of language use that influence international scholars' writing in rather unique and complex ways (Mauranen et al., 2010b). Thus, it is plausible to suggest that "universals" of ELF can best be understood when the factors that contribute to its hybridization are taken into account. Hence, this dissertation first focused on identifying constructions that documented the "universal" differences between edited and unedited texts by ELF writers from nine similectal and two broad disciplinary backgrounds and edited texts by ENL writers from matching disciplinary backgrounds. While the findings of these cross-corpus analyses were reported in the seventh chapter, the eighth chapter explored the complex links between the diverse disciplinary and similectal backgrounds represented in the use of the target constructions. As disciplinary and similectal backgrounds of the writers were found to interact with the general profile of the construction use across the three study corpora, it is necessary that any claims about general characteristics of ELF based on this study take into account both levels of the analysis. Therefore, this chapter tackles the question of whether there are ELF 'universals' that transcend different background variables regarding the findings on the use of the target constructions. It is however, crucial to note that this is not a chapter aimed to argue for or against the existence of written academic ELF as a distinct variety of English. Instead, the chapter attempts to add to the current view of academic writing by international scholars from an ELF perspective.

9.1 An Overview of the Findings in Search of ELF 'Universals'

Before discussing the issue of whether ELF universals can be identified from the findings of the current study, the following subsections provide a concise summary of the results to give an overview of the two major levels of analysis: the whole-corpus level, and the similectal and disciplinary level.

9.1.1 Formal Properties of the Target Constructions

As established characteristics of academic writing, nominal and passive constructions were found to be distinctive features of written academic ELF, with the exception of the key ENL constructions 'N *to* (NP)' and 'N *for* (NP)' (see Table 9.1). While the follow-up analyses of the interactions across the target background variables revealed no significant ELF-ENL differences, the variation across "corpus:similect," and "similect:discipline" subsets revealed considerable variation within the study corpora. Among the nominal constructions, 'N *of* (NP)' and 'N *for* (NP)' yielded the majority of the variation across the similectal and disciplinary groups. Subsets of both ELF and ENL corpora are quite varied and in parallel ways. However, the number of significant differences between ELF_Unedited subsets is slightly higher than for ELF_Edited. As for the findings from the "similect:discipline" interaction, the fluctuation of construction frequencies across SCI subsets, including ELF writers, is more remarkable that that in SSH subsets. The only significant findings for 'N *to* (NP)' were present across "corpus:discipline" subsets. However, the finding that the construction is more frequent in SSH than in SCI texts applies to all three corpora. The SSH-dominance in the usage of this construction can be related to the lack of similectal variation without the factor of similectal differences in the use of this construction.

Among the two complex nominal constructions with multiple prepositions, no significant differences were found for 'N *of* NP *in* (NP)', and 'N *of* NP of (NP)' was not investigated across subsets due to its high correlation with 'N *of* (NP).' The lack of cross-subset differences for 'N *of* NP *in* (NP)' can also be related to the fact that it was only a key construction in ELF_Edited but not in ELF_Unedited. Besides, as no other complex nominal patterns with multiple prepositions were identified as key, the

findings of the current study could provide only limited support for previous research that highlighted the commonality of these patterns in ELF (e.g., Wu et al., 2020; Yilmaz & Römer, 2020).

As for the passives, similarly, no significant differences were identified across "corpus:discipline" subsets. *In*-passives showed the highest degree of variation, while *on*-passives showed the least variation. This contrast is in line with the high lexical and semantic diversity in the former as opposed to the latter. A difference between edited and unedited subsets of ELF was observed only in the case of *in*-passives, usage of which yielded a larger number of differences among unedited ELF texts. However, a finding similar to those for nominal constructions is the greater variation among SCI than in SSH texts in ELF.

Table 9.1 Summary of Statistical Analyses for 'N PREP (NP)' and '(NP) V-be *V*-ed *PREP (NP)' Constructions*

Cx	**Findings from the analysis of differences across the corpora (Chapter 7)**	**Findings from the analysis of differences across corpora, similects, disciplines (chapter 8): Differences from the whole-corpus patterns**		
		Corpus:Similect	**Corpus:Discipline**	**Similect:Discipline**
N *of* (NP)	ELF_Unedited & ELF_Edited > ENL_Edited	**- Within ELF_Unedited:** IT, PO, RU, SP & SW > CH CZ, PO, IT, RU, SP & SW > FI **- Within ELF_Edited:** IT, PO, RU, SP & SW > CH & FI **- Within ENL_Edited:** IT, PO, RU & SP >CH PO & SP > FI SP > CZ	NA	**- Within SCI:** IT, PO, RU & SP > CH CZ, FR, IT, PO, RU, SP & SW > FI **- Within SSH:** IT, PO, RU & SW > CH IT & RU > FI IT > CZ **- Cross-disciplinary:** FI (SSH > SCI)
N *of* NP *in* (NP)	ELF_Edited > ENL_Edited	NA	NA	NA
N *of* NP *of* (NP)	ELF_Unedited & ELF_Edited > ENL_Edited	NA	NA	NA
N *for* (NP)	ENL_Edited > ELF_Unedited & ELF_Edited	**- Within ELF_Unedited:** IT, RU & PO > FI PO, RU & SP > CZ **- Within ELF_Edited:** RU & SW > FI RU & SW > CZ **- Within ENL_Edited:** IT, PO, RU, SP & SW > FI IT > CH RU > CZ	NA	**- Within SCI:** RU > CZ & FI
N *to* (NP)	ENL_Edited > ELF_Unedited & ELF_Edited	NA	**Cross-disciplinary:** ELF_Unedited, ELF_Edited & ENL_Edited (SSH> SCI)	NA
(NP) V-*be* V-ed *by* (NP)	ELF_Unedited & ELF_Edited > ENL_Edited	**- Within ELF_Unedited:** PO > CH & FI	NA	**Within SCI:** CZ > CH & FI FR, IT, PO & SW> FI **Within SSH:** RU > CH

(NP) V-*be* V-*ed in* (NP)	ELF_Unedited & ELF_Edited > ENL_Edited	**- Within ELF_Unedited**: PO > CZ & FI SW > CH, CZ & FI **- Within ELF_Edited:** FR, PO & SW > FI & CH **- Within ENL_Edited:** PO > CZ & FI SW > FI	NA	**- Within SCI:** CH, FR, IT, PO, SP & SW > FI SW > CZ & RU **- Within SSH:** PO > CH & CZ CZ & SW > CH
(NP) V-*be* V-*ed on* (NP)	ELF_Unedited & ELF_Edited > ENL_Edited	**Within ELF_Edited:** SW > CH	NA	**Within SCI:** SW > CH, CZ & FI

The findings for the 'PREP N,' 'ADJ PREP (NP),' and 'DET PREP (NP)' constructions were summarized separately as the results pointed to findings distinct from those for nominal and passive constructions. These findings included the differences between ELF and ENL, as well as edited and unedited ELF at the corpus and subset levels.t At the subset-level, the "corpus:discipline" interaction in particular, revealed interesting patterns of variation in the use of 'PREP N,' 'ADJ PREP (NP),' and 'DET PREP (NP)' constructions.

Among these constructions, '*after* N,' '*in* NN *with* (NP),' '*in order to* V,' and 'ADJ-SUPERL *of* (NP)' were identified as key constructions in ELF. '*Across* N-s,' '*within* N,' '*through* N,' 'ADJ *of* (NP),' and 'ADJ *at* (NP)' were found to be key in ENL. '*Over* N' and 'DET *of* (NP)' were identified as key in both ENL and ELF_Edited as opposed to ELF_Unedited. Besides, the only ELF-ENL differences involving similects in the study were found for in '*in order to* V' and '*within* N' constructions.

Other ELF-ENL differences came from the comparisons of "corpus:discipline" subsets for the usage of '*over* N,' '*within* N,' 'ADJ *of* (NP)' and 'ADJ *at* (NP).' All these findings pointed to the significantly higher frequencies of these constructions in SCI texts by ENL than ELF writers. The significantly higher frequencies of '*over* N,' '*within* N,' and 'ADJ *of* (NP)' in SSH than in SCI texts by ELF writers, also confirmed this finding. As these disciplinary differences across the study corpora comprise the majority of the ELF-ENL findings for the analysis of disciplinary and similectal variation, it is plausible to argue that the limited use of these low-frequency constructions is a defining characteristic of ELF writing in SCI disciplines as opposed to their higher frequencies in ENL writing. As argued by Flowerdew (2019), these findings, along with previous research on specific SCI disciplines (Martinez,

2018; Rozycki & Johnson, 2013; Tribble, 2017, 2019) indicate that SCI writing accommodates ELF features to a greater extent than in SSH writing (see Table 9.2).

Table 9.2 Summary of Statistical Analyses for 'PREP N', 'ADJ PREP (NP)', and 'DET PREP (NP)' Constructions

Cx	**Findings from the analysis of differences across the corpora (Chapter 7)**	**Findings from the analysis of differences across corpora, similects, disciplines (chapter 8): Differences from the whole-corpus patterns**		
		Corpus:Similect	**Corpus:Discipline**	**Similect:Discipline**
across N-s	ENL_Edited > ELF_Unedited & ELF_Edited	NA	**- Within SCI & SSH:** ENL_Edited > ELF_Unedited & ELF_Edited	NA
after N	ELF_Unedited & ELF_Edited > ENL_Edited	NA	NA	NA
over N	ENL_Edited > ELF_Unedited & ELF_Edited ELF_Edited > ELF_Unedited	NA	**- Within SCI:** ENL_Edited > ELF_Unedited & ELF_Edited **- Cross-disciplinary:** ELF_Unedited (SSH > SCI)	NA
within N	ENL_Edited > ELF_Unedited	**- Within ELF_Unedited**: CZ > FI	**- Within SCI:** ENL_Edited > ELF_Unedited **- Cross-disciplinary:** ELF_Unedited (SSH > SCI)	**- ELF_ENL:** SCI **(**EN_CZ > CZ)
through N	ENL_Edited > ELF_Edited	NA	**- Cross-disciplinary:** ELF_Unedited, ELF_Edited & ENL_Edited (SSH >SCI)	**- Within SCI:** CH & FR > FI FR > CZ SP > FR **- Cross-disciplinary:** CZ (SSH > SCI) FI (SSH > SCI)
in order to V	ELF_Unedited & ELF_Edited > ENL_Edited	**- Within ELF_Edited**: SP > FI **- ENL - ELF:** FR_Unedited > EN_FR	NA	**- Within SCI:** RU > CH, CZ & FI SP & SW > FI **- Cross-disciplinary:** CZ (SSH > SCI)
in NN *with* (NP)	ELF_Unedited & ELF_Edited > ENL_Edited	NA	NA	NA
ADJ *of* (NP)	ENL_Edited > ELF_Unedited & ELF_Edited	NA	**- Within SCI:** ENL_Edited > ELF_Unedited & ELF_Edited **- Cross-disciplinary:** ELF_Unedited & ELF_Edited (SSH > SCI)	NA
ADJ-SUPERL *of* (NP)	ELF_Unedited & ELF_Edited > ENL_Edited	**- Within ELF_Edited:** RU > CH	NA	NA
ADJ at (NP)	ENL_Edited > ELF_Unedited	NA	**- Within SCI:** ENL_Edited > ELF_Unedited & ELF_Edited	NA
DET *of* (NP)	ENL_Edited > ELF_Unedited & ELF_Edited ELF_Edited > ELF_Unedited	NA	**- Within SCI & SSH:** ENL_Edited > ELF_Edited	NA

Overall, the general discrepancy between the findings for nominal and passive constructions on the one hand, and 'PREP N,' 'ADJ PREP (NP),' and 'DET PREP (NP)' on the other, is remarkable. The usage of high-frequency nominal and passive constructions interacts with the similectal backgrounds of

the authors and editing status of the texts in ELF corpora to a limited extent, which does not lead to ELF-ENL differences at the level of subsets. In the case of the lower frequency 'PREP N,' 'ADJ PREP (NP,' and 'DET PREP (NP)' constructions, variation across individual similects plays an even more limited role. These constructions vary across disciplinary subsets of the corpora. Similar disciplinary and similectal influences of written academic ELF were also reported in Murillo (2018) and Murillo-Ornat (2019). This indicates that it is important to account for similectal and disciplinary variation before arguing for 'universals' of written academic ELF.

9.1.2 Semantic Properties of the Target Constructions

The analysis of the lexical items in the variable slots of the key constructions, including the head nouns in 'N PREP (NP)', main verbs in '(NP) V-*be* V-*ed by* (NP),' nouns in 'PREP (NP),' adjectives in 'ADJ PREP (NP),' and determiners in 'DET PREP N' followed the analysis of the constructions' formal properties. The focus was placed on the proportions of the semantic domains for the lexical items that passed the collocational strength threshold. This phase of the analysis enabled a closer and more qualitative look into usage differences across the corpora and their similectal and disciplinary subsets from a semantic and functional perspective.

A wide array of differences across corpora and their subsets were reported in terms of the semantic preferences observed in the constructions. However, the abstract-concrete distinction is of particular relevance for the discussions of 'universals' this chapter as it pointed to the salient patterns of conceptualization in the data at a broader level (Hoffmann, 2004, p. 191) . Of the 19 focal constructions, ten are key in ELF, and the other nine are key in ENL writing. With the exception of 'ADJ-SUPERL *of* (NP)' for ELF and 'ADJ *at* (NP)' for ENL, all the key constructions in ELF writing primarily denote abstractions. In contrast, all constructions that are key in ENL, except for 'N *to* (NP)' and 'N *for* (NP),' refer to concrete entities. When the constructions are grouped together, it can be seen that the majority of items semantically identified for 'N PREP (NP)' and '(NP) V-*be* V-ed *by* (NP)' are abstract in all three corpora, and that ELF writers use abstract meanings more frequently than their ENL counterparts regarding the use of the former. In 'DET PREP (NP),' higher proportions of the concrete (quantifying)

than abstract (referential) domain is observed in all corpora, with ELF_Edited having the highest proportion of latter. The proportions of 'PREP N' constructions, however, show a contrast where the abstract domain is more frequent than the concrete domain only in the ELF corpora. This contrast is also in line with the significantly higher frequencies of *'across* N-s*,' 'over N,' 'through N,'* and *'within N'* in ENL_Edited, where concrete meanings dominate. The higher proportion of abstract meanings in these constructions, as well as the exclusively discursive key ELF constructions *'in* NN *with* (NP)*'* and *'in order to* V,' differentiate ELF from ENL writing (see Table 9.3).

Table 9.3 Proportions of Abstract and Concrete Semantic Domains in Cx Groups across Corpora

Cx	**N PREP (NP)**		**PREP N**		**(NP) V-*be* V-ed PREP (NP)**		**ADJ PREP (NP)**		**DET PREP (NP)**	
Domain	**Abstract**	**Concrete**	**Abstract**	**Concrete**	**Abstract**	**Concrete**	**Abstract**	**Concrete**	**Abstract**	**Concrete**
ELF_Unedited	69.51	30.49	56.05	43.95	98.23	1.77	36.92	63.08	24.46	75.54
ELF_Edited	67.11	32.89	55.85	44.15	98.88	1.12	40.84	59.16	34.53	65.47
ENL_Edited	63.89	36.11	42.29	57.71	98.02	1.98	39.96	60.04	23.75	76.25
Total	66.60	33.40	50.85	49.15	98.41	1.59	39.57	60.43	27.59	72.41

Note. Sum of percentages for semantic domains for each subset equals all Cx collocates included in semantic analysis.

However, while these general abstract-to-concrete ratios hold true for all "similect:discipline" subsets in the corpora regarding the use of 'N PREP (NP)' and '(NP) V-*be* V-ed *by* (NP),' exceptions exist for the 'PREP N,' 'ADJ PREP (NP),' and 'DET PREP (NP)' constructions. With the exceptions of 'DET PREP (NP)' in the ENL subsets and 'PREP N' constructions in SSH writing subset of ELF_Edited, there is considerable variation across "similect:discipline" subsets across corpora. in terms of the choice of one domain over the other. This summative overview of the variation is also congruent with the findings from statistical tests indicating that there is a higher degree of variation in SCI subsets by ELF writers (see Table 9.4).

Table 9.4 Differences between Corpora and "Similect:Discipline" Subsets in Ratio of Abstract and Concrete Domains

	PREP N		**ADJ PREP (NP)**		**DET PREP (NP)**	
Ratio across Corpora	Abstract > concrete in ELF Concrete > abstract in ENL		Concrete > abstract		Concrete > abstract	
Corpus:Discipline	**Ratio difference**	**Similect**	**Ratio difference**	**Similect**	**Ratio difference**	**Similect**
ELF_Unedited_SCI	Concrete > abstract	CZ, FI, SW	Abstract > concrete	IT	Abstract > concrete	IT, CH, RU
ELF_Unedited_SSH		SW, FR		CH, FI, IT, SP, SW		CH
ELF_Edited_SCI		CZ, FI, FR, PO, RU		CZ, SW		CH, SW, FR
ELF_Edited_SSH		n/a		CH, CZ, SP		CH
ENL_Edited_SCI	Abstract > concrete	SP, SW		FR, IT, SW		n/a
ENL_Edited_SSH		CH, PO, RU, FI		FR, CZ		n/a

Note. No ratio differences were found for 'N PREP (NP)' and '(NP) V-*be* V-ed PREP (NP)' constructions.

As the identification of the key constructions in question was an exhaustive process of looking for the items most representative of ELF-ENL differences, it is fair to argue that the constructional usages of ELF and ENL writers show remarkable differences in the types of meanings they express. However, the nature and degree of differences vary in relation to the usage dynamics of the constructions. In the case of high-frequency sequences that have highly conventional formal and functional properties in academic writing, such as nominals and passives, ELF-ENL differences are limited to the higher frequency of the generally preferred semantic domain. This finding, however, is not as straightforward as it may seem at first. As much as an expanding use of abstractions has been documented in academic writing (Biber & Gray, 2011; Parkinson & Musgrave, 2014), the even higher proportions of abstract meanings in ELF writing can also be attributed to an interaction of the "lingua franca factor" (Firth, 2009) with academic writing conventions. Similarly, following his analysis of the use of modals in ELF over time, Laitinen (2016) concluded that "lingua franca uses polarize the diffusion of change" (p. 176). This polarizing effect is evident in almost all 'N PREP (NP)' constructions where certain subsets have much higher frequencies of abstract head nouns in the ELF corpora than in ENL_Edited. This variation can be a sign of the varying levels with which ELF writers adopt current trends by.

As for the less frequent items, 'PREP N' 'ADJ PREP (NP),' 'DET PREP (NP),' the semantic differences go beyond the polarizing effect in the existing trend and entail contrasting choices, as well. This difference can be explained in several interconnected ways. From a usage-based perspective, it makes sense to find that second language writers rely on commonly encountered constructions with widely known uses (Carey, 2013; Ellis, 2012; Tomasello, 2003), and that there is substantial variation in the use of low-frequency items with also less prototypical formal and functional features (Ellis, 2002; Ellis et al., 2016). Besides, although Biber and Gray (2016) reported an increase in the use of less frequent prepositions in academic writing, these prepositions were not studied extensively. As they also have not been the focus of research in written academic ELF either, further evidence is needed to better understand both the formal and functional properties of these 'PREP N' constructions.

9.2 Bringing It All Together

As O'Neil (2018) explains, English academic writing is the global medium of written scientific communication involving "a wide range of speakers, topics, and registers, which (again) helps explain why ELF cannot be adequately defined as a particularized formal variety" (p. 148). Having abandoned the varietal perspective inspired by World Englishes research long ago (Jenkins, 2015), ELF research today also embraces the diversity inherent in ELF, also known as the "lingua franca factor" (Firth, 2009). However, as most of the early ELF research focused on the analysis of oral interaction, we have come to know only recently, that this conceptualization of ELF may not entirely hold true for academic writing. As researchers have discussed repeatedly, common perceptions and practices of English academic writing that align closely with standard language ideologies, help maintain stability to a greater extent than in speaking (Casanave, 2017; Crowley, 2003; Laitinen et al., 2018; McKinley & Rose, 2018; O'Neil, 2018). Findings from recent research on 'universal' features of ELF also support this view by showing that written academic ELF reflects the complexity of factors that shape it such as nativelike norms, disciplinary and generic conventions, as well as the diverse backgrounds and bi/multilingual repertoires of its users (e.g., Laitinen, 2018, 2020; Laitinen et al., 2018; Martinez, 2018; Pérez-Llantada, 2013; Tribble, 2017, 2019; Wu et al., 2020; Yilmaz & Römer, 2020). As this relatively new line of research is of immense importance for a better understanding of the current, highly diversified sphere of English academic communication, this chapter also aimed to uncover the 'universal' usage dynamics of ELF in research papers by ELF and ENL writers. In order to be able to argue for features of written academic ELF that transcend the complex patterns of variation identified in this study, I synthesized the two sets of findings reported in Chapters 7 and 8; that is, the results of the whole-corpus as well as the similectal and disciplinary analysis of key constructions within and across the study corpora.

The findings from the whole-corpus approach enabled the identification of the target constructions, frequency counts of which pointed to the statistically most remarkable differences between

ELF and ENL academic writing. The key constructions found in the ELF_Unedited and ELF_Edited largely overlapped in that they similarly included high-frequency prepositions (‘*of,’ ‘in,’ ‘on,’ ‘by,’ ‘after’*) coupled with content words commonly used in academic writing (e.g., *type, effect, characterize, study*, etc.). These constructions were generally found to convey abstract meanings instrumental in ELF writers’ discourse construction, including various functions such as denoting reference, labeling, research activities, and argumentation. In contrast, the constructions key in ENL_Edited employed relatively less frequent prepositions (‘*to,’ ‘for,’ ‘across,’ ‘at,’ ‘within,’ ‘over,’ ‘through’)* that generally attracted content words found to denote not only abstract concepts similar to ELF, but also a more prevalent use of concrete and quantifiable entities (e.g., *number, time, site, locate, etc.*). These findings provided substantial evidence for the preference of ELF academic writers for conventional and high-frequency lexicogrammatical sequences (Bondi & Borelli, 2018; Carey, 2013; Vitali & Bondi, 2020; Yilmaz & Römer, 2020; Zapletalová, 2018). From a sociolinguistic perspective on the evolution of contact languages like ELF, this finding is an important sign of simplification in that ELF writers make use of a small number of highly frequent sequences in a rather regularized and semantically transparent manner (Mauranen, 2019; Trudgill, 2016). The fact that the key constructions in ENL are generally much less frequent and are used to express a wider array of functions is also in line with this conclusion.

The other prominent finding that can be argued to be an ELF ‘universal’ is the greater degree of variation in lower-frequency constructions and their more specified dynamics of usage. Carey (2013) similarly found that approximations, usage instances that partially differed from those in ENL, were generally observed among low-frequency patterns. While no such approximations were identified in this study, distinct usage preferences were observed especially for the constructions less commonly used by ELF writers. For instance, not only were ‘PREP N,’ ‘ADJ PREP (NP),’ and ‘DET PREP (NP)’ less frequent in ELF than in ENL, these constructions also yielded the only contrasting proportions of semantic domain preferences. However, the fact that the individual similectal backgrounds were not as influential in these differences as they were when it came to the nominal and passive constructions signals that these can be considered “collective ELF-specific trends” (Murillo-Ornat, 2019, p. 39).

The findings also pointed to differences between ELF and ENL writers in terms of their expression of authorial voice and stance. Some relevant distinct ELF features include: high frequency of abstractions in general labeling nouns that might be perceived as vague and imprecise (Cutting, 2007, 2012; Ruzaitė, 2007); common use of passives, generally associated with less personally involved authorial voice (Baratta, 2009; Hyland & Jiang, 2019; Pérez-Llantada, 2014; Seoane & Hundt, 2018); and the choice of '*most of* (NP)' in ELF over '*many/much of* (NP)' in ENL that communicates a boosting effect (Li & Wharton, 2012). As discussed earlier in Chapters 7 and 8, these ELF vs. ENL differences are becoming more complex as academic language was reported to become more personal with a decrease in passives, and also more precise with the increase in quantifying nouns (Biber & Gray, 2016; Hyland & Jiang, 2019). In addition, previous studies also found differences in evaluative and stance expressions used by ELF writers in comparison to ENL writers (Bernardini et al., 2010; Bondi & Borelli, 2018; Farley, 2018; Lafuente-Millán, 2018; Mur-Dueñas, 2015, 2016, 2018a, 2018b; Palumbo, 2015). Such differences have generally been attributed to the blending of local and global discursive conventions, which results in hybridized expressions of evaluative stance and authorial voice (Mauranen et al., 2010b). While such hybridizing influences were also observed in specific subsets (e.g., those including Chinese and Romance ELF texts), neither this study nor the previous research enables a comprehensive understanding of the language of evaluation and stance expression in ELF. Thus, further research is needed to arrive at definitive conclusions.

The overall findings on 'universals' generally confirm previous research findings on ELF academic writing. However, as also discussed in the literature review chapters, most of these studies have a limited scope in terms of the size of corpora, as well as linguistic and statistical analysis. Thus, I would argue that the comprehensive analysis of forms and meanings of constructions derived solely from usage data and robust statistical tests used to quantify the dynamics of written academic ELF in this study have provided extensive support for claims made in previous research. Lastly, although several 'universals' were identified for groups of constructions in earlier studies, the complex web of links between constructional usage and factors of nativeness, discipline, similect, and editing which were highlighted in

this study once again confirms the existence of the broader ELF ‘universal’: the “lingua franca factor” (Firth, 2009). Apart from the different types of generalizations at different levels, the complexity inherent in ELF, and the outcome of its being an “an exceptionally complex form of language contact” (Mauranen, 2018a, p. 39), manifested itself consistently across the findings.

The next chapter concludes this dissertation with a brief summary of the study, including its strengths and limitations, followed by the discussion of implications for future research and pedagogy.

10 DISCUSSION AND CONCLUSION

This chapter starts with a brief review of the motivations, design, and findings of this dissertation project. It then discusses the limitations of the study as well as implications for future research and academic writing pedagogy.

10.1 Summary of the study

ELF research emerged out of the need to accommodate the current global reality of the English language as the common medium of communication for an ever-growing number of nonnative speakers and a wide variety of purposes. Scientific communication, primarily carried out via writing, is also an inherently global phenomenon where the use of a common language can be considered essential for dialogue and progress (Mauranen, 2018a). Accordingly, ELF perspectives on the study of academic writing have been increasingly influential in recent years. The WrELFA (2015) project, in particular, has initiated widespread explorations of second language use in the writings of international scholars, by bringing together the strengths of different research strands to contribute to a renewed understanding of English academic discourse. However, ENL-oriented standard language ideologies, traditional conceptions of writtenness, and highly institutionalized regulating practices still dominate academic writing practices (Horner, 2017; Hynninen & Solin, 2017; Turner, 2018). As Marlina and Xu (2018) explain:

> [T]he pluralistic conceptualization of English and of writing in English still has not yet been welcomed by tertiary educators, language practitioners, and academic "literacy brokers" such as proof-readers or publishing editors who, driven by the native-speakerist ideologies, still insist on conformity to conventional expectations of academic writing as well as NES varieties of English (p. 10).

Hence, this study was primarily motivated by the need to improve our understanding of academic writing from an ELF perspective that embraces the complexities of second language use in academic contexts. Three corpora, ELF_Unedited, ELF_Edited, and ENL_Edited, were analyzed to identify the distinctive

lexicogrammatical features of edited and unedited writing in hard and soft sciences by ELF scholars from different similectal backgrounds and by ENL writers.

With the aim of offering theoretically and empirically strong implications for the field, several other linguistic and methodological motivations also influenced the research design. A usage-based constructionist approach to the analysis of multi-word phraseological patterns, defined as pairings of form and meaning essential for language acquisition and use (e.g., Ellis, 1996, 2002, 2003; Goldberg, 1995, 2003, 2006), was employed as the theoretical lens of the study. Regarding its methodology, the "key function words to constructions" (KFWs to Cxs) approach adapted from Groom (2007, 2009, 2010) enabled the systematic analysis of constructions by relying solely on frequency differences across the study corpora. Robust corpus analytic and statistical methods such as directional association and dispersion measures (Ellis, 2006; Gries, 2008), and bootstrapped nonparametric analyses of variance (D. s. Dobler, S. Friedrich, & M. Pauly, 2019; Labouriau, 2020) were used to ensure the validity of the findings. Besides, results were interpreted by making connections with research from a wide array of fields, including academic writing, ELF, world Englishes, cognitive linguistics, and second language acquisition. The research questions addressed were:

4) RQ 1: What are the dominant constructions distinctive of unedited ELF, edited ELF, and edited ENL??
    a) How do distinctive constructions in ELF compare to those in ENL?
    b) How do distinctive constructions in unedited ELF compare to those in edited ELF?
5) How do the identified sets of constructions distinctive of unedited ELF, edited ELF, and edited ENL compare statistically in terms of the background variables of editing status, similects, and disciplinary categories?
6) Based on the answers to research questions 1 and 2, is it possible to determine the constructional profile of "universal ELF," that is, a common and generalizable set of constructions in written academic ELF with little to no influence of the above-mentioned background variables?

To address RQ1, the "KFWs to Cxs" method was used. The identified key prepositions enabled the extraction of grammatical sequences most distinctive of each corpus in relation to the other two. 19 key constructions that largely documented ELF-ENL differences were identified. These included ten nominal and passive constructions with highly frequent prepositions in ELF, and nine constructions around less frequent prepositions used with nouns as objects, adjectives, and determiners in ENL. The key constructions in ELF generally denoted abstract meanings and processes commonly found in academic discourse. On the other hand, those in ENL made use of concrete meanings for specifying quantities and properties of entities. The only remarkable differences found between unedited and edited ELF were the even scarcer usage of low-frequency constructions in the former. Thus, the exhaustive quantitative and qualitative analyses allowed me to identify a comprehensive list of ELF-ENL differences showing with limited difference between edited and unedited ELF writing.

As for answering RQ2, the statistical analyses of similects and disciplines within and across corpora revealed intriguing patterns of variation. The high-frequency nominal and passive constructions were found to vary across similects within the ELF corpora, especially in the SCI subsets. However, no significant ELF vs. ENL differences were found across matching disciplinary and similectal subsets. The only ELF vs. ENL differences across subsets were identified in the use of the constructions that include nominal complements, predicative adjectives, and determiners used with low-frequency prepositions. These differences were generally found between disciplinary subsets of the matching corpora, and to a much smaller extent, across similectal subsets. A closer look into the proportions of semantic and abstract domains also demonstrated that, in line with the results from the statistical tests, contrasting semantic preferences of ELF and ENL writers were only present in the second group of low-frequency constructions. While abstract meanings generally had higher proportions in ELF for nominal and passive constructions, both ELF and ENL texts similarly had an overwhelming proportion of head nouns with abstract meanings. For the second group of low-frequency constructions, ELF writers were found to use higher proportions of abstract meanings even when ENL writers preferred the more concrete meanings.

The findings indicated that, beyond the cross-corpus differences, there were also different types of similectal and disciplinary variations in the use of the target constructions.

Lastly, for RQ3, the findings from the earlier analyses that helped me address RQ1 and RQ 2 were taken together and examined for potential "universals"; that is, for general characteristics of written academic ELF that transcend the observed similectal and disciplinary differences. The identified ELF "universals" included: the common use of conventional academic constructions such as nominals and passives with high-frequency prepositions, as well as abstract meanings for general and discursive purposes; limited use of constructions around low-frequency prepositions with remarkable variation especially in SCI writing; and distinctive expressions of evaluation and voice. All nominal and passive constructions, along with '*across* N-s,' '*after* N,' '*through* N,' '*in NN with* (NP),' and 'ADJ-SUPERL *of* (NP)' were found to only yield ELF vs. ENL differences at the whole-corpus level. Hence, these constructions were considered indicators of "universal" differences between ELF and ENL. In contrast, the remainder of the key ENL constructions pointed to significant differences between disciplinary and similectal groups across ELF and ENL. Comparisons of the SCI texts by ELF writers with those by ENL writers resulted in the majority of these differences, where contrasting semantic preferences in ELF and ENL were also observed. Thus, for low-frequency items not particularly known as features of academic writing, the "lingua franca factor" (Firth, 2009) seems to be the main "universal" for ELF. To put it in another way, the primary general characteristic of usage for these constructions is the complexity and variation inherent across similectal and disciplinary groups. Overall, from a methodological point of view, bringing together the findings from both the corpus- and the subset-level analyses, were instrumental in arguing for "universals," or the lack thereof, in ELF writing,

10.2 Implications for Research

As Mauranen (2012, 2018b) explains, much of the complexity of ELF stems from the fact that it is a contact language that involves the coming together of second language users from different linguistic and cultural backgrounds. Thus, as argued by Murillo (2018) and Murillo-Ornat (2019), it is possible to identify "universal" as well as group-specific characteristics of ELF relating to similectal and disciplinary

backgrounds of the writers. As was also found in this study, this the "universals" generally point to usage in line with academic conventions, yet with partially different formal and functional preferences. Despite these differences, considerable evidence supports the existence of simplification processes similar to those found in the evolution of post-colonial varieties of English (Mauranen, 2019; Trudgill, 2016). Such similarities with remarkably more stable and uniform varieties can be attributed to common characteristics of second language acquisition in adulthood (Mauranen, 2018b; Schneider, 2012). However, zooming in on data by writing from individual disciplinary and similectal groups, generally enables a closer look into the hybridity of ELF. This hybridity is primarily due to the blending of different styles in the writers' multilingual and multicultural repertoires (Mauranen et al., 2010b). As contradictory as they may seem, these two levels of analysis, corpus- and subset-levels, actually complement each other. For instance, while Carey's (2013) whole-corpus approach showed that there was more variation among lower-frequency patterns, the present study went a step further to document the types and degrees of variation among subsets for the relatively infrequent constructions. Further, in addition to the non-significant differences among specific ELF vs. ENL subsets in the use of nominal and passive constructions, remarkable fluctuation was found across similectal groups especially in the SCI subsets of the corpora. Such noticeable differences across ELF subsets might well lead to significant differences for the analyses of different datasets, similects, or disciplines. Therefore, when examining "universals" of ELF, it is of vital importance to ensure the uniform distribution of target the lexicogrammatical items across texts by writers from different backgrounds.

It should also be noted that the variation across disciplinary and similectal subsets in the use of the constructions was found despite the corpus-level thresholds of range and frequency, and analysis of keyness through multiple tests that took into account construction frequencies as well as their dispersions. While the thresholds were not set for each disciplinary and similectal subset, the target constructions had occurrences across both disciplinary groups, and in at least the majority of the similects as in Martinez (2018). However, despite the common use of the target constructions, variation was still identified even in the use of high-frequency constructions that led to the argument for ELF "universals". This finding is also

in line with the multiple perspectives that are argued to shape the general characteristics of ELF in Chapter 5. In the four-perspective model adapted from Mauranen (2012, 2018b), I argued that the macrosocial perspective on ELF included the similectal and disciplinary backgrounds of writers as two pivotal social factors that contributed to the "universals" of ELF. However, the "universal" perspective placed at the center of the model was conceptualized as a set of general characteristics that still showed a certain degree of variability (see Figure 5.1). Therefore, it is plausible to suggest that definite claims on "universals" of ELF should be made only when the variables likely to cause variation beyond ELF vs. ENL differences are controlled for.

Throughout Chapters 7, 8, and 9, connections of the findings with studies on general characteristics of (Biber, 2006b; Biber et al., 1999; Hyland, 2008b; Liu, 2012; Simpson-Vlach & Ellis, 2010), as well as the ongoing change (Biber & Gray, 2011, 2016; Hyland & Jiang, 2019; Seoane, 2006, 2013) in academic discourse were made. Although these studies helped me interpret the findings from multiple perspectives, they also highlighted some challenges. For instance, Tribble (2015) argues that criticizing EAP approaches as being ENL-normative is not fair as EAP research on characteristics of academic discourse uses corpora of expert writing and does not take into account the nativeness of the writers (p. 444). The findings of this study, as well as many others (e.g., Marti et al., 2019; Martinez, 2018; Pan et al., 2016; Pérez-Llantada, 2014), however, consistently showed that edited research papers by expert writers published on respectable platforms still included features indicative of the hybridity of ELF. Hence, instead of ignoring this important factor, future research should openly acknowledge the inclusion of ELF writing as an integral part of written scientific communication around the world today (Horner, 2017).

Similarly, in the case of language change in academic discourse, I argued that the variation in nominal and passive constructions among ELF subsets could be related to differences between the subsets in terms of the pace and degree of language change. While Laitinen (2016, 2020) reported similar findings for the use of modals in ELF academic writing, this interpretation of the current findings is promising yet tentative. Therefore, it is of vital importance for future ELF research to longitudinally

investigate the effects of language change comparatively in ELF and ENL. Factoring in the role of longitudinal trends could reveal additional interesting layers of complexity in ELF.

Another implication of the study has to do with the corpus methodology generally used in written academic ELF research. Many studies to date have only investigated a small number of select items in written academic ELF (e.g., contiguous sequences with high frequencies or known functions such as expressing evaluation). The inductive approach adopted in this study, on the other hand, led to an exhaustive analysis of written academic ELF thanks to the flexibility of CxG that accommodates the analysis of form-meaning pairs of different kinds (Groom, 2019). In addition, previous ELF academic writing research generally employed either no statistical analyses or simple tests such as chi-square or log-likelihood ratio. As these tests rely only on aggregate frequencies without taking dispersion into account, they are likely to point to misleading results in the analyses of corpora with high internal variability (Brezina & Meyerhoff, 2014; Gries, 2008; Lijffijt et al., 2016; Paquot & Bestgen, 2009). Regarding the analysis of ELF corpora that include small and differently-sized subsets of texts by writers from of a large number of different backgrounds, more robust tests were considered necessary to be able to make valid claims about the nature of ELF. Consequently, this study used a wide range of measures of distribution, dispersion, collocational strength, predictability of lexical fillers in variable slots of the construction, as well as significance testing for group differences via log-likelihood and bootstrapped MANOVAs. Besides, so as not to neglect the meaning aspect of the constructional analysis, common lexical items in the variable slots of constructions; that is, those strongly attracted by the constructional frame, were also carefully analyzed and categorized into semantic domains. As every applied measure revealed a different layer of complexity in the data, I would argue that research efforts on written academic ELF could benefit immensely from comprehensive mixed-method analysis appropriate for such complex data.

10.3 Implications for Academic Writing Pedagogy

If complexity is such a defining feature of ELF academic writing, then there is a dire need to adapt academic writing instruction to the dynamics of written scientific communication today. Along with the

multiplicity of factors that shape ELF such as the different similectal and cultural backgrounds of nonnative users (Mauranen, 2012, 2018b), academic writing is also known to be one of the most complex and varied registers (Biber, 2006b, 2009; Biber & Conrad, 2019; Biber et al., 2002). As Horner (2011) argues, an ELF perspective could offer both native and nonnative students the benefit of familiarizing themselves with a dynamic view of academic writing that accommodates diversity in scientific communication, with the ultimate goal of effective communication. Adopting an ELF perspective for writing pedagogy could be done by integrating findings from ELF research on academic writing into instructional materials (Björkman, 2011). Similar to what researchers have suggested for data-driven learning, instructors could either design corpus-based materials or hands-on activities for students to explore ELF usage themselves (Römer, 2011a).

The selection of linguistic features for writing instruction is also another area for which this study has implications. The current study documented the links between high frequency, even dispersion, and semantic generality in the use of constructions. A similar link was also found between these factors by Hilpert and Correia Saavedra (2017). Durrant (2018) also adds collocational strength to the list of criteria for inclusion of sequences to teach. Thus, corpus-informed instructional materials that follow a discursive approach to the introduction of lexicogrammatical sequences, such as Strauss et al. (2018), could be of great help for students to understand the formal and functional features of constructional usage in an integrated fashion. In addition, as also shown in Chapter 7 on how the meanings of target constructions were analyzed, prepositions are highly polysemous words, which also manifests itself in the wide variety of meanings they denote as part of constructions. Following Tyler (2012), I would argue that a cognitive linguistic approach that brings to the attention the network of senses that ubiquitous words like prepositions denote could allow language learners to greatly expand their usage repertoire.

10.4 Limitations

Despite the multiple measures taken to ensure its validity and generalizability, the study also has certain limitations that need to be considered in order to better contextualize the findings and inform future investigations of written academic ELF. First, only nine similects, most of which are from European

contexts, were analyzed. Future research should investigate texts from other similectal groups, including speakers of first languages from different language families. Second, the analysis of similectal and disciplinary variation was based on small subsets of unequal sizes within the corpora. Larger corpora could reveal a more detailed description of hybridity in ELF. Third, previous studies that included unedited, as well as edited ENL writing revealed interesting differences between ENL and ELF writing of individual similectal groups (e.g., Marti et al., 2019; Rowley-Jolivet & Carter-Thomas, 2014). Thus, including a corpus of unedited ENL writing, if it had been feasible within the scope of this study, could have enabled essential insights into the role of editing from a comparative perspective. Fourth, when investigating a complex phenomenon like ELF, a textographic approach (Swales, 1998) could allow researchers to gain deeper insights into how ELF texts are constructed and what factors shape writers' discursive practices by shedding light on the writing processes of international scholars.

Lastly, as was discussed throughout the study in the interpretation of the results, academic writing trends continue to change, and research shows that the impact of these changes on ELF adds another layer of complexity to the phenomenon (Laitinen, 2016, 2020; Martinez, 2018). Therefore, a longitudinal perspective on ELF could be useful in exploring how certain "universals" manifest themselves over time. For instance, a longitudinal analysis of abstract meanings commonly found in ELF could help us understand whether ELF writers follow the general academic writing trend of semantic expansion in abstract meanings.

REFERENCES

Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes, 31*(2), 81-92.

Alptekin, C. (2011). Beyond ENL norms in ELF use. *Australian Review of Applied Linguistics, 34*(2), 148-165. doi:10.1075/aral.34.2.02alp

Alptekin, C. (2013). English as a lingua franca through a usage-based perspective: merging the social and the cognitive in language use. *Language, culture and curriculum, 26*(2), 197-207. doi:10.1080/07908318.2013.810224

Anderson, L. (2010). Standards of acceptability in English as an academic lingua franca: Evidence from a corpus of peer-reviewed working papers by international scholars. In R. J. Cagliero, J. (Ed.), *Discourses, communities, and global englishes* (pp. 115-143). Bern: Peter Lang.

Ansarifar, A., Shahriari, H., & Pishghadam, R. (2018). Phrasal complexity in academic writing: A comparison of abstracts written by graduate students and expert writers in applied linguistics. *Journal of English for Academic Purposes, 31*, 58-71.

Archer, D. (2009). *What's in a word-list?: investigating word frequency and keyword extraction*: Ashgate Publishing, Ltd.

Baayen, R. H. (2013). Multivariate Statistics. In R. Podesva & D. Sharma (Eds.), *Research methods in linguistics* (pp. 337-372). Cambridge: Cambridge University Press.

Baayen, R. H., Milin, P., & Ramscar, M. (2016). Frequency in lexical processing. *Aphasiology, 30*(11), 1174-1220.

Baird, R., Baker, W., & Kitazawa, M. (2014). The complexity of ELF. *Journal of English as a Lingua Franca, 3*(1), 171-196.

Baker, P. (2004a). Querying keywords: Questions of difference, frequency, and sense in keywords analysis. *Journal of English Linguistics, 32*(4), 346-359.

Baker, P. (2004b). 'Unnatural Acts': Discourses of homosexuality within the House of Lords debates on gay male law reform. *Journal of sociolinguistics, 8*(1), 88-106.

Baker, P. (2006). *Using corpora in discourse analysis*. London: Continuum.

Baratta, A. M. (2009). Revealing stance through passive voice. *Journal of Pragmatics, 41*(7), 1406-1421.

Bardi, M., & Muresan, L. (2014). Changing research writing practices in Romania: Perceptions and attitudes. In K. Benett (Ed.), *The semiperiphery of academic writing: Discourses, communities and practices* (pp. 121-147). London: Palgrave Macmillan.

Becher, Y. T., Paul R. (2001). *Academic tribes and territories: Intellectual enquiry and the cultures of disciplines.* . Bristol: Society for Research into Higher Education & Open University Press.

Belcher, D. (2007). Seeking acceptance in an English-only research world. *Journal of Second Language Writing, 16*(1), 1-22.

Benelhadj, F. (2019). Discipline and genre in academic discourse: Prepositional Phrases as a focus. *Journal of Pragmatics, 139*, 190-199.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, 57*, 289-300.

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*, 1165-1188.

Bennett, K. (2010). Academic discourse in Portugal: A whole different ballgame? *Journal of English for Academic Purposes, 9*(1), 21-32.

Bennett, K. (2014). The erosion of Portuguese historiographic discourse. In *The semiperiphery of academic writing* (pp. 13-38). London: Springer.

Bennett, K., & Muresan, L. (2016). Rhetorical incompatibilities in academic writing: English versus the Romance cultures. *Synergy, 12*(1), 95-119.

Benson, M., Benson, E., & Ilson, R. F. (2010). *The BBI combinatory dictionary of English: Your guide to collocations and grammar* (3 ed.). Amsterdam: John Benjamins.

Berkenkotter, C., & Huckin, T. N. (1995). *Genre knowledge in disciplinary communication: Cognition/culture/power*. New Jersey: Lawrence Erlbaum Associates.

Bernardini, S., Ferraresi, A., & Gaspari, F. (2010). Institutional academic English in the European context: A web-as-corpus approach to comparing native and non-native language. In A. L. Lopez & R. C. Jimenez (Eds.), *Professional English in the European context: The EHEA challenge* (pp. 27-53). Bern: Peter Lang.

Bestgen, Y. (2018). Getting rid of the Chi-square and Log-likelihood tests for analysing vocabulary differences between corpora. *Quaderns de Filologia - Estudis Lingüístics, 22*(22). doi:10.7203/qf.22.11299

Bhatia, V. K. (2002a). Applied genre analysis: A multi-perspective model. *Ibérica: Revista de la Asociación Europea de Lenguas para Fines Específicos (AELFE)*(4), 3-19.

Bhatia, V. K. (2002b). A generic view of academic discourse. In J. F. (Ed.) (Ed.), *Academic Discourse* (pp. 31-49). New York: Routledge.

Biber, D. (1985). Investigating macroscopic textual variation through multifeature/multidimensional analyses. *Linguistics, 23*(2), 337-360.

Biber, D. (1986). On the investgation of spoken/written differences. *Studia Linguistica, 40*(1), 1-21.

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, D. (2006a). Stance in spoken and written university registers. *Journal of English for Academic Purposes, 5*(2), 97-116.

Biber, D. (2006b). *University language: A corpus-based study of spoken and written registers* (Vol. 23). Amsterdam: John Benjamins Publishing.

Biber, D. (2009). A corpus-driven approach to formulaic language in English. *International Journal of Corpus Linguistics, 14*(3), 275-311. doi:10.1075/ijcl.14.3.08bib

Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. *Language and Computers, 26*, 181-190.

Biber, D., & Conrad, S. (2019). *Register, genre, and style* (2nd ed.). Cambridge: Cambridge University Press.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at…: Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*(3), 371-405.

Biber, D., Conrad, S., Reppen, R., Byrd, P., & Helt, M. (2002). Speaking and writing in the university: A multidimensional comparison. *TESOL quarterly, 36*(1), 9-48.

Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes, 9*(1), 2-20.

Biber, D., & Gray, B. (2011). Grammatical change in the noun phrase: The influence of written language use. *English Language & Linguistics, 15*(2), 223-250.

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

Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL quarterly, 45*(1), 5-35. doi:10.5054/tq.2011.244483

Biber, D., Gray, B., & Staples, S. (2016). Predicting Patterns of Grammatical Complexity Across Language Exam Task Types and Proficiency Levels. *Applied Linguistics, 37*(5), 639-668. doi:10.1093/applin/amu059

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

Björkman, B. (2008a). English as the lingua franca of engineering: The morphosyntax of academic speech events. *Nordic Journal of English Studies, 7*(3), 103-122.

Björkman, B. (2008b). 'So where we are?'Spoken lingua franca English at a technical university in Sweden. *English today, 24*(2), 35-41.

Björkman, B. (2009). From code to discourse in spoken ELF. In A. Mauranen & E. Ranta (Eds.), *English as a lingua franca: Studies and findings* (pp. 225-251). Cambridge: Cambridge Scholars Publishing.

Björkman, B. (2011). English as a lingua franca in higher education: Implications for EAP. *Ibérica: Revista de la Asociación Europea de Lenguas para Fines Específicos (AELFE)*(22), 79-100.

Björkman, B. (2013). *English as an academic lingua franca: An investigation of form and communicative effectiveness* (Vol. 3). New York: Walter de Gruyter.

Bondi, M., & Borelli, C. (2018). Publishing in English: ELF writers, textual voices and metadiscourse. In P. M.-D. J. Ŝinkūienė (Ed.), *Intercultural perspectives on research writing* (Vol. John Benjamins, pp. 217-235): Amsterdam.

Bondi, M., & Morelli, C. (2018). Publishing in English: ELF writers, textual voices and metadiscourse. In P. Mur-Dueñas & J. Ŝinkūienė (Eds.), *Intercultural perspectives on research writing* (pp. 217-235). Amsterdam: John Benjamins.

Boyle, R. (2011). Patterns of change in English as a lingua franca in the UAE. *International Journal of Applied Linguistics, 21*(2), 143-161.

Breiteneder, A. (2005). The naturalness of English as a European lingua franca: the case of the 'third person–s'. *Vienna English Working Papers, 14*(2), 3-26.

Breiteneder, A. (2009). English as a lingua franca in Europe: An empirical perspective. *World Englishes, 28*(2), 256-269.

Brett, D., & Pinna, A. (2015). Patterns, fixedness and variability: Using PoS-grams to find phraseologies in the language of travel journalism. *Procedia-Social and Behavioral Sciences, 198*, 52-57.

Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge: Cambridge University Press.

Brezina, V., McEnery, T., & Wattam, S. (2015). Collocations in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics, 20*(2), 139-173.

Brezina, V., & Meyerhoff, M. (2014). Significant or random?: A critical review of sociolinguistic generalisations based on large corpora. *International Journal of Corpus Linguistics, 19*(1), 1-28. doi:10.1075/ijcl.19.1.01bre

Brezina, V., Timperley, M., & McEnery, T. (2018). #LancsBox v. 4.5. Retrieved from http://corpora.lancs.ac.uk/lancsbox

Bruce, I. J. (2015). Use of cognitive genres as textual norms in academic. *Bulletin VALS-ASLA*, 161-175.

Bybee, J. L. (2006). From usage to grammar: The mind's response to repetition. *Language*, 711-733.

Bybee, J. L. (2008). Usage-based grammar and second language acquisition. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and SLA* (pp. 226-246). London: Routledge.

Bybee, J. L., & Hopper, P. J. (2001). Introduction to the frequency and the emergence of linguistic structure. In J. L. Bybee & P. J. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (Vol. 45, pp. 1-26). Amsterdam: John Benjamins.

Bychkovska, T., & Lee, J. (2017). At the same time: Lexical bundles in L1 and L2 university student argumentative writing. *Journal of English for Academic Purposes, 30*, 38-52.

Canagarajah, S. (2012). *Translingual practice: Global Englishes and cosmopolitan relations*. London: Routledge.

Carey, R. (2013). On the other side: Formulaic organizing chunks in spoken and written academic ELF. *Journal of English as a Lingua Franca, 2*(2), 207-228.

Casanave, C. P. (2017). *Controversies in second language writing: Dilemmas and decisions in research and instruction* (2nd ed.). Ann Arbor: University of Michigan Press.

Chafe, W. (1982). Integration and involvement in speaking, writing, and oral literature. In D. Tannen (Ed.), *Spoken and written language: Exploring orality and literacy* (pp. 35-54). Norwood, NJ: Ablex.

Charles, M. (2013). English for academic purposes. In B. S. Paltridge, Sue (Ed.), *Handbook of English for specific purposes* (pp. 137-154). Wiley-Blackwell: Malden, MA.

Chen, L., & Chung, S.-F. (2018). *Investigating the English ADJECTIVE OF Construction in Academic Writing.* Paper presented at the Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation.
Chen, Y., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology, 14*(2), 30-49.
Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge: MIT Press.
CiteScore™. (2019). Retrieved December 2019, from Scopus www.scopus.com/sources
Cogo, A. (2009). Accommodating difference in ELF conversations: A study of pragmatic strategies. In A. Mauranen & E. Ranta (Eds.), *English as a lingua franca: Studies and findings* (pp. 254-273). Cambridge: Cambridge Scholars Publishing.
Cogo, A., & Dewey, M. (2006). Efficiency in ELF communication: From pragmatic motives to lexico-grammatical innovation. *Nordic Journal of English Studies, 5*(2), 59-93.
Cogo, A., & Dewey, M. (2012). *Analysing English as a lingua franca: A corpus-driven investigation*. London: Continuum
Collins, P. (2008). The progressive aspect in World Englishes: A corpus-based study. *Australian Journal of Linguistics, 28*(2), 225-249.
Connor, U. (2004). Intercultural rhetoric research: Beyond texts. *Journal of English for Academic Purposes, 3*(4), 291-304.
Cook, V. (1999). Going beyond the native speaker in language teaching. *TESOL quarterly, 33*(2), 185-209.
Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*(4), 397-423.
Cortes, V. (2013). The purpose of this study is to: Connecting lexical bundles and moves in research article introductions. *Journal of English for Academic Purposes, 12*(1), 33-43.
Crossley, S. A., & McNamara, D. S. (2011). Shared features of L2 writing: Intergroup homogeneity and text classification. *Journal of Second Language Writing, 20*(4), 271-285.
Crossley, S. A., Salsbury, T., & McNamara, D. (2010). The development of polysemy and frequency use in English second language speakers. *Language Learning, 60*(3), 573-605.
Crowley, T. (2003). *Standard English and the politics of language*. London: Palgrave Macmillan.
Crystal, D. (2003). *English as a global language* (2nd ed.). Cambridge: Cambridge University Press.
Curry, M. J., & Lillis, T. (2019). Unpacking the lore on multilingual scholars publishing in English: A discussion paper. *Publications, 7*(2).
Cutting, J. (2007). *Vague language explored*: Springer.
Cutting, J. (2012). Vague language in conference abstracts. *Journal of English for Academic Purposes, 11*(4), 283-293.
Dang, T. N. Y., & Webb, S. (2016). Making an essential word list for beginners. In I. S. P. Nation (Ed.), *Making and using word lists for language learning and testing* (pp. 153-167). Amsterdam: John Benjamins.
Danielsson, P. (2007). What constitutes a unit of analysis in language? *Linguistik Online, 31*(2), 17-24.
De Knop, S., & Gilquin, G. (2016). *Applied construction grammar*. Berlin: DeGruyter.
Deshors, S. C. (2019). English as a Lingua Franca: A random forests approach to particle placement in multi-speaker interactions. *International Journal of Applied Linguistics*, 1-18.
Dirven, R. (1993). Dividing up physical and mental space into conceptual categories by means of English prepositions. In C. Zelinsky-Wibbelt (Ed.), *The semantics of prepositions: From mental processing to natural language processing* (pp. 73-97). Berlin: Walter de Gruyter.
Dobler, D., Friedrich, S., & Pauly, M. (2019). Nonparametric MANOVA in meaningful effects. *Annals of the Institute of Statistical Mathematics*, 1-26.
Dobler, D. s., Friedrich, S., & Pauly, M. (2019). Nonparametric MANOVA in meaningful effects. *Annals of the Institute of Statistical Mathematics*, 1-26. doi:https://doi.org/10.1007/s10463-019-00717-3
Dontcheva-Navratilova, O. (2014). The changing face of Czech academic discourse. In K. Bennett (Ed.), *The semiperiphery of academic writing* (pp. 39-61). London: Springer.

Douglas Fir Group. (2016). A transdisciplinary framework for SLA in a multilingual world. *Modern Language Journal, 100*(S1), 19-47.

Durrant, P. (2015). Lexical bundles and disciplinary variation in university students' writing: Mapping the territories. *Applied Linguistics, 38*(2), 165-193. doi:10.1093/applin/amv011

Durrant, P. (2018). Formulaic language in English for Academic Purposes. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding Formulaic Language: A Second Language Acquisition Perspective* (pp. 211-227). London: Routledge.

Edwards, A. (2017). ICE Age 3: The Expanding Circle. *World Englishes, 36*(3), 404-426. doi:10.1111/weng.12279

Edwards, A., & Laporte, S. (2015). Outer and Expanding Circle Englishes: The competing roles of norm orientation and proficiency levels. *English World-Wide, 36*(2), 135-169.

Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition, 18*(1), 91-126.

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition, 24*(2), 143-188.

Ellis, N. C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. In C. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 33-68). Oxford: Blackwell.

Ellis, N. C. (2006). Language acquisition as rational contingency learning. *Applied Linguistics, 27*(1), 1-24. doi:10.1093/applin/ami038

Ellis, N. C. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics, 32*, 17-44.

Ellis, N. C., & Ferreira-Junior, F. (2009a). Construction learning as a function of frequency, frequency distribution, and function. *Modern Language Journal, 93*(3), 370-385.

Ellis, N. C., & Ferreira-Junior, F. (2009b). Constructions and their acquisition: Islands and the distinctiveness of their occupancy. *Annual Review of Cognitive Linguistics, 7*(1), 188-221.

Ellis, N. C., & Larsen-Freeman, D. (2009). Constructing a second language: Analyses and computational simulations of the emergence of linguistic constructions from usage. *Language Learning, 59*, 90-125.

Ellis, N. C., O'Donnell, M. B., & Römer, U. (2014). The processing of verb-argument constructions is sensitive to form, function, frequency, contingency and prototypicality. *Cognitive Linguistics, 25*(1), 55-98.

Ellis, N. C., Römer, U., & O'Donnell, M. B. (2016). *Usage-based aproaches to language acquisition and ğrocessing: cognitive and corpus investigations of Construction Grammar* Malden, MA: Wiley.

Ellis, N. C., & Wulff, S. (2015a). Second language acquisition. In E. Dąbrowska & D. Divjak (Eds.), *Handbook of cognitive linguistics* (pp. 409-431). Berlin: DeGruyter Mouton.

Ellis, N. C., & Wulff, S. (2015b). Usage–based approaches to SLA. In P. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 75-93). Florence: Taylor and Francis.

Ellis, N. C., & Wulff, S. (2019). Cognitive approaches to second language acquisition. In C. C. U. P. d. J. Schwieter & A. Benati (Eds.) (Ed.), *The Cambridge handbook of language learning (Cambridge handbooks in language and linguistics)* (pp. 41-61). Cambridge: Cambridge University Press.

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text-Interdisciplinary Journal for the Study of Discourse, 20*(1), 29-62.

Esfandiari, R., & Barbary, F. (2017). A contrastive corpus-driven study of lexical bundles between English writers and Persian writers in psychology research articles. *Journal of English for Academic Purposes, 29*, 21-42.

Eskildsen, S. W. (2012). L2 negation constructions at work. *Language Learning, 62*(2), 335-372.

Farley, A. F. (2018). NNES RAs: How ELF RAs inform literacy brokers and English for research publication instructors. *Journal of English for Academic Purposes, 33*, 69-81.

Ferraresi, A., & Bernardini, S. (2015). Institutional academic English and its phraseology: Native and lingua franca perspectives. In P. Thompson & D. Giuliana (Eds.), *English for Academic Purposes: Approaches and Implications* (pp. 225-244). Cambridge: Cambridge Scholars Publishing.

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R* (Vol. London): Sage publications.

Fillmore, C. J. (1985). *Syntactic intrusions and the notion of grammatical construction.* Paper presented at the Annual Meeting of the Berkeley Linguistics Society.

Fillmore, C. J. (1988). *The mechanisms of "Construction Grammar".* Paper presented at the Annual Meeting of the Berkeley Linguistics Society.

Fillmore, C. J., Kay, P., & O'Connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions: The case of let alone. *Language*, 501-538.

Firth, A. (1990). 'Lingua franca'negotiations: towards an interactional approach. *World Englishes, 9*(3), 269-280.

Firth, A. (1996). The discursive accomplishment of normality: On 'lingua franca' English and conversation analysis. *Journal of Pragmatics, 26*(2), 237-259.

Firth, A. (2009). The lingua franca factor. *Intercultural Pragmatics, 6*(2), 147-170.

Firth, A., & Wagner, J. (1997). On discourse, communication, and (some) fundamental concepts in SLA research. *Modern Language Journal, 81*(3), 285-300.

Fløttum, K., Dahl, T., & Kinn, T. (2006). *Academic voices: Across languages and disciplines* (Vol. 148): John Benjamins Publishing.

Flowerdew, J. (2000). Discourse community, legitimate peripheral participation, and the nonnative-English-speaking scholar. *TESOL quarterly, 34*(1), 127-150.

Flowerdew, J. (2008). Scholarly writers who use English as an additional language: What can Goffman's "stigma" tell us? *Journal of English for Academic Purposes, 7*(2), 77-86.

Flowerdew, J. (2013). English for research publication purposes. In B. Paltridge & S. Starfield (Eds.), *The handbook of English for specific purposes* (pp. 301-321). Chichester: Wiley-Blackwell.

Flowerdew, J. (2015). Some thoughts on English for research publication purposes (ERPP) and related issues. *Language Teaching, 48*(2), 250-262.

Flowerdew, L. (2015). Adjusting pedagogically to an ELF world: An ESP perspective. In Y. Bayyurt & S. Akcan (Eds.), *Current Perspectives on Pedagogy for English as a Lingua Franca* (pp. 13-34). Berlin: De Gruyter Mouton.

Flowerdew, L. (2019). English as a lingua franca and learner English in disciplinary writing: A corpus perspective. In K. Hyland & L. L. C. Wong (Eds.), *Specialised English: New Directions in ESP and EAP Research and Practice*. London: Routledge.

Francis, G., Hunston, S., & Manning, E. (1996). *Collins COBUILD grammar patterns: Verbs*. London: Harper Collins.

Francis, G., Hunston, S., & Manning, E. (1998). *Collins COBUILD grammar patterns: Nouns and adjectives*. London: Harper Collins.

Friedrich, S., Dobler, D., & Pauly, M. (2019). rankMANOVA: Rank-based tests for multivariate data in nonparametric factorial designs. Retrieved from http://github.com/smn74/rankMANOVA

Gabrielatos, C., & Marchi, A. (2012, September). *Keyness: Appropriate metrics and practical issues.* Paper presented at the Corpus-assisted Discourse Studies (CADS) International Conference, Bologna, Italy.

Giles, H., & Marlow, M. L. (2011). Theorizing language attitudes existing frameworks, an integrative model, and new directions1. *Annals of the International Communication Association, 35*(1), 161-197.

Gilquin, G. (2012). Lexical infelicity in causative constructions. Comparing native and learner collostructions. In J. L. R. v. Waldenfels (Ed.), *Analytical causatives: From 'give' and 'come' to 'let' and 'make'* (pp. 41-64). Munich: Lincom Europa.

Gilquin, G. (2015). The use of phrasal verbs by French-speaking EFL learners: A constructional and collostructional corpus-based approach. *Corpus Linguistics and Linguistic Theory, 11*(1), 51-88.

Gilquin, G. (2018). Exploring the spoken learner English constructicon. In R. Alonso Alonso (Ed.), *Speaking in a Second Language* (pp. 127-115).

Gilquin, G., & De Knop, S. (2016). Exploring L2 constructionist approaches. In S. De Knop & G. Gilquin (Eds.), *Applied Construction Grammar* (Vol. 32, pp. 3-17). Berlin: DeGruyter.

Gledhill, C. J. (2000a). *Collocations in science writing* (Vol. 22). Tübingen: Gunter Narr Verlag.

Gledhill, C. J. (2000b). The discourse function of collocation in research article introductions. *English for Specific Purposes, 19*(2), 115-135.

Gledhill, C. J. (2011). The 'lexicogrammar'approach to analysing phraseology and collocation in ESP texts. *ASp. la revue du GERAS*(59), 5-23.

Gnutzmann, C. (2000). Lingua franca. In M. Byram (Ed.), *The Routledge encyclopedia of language teaching and learning* (pp. 356–359). London: Routledge.

Goldberg, A. E. (1992). The inherent semantics of argument structure: The case of the English ditransitive construction. *Cognitive Linguistics (includes Cognitive Linguistic Bibliography), 3*(1), 37-74.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Goldberg, A. E. (2003). Constructions: A new theoretical approach to language. *Trends in cognitive sciences, 7*(5), 219-224.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press on Demand.

Goldberg, A. E. (2013). Constructionist approaches. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford handbook of construction grammar* (pp. 15-31). Oxford: Oxford University Press.

Goldberg, A. E. (2019). *Explain me this: Creativity, competition, and the partial productivity of constructions*. Princeton: Princeton University Press.

Granger, S., Dagneaux, E., & Meunier, F. (2002). *The international corpus of learner English: Handbook and CD-ROM*. Louvain-la-Neuve: Presses Universitaires de Louvain.

Granger, S., Dagneaux, E., Meunier, F., & Paquot, M. (2009). *The International Corpus of Learner English: Handbook and CD-ROM (version 2)*. Louvain-la-Neuve: Presses Universitaires de Louvain.

Gray, B. (2013). More than discipline: uncovering multi-dimensional patterns of variation in academic research articles. *Corpora, 8*(2), 153-181.

Gray, B. (2015). *Linguistic variation in research articles: When discipline tells only part of the story*. Amsterdam: John Benjamins.

Gray, B., & Biber, D. (2013). Lexical frames in academic prose and conversation. *International Journal of Corpus Linguistics, 18*(1), 109-136.

Greenbaum, S., & Nelson, G. (1996). The international corpus of English (ICE) project. *World Englishes, 15*(1), 3-15.

Gries, S. T. (2008). Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics, 13*(4), 403-437.

Gries, S. T. (2010). Dispersions and adjusted frequencies in corpora: further explorations. In S. W. Stefan Th. Gries, and Mark Davies (Ed.), *Corpus-linguistic applications: Current studies, new directions* (pp. 197-212). Leiden: Brill Rodopi.

Gries, S. T. (2013). 50-something years of work on collocations: What is or should be next…. *International Journal of Corpus Linguistics, 18*(1), 137-166.

Gries, S. T. (2020). *Ten Lectures on Corpus Linguistics with R: Applications for Usage-Based and Psycholinguistic Research*. Leiden: Brill.

Gries, S. T., & Adelman, A. S. (2014). Subject realization in Japanese conversation by native and non-native speakers: Exemplifying a new paradigm for learner corpus research. In J. Romero-Trillo (Ed.), *Yearbook of corpus linguistics and pragmatics 2014: New empirical and theoretical paradigms* (pp. 35–54). Cham: Springer.

Gries, S. T., & Deshors, S. C. (2014). Using regressions to explore deviations between corpus data and a standard/target: Two suggestions. *Corpora, 9*(1), 109-136.
Gries, S. T., & Ellis, N. C. (2015). Statistical measures for usage-based linguistics. *Language Learning, 65*(S1), 228-255.
Groom, N. W. (2007). *Phraseology and epistemology in humanities writing: a corpus-driven study*. (Doctor of Philosophy). The University of Birmingham, Birmingham.
Groom, N. W. (2009). Phraseology and epistemology in academic book reviews: A corpus-driven analysis of two humanities disciplines. In K. Hyland & G. Diani (Eds.), *Academic evaluation: Review genres in university settings* (pp. 122-139). London: Palgrave Macmillan.
Groom, N. W. (2010). Closed-class keywords and corpus-driven discourse analysis. In M. Bondi & M. Scott (Eds.), *Keyness in texts* (pp. 59-78). Amsterdam: John Benjamins.
Groom, N. W. (2018). *Construction grammar and academic discourse analysis*. Paper presented at the American Association for Applied Linguistics Conference, Chicago, US.
Groom, N. W. (2019). Construction Grammar and the corpus-based analysis of discourses. *International Journal of Corpus Linguistics, 24*(3), 291-323. doi:10.1075/ijcl.00014.gro
Hall, C. J. (2012). Cognitive contributions to plurilithic views of English and other languages. *Applied Linguistics, 34*(2), 211-231.
Hall, C. J. (2017). Cognitive perspectives on English as a lingua franca. In J. Jenkins, W. Baker, & M. Dewey (Eds.), *The Routledge Handbook of English as a Lingua Franca* (pp. 74-84): Routledge.
Hall, C. J., Joyce, J., & Robson, C. (2017). Investigating the lexico-grammatical resources of a non-native user of English: The case of can and could in email requests. *Applied Linguistics Review, 8*(1), 35-59.
Halliday, M. A. K. (1989). *Spoken and written language*. Oxford: Oxford Uniersity Press.
Halliday, M. A. K. (1994). *An introduction to functional grammar*. London: Edward Arnold.
Hanauer, D. I., Sheridan, C. L., & Englander, K. (2019). Linguistic injustice in the writing of research articles in English as a second language: Data from Taiwanese and Mexican researchers. *Written Communication, 36*(1), 136-154.
Hartse, J. H., & Kubota, R. (2014). Pluralizing English? Variation in high-stakes academic texts and challenges of copyediting. *Journal of Second Language Writing, 24*, 71-82.
Hilferty, J. (1999). Through as a means to metaphor. In L. E. Stadler, C. (Ed.), *Issues in cognitive linguistics* (pp. 347-366). Berlin: Mouton De Gruyter.
Hilpert, M. (2014). *Construction grammar and its application to English*. Edinburgh: Edinburgh University Press.
Hilpert, M., & Correia Saavedra, D. (2017). Why are grammatical elements more evenly dispersed than lexical elements? Assessing the roles of text frequency and semantic generality. *Corpora, 12*(3), 369-392. doi:10.3366/cor.2017.0125
Hoey, M. (2005). *Lexical priming: A new theory of words and language*. London: Routledge.
Hoffmann, S. (2004). Are low-frequency complex prepositions grammaticalized. In H. M. Lindquist, C. (Ed.), *Corpus approaches to grammaticalization in English* (pp. 171-210). Amsterdam: John Benjamins.
Hoffmann, S. (2005). *Grammaticalization and English complex prepositions: A corpus-based study*. London: Routledge.
Hoffmann, T. (2014). The cognitive evolution of Englishes: The role of constructions in the dynamic model. In S. Buschfeld, T. Hoffmann, M. Huber, & A. Kautzsch (Eds.), *The evolution of Englishes: The dynamic model and beyond* (pp. 160-180). Amsterdam: John Benjamins.
Hoffmann, T., & Bergs, A. (2018). A Construction Grammar approach to genre. *Cognitextes, 18*, 1-21.
Horner, B. (2011). Writing English as a lingua franca. In A. Archibald, A. Cogo, & J. Jenkins (Eds.), *Latest trends in ELF research* (pp. 299-311). Newcastle Upon Tyne: Cambridge Scholars Publishing.

Horner, B. (2017). Written academic English as a lingua franca. In J. Jenkins, W. Baker, & M. Dewey (Eds.), *The Routledge Handbook of English as a Lingua Franca* (pp. 413-426). London: Routledge.

House, J. (1999). Misunderstanding in intercultural communication: Interactions in English as a lingua franca and the myth of mutual intelligibility. In C. Gnutzmann (Ed.), *Teaching and learning English as a global language* (pp. 73-89). Tübingen: Stauffenburg.

House, J. (2013). Developing pragmatic competence in English as a lingua franca: Using discourse markers to express (inter) subjectivity and connectivity. *Journal of Pragmatics, 59*, 57-67.

Hülmbauer, C. (2013). From within and without: The virtual and the plurilingual in ELF. *Journal of English as a Lingua Franca, 2*(1), 47-73.

Hultgren, A. K. (2019). English as the Language for Academic Publication: on Equity, Disadvantage and 'Non-Nativeness' as a Red Herring. *Publications, 7*(2). doi:10.3390/publications7020031

Hundt, M., G., S., & Seoane, E. (2016). The use of the be-passive in academic Englishes: Local vs global usage in an international language. *Corpora, 11*(1), 29-61.

Hunston, S. (2010). Starting with the small words: Patterns, lexis and semantic sequences. In U. Römer & R. Schulze (Eds.), *Patterns, meaningful units and specialized discourses* (pp. 7-30).

Hunston, S. (2019). Patterns, constructions, and applied linguistics. *International Journal of Corpus Linguistics, 24*(3), 324-353. doi:10.1075/ijcl.00015.hun

Hunston, S., & Francis, G. (2000). *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins.

Hunston, S., & Su, H. (2017). Patterns, constructions, and local grammar: A case study of 'evaluation'. *Applied Linguistics*.

Hyland, K. (1999). Academic attribution: Citation and the construction of disciplinary knowledge. *Applied Linguistics, 20*(3), 341-367.

Hyland, K. (2008a). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics, 18*(1), 41-62.

Hyland, K. (2008b). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*(1), 4-21.

Hyland, K. (2009). English for professional academic purposes: Writing for scholarly publication. In D. Belcher (Ed.), *English for specific purposes in theory and practice*. Ann Arbor, MI: University of Michigan Press.

Hyland, K. (2016a). Academic publishing and the myth of linguistic injustice. *Journal of Second Language Writing, 31*, 58-69.

Hyland, K. (2016b). General and specific EAP. In K. Hyland & P. Shaw (Eds.), *The Routledge handbook of English for academic purposes* (pp. 41-53). London: Routledge.

Hyland, K., & Jiang, K. (2019). *Academic discourse and global publishing: Disciplinary persuasion in changing times*. London: Routledge.

Hyland, K., & Shaw, P. (2016). The Routledge handbook of English for academic purposes. In K. Hyland & P. Shaw (Eds.), *Introduction* (pp. 1-13). London: Routledge.

Hynninen, N., & Kuteeva, M. (2017). "Good" and "acceptable" English in L2 research writing: Ideals and realities in history and computer science. *Journal of English for Academic Purposes, 30*, 53-65.

Hynninen, N., & Solin, A. (2017). Language norms in ELF In J. B. Jenkins, Will; Dewey, Martin (Ed.), *The Routledge Handbook of English as a Lingua Franca* (pp. 267-278). London: Routledge.

Ingvarsdóttir, H., & Arnbjörnsdóttir, B. (2013). ELF and academic writing: A perspective from the expanding circle. *Journal of English as a Lingua Franca, 2*(1), 123-145.

Jackendoff, R. (1995). The boundaries of the lexicon. In M. Everart, E. Van der Linden, & A. S. Schenk, R. (Eds.), *Idioms: Structural and psychological perspectives* (pp. 133-166). Hillsdale: Lawrence Erlbaum.

Jenkins, J. (2000). *The phonology of English as an international language*. Oxford: Oxford University Press.

Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied Linguistics, 23*(1), 83-103.
Jenkins, J. (2006). Current perspectives on teaching World Englishes and English as a Lingua Franca. *TESOL quarterly, 40*(1), 157-181. doi:Doi 10.2307/40264515
Jenkins, J. (2007). *English as a lingua franca: Attitude and identity*. Oxford: Oxford University Press.
Jenkins, J. (2014). *English as a Lingua Franca in the international university: the politics of academic English language policy*. Abingdon: Routledge.
Jenkins, J. (2015). Repositioning English and multilingualism in English as a Lingua Franca. *Englishes in Practice, 2*(3), 49-85.
Jenkins, J. (2017a). ELF and WE: competing or complementing paradigms? In E. L. P. Low, A. (Ed.), *World Englishes: rethinking paradigms* (pp. 12-28): Routledge.
Jenkins, J. (2017b). Not English but English-within-multilingualism. In S. W. Coffey, U. (Ed.), *New Directions for Research in Foreign Language Education* (pp. 63-78). New York: Routledge.
Jenkins, J., & Mauranen, A. (2019). *Linguistic Diversity on the EMI Campus: Insider accounts of the use of English and other languages in universities within Asia, Australasia, and Europe*. Abingdon: Routledge.
Jiajin, X., Liang, M., & Jia, Y. (2012). BFSU PowerConc 1.0. In: National Research Centre for Foreign Language Education, Beijing Foreign Studies University.
Jordan, B., & Fuller, N. (1975). On the non-fatal nature of trouble: Sense-making and trouble-managing in lingua franca talk. *Semiotica, 13*(1), 11-32.
Kachru, B. B. (1987). *The past and prejudice: Toward de-mythologizing the English canon*. Amsterdam: John Benjamins.
Kaplan, R. B. (1966). Cultural thought patterns in inter-cultural education. *Language Learning, 16*, 1-20.
Kecskes, I. (2007). Formulaic language in English lingua franca. In I. Kecskes & L. R. Horn (Eds.), *Explorations in pragmatics: Linguistic, cognitive and intercultural aspects* (pp. 191–218). Berlin & New York: Mouton de Gruyter.
Kecskes, I. (2015). Is the idiom principle blocked in Bilingual L2 production? In R. H. A. Cieslicka (Ed.), *Bilingual figurative language processing* (pp. 28–53). Cambridge: Cambridge University Press.
Kecskes, I. (2019). *English as a Lingua Franca: The Pragmatic Perspective*. Cambridge: Cambridge University Press.
Klégr, A. (1997). English complex prepositions of the prepositional phrase type. *Prague Studies in English, 22*(5), 51-78.
Krishnamurthy, R., & Kosem, I. (2007). Issues in creating a corpus for EAP pedagogy and research. *Journal of English for Academic Purposes, 6*(4), 356-373. doi:10.1016/j.jeap.2007.09.003
Labouriau, R. (2020). postHoc: Tools for Post-Hoc Analysis. (Version R package version 0.1.3.). Retrieved from https://CRAN.R-project.org/package=postHoc
LaFlair, G. T., Egbert, J., & Plonsky, L. (2015). A practical guide to bootstrapping descriptive statistics, correlations, t tests, and ANOVAs. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research.* (pp. 46-77). London: Routledge.
Lafuente-Millán, E. (2018). Evaluation in research article introductions in the Social Sciences written by English as a lingua franca and English native users. In P. Mur-Dueñas & J. Ŝinkūienė (Eds.), *Intercultural perspectives on research writing* (pp. 255-275). Amsterdam: John Benjamins.
Laitinen, M. (2016). Ongoing changes in English modals. In O. Timofeeva, S. Chevalier, A. Gardner, & A. Honkapohja (Eds.), *New approaches to English linguistics: Building bridges* (Vol. 177, pp. 175-196). Amsterdam: John Benjamins.
Laitinen, M. (2018). Placing ELF among the varieties of English: Observations from typological profiling. In S. Deshors (Ed.), *Modeling World Englishes: Assessing the interplay of emancipation and globalization of ESL varieties* (pp. 109-132). Amsterdam: John Benjamins.
Laitinen, M. (2020). Empirical perspectives on English as a lingua franca (ELF) grammar. *World Englishes, n/a*(n/a). doi:10.1111/weng.12482

Laitinen, M., Levin, M., & Lakaw, A. (2018). Charting new sources of ELF data: A Multi-genre corpus approach. In C. Suhr, T. Nevalainen, & I. Taavitsainen (Eds.), *From data to evidence in English language lesearch* (pp. 326-350). Leiden: Brill.

Lakoff, G. (1984). There-constructions: A case study in grammatical construction theory. *Berkeley Cognitive Science report no. 18*.

Langacker, R. W. (1991). *Concept, image, and symbol*. Berlin: De Gruyter.

Levshina, N. (2015). *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam: John Benjamins.

Li, T., & Wharton, S. (2012). Metadiscourse repertoire of L1 Mandarin undergraduates writing in English: A cross-contextual, cross-disciplinary study. *Journal of English for Academic Purposes, 11*(4), 345-356.

Liardet, C. L., Black, S., & Bardetta, V. S. (2019). Defining formality: Adapting to the abstract demands of academic discourse. *Journal of English for Academic Purposes, 38*, 146-158. doi:10.1016/j.jeap.2019.02.007

Lijffijt, J., Nevalainen, T., Säily, T., Papapetrou, P., Puolamäki, K., & Mannila, H. (2016). Significance testing of word frequencies in corpora. *Literary and Linguistic Computing, 31*(2), 374-397.

Lillis, T. (2013). *Sociolinguistics of writing*: Edinburgh University Press.

Lillis, T., & Curry, M. J. (2006). Professional academic writing by multilingual scholars: Interactions with literacy brokers in the production of English-medium texts. *Written Communication, 23*(1), 3-35.

Lillis, T., & McKinney, C. (2013). The sociolinguistics of writing in a global context: Objects, lenses, consequences. *Journal of sociolinguistics, 17*(4), 415-439.

Lippi-Green, R. (2012). *English with an accent: Language, ideology and discrimination in the United States*. London: Routledge.

Liu, D. (2012). The most frequently-used multi-word constructions in academic written English: A multi-corpus study. *English for Specific Purposes, 31*(1), 25-35.

Lorés-Sanz, R. (2016). ELF in the making? simplification and hybridity in abstract writing. *Journal of English as a Lingua Franca, 5*(1), 53-81.

Lorés-Sanz, R. (2018). Hybrid rhetorical structure in English Sociology research article abstracts: The ambit of ELF and translation. In P. Mur-Dueñas & J. Ŝinkūienė (Eds.), *Intercultural perspectives on research writing* (pp. 175-194). Amsterdam: John Benjamins.

Lorés-Sanz, R., & Murillo-Ornat, S. (2018). International and local contexts of scientific publication: Rhetorical-discursive analysis of sociology abstracts in English. *Estudios de Linguistica-Universidad de Alicante (ELUA)*(32), 331-348.

Lu, X., Kisselev, O., Yoon, J., & Amory, M. D. (2018). Investigating effects of criterial consistency, the diversity dimension, and threshold variation in formulaic language research: Extending the methodological considerations of O'Donnell et al.(2013). *International Journal of Corpus Linguistics, 23*(2), 158-182.

Lu, X., Yoon, J., & Kisselev, O. (2018). A phrase-frame list for social science research article introductions. *Journal of English for Academic Purposes, 36*, 76-85.

Luzón, M.-J. (2018). Features of online ELF in research group blogs written by multilingual scholars. *Discourse, Context & Media, 24*, 24-32.

MacKenzie, I. (2014). *English as a lingua franca: Theorizing and teaching English*. London: Routledge.

Marlina, R., & Xu, Z. (2018). English as a lingua franca. In J. Liontas (Ed.), *The TESOL Encyclopedia of English Language Teaching* (pp. 1-13). London: Wiley-BlackWell.

Marti, L., Yilmaz, S., & Bayyurt, Y. (2019). Reporting research in applied linguistics: The role of nativeness and expertise. *Journal of English for Academic Purposes, 40*, 98-114.

Martin, J. R., & Halliday, M. A. K. (1993). *Writing science: Literacy and discursive power*. London: Routledge.

Martinez-Garcia, M. T., & Wulff, S. (2012). Not wrong, yet not quite right: Spanish ESL students' use of gerundial and infinitival complementation. *International Journal of Applied Linguistics, 22*(2), 225-244.

Martinez, R. (2018). "Specially in the last years…": Evidence of ELF and non-native English forms in international journals. *Journal of English for Academic Purposes, 33*, 40-52.

Mauranen, A. (1993). Contrastive ESP rhetoric: Metatext in Finnish-English economics texts. *English for Specific Purposes, 12*(1), 3-22.

Mauranen, A. (2003). The corpus of English as lingua franca in academic settings. *TESOL quarterly, 37*(3), 513-527. doi:Doi 10.2307/3588402

Mauranen, A. (2005). English as a Lingua Franca - An unknown language? In G. Cortese & A. Duszak (Eds.), *Identity, community, discourse: English in intercultural settings* (Vol. 269293, pp. 269–293). Frankfurt: Peter Lang.

Mauranen, A. (2006a). A rich domain of ELF-the ELFA corpus of academic discourse. *Nordic Journal of English Studies, 5*(2), 145-159.

Mauranen, A. (2006b). Speaking the discipline: discourse and socialisation in ELF and L1 English. In K. Hyland & M. Bondi (Eds.), *Academic discourse across disciplines* (Vol. 271-294). Bern: Peter Lang.

Mauranen, A. (2007). Hybrid voices: English as the lingua franca of academics. In K. Flottum, T. Dahl, & T. Kinn (Eds.), *Language and discipline perspectives on academic discourse*. Cambridge: Cambridge Scholars Press.

Mauranen, A. (2009). Chunking in ELF: Expressions for managing interaction. *Intercultural Pragmatics, 6*(2), 217-233. doi:10.1515/Iprg.2009.012

Mauranen, A. (2011). Learners and users - Who do we want corpus data from? *Taste for Corpora: In Honour of Sylviane Granger, 45*, 155-171. doi:BOOK_DOI 10.1075/scl.45

Mauranen, A. (2012). *Exploring ELF: Academic English shaped by non-native speakers*. Cambridge: Cambridge University Press.

Mauranen, A. (2013). Hybridism, edutainment, and doubt: Science blogging finding its feet. *Nordic Journal of English Studies, 12*(1), 7-36.

Mauranen, A. (2015). What is going on in academic ELF? Findings and implications. . In P. Vettorel (Ed.), *New frontiers in teaching and learning English* (pp. 31-52). Newcastle upon Tyne: Cambridge Scholars Publishing.

Mauranen, A. (2016). *How do we write in English as a Lingua Franca?* Paper presented at the Centre for Global Englishes (CGE), University of Southampton, UK.

Mauranen, A. (2017). Conceptualising ELF. In J. Jenkins, W. Baker, & M. Dewey (Eds.), *The Routledge Handbook of English as a Lingua Franca* (pp. 7-24). London: Routledge.

Mauranen, A. (2018a). English as a lingua franca in written discourse. In J. I. Liontas (Ed.), *The TESOL Encyclopedia of English Language Teaching* (pp. 1-7).

Mauranen, A. (2018b). Second language acquisition, world Englishes, and English as a lingua franca (ELF). *World Englishes, 37*(1), 106-119. doi:10.1111/weng.12306

Mauranen, A. (2019). Academically speaking: Englsih as the lingua franca. In K. Hyland & L. L. C. Wong (Eds.), *Specialised English: New Directions in ESP and EAP Research and Practice* (pp. 9-21). London: Routledge.

Mauranen, A., Hynninen, N., & Ranta, E. (2010a). English as an academic lingua franca: The ELFA project. *English for Specific Purposes, 29*(3), 183-190.

Mauranen, A., Hynninen, N., & Ranta, E. (2016). English as the academic lingua franca. In K. Hyland & P. Shaw (Eds.), *The Routledge handbook of English for academic purposes* (pp. 68-79). London: Routledge.

Mauranen, A., Pérez-Llantada, C., & Swales, J. (2010b). Academic Englishes: A standardized knowledge? In A. Kirkpatrick (Ed.), *The Routledge handbook of World Englishes* (pp. 656-674). London: Routledge.

Mauranen, A., & Ranta, E. (2008). English as an academic lingua franca–the ELFA project. *Nordic Journal of English Studies, 7*(3), 199-202.

McIntosh, K., Connor, U., & Gokpinar-Shelton, E. (2017). What intercultural rhetoric can bring to EAP/ESP writing studies in an English as a lingua franca world. *Journal of English for Academic Purposes, 29*, 12-20.

McKinley, J., & Rose, H. (2018). Conceptualizations of language errors, standards, norms and nativeness in English for research publication purposes: An analysis of journal submission guidelines. *Journal of Second Language Writing, 42*, 1-11.

Melchers, G., Shaw, P., & Sundkvist, P. (2019). *World Englishes*. London: Routledge.

Milroy, J. (2006). The ideology of the standard language In C. Llamas, L. Mullany, & P. Stockwell (Eds.), *The Routledge companion to sociolinguistics* (pp. 133-139). Abingdon: Routledge.

Milroy, J., & Milroy, L. (2012). *Authority in language: Investigating standard English* (4th ed.). London: Routledge.

Mollin, S. (2006). *Euro-English: Assessing variety status*. Tübingen: Narr.

Morley, J. (2018). *Academic phrasebank*. Manchester: University of Manchester.

Mortensen, J. (2013). Notes on English used as a lingua franca as an object of study. *Journal of English as a Lingua Franca, 2*(1), 25-46.

Mukherjee, J., & Gries, S. T. (2009). Collostructional nativisation in new Englishes: Verb-construction associations in the International Corpus of English. *English World-Wide, 30*(27-51).

Mur-Dueñas, P. (2013). Spanish scholars' research article publishing process in English-medium journals: English used as a lingua franca? *Journal of English as a Lingua Franca, 2*(2), 315-340.

Mur-Dueñas, P. (2015). Looking into ELF variants: A study of evaluative it-clauses in research articles. *ESP Today, 3*(2), 160-179.

Mur-Dueñas, P. (2016). Modal hedging verbs in English as a Lingua Franca (ELF) Business Management research articles. *Kalbotyra*(69), 153-178.

Mur-Dueñas, P. (2018a). Exploring ELF manuscripts: An analysis of the anticipatory it pattern with an interpersonal function. In P. Mur-Dueñas & J. Ŝinkūienė (Eds.), *Intercultural perspectives on research writing* (pp. 277-297). Amsterdam: John Benjamins.

Mur-Dueñas, P. (2018b). Promotional strategies in academic writing: Statements of contribution in Spanish and ELF research articles. In J. Pelclová & W. Lu (Eds.), *Persuasion in public discourse: Cognitive and functional perspectives* (pp. 259-277). Amsterdam: John Benjamins.

Murillo-Ornat, S. (2019). Reformulation and its markers in unpublished research articles: Some evidence on the rhetorical patterns of written academic ELF. *Revista de Lenguas para Fines Específicos, 25*(2), 26-43.

Murillo, S. (2018). Not the same, but how different? Comparing the use of reformulation markers in ELF and in ENL research articles. In P. Mur-Dueñas & J. Ŝinkūienė (Eds.), *Intercultural perspectives on research writing* (pp. 237–253). Amsterdam: John Benjamins.

O'Donnell, M. B., Römer, U., & Ellis, N. C. (2013). The development of formulaic sequences in first and second language writing Investigating effects of frequency, association, and native norm. *International Journal of Corpus Linguistics, 18*(1), 83-108.

O'Neil, D. (2018). English as the lingua franca of international publishing. *World Englishes, 37*(2), 146-165.

Ortega, L. (2018). Ontologies of language, second language acquisition, and world Englishes. *World Englishes, 37*(1), 64-79.

Ortega, L. (2019). SLA and the study of equitable multilingualism. *The Modern Language Journal, 103*(S1), 23-38. doi:10.1111/modl.12525

Osmond, M. (1997). The prepositions we use in the construal of emotion: Why do we say fed up with but sick and tired of? In S. D. Rinemeier, R. (Ed.), *The language of emotions: Conceptualization, expression, and theoretical foundation* (pp. 111-133). Amsterdam: John Benjamins.

Östman, J.-O. (2005). Construction discourse: A prolegomenon. In J.-O. Östman & M. Fried (Eds.), *Construction grammars: Cognitive grounding and theoretical extensions* (Vol. 3, pp. 121-146). Amsterdam: John Benjamins Publishing.

Pakir, A. (2019). The Kachruvian paradigm and its range and depth of influence. *World Englishes, 38*(1-2), 174-184.

Palumbo, G. (2015). Studying ELF institutional web-based communication by universities: Comparison and contrast with English native texts. In P. Thompson & D. Giuliana (Eds.), *English for Academic Purposes: Approaches and Implications* (pp. 245-261). Cambridge: Cambridge Scholars Publishing.

Palumbo, G. (2017). Notes on investigating the native vs non-native distinction in written academic English. In G. Palumbo (Ed.), *Testi, corpora, confronti interlinguistici: approcci qualitativi e quantitativi*. Trieste: EUT Edizioni Università di Trieste.

Pan, F., Reppen, R., & Biber, D. (2016). Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in Telecommunications research journals. *Journal of English for Academic Purposes, 21*, 60-71.

Paquot, M. (2019). The phraseological dimension in interlanguage complexity research. *Second language research, 35*(1), 121-145.

Paquot, M., & Bestgen, Y. (2009). Distinctive words in academic writing: A comparison of three statistical tests for keyword extraction. In A. H. Jucker, D. Schreier, & M. Hundt (Eds.), *Corpora: Pragmatics and discourse* (pp. 247-269). Amsterdam & New York: Brill Rodopi.

Parkinson, J., & Musgrave, J. (2014). Development of noun phrase complexity in the writing of English for Academic Purposes students. *Journal of English for Academic Purposes, 14*, 48-59.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and Communication* (pp. 191-226). New York: Longman.

Pérez-Llantada, C. (2012). *Scientific discourse and the rhetoric of globalization: The impact of culture and language*. London: Contnuum.

Pérez-Llantada, C. (2013). 'Glocal' rhetorical practices in academic writing: An intercultural rhetoric approach to L2 English discoursal hybridisation. *European Journal of English Studies, 17*(3), 251-268.

Pérez-Llantada, C. (2014). Formulaic language in L1 and L2 expert academic writing: Convergent and divergent usage. *Journal of English for Academic Purposes, 14*, 84-94.

Phillips, M. (1989). *Lexical structure of text (Discourse analysis monographs no.12)*. Birmingham: University of Birmingham.

Phillipson, R. (1992). *Linguistic Imperialism*. Oxford: Oxford University Press.

Pinna, A. D., & Brett, D. (2018). Constance and variability: Using PoS-grams to find phraseologies in the language of newspapers. In J. Kopaczyk & J. Tyrkkö (Eds.), *Applications of Pattern-driven Methods in Corpus Linguistics* (Vol. Applications of pattern-driven methods in corpus linguistics, pp. 107-130).

Pirc, A. M. (2013). Construction grammar and" non-native discourse". *Theory & Practice in English Studies, 6*(1).

Pirc, A. M. (2014, 30-31 May). *The constructionist modelling of "non-native discourse"*. Paper presented at the 16th Rencontres Jeunes Chercheurs in Linguistics : Models and modeling in language sciences, Paris, France.

Pitzl, M. L. (2012). Creativity meets convention: Idiom variation and remetaphorization in ELF. *Journal of English as a Lingua Franca, 1*(1), 27-55.

Pitzl, M. L. (2018). *Creativity in English as a lingua franca: Idiom and metaphor*. Boston: Mouton de Gruyter.

Pitzl, M. L., Breiteneder, A., & Klimpfinger, T. (2008). A world of words: Processes of lexical innovation in VOICE. *VIEWZ, 17*(2), 21-46.

Prodromou, L. (2008). *English as a lingua franca: A corpus-based analysis*. London: A&C Black.

Qin, J. (2014). Use of formulaic bundles by non-native English graduate writers and published authors in applied linguistics. *System, 42*, 220-231.
Quirk, R., Greenbaum, S., & Leech, G. S., J. (1985). *A comprehensive grammar of the English language*. Harlow: Longman.
R Core Development Team. (2019). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/
Radden, G., & Dirven, R. (2007). *Cognitive English grammar* (Vol. 2). Amsterdam: John Benjamins.
Radden, G., & Matthis, E. (2002). Why similar to but different from. In H. Cuyckens & G. Radden (Eds.), *Perspectives on prepositions* (pp. 231-255). Berlin: De Gruyter.
Ranta, E. (2006). The 'attractive' progressive -Why use the -ing form in English as a lingua franca? *Nordic Journal of English Studies, 5*(2), 95-116.
Ranta, E. (2009). Syntactic features in spoken ELF–Learner language or spoken grammar. In A. Mauranen & E. Ranta (Eds.), *English as a lingua franca: Studies and findings* (pp. 84-106). Cambridge: Cambridge Scholars Publishing.
Ranta, E. (2017). Grammar in ELF. In J. Jenkins, W. Baker, & M. Dewey (Eds.), *The Routledge Handbook of English as a Lingua Franca* (pp. 244-254). London: Routledge.
Rauh, G. (1993). On the grammar of lexical and non-lexical prepositions in English. In C. Zelinsky-Wibbelt (Ed.), *The semantics of prepositions: From mental processing to natural language processing* (pp. 99-150). Berlin: De Gruyter.
Redeker, G. (1984). On differences between spoken and written language. *Discourse Processes, 7*(1), 43-55.
Richards, J. C., & Schmidt, R. (2010). *Longman dictionary of language teaching and applied linguistics*. Harlow: Longman.
Römer, U. (2009a). English in academia: Does nativeness matter? *Anglistik: International Journal of English Studies, 20*(2), 89-100.
Römer, U. (2009b). The inseparability of lexis and grammar: Corpus linguistic perspectives. *Annual Review of Cognitive Linguistics, 7*(1), 140-162.
Römer, U. (2010). Establishing the phraseological profile of a text type: The construction of meaning in academic book reviews. *English Text Construction, 3*(1), 95-119.
Römer, U. (2011a). Corpus research applications in second language teaching. *Annual Review of Applied Linguistics, 31*, 205.
Römer, U. (2011b). Observations on the phraseology of academic writing: Local patterns–local meanings. In T. Herbst, S. Faulhaber, & F. Uhrig (Eds.), *The phraseological view of language: A tribute to John Sinclair* (pp. 211-227). Berlin: De Gruyter Mouton.
Römer, U., O'Donnell, M. B., & Ellis, N. C. (2014). Second language learner knowledge of verb-argument constructions: Effects of language transfer and typology. *Modern Language Journal, 98*(4), 952-975.
Römer, U., Roberson, A., O'Donnell, M. B., & Ellis, N. C. (2014). Linking learner corpus and experimental data in studying second language learners' knowledge of verb-argument constructions. *ICAME Journal, 38*(1), 115-135.
Römer, U., Skalicky, S. C., & Ellis, N. C. (2018). Verb-argument constructions in advanced L2 English learner production: Insights from corpora and verbal fluency tasks. *Corpus Linguistics and Linguistic Theory*. doi:https://doi.org/10.1515/cllt-2016-0055.
Ronan, P., & Schneider, G. (2015). Determining light verb constructions in contemporary British and Irish English. *International Journal of Corpus Linguistics, 20*(3), 326-354.
Rowley-Jolivet, E. (2017). English as a Lingua Franca in research articles: the SciELF corpus. *ASp. la revue du GERAS*(71), 145-158.
Rowley-Jolivet, E., & Carter-Thomas, S. (2014). Citation practices of expert French writers of English: Issues of attribution and stance. In A. Łyda & K. Warchał (Eds.), *Occupying niches: Interculturality, cross-culturality and aculturality in academic research* (pp. 17-34). Cham: Springer.

Rozycki, W., & Johnson, N. H. (2013). Non-canonical grammar in Best Paper award winners in engineering. *English for Specific Purposes, 32*(3), 157-169.
Ruppenhofer, J., & Michaelis, L. A. (2010). A constructional account of genre-based argument omissions. *Constructions and Frames, 2*, 158-184.
Ruzaitė, J. (2007). Vague references to quantities as a face-saving strategy in teacher-student interaction. *Lodz papers in pragmatics, 3*(1), 157-178.
Samarin, W. (1987). Lingua franca. In U. Ammon, N. Dittmar, & K. Mattheier (Eds.), *Sociolinguistics: An International handbook of the science of language and society* (pp. 371-374). Berlin: Walter de Gruyter.
Schmid, H.-J. (1998). *Constant and ephemeral hypostatization: thing, problem and other shell nouns*. Paper presented at the 16th International Congress of Linguists, Paris.
Schmid, H.-J. (2000). *English abstract nouns as conceptual shells: From corpus to cognition* (Vol. 34). Berlin: Walter de Gruyter.
Schmid, P. (1994). *Probabilistic part-of-speech tagging using decision trees.* Paper presented at the International Conference on New Methods in Language Processing, , Manchester, UK.
Schneider, E. W. (2007). *Postcolonial English: Varieties around the world*. Cambridge: Cambridge University Press.
Schneider, E. W. (2012). Exploring the interface between World Englishes and second language acquisition–and implications for English as a lingua franca. *Journal of English as a Lingua Franca, 1*(1), 57-91.
Schneider, U. (2018). ΔP as a measure of collocation strength. *Corpus Linguistics and Linguistic Theory*. doi:https://doi.org/10.1515/cllt-2017-0036
SCImago. (2019). Retrieved December 2019, from Scopus http://www.scimagojr.com
Scott, M. (1997). PC analysis of key words—and key key words. *System, 25*(2), 233-245.
Scott, M., & Tribble, C. (2006). *Textual patterns: Key words and corpus analysis in language education*. Amsterdam: John Benjamins.
Seidlhofer, B. (2001). Closing a conceptual gap: The case for a description of English as a lingua franca. *International Journal of Applied Linguistics, 11*(2), 133-158.
Seidlhofer, B. (2002). Habeas corpus and divide et impera: 'Global English' and applied linguistics. In K. Spelman Miller & P. Thompson (Eds.), *Unity and diversity in language use* (pp. 198–217). London: Continuum.
Seidlhofer, B. (2004). Research perspectives on teaching English as a lingua franca. *Annual Review of Applied Linguistics, 24*, 209-239. doi:10.1017/s0267190504000145
Seidlhofer, B. (2009a). Accommodation and the idiom principle in English as a lingua franca. *Intercultural Pragmatics, 6*(2), 195-215.
Seidlhofer, B. (2009b). Common ground and different realities: World Englishes and English as a lingua franca. *World Englishes, 28*(2), 236-245.
Seidlhofer, B. (2009c). Orientations in ELF research: Formand function. In A. Mauranen & E. Ranta (Eds.), *English as a lingua franca: Studies and findings* (pp. 37-59). Cambridge: Cambridge Scholars Publishing.
Seidlhofer, B. (2011). *Understanding English as a lingua franca*. Oxford: Oxford University Press.
Seidlhofer, B. (2017). Standard English and the dynamics of ELF variation. In J. B. Jenkins, Will; Dewey, Martin (Ed.), *The Routledge Handbook of English as a Lingua Franca* (pp. 85-100). London: Routledge.
Seidlhofer, B., Breiteneder, A., & Pitzl, M. L. (2006). English as a lingua franca in Europe: Challenges for applied linguistics. *Annual Review of Applied Linguistics, 26*, 3-34.
Seidlhofer, B., & Widdowson, H. (2017). Competence, capability and virtual language. *Lingue e Linguaggi, 24*, 23-36.
Seoane, E. (2006). Changing styles: On the recent rvolution of scientific British and American English. In C. K. Dalton-Puffer, D., N. Ritt, & H. Schendl (Eds.), *Syntax, Style and Grammatical Norms: English from 1500-2000*. Bern: Peter Lang.

Seoane, E. (2013). On the conventionalisation and loss of pragmatic function of the passive in Late Modern English scientific discourse. *Journal of Historical Pragmatics, 14*(1), 70-99. doi:https://doi.org/10.1075/jhp.14.1.03seo

Seoane, E., & Hundt, M. (2018). Voice alternation and authorial presence: Variation across disciplinary areas in Academic English. *Journal of English Linguistics, 46*(1), 3-22.

Shchemeleva, I. (2018). "It Seems Plausible to Maintain that…": Clusters of Epistemic Stance Expressions in Written Academic Elf Texts. *ESP Today, 7*(1), 24-43. doi:10.18485/esptoday.2019.7.1.2

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics, 31*(4), 487-512.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sinclair, J. (1996). *Collins COBUILD English Guides 1: Prepositions*. London: Harper Collins.

Siyanova-Chanturia, A., & Pellicer-Sánchez, A. (2018). Formulaic language: Setting the scene. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding Formulaic Language: A Second Language Acquisition Perspective* (pp. 1-15). London: Routledge.

Solin, A., & Hynninen, N. (2018). Regulating the language of research writing: disciplinary and institutional mechanisms. *Language and Education, 32*(6), 494-510. doi:10.1080/09500782.2018.1511727

Staples, S., Egbert, J., Biber, D., & Gray, B. (2016). Academic writing development at the university level: Phrasal and clausal complexity across level of study, discipline, and genre. *Written Communication, 33*(2), 149-183.

Stefanowitsch, A., & Gries, S. T. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics, 8*(2), 209-243.

Strauss, S., Feiz, P., & Xiang, X. (2018). *Grammar, meaning, and concepts: A discourse-based approach to English grammar*. London: Routledge.

Stubbs, M. (2007). An example of frequent English phraseology: Distributions, structures and functions. In R. Facchinetti (Ed.), *Corpus linguistics 25 years on* (pp. 89-105). Amsterdam: Brill Rodopi.

Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Swales, J. (1998). Textography: Toward a contextualization of written academic discourse. *Research on language and social interaction, 31*(1), 109-121.

Tai, H.-Y. (2019). Investigation into Tense-Aspect Patterns of Chinese ELF Academic Writing—a Variationist's Approach. *English Teaching & Learning, 43*(3), 277-295.

Talmy, L. (2000). *Toward a cognitive semantics.* (Vol. 1). Cambridge: MIT Press.

Tåqvist, M. K. (2018). "A wise decision": Pre-modification of discourse-organising nouns in L2 writing. *Journal of Second Language Writing, 41*, 14-26.

Thomas, S., & Hawes, T. P. (1994). Reporting verbs in medical journal articles. *English for Specific Purposes, 13*(2), 129-148.

Thompson, G., & Yiyun, Y. (1991). Evaluation in the reporting verbs used in academic papers. *Applied Linguistics, 12*(4), 365-382.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge: Harvard University Press.

Tomasello, M. (2005). Beyond formalities: The case of language acquisition. *The Linguistic Review, 22*(2-4), 183-197.

Tomasello, M. (2011). Language development. In U. Goswami (Ed.), *The Wiley-Blackwell handbook of childhood cognitive development* (2 ed., Vol. 32, pp. 239-257). Oxford: Wiley-Blackwell.

Tribble, C. (2015). Writing academic English further along the road. What is happening now in EAP writing instruction? *Elt Journal, 69*(4), 442-462.

Tribble, C. (2017). ELFA vs. genre: A new paradigm war in EAP writing instruction? *Journal of English for Academic Purposes, 25*, 30-44.

Tribble, C. (2019). Expert, native or lingua franca? Paradigm choices in novice academic writer support. In H. Habibie & K. Hyland (Eds.), *Novice writers and scholarly publication: Authors, mentors, gatekeepers* (pp. 53-75). Cham: Palgrave Macmillan.

Trudgill, P. (2016). ELF and new-dialect formation. In M. Pitzl & R. Osimk-Teasdale (Eds.), *English as a Lingua Franca: Perspectives and Prospects: Contributions in Honour of Barbara Seidlhofer* (pp. 115-121). Berlin: de Gruyter Mouton.

Turner, J. (2018). *On writtenness: The cultural politics of academic writing*. New York: Bloomsbury Publishing.

Tyler, A. (2012). *Cognitive linguistics and second language learning: Theoretical basics and experimental evidence*: Routledge.

Tyler, A., & Evans, V. (2003). *The semantics of English prepositions: Spatial scenes, embodied meaning, and cognition*. Cambrridge: Cambridge University Press.

Valenzuela Manzanares, J., & Rojo López, A. M. (2008). What can language learners tell us about constructions? In D. R. S. De Knop & T, T. ((Ed.), *Cognitive approaches to pedagogical grammar: A volume in honour of René Dirven* (Vol. 9, pp. 197-230). Berlin: Walter de Gruyter.

Vassileva, I. (2000). *Who is the author?: A contrastive analysis of authorial presence in English, German, French, Russian, and Bulgarian academic discourse*. Sankt Augustin: Asgard Verlag.

Vassilieva, I. (1998). Who am I/who are we in academic writing? *International Journal of Applied Linguistics, 8*(2), 163-190.

Vetchinnikova, S. (2015). Usage-based recycling or creative exploitation of the shared code? The case of phraseological patterning. *Journal of English as a Lingua Franca, 4*(2), 223-252.

Vetchinnikova, S. (2017). On the relationship between the cognitive and the communal: A complex systems perspective. In M. Filppula, J. Klemola, A. Mauranen, & S. Vetchinnikova (Eds.), *Changing English: Global and local perspectives* (pp. 277-310). Berlin: De Gruyter.

Vetchinnikova, S. (2019). *Phraseology and the advanced Language Learner*. Cambridge: Cambridge University Press.

Vitali, M., & Bondi, M. (2020). Analysing academic ELF in economics. In S. Maci & M. Sala (Eds.), *Communicating English in specialised domains: A festschrift for Maurizio Gotti* (pp. 208-224). Cambridge: Cambridge Scholars Publishing.

VOICE. (2009). The Vienna-Oxford International Corpus of English. Retrieved 19 October 2019 http://voice.univie.ac.at/

Wahl, A. (2015). Intonation unit boundaries and the storage of bigrams: Evidence from bidirectional and directional association measures. *Review of Cognitive Linguistics., 13*(1), 191-219.

Wang, Y. (2017). Lexical bundles in spoken academic ELF. *International Journal of Corpus Linguistics, 22*(2), 187-211.

Wang, Y. (2018). Formulaic sequences signalling discourse organisation in ELF academic lectures: a disciplinary perspective. *Journal of English as a Lingua Franca, 7*(2), 355-376.

Whiteside, K., & Wharton, S. (2019). Semantic patterning of grammatical keywords in undergraduate academic writing from two close disciplines. *Journal of English for Academic Purposes, 39*, 1-20.

Widdowson, H. (1997). EIL, ESL, EFL: Global issues and local interests. *World Englishes, 16*(1), 135-146.

Widdowson, H. (2017). The cultural and creative use of English as a Lingua Franca. *Lingue e Linguaggi, 21*, 275-281.

Widdowson, H. (2019). Creativity in English. *World Englishes, 38*(1-2), 312-318.

Wong, M. L.-Y. (2014). Verb-preposition constructions in Hong Kong English: A cognitive semantic account. *Linguistics: An Interdisciplinary Journal of the Language Sciences, 52*(3), 603-635.

Wood, A. (2001). International scientific English: The language of research scientists around the world. In J. Flowerdew & M. Peacock (Eds.), *Research perspectives on English for academic purposes* (pp. 71-83). Cambridge: Cambridge University Press.

Wray, A. (2002). Formulaic language and the lexicon. In. Cambridge: Cambridge University Press.

Wray, A. (2018). Concluding question: Why don't second language learners more proactively target formulaic sequences? In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding Formulaic Language: A Second Language Acquisition Perspective* (pp. 248-269). London: Routledge.

WrELFA. (2015). The Corpus of Written English as a Lingua Franca in Academic Settings. http://www.helsinki.fi/elfa/wrelfa.html

Wu, X., Mauranen, A., & Lei, L. (2020). Syntactic complexity in English as a lingua franca academic writing. *Journal of English for Academic Purposes*, *43*, 100798. doi:https://doi.org/10.1016/j.jeap.2019.100798

Wulff, S. (2018). Acquisiton of formulaic language from a usage-based perspective. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding Formulaic Language: A Second Language Acquisition Perspective* (pp. 20-37). London: Routledge.

Wulff, S., & Gries, S. T. (2009). Corpus-driven methods for assessing complexity and accuracy in learner production. In P. Robinson (Ed.), *Researching task complexity: task demands, task-based language learning and performance* (pp. 61-87). Amsterdam: John Benjamins.

Wulff, S., & Gries, S. T. (2011). Corpus-driven methods for assessing accuracy in learner production. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 61-88). Amsterdam: John Benjamins.

Wulff, S., & Gries, S. T. (2019). Particle placement in learner language. *Language Learning, 69*(4), 873-910.

Wulff, S., Römer, U., & Swales, J. (2012). Attended/unattended this in academic student writing: Quantitative and qualitative perspectives. *Corpus Linguistics and Linguistic Theory*, *8*(1), 129-157.

Xiao, R., & McEnery, T. (2006). Collocation, semantic prosody, and near synonymy: A cross-linguistic perspective. *Applied Linguistics*, *27*(1), 103-129.

Yilmaz, S., & Römer, U. (2020). A Corpus-based exploration of constructions in written academic English as a lingua franca. In U. Römer, V. Cortes, & E. Friginal (Eds.), *Advances in corpus-based research on academic writing: Effects of discipline, register, and writer expertise* (pp. 59-88). Amsterdam: John Benjamins.

Zapletalová, G. (2018). Exploring formulaic clusters in L2 English expert writing: Academic ELF discourse. In C. Haase, N. Orlova, & J. Head (Eds.), *The Foundations and Versatility of English Language Teaching (ELT)* (pp. 15-32).

Zipf, G. K. (1935). *The psycho-biology of language: An introduction to dynamic philology.* . Cambridge: MIT Press.

APPENDICES

Appendix A Corpora across Disciplines

Appendix A.1 Disciplines in SCI: Natural Sciences

Domain	Discipline	L1	No. of texts	ELF_ Unedited	ELF_ Edited	ENL_ Edited
Natural sciences	Agriculture & forestry	Portuguese	1	2,189	2,853	4,722
		Finnish	2	9,045	12,279	11,984
	Applied mathematics	Spanish	1	4,417	5,183	4,965
	Applied physics	French	1	5,034	4,923	4,508
	Chemical engineering	Chinese	1	3,713	6,329	16,003
		Swedish	5	19,941	29,247	34,677
	Chemistry	French	1	6,608	3,921	7,228
		Italian	2	10,262	6,978	8,783
		Spanish	1	4,569	5,618	6,819
	Ecology	Finnish	1	4,628	4,164	4,126
	Economy	Finnish	1	4,804	6,188	5,233
		Italian	2	10,314	14,479	16,701
		Portuguese	1	2,942	5,693	7,608
		Russian	6	27,803	39,681	40,319
	Entomology	Czech	4	13,282	16,893	31,605
	Environmental engineering	French	2	11,917	10,824	11,462
	Environmental sciences	Chinese	2	6,259	9,436	12,450
		Finnish	1	9,742	8,489	12,166
		Swedish	1	3,449	4,625	1,950
		French	1	5,813	7,801	9,969
	Food technology	Spanish	1	4,223	2,789	7,452
	Earth sciences	Finnish	1	6,948	9,634	9,789
		Czech	6	32,305	50,697	39,271
		French	1	5,597	7,021	5,696
	Information sciences	Chinese	1	3,659	5,611	9,491
		Russian	1	5,172	5,935	6,439
	Materials sciences	French	1	6,164	5,061	6,317
		Portuguese	2	3,889	4,745	11,074
	Mechanical engineering	Chinese	3	10,319	15,904	29,343
		French	1	3,687	3,350	4,630
		Spanish	2	10,520	14,280	9,291
	Statistics	Portuguese	2	8,789	8,268	13,474
Total			**59**	**268,003**	**338,899**	**405,545**

Appendix A.2 Disciplines in SCI: Medicine

Domain	Discipline	L1	No. of texts	ELF_ Unedited	ELF_ Edited	ENL_ Edited
Medicine	Pharmacology	Chinese	3	15,342	11,572	18,761
	Neurology		1	2,551	5,052	4,737
	Otolaryngology	Finnish	1	2,555	2,270	2,750
	Health sciences		1	2,740	4,358	3,984
	Dentistry		1	2,389	4,725	3,334
	Ophthalmology		1	2,263	2,821	3,419
	Otorhinolaryngology		1	2,123	3,136	3,631
	Psychiatry		4	18,873	19,435	13,167
	Virology	Italian	1	3,556	3,672	2,964
	Gastroenterology		1	3,733	3,302	3,843
	Sports medicine	Spanish	1	4,760	4,799	3,760
Total			**16**	**60,885**	**65,142**	**64,350**

Appendix A.3 Disciplines in SSH: Social Sciences

Domain	Discipline	L1	No. of texts	ELF_ Unedited	ELF_ Edited	ENL_ Edited
Social Sciences	Cognitive psychology	French	1	8,009	6,514	9,823
	Economy	Czech	6	21,290	34,649	42,679
		Finnish	2	11,308	15,154	16,703
		Italian	1	3,128	7,564	3,495
		Russian	1	3,621	6,550	5,851
		Spanish	1	5,523	4,423	8,052
	Educationalsciences	Chinese	10	42,321	60,218	66,067
		Finnish	2	9,547	10,253	13,330
		Portuguese	2	12,203	12,105	13,142
		Swedish	2	9,039	17,331	15,664
	Information sciences	French	1	8,864	5,702	9,580
	Law	Spanish	1	8,595	3,130	5,601
	Political history	Russian	1	6,262	8,747	7,743
	Psychology	Swedish	1	5,371	5,809	5,339
	Science didactics	Spanish	1	5,815	3,804	6,398
	Social anthropology	Russian	1	7,460	9,635	5,529
	Social policy	Finnish	2	10,531	9,691	12,403
	Sociology	Italian	1	5,699	6,206	8,074
		Portuguese	1	4,790	8,790	7,692
		Swedish	1	6,675	7,828	5,817
		Russian	3	17,683	22,647	16,161
	Urban design	Spanish	1	5,832	8,067	3,236
Total			**42**	**213,734**	**266,750**	**285,143**

Appendix A.4 Disciplines in SSH: Humanities

Domain	Discipline	L1	No. of texts	ELF_ Unedited	ELF_ Edited	ENL_ Edited
Humanities	Cultural anthropology	Finnish	1	6,416	7,311	9,548
	History	Italian	1	3,398	4,677	3,521
		Swedish	1	6,195	5,344	9,323
	Literature	French	1	3,959	6,188	4,771
		Spanish	1	3,531	4,776	5,960
		Swedish	1	6,036	4,521	3,202
		Czech	1	5,079	5,076	9,058
	Philosophy	Italian	1	6,460	7,282	7,546
	Theatre	Finnish	1	7,112	6,245	6,201
	Linguistics	Czech	2	13,020	13,620	13,741
		Finnish	1	4,450	7,952	7,408
		French	5	23,392	35,657	49,183
		Italian	1	8,137	9,610	3,921
		Portuguese	3	18,175	16,293	20,599
		Spanish	1	4,296	6,834	6,496
Total			**22**	**119,656**	**141,386**	**160,478**

Appendix B List of Articles Included in the ELF_Edited Corpus

Adam, M. (2012). Existential there-construction as a means of presentation in narrative (A corpus-based syntactic-semantic analysis). *Linguistica Pragensia, 1*, 1-17.

Ali, S., Russo, M. A., & Margraf, L. (2009). Biliary rhabdomyoscarcoma mimicking choledochal cyst. *Journal of Gastrointestinal and Liver Diseases, 18*(1), 95-97.

Antipov, E. A., & Pokryshevskaya, E. B. (2017). Are box office revenues equally unpredictable for all movies? Evidence from a Random forest-based model. *Journal of Revenue and Pricing Management, 16*(3), 295-307. doi:10.1057/s41272-016-0072-y

Aragão, R. M. L. d. (2016). The IMRD overflow phenomenon: form and function of Japanese-language journal documents. *Revista de Lingüística y Lenguas Aplicadas, 11*(1). doi:10.4995/rlyla.2016.4652

Aronova, E. (2011). The politics and contexts of Soviet science studies (Naukovedenie): Soviet philosophy of science at the crossroads. *Studies in East European Thought, 63*(3), 175-202. doi:10.1007/s11212-011-9146-y

Audier, P., Rabat, H., Leroy, A., & Hong, D. (2014). Experimental investigation of a surface DBD plasma actuator at atmospheric pressure in different N2/O2gas mixtures. *Plasma Sources Science and Technology, 23*(6). doi:10.1088/0963-0252/23/6/065045

Aussenac-Gilles, N., & Sörgel, D. (2005). Supervised text analysis for ontology and terminology engineering *Applied Ontology, 1*(1), 35-46.

Autio, T. J., Koskenkorva, T., Narkio, M., Leino, T. K., Koivunen, P., & Alho, O. P. (2015). Diagnostic accuracy of history and physical examination in bacterial acute rhinosinusitis. *Laryngoscope, 125*(7), 1541-1546. doi:10.1002/lary.25247

Ayravainen, L., Leirisalo-Repo, M., Kuuliala, A., Ahola, K., Koivuniemi, R., Meurman, J. H., & Heikkinen, A. M. (2017). Periodontitis in early and chronic rheumatoid arthritis: a prospective follow-up study in Finnish population. *BMJ Open, 7*(1), e011916. doi:10.1136/bmjopen-2016-011916

Barrière, C. (2001). Investigating the causal relation in informative texts. *Terminology, 7*(2), 135-154.

Bockova, N., & Zizlavsky, O. (2016). The innovation and financial performance of a company: A study from Czech manufacturing industry. *Transformations in Business & Economics, 15*(3), 156-175.

Boula de Mareüil, P., Woehrling, C., & Adda-Decker, M. (2013). Contribution of automatic speech processing to the study of Northern/Southern French. *Language Sciences, 39*, 75-82. doi:10.1016/j.langsci.2013.02.006

Buss, L. (2016). Beliefs and practices of Brazilian EFL teachers regarding pronunciation. *Language Teaching Research, 20*(5), 619-637. doi:10.1177/1362168815574145

Buyantueva, R. (2018). LGBT rights activism and homophobia in Russia. *Journal of Homosexuality, 65*(4), 456-483. doi:10.1080/00918369.2017.1320167

Carlsson, H., Nordström, E., Bohlin, A., Petersson, P., Wu, Y., Collin, R., . . . Bai, X.-S. (2014). Large eddy simulations and rotational CARS/PIV/PLIF measurements of a lean premixed low swirl stabilized flame. *Combustion and Flame, 161*(10), 2539-2551. doi:10.1016/j.combustflame.2014.03.017

Carlsson, H., Nordström, E., Bohlin, A., Wu, Y., Zhou, B., Li, Z., . . . Bai, X.-S. (2015). Numerical and experimental study of flame propagation and quenching of lean premixed turbulent low swirl flames at different Reynolds numbers. *Combustion and Flame, 162*(6), 2582-2591. doi:10.1016/j.combustflame.2015.03.007

Cazzola, R., Russo-Volpe, S., Cervato, G., & Cestaro, B. (2003). Biochemical assessments of oxidative stress, erythrocyte membrane fluidity and antioxidant status in professional soccer players and sedentary controls. *European journal of clinical investigation, 33*(10), 924-930.

Cevolani, G. (2014). Truth approximation, belief merging, and peer disagreement. *Synthese, 191*(11), 2383-2401. doi:10.1007/s11229-014-0486-2

Chang, C.-F. (2012). Peer review via three modes in an EFL writing course. *Computers and Composition, 29*(1), 63-78.

Chen, M.-L. (2012). Effects of integrating children's literature and DVD films into a college EFL class. *English Teaching: Practice and Critique, 11*(4), 88-98.

Col, G., Aptekman, J., Girault, S., & Poibeau, T. (2012). Gestalt compositionality and instruction-based meaning construction. *Cognitive Processing, 13*(2), 151-170. doi:10.1007/s10339-011-0431-y

Collado, F. J. (2009). Preliminary design of surrounding heliostat fields. *Renewable Energy, 34*(5), 1359-1363. doi:10.1016/j.renene.2008.09.003

de Freitas, M. E., Bertero, C. O., Fleury, M. T. L., Mariotto, F. L., & Silva, A. L. (2016). Process of internationalization of business schools in Latin America: the case of EAESP/FGV, Brazil. *Tertiary Education and Management, 22*(4), 267-286. doi:10.1080/13583883.2016.1222630

De Oliveira, J. M., & Pagano, A. S. (2006). The research article and the science popularization article: a probabilistic functional grammar perspective on direct discourse representation. *Discourse studies, 8*(5), 627-646.

Deligant, M., Podevin, P., Descombes, G., Vidal, F., Lefebvre, A., & Lamquin, T. (2010). Computational fluid dynamics calculations of turbocharger's bearing losses. *SAE International Journal of Engines, 3*(2), 103-114.

Dolný, A., & Harabiš, F. (2012). Underground mining can contribute to freshwater biodiversity conservation: Allogenic succession forms suitable habitats for dragonflies. *Biological Conservation, 145*(1), 109-117. doi:10.1016/j.biocon.2011.10.020

Dolný, A., Harabiš, F., Bárta, D., Lhota, S., & Drozd, P. (2012). Aquatic insects indicate terrestrial habitat degradation: changes in taxonomical structure and functional diversity of dragonflies in tropical rainforest of East Kalimantan. *Tropical Zoology, 25*(3), 141-157. doi:10.1080/03946975.2012.717480

Drozdova, M., Sipos, J., & Drozd, P. (2013). Key factors affecting the predation risk on insects on leaves in temperate floodplain forest. *European Journal of Entomology, 110*(3), 469-476. doi:10.14411/eje.2013.062

Farias de Souza, L. M. (2013). Interlingual re-instantiation – a new systemic functional perspective on translation. *Text & Talk, 33*(4-5). doi:10.1515/text-2013-0026

Filimonov, V., & Sornette, D. (2012). Quantifying reflexivity in financial markets: toward a prediction of flash crashes. *Phys Rev E Stat Nonlin Soft Matter Phys, 85*(5 Pt 2), 056108. doi:10.1103/PhysRevE.85.056108

Gabriel, S. B., Panaino, J. V. P., Santos, I. D., Araujo, L. S., Mei, P. R., de Almeida, L. H., & Nunes, C. A. (2012). Characterization of a new beta titanium alloy, Ti–12Mo–3Nb, for biomedical applications. *Journal of Alloys and Compounds, 536*, S208-S210. doi:10.1016/j.jallcom.2011.11.035

Gadet, F., & Ludwig, R. (2014). Introduction: French Language(s) in Contact Worldwide. *Journal of Language Contact, 7*(1), 3-35. doi:10.1163/19552629-00701002

Galia, T., & Hradecký, J. (2014). Channel-reach morphology controls of headwater streams based in flysch geologic structures: An example from the Outer Western Carpathians, Czech Republic. *Geomorphology, 216*, 1-12. doi:10.1016/j.geomorph.2014.03.026

Galia, T., & Hradecký, J. (2014). Morphological patterns of headwater streams based in flysch bedrock: Examples from the Outer Western Carpathians. *Catena, 119*, 174-183. doi:10.1016/j.catena.2014.02.013

Galia, T., & Škarpich, V. (2016). Do the coarsest bed fractions and stream power record contemporary trends in steep headwater channels? *Geomorphology, 272*, 115-126. doi:10.1016/j.geomorph.2015.07.047

Gallice, A. (2018). Social status, preferences for redistribution and optimal taxation: a survey. *Economics: The Open-Access, Open-Assessment E-Journal*. doi:10.5018/economics-ejournal.ja.2018-55

Gil-Lacruz, A. I., & Marcuello, C. (2012). Voluntary work in Europe: Comparative analysis among countries and welfare systems. *Social Indicators Research, 114*(2), 371-382. doi:10.1007/s11205-012-0150-5

Gogo, S., Laggoun-Défarge, F., Merzouki, F., Mounier, S., Guirimand-Dufour, A., Jozja, N., . . . Défarge, C. (2015). In situ and laboratory non-additive litter mixture effect on C dynamics of Sphagnum rubellum and Molinia caerulea litters. *Journal of Soils and Sediments, 16*(1), 13-27. doi:10.1007/s11368-015-1178-3

Guimón, J., & Salazar-Elena, J. C. (2015). Collaboration in innovation between foreign subsidiaries and local universities: Evidence from Spain. *Industry & Innovation, 22*(6), 445-466. doi:10.1080/13662716.2015.1089034

Harabis, F., Dolny, A., Helebrandova, J., & Ruskova, T. (2015). Do egg parasitoids increase the tendency of Lestes sponsa (Odonata: Lestidae) to oviposit underwater? *European Journal of Entomology, 112*(1), 63-68. doi:10.14411/eje.2015.017

He, Z., Hu, M., Zha, Y. H., Li, Z. C., Zhao, B., Yu, L. L., . . . Qian, Y. (2014). Piracetam ameliorated oxygen and glucose deprivation-induced injury in rat cortical neurons via inhibition of oxidative stress, excitatory amino acids release and P53/Bax. *Cellular and Molecular Neurobiology, 34*(4), 539-547. doi:10.1007/s10571-014-0037-x

Hertzberg, F. (2014). Double gestures of inclusion and exclusion. Notions of learning outcomes, autonomy, and informed choices in Swedish educational and vocational guidance. *International Journal of Qualitative Studies in Education, 28*(10), 1203-1228. doi:10.1080/09518398.2014.974723

Hiilamo, H. (2012). Rethinking the role of church in a socio-democratic welfare state. *International Journal of Sociology and Social Policy, 32*(7/8), 401-414. doi:10.1108/01443331211249048

Hoikkala, L., Tammert, H., Lignell, R., Eronen-Rasimus, E., Spilling, K., & Kisand, V. (2016). Autochthonous dissolved organic matter drives bacterial community composition during a bloom of Filamentous Cyanobacteria. *Frontiers in Marine Science, 3*. doi:10.3389/fmars.2016.00111

Hokkanen, M. (2012). Imperial networks, colonial bioprospecting and Burroughs Wellcome & Co.: The Case of Strophanthus Kombefrom Malawi (1859–1915). *Social History of Medicine, 25*(3), 589-607. doi:10.1093/shm/hkr167

Hui, Z., Li, J., Xu, Q., Song, C., Zhang, J., Wu, F., & Zhao, Z. (2011). Miocene vegetation and climatic changes reconstructed from a sporopollen record of the Tianshui Basin, NE Tibetan Plateau. *Palaeogeography, Palaeoclimatology, Palaeoecology, 308*(3-4), 373-382.

Hung, R. (2015). A Critique of Confucian learning: On learners and knowledge. *Educational Philosophy and Theory, 48*(1), 85-96. doi:10.1080/00131857.2015.1084220

Iakovleva, T., Bay-Larsen, I., Kharitonova, G., & Didyk, V. (2012). Entrepreneurship and sustainability in nature-based tourism: The role of institutional profiles in Northern Norway and Northwest Russia. *Journal of Small Business & Entrepreneurship, 25*(4), 433-450. doi:10.1080/08276331.2012.10593583

Jiang, W. L., Fu, F. H., Zheng, S. G., Zhang, D. L., Zhu, H. B., & Jian, H. (2010). 8-O-acetyl shanzhiside methylester attenuates apoptosis and ameliorates mitochondrial energy metabolism in rat cortical neurons exposed to oxygen-glucose deprivation. *European Journal of Pharmacology, 629*(1-3), 20-24. doi:10.1016/j.ejphar.2009.11.065

Júnior, A. M. J., Reis, G. S., & Balancin, O. (2011). Influence of the microstructure on the plastic behaviour of duplex stainless steels. *Materials Science and Engineering: A, 528*(6), 2259-2264.

Junqueira, L. (2013). A genre-based investigation of applied linguistics book reviews in English and Brazilian Portuguese. *Journal of English for Academic Purposes, 12*(3), 203-213. doi:10.1016/j.jeap.2013.05.001

Juutinen, A., Kosenius, A.-K., & Ovaskainen, V. (2014). Estimating the benefits of recreation-oriented management in state-owned commercial forests in Finland: A choice experiment. *Journal of Forest Economics, 20*(4), 396-412. doi:10.1016/j.jfe.2014.10.003

Karlsson, H. K., Drabo, P., & Svensson, H. (2019). Precipitating non-aqueous amine systems for absorption of carbon dioxide using 2-amino-2-methyl-1-propanol. *International Journal of Greenhouse Gas Control, 88*, 460-468. doi:10.1016/j.ijggc.2019.07.001

Khmelnitskaya, M. (2014). Russian housing finance policy: state-led institutional evolution. *Post-Communist Economies, 26*(2), 149-175. doi:10.1080/14631377.2014.904104

Kopečková, L. (2012). Language policy in Georgia with focus on non-Georgian minorities. *Annual of Language & Politics and Politics of Identity, 6*(6), 111-127.

Koponen, T., Honkasalo, M.-L., & Rautava, P. (2017). Cultural plan model: integrating cultural and creative activities into care units for the elderly. *Arts & Health, 10*(1), 65-71. doi:10.1080/17533015.2017.1315436

Kormina, J., & Luehrmann, S. (2018). The social nature of prayer in a church of the unchurched: Russian Orthodox Christianity from its edges. *Journal of the American Academy of Religion, 86*(2), 394-424. doi:10.1093/jaarel/lfx055

Kröger, M. (2012). Neo-mercantilist capitalism and post-2008 cleavages in economic decision-making power in Brazil. *Third World Quarterly, 33*(5), 887-901. doi:10.1080/01436597.2012.674703

Krzemien, M., Maillart, C., Parisse, C., & Leroy, S. (2019). Impact of processing load on analogical mapping with visual sequences in children with developmental language disorders (DLD). *International Journal of Language & Communication Disorders, 54*(3), 417-429. doi:10.1111/1460-6984.12446

Kumpulainen, S. (2012). From sweat and tears towards sweat and harmony. *Theatre, Dance and Performance Training, 3*(2), 229-241. doi:10.1080/19443927.2012.691895

Lacour, S., Chinese, T., Alkadee, D., Perilhon, C., & Descombes, G. (2012). Energy and environmental balance of biogas for dual-fuel mobile applications. *Renewable and Sustainable Energy Reviews, 16*(3), 1745-1753. doi:10.1016/j.rser.2011.11.029

Lan, Y.-J., Sung, Y.-T., & Chang, K.-E. (2013). From particular to popular: Facilitating EFL reading through mobile-supported cooperation. *Language Learning & Technology, 17*(3), 23-38.

Larivée, A. (2012). Eros Tyrannos: Alcibiades as the model of the tyrant in book IX of the republic. *The International Journal of the Platonic Tradition, 6*(1), 1-26. doi:10.1163/187254712x619575

Lenart, J., Pánek, T., & Dušek, R. (2014). Genesis, types and evolution of crevice-type caves in the flysch belt of the Western Carpathians (Czech Republic). *Geomorphology, 204*, 459-476. doi:10.1016/j.geomorph.2013.08.025

Li, S., He, Z., Guo, L., Huang, L., Wang, J., & He, W. (2010). Behavioral alterations associated with a down regulation of HCN1 mRNA in hippocampal cornus ammon 1 region and neocortex after chronic incomplete global cerebral ischemia in rats. *Neuroscience, 165*(3), 654-661. doi:10.1016/j.neuroscience.2009.10.053

Li, Y., Yu, M., Zhao, B., Wang, Y., Zha, Y., Li, Z., . . . He, Z. (2018). Clonidine preconditioning improved cerebral ischemia-induced learning and memory deficits in rats via ERK1/2-CREB/ NF-kappaB-NR2B pathway. *European Journal of Pharmacology, 818*, 167-173. doi:10.1016/j.ejphar.2017.10.041

Liettu, A., Saavala, H., Hakko, H., Joukamaa, M., & Rasanen, P. (2012). Weapons used in serious violence against a parent: retrospective comparative register study. *Scandinavian Journal of Public Health, 40*(6), 563-570. doi:10.1177/1403494812455731

Lo, Y. Y., & Lo, E. S. C. (2014). A meta-analysis of the effectiveness of English-medium education in Hong Kong. *Review of Educational Research, 84*(1), 47-73. doi:10.3102/0034654313499615

Lombardi, D., & Siklos, P. L. (2016). Benchmarking macroprudential policies: An initial assessment. *Journal of Financial Stability, 27*, 35-49. doi:10.1016/j.jfs.2016.08.007

Lundgren, L., Skare, L., & Liden, C. (2006). Measuring dust on skin with a small vacuuming sampler: A comparison with other sampling techniques. *Annals of Occupational Hygiene, 50*(1), 95-103. doi:10.1093/annhyg/mei055

Lundmark, C., & Jonsson, G. (2013). Prospects for learning in river management: exploring the initial implementation of the Water Framework Directive in a Swedish river basin. *Environmental Education Research, 20*(2), 161-176. doi:10.1080/13504622.2013.780585

Lysell, R. (2018). Shelley's Prometheus unbound in the light of contemporary concepts of tragedy. *European Romantic Review, 29*(1), 25-35. doi:10.1080/10509585.2018.1417017

Macedo de Morais, R., Kazan, S., Inês Dallavalle de Pádua, S., & Lucirton Costa, A. (2014). An analysis of BPM lifecycles: from a literature review to a framework proposal. *Business Process Management Journal, 20*(3), 412-432. doi:10.1108/bpmj-03-2013-0035

Makinen, M., Marttunen, M., Komulainen, E., Terevnikov, V., Puukko-Viertomies, L. R., Aalberg, V., & Lindberg, N. (2015). Development of self-image and its components during a one-year follow-up in non-referred adolescents with excess and normal weight. *Child and Adolescent Psychiatry and Mental Health, 9*, 5. doi:10.1186/s13034-015-0038-7

Marin-Puyalto, J., Gomez-Cabello, A., Gonzalez-Aguero, A., Gomez-Bruton, A., Matute-Llorente, A., Casajus, J. A., & Vicente-Rodriguez, G. (2018). Is vibration training good for your bones? An overview of systematic reviews. *BioMed Research International, 2018*, 5178284. doi:10.1155/2018/5178284

Mazzotta, E., Ursini, T., Agostinone, A., Di Nicola, A. D., Polilli, E., Sozio, F., . . . Parruti, G. (2015). Prevalence and predictors of low bone mineral density and fragility fractures among HIV-infected patients at one Italian center after universal DXA screening: sensitivity and specificity of current guidelines on bone mineral density management. *AIDS Patient Care STDS, 29*(4), 169-180. doi:10.1089/apc.2014.0205

Mistro, J. C., Fazuoli, L. C., Gallo, P. B., de Oliveira, A. C. B., Toma-Braghini, M., & Silvar, M. B. (2007). Estimates of genetic parameters in Arabic coffee derived from the Timor hybrid. *Crop Breeding and Applied Biotechnology, 7*(2), 141-147.

Monclús, F.-J. (2003). The Barcelona model: and an original formula? From 'reconstruction' to strategic urban projects (1979–2004). *Planning Perspectives, 18*(4), 399-421. doi:10.1080/0266543032000117514

Moreno, A. I., Rey-Rocha, J., Burgess, S., López-Navarro, I., & Sachdev, I. (2012). Spanish researchers' perceived difficulty writing research articles for English-medium journals: The impact of proficiency in English versus publication experience. *Ibérica: Revista de la Asociación Europea de Lenguas para Fines Específicos, 24*, 157-183.

Müller, C. J., Lidin, S., de Debiaggi, S. R., Deluque Toro, C. E., & Guillermet, A. F. (2012). Synthesis, structural characterization, and ab initio study of Cu(5+delta)In(2+x)Sb(2-x): a new B8-related structure type. *Inorganic Chemistry, 51*(20), 10787-10792. doi:10.1021/ic3011794

Müller, L. (2012). The forgotten age of Swedish shipping: The eighteenth century. *International Journal of Maritime History, 14*(2), 1-18.

Muzzioli, S. (2012). The forecasting performance of corridor implied volatility in the Italian market. *Computational Economics, 41*(3), 359-386. doi:10.1007/s10614-012-9343-x

Mynttinen, S., Logrén, J., Särkkä-Tirkkonen, M., & Rautiainen, T. (2015). Perceptions of food and its locality among Russian tourists in the South Savo region of Finland. *Tourism Management, 48*, 455-466. doi:10.1016/j.tourman.2014.12.010

Nebout, P., Cagnon, B., Delpeux, S., Di Giusto, A., & Chedeville, O. (2016). Comparison of the Eeficiency of adsorption, ozonation, and ozone-activated carbon coupling for the removal of pharmaceuticals from water. *Journal of Environmental Engineering, 142*(2). doi:10.1061/(asce)ee.1943-7870.0001042

Nieminen, K. (2019). Eroding the protection against discrimination: The procedural and de-contextualized approach to S.A.S. v France. *International Journal of Discrimination and the Law, 19*(2), 69-88. doi:10.1177/1358229119838457

Norling, M., Sandberg, A., & Almqvist, L. (2015). Engagement and emergent literacy practices in Swedish preschools. *European Early Childhood Education Research Journal, 23*(5), 619-634. doi:10.1080/1350293x.2014.996423

Nummi, P., & Holopainen, S. (2014). Whole-community facilitation by beaver: ecosystem engineer increases waterbird diversity. *Aquatic Conservation: Marine and Freshwater Ecosystems, 24*(5), 623-633. doi:10.1002/aqc.2437

Nykänen, J. (2013). Identity, narrative and frames: Assessing Turkey's Kurdish initiatives. *Insight Turkey, 15*(2), 85-101.

Ozhegov, E. M., & Sidorovykh, A. S. (2017). Heterogeneity of sellers in housing market: Difference in pricing strategies. *Journal of Housing Economics, 37*, 42-51. doi:10.1016/j.jhe.2017.03.002

Pallaskorpi, S., Suominen, K., Ketokivi, M., Valtonen, H., Arvilommi, P., Mantere, O., . . . Isometsa, E. (2017). Incidence and predictors of suicide attempts in bipolar I and II disorders: A 5-year follow-up study. *Bipolar Disorders, 19*(1), 13-22. doi:10.1111/bdi.12464

Pan, S.-Y. (2007). Intertwining of academia and officialdom and university autonomy: Experience from Tsinghua University in China. *Higher Education Policy, 20*(2), 121-144. doi:10.1057/palgrave.hep.8300149

Pánek, T., & Klimeš, J. (2016). Temporal behavior of deep-seated gravitational slope deformations: A review. *Earth-Science Reviews, 156*, 14-38. doi:10.1016/j.earscirev.2016.02.007

Pena, B. M., & Gil Quilez, M. J. (2001). The importance of images in astronomy education. *International Journal of Science Education, 23*(11), 1125-1135. doi:10.1080/09500690110038611

Peterková, J., & Zapletalová, Š. (2018). Evaluation of the usability of selected innovation concepts for managing innovation activities. *E+M Ekonomie a Management, 21*(4), 141-158. doi:10.15240/tul/001/2018-4-010

Petit, M., Nemo, F., & Létang, C. (2016). Prosodic constraints on pragmatic interpretation: a new chapter in linguistic pragmatics. *Lodz Papers in Pragmatics, 12*(1), 103-121.

Petrini, R., Sansone, L., Slejko, F. F., Buccianti, A., Marcuzzo, P., & Tomasi, D. (2015). The (87)Sr/(86)Sr strontium isotopic systematics applied to Glera vineyards: a tracer for the geographical origin of the Prosecco. *Food Chemistry, 170*, 138-144. doi:10.1016/j.foodchem.2014.08.051

Platero, R. (2007). Love and the state: Gay marriage in Spain. *Feminist Legal Studies, 15*(3), 329-340. doi:10.1007/s10691-007-9064-z

Pokryshevskaya, E., & Antipov, E. (2014). A comparison of methods used to measure the importance of service attributes. *International Journal of Market Research, 56*(3), 283-296.

Pons-Branchu, E., Ayrault, S., Roy-Barman, M., Bordier, L., Borst, W., Branchu, P., . . . Dumont, E. (2015). Three centuries of heavy metal pollution in Paris (France) recorded by urban speleothems. *Science of the Total Environment, 518-519*, 86-96. doi:10.1016/j.scitotenv.2015.02.071

Procházka, O. (2018). A chronotopic approach to identity performance in a Facebook meme page. *Discourse, Context & Media, 25*, 78-87. doi:10.1016/j.dcm.2018.03.010

Qiao, L., Liu, J., Zheng, G., & Xiong, Z. (2009). Augmentation of XNAV system to an ultraviolet sensor-based satellite navigation system. *IEEE Journal of Selected Topics in Signal Processing, 3*(5), 777-785. doi:10.1109/jstsp.2009.2028380

Ren, W., Chen, Y.-s., & Lin, C.-Y. (2016). University students' perceptions of ELF in mainland China and Taiwan. *System, 56*, 13-27. doi:10.1016/j.system.2015.11.004

Ribeiro, A. G., Baquero, O. S., Freitas, C. U., Chiaravalloti Neto, F., Cardoso, M. R. A., Latorre, M., & Nardocci, A. C. (2019). Bayesian modeling of hematologic cancer and vehicular air pollution among young people in the city of Sao Paulo, Brazil. *International Journal of Environmental Health Research*, 1-11. doi:10.1080/09603123.2019.1608916

Riffault-Valois, L., Blanchot, L., Colas, C., Destandau, E., Pasquier, L., Andre, P., & Elfakir, C. (2017). Molecular fingerprint comparison of closely related rose varieties based on UHPLC-HRMS analysis and chemometrics. *Phytochemical Analysis, 28*(1), 42-49. doi:10.1002/pca.2641

Rizzi, M. (2013). Eusebius on Constantine and Nicaea: Intentions and Omissions. *Annali di Scienze Religiose, 6*, 209-223. doi:10.1484/j.Asr.1.103465

Rozanova, N. (2010). Competitive Interactions in the Information Technology Sector. *Problems of Economic Transition, 52*(10), 44-64. doi:10.2753/pet1061-1991521003

Rozenbaum, O., Blanchard, C., & De Sousa Meneses, D. (2019). Determination of high-temperature radiative properties of porous silica by combined image analysis, infrared spectroscopy and numerical simulation. *International Journal of Thermal Sciences, 137*, 552-559. doi:10.1016/j.ijthermalsci.2018.12.025

Sablin, I., & Sukhan, D. (2018). Regionalisms and imperialisms in the making of the Russian Far East, 1903–1926. *Slavic Review, 77*(2), 333-357. doi:10.1017/slr.2018.126

Sandman, N., Valli, K., Kronholm, E., Ollila, H. M., Revonsuo, A., Laatikainen, T., & Paunio, T. (2013). Nightmares: Prevalence among the Finnish general adult population and war veterans during 1972-2007. *Sleep, 36*(7), 1041-1050. doi:10.5665/sleep.2806

Sanz-Penella, J. M., Laparra, J. M., & Haros, M. (2014). Impact of alpha-amylase during breadmaking on in vitro kinetics of starch hydrolysis and glycaemic index of enriched bread with bran. *Plant Foods for Human Nutrition, 69*(3), 216-221. doi:10.1007/s11130-014-0436-7

Šebestová, J., & Nováková, K. (2013). Dynamic strategy for sustainable business development: mania or hazard? *Amfiteatru Economic Journal, 15*(34), 442-454.

Seppälä, S. (2015). An ontological framework for modeling the contents of definitions. *Terminology, 21*(1), 23-50. doi:10.1075/term.21.1.02sep

Šilhán, K., Galia, T., & Škarpich, V. (2016). Detailed spatio-temporal sediment supply reconstruction using tree roots data. *Hydrological Processes, 30*(22), 4139-4153. doi:10.1002/hyp.10923

Soininen, J. (2016). Spatial structure in ecological communities - a quantitative analysis. *Oikos, 125*(2), 160-166. doi:10.1111/oik.02241

Song, S., Xu, L., Zhang, H., & Bai, Y. (2015). Novel X-ray communication-based XNAV augmentation method using X-ray detectors. *Sensors (Basel), 15*(9), 22325-22342. doi:10.3390/s150922325

Sperandei, S., Bastos, L. S., Ribeiro-Alves, M., & Bastos, F. I. (2018). Assessing respondent-driven sampling: A simulation study across different networks. *Social Networks, 52*, 48-55. doi:10.1016/j.socnet.2017.05.004

Su, Z., Wang, H., Yao, P., Huang, Y., & Qin, Y. (2017). Back-stepping based anti-disturbance flight controller with preview methodology for autonomous aerial refueling. *Aerospace Science and Technology, 61*, 95-108. doi:10.1016/j.ast.2016.11.028

Sun, P., Li, Z., Wang, S., & Yin, X. (2018). Performance enhancement of polybenzimidazole based high temperature proton exchange membranes with multifunctional crosslinker and highly sulfonated polyaniline. *Journal of Membrane Science, 549*, 660-669. doi:10.1016/j.memsci.2017.10.053

Suutarinen, T. (2015). Local natural resource curse and sustainable socio-economic development in a Russian mining community of Kovdor. *Fennia – International Journal of Geography, 193*(1), 99-116. doi:10.11143/45316

Svensson, H., Hulteberg, C., & Karlsson, H. T. (2013). Heat of absorption of CO2 in aqueous solutions of N-methyldiethanolamine and piperazine. *International Journal of Greenhouse Gas Control, 17*, 89-98. doi:10.1016/j.ijggc.2013.04.021

Tahvonen, O., & Rämö, J. (2016). Optimality of continuous cover vs. clear-cut regimes in managing forest resources. *Canadian Journal of Forest Research, 46*(7), 891-901. doi:10.1139/cjfr-2015-0474

Tan, C. (2014). Education policy borrowing and cultural scripts for teaching in China. *Comparative Education, 51*(2), 196-211. doi:10.1080/03050068.2014.966485

Tenca, A., Jaakkola, T., Farkkila, M., Arola, J., & Kolho, K. L. (2019). Impact of paediatric onset primary sclerosing cholangitis on clinical course and outcome of inflammatory bowel disease: a case-control population-based study in Finland. *Scandinavian Journal of Gastroenterology, 54*(8), 984-990. doi:10.1080/00365521.2019.1648547

Torner, S., & Arias-Badia, B. (2019). Visual networks as a means of representing collocational information in electronic dictionaries. *International Journal of Lexicography, 32*(3), 270-295. doi:10.1093/ijl/ecz009

Tresaco, E., Elipe, A., & Riaguas, A. (2011). Computation of families of periodic orbits and bifurcations around a massive annulus. *Astrophysics and Space Science, 338*(1), 23-33. doi:10.1007/s10509-011-0925-1

Tuovinen, E.-L., Saarni, S. E., Kinnunen, T. H., Ollila, H., Ruokolainen, O., Patja, K., . . . Korhonen, T. (2018). Weight concerns as a predictor of smoking cessation according to nicotine dependence. *Nordic Studies on Alcohol and Drugs, 35*(5), 344-356. doi:10.1177/1455072518800217

Tuuminen, R., & Loukovaara, S. (2016). Statin medication in patients with epiretinal membrane is associated with low intravitreal EPO, TGF-beta-1, and VEGF levels. *Clinical Ophthalmology, 10*, 921-928. doi:10.2147/OPTH.S105686

Uría, J. (2010). Latin grammarians echoing the Greeks: The doctrine of proper epithets and the adjective. *Philologus, 154*(1), 97-118.

Varriale, S. (2015). Beyond distinction: Theorising cultural evaluation as a social encounter. *Cultural Sociology, 10*(2), 160-177. doi:10.1177/1749975515596447

Vechtomova, O., & Hao, Z. (2009). Articulating complex information needs using query templates. *Journal of Information Science, 35*(4), 439-452. doi:10.1177/0165551508101863

Vozyanov, A. (2018). Solution into problem: Ukrainian marshrutka and Romanian maxi-taxi at the fall of planning paradigms after 1990. *The Journal of Transport History, 39*(1), 25-40. doi:10.1177/0022526618757086

Wiksten, J., Blomgren, K., Roine, R. P., Sintonen, H., & Pitkaranta, A. (2013). Effect of tonsillectomy on health-related quality of life and costs. *Acta Otolaryngologica, 133*(5), 499-503. doi:10.3109/00016489.2012.728716

Wittbom, E. E. (2015). Management control for gender mainstreaming – a quest of transformative norm breaking. *Journal of Accounting & Organizational Change, 11*(4), 527-545. doi:10.1108/jaoc-08-2012-0069

Xie, X. (2009). Why are students quiet? Looking at the Chinese context and beyond. *ELT Journal, 64*(1), 10-20. doi:10.1093/elt/ccp060

Yang, L., & Gao, S. (2013). Beliefs and practices of Chinese university teachers in EFL writing instruction. *Language, Culture and Curriculum, 26*(2), 128-145. doi:10.1080/07908318.2013.794817

Zapata, J., Lopez, R., Herrero, P., & Ferreira, V. (2012). Multiple automated headspace in-tube extraction for the accurate analysis of relevant wine aroma compounds and for the estimation of their relative liquid-gas transfer rates. *Journal of Chromatography A, 1266*, 1-9. doi:10.1016/j.chroma.2012.10.015

Zapletalová, Š. (2015). Models of Czech companies' internationalization. *Journal of International Entrepreneurship, 13*(2), 153-168. doi:10.1007/s10843-015-0145-9

Zeng, H., & Zhang, J. (2016). Design of impulsive earth-moon halo transfers: Lunar proximity and direct options. *Astrophysics and Space Science, 361*(10). doi:10.1007/s10509-016-2888-8

Zhu, J., He, Y., Wang, J., Qiao, Z., Wang, Y., Li, Z., & Huang, M. (2017). Impact of aeration disturbances on endogenous phosphorus fractions and their algae growth potential from malodorous river sediment. *Environmental Science and Pollution Research, 24*(9), 8062-8070. doi:10.1007/s11356-017-8471-1

Zizlavsky, O. (2016). Innovation performance measurement: research into Czech business practice. *Economic Research-Ekonomska Istraživanja, 29*(1), 816-838. doi:10.1080/1331677x.2016.1235983

Appendix C List of Articles Included in the ENL_Edited Corpus

Allen, M. B., & Armstrong, H. A. (2012). Reconciling the intertropical convergence zone, Himalayan/Tibetan tectonics, and the onset of the Asian monsoon system. *Journal of Asian Earth Sciences, 44*, 36-47. doi:10.1016/j.jseaes.2011.04.018

Anderson, K. D., Pines, D. J., & Sheikh, S. I. (2015). Validation of pulsar phase tracking for spacecraft navigation. *Journal of Guidance, Control, and Dynamics, 38*(10), 1885-1897. doi:10.2514/1.G000789

Armitage, D., Dzyundzyak, A., Baird, J., Bodin, Ö., Plummer, R., & Schultz, L. (2018). An approach to assess learning conditions, effects and outcomes in environmental governance. *Environmental Policy and Governance, 28*(1), 3-14. doi:10.1002/eet.1781

Aronson, B., & Laughter, J. (2016). The theory and practice of culturally relevant education. *Review of Educational Research, 86*(1), 163-206. doi:10.3102/0034654315582066

Ayoub, P. M. (2014). With arms wide shut: threat perception, Norm reception, and mobilized resistance to LGBT rights. *Journal of Human Rights, 13*(3), 337-362. doi:10.1080/14754835.2014.919213

Bailey, A. S., Hughes, R. W., Hubberstey, P., Ritter, C., Smith, R. I., & Gregory, D. H. (2011). New ternary and quaternary barium nitride halides; synthesis and crystal chemistry. *Inorganic Chemistry, 50*(19), 9545-9553. doi:10.1021/ic201264u

Bartel, R. A., Haddad, N. M., & Wright, J. P. (2010). Ecosystem engineers maintain a rare species of butterfly and increase plant diversity. *Oikos, 119*(5), 883-890. doi:10.1111/j.1600-0706.2009.18080.x

Bartley, L., & Hidalgo-Tenorio, E. (2015). Constructing perceptions of sexual orientation: A corpus-based critical discourse analysis of transitivity in the Irish press. *Estudios Irlandeses, 10*(10), 14-34.

Bergmann, M. (2012). Rational disagreement after full disclosure. *Episteme, 6*(3), 336-353. doi:10.3366/e1742360009000756

Beverland, I. J., Robertson, C., Yap, C., Heal, M. R., Cohen, G. R., Henderson, D. E. J., . . . Agius, R. M. (2012). Comparison of models for estimation of long-term exposure to air pollution in cohort studies. *Atmospheric Environment, 62*, 530-539.

Bradshaw, M. (2013). Observations on the geographical dimensions of Russia's resource abundance. *Eurasian Geography and Economics, 47*(6), 724-746. doi:10.2747/1538-7216.47.6.724

Breen, M., Dilley, L. C., Kraemer, J., & Gibson, E. (2012). Inter-transcriber reliability for two systems of prosodic annotation: ToBI (Tones and Break Indices) and RaP (Rhythm and Pitch). *Corpus Linguistics and Linguistic Theory, 8*(2). doi:10.1515/cllt-2012-0011

Brown, K. D. (2013). Language policy and education: Space and place in multilingual Post-Soviet states. *Annual Review of Applied Linguistics, 33*, 238-257. doi:10.1017/s0267190513000093

Brust, T. B., Cayabyab, F. S., & MacVicar, B. A. (2007). C-Jun N-terminal kinase regulates adenosine A1 receptor-mediated synaptic depression in the rat hippocampus. *Neuropharmacology, 53*(8), 906-917. doi:10.1016/j.neuropharm.2007.09.001

Bunton, D. (2005). The structure of PhD conclusion chapters. *Journal of English for Academic Purposes, 4*(3), 207-224.

Cassidy, Y. M., McSorley, E. M., & Allsopp, P. J. (2018). Effect of soluble dietary fibre on postprandial blood glucose response and its potential as a functional food ingredient. *Journal of Functional Foods, 46*, 423-439. doi:10.1016/j.jff.2018.05.019

Chetty, S., & Campbell-Hunt, C. (2004). A strategic approach to internationalization: a traditional versus a “born-global” approach. *Journal of International Marketing, 12*(1), 57-81.

Chiverrell, R. C., Foster, G. C., Marshall, P., Harvey, A. M., & Thomas, G. S. P. (2009). Coupling relationships: Hillslope–fluvial linkages in the Hodder catchment, NW England. *Geomorphology, 109*(3-4), 222-235. doi:10.1016/j.geomorph.2009.03.004

Clay, A., & Tansley, G. D. (2011). Exploration of a simple, low cost, micro gas turbine recuperator solution for a domestic combined heat and power unit. *Applied Thermal Engineering, 31*(14-15), 2676-2684. doi:10.1016/j.applthermaleng.2011.04.037

Coffin, C., & O'Halloran, K. (2006). The role of appraisal and corpora in detecting covert evaluation. *Functions of Language, 13*(1), 77-110.

Coleman, J. M., McTigue, E. M., & Smolkin, L. B. (2017). Elementary teachers' use of graphical representations in science teaching. *Journal of Science Teacher Education, 22*(7), 613-643. doi:10.1007/s10972-010-9204-1

Cowling, M., Liu, W., & Ledger, A. (2012). Small business financing in the UK before and during the current financial crisis. *International Small Business Journal: Researching Entrepreneurship, 30*(7), 778-800. doi:10.1177/0266242611435516

Culver, K., Guilloteau, R., & Hue, C. (2011). Hard nodes in soft surroundings: A 'dream of islands' strategy for urban sustainability. *Development, 54*(3), 336-342. doi:10.1057/dev.2011.58

Daniel, E., Kodwani, D., & Datta, S. (2009). The impact of ICT-enabled offshoring announcements on share prices. *Journal of Enterprise Information Management, 22*(3), 241-256. doi:10.1108/17410390910949706

Davis, L. S. (2018). Political economy of growth with a taste for status. *Journal of Public Economics, 168*, 35-46. doi:10.1016/j.jpubeco.2018.10.002

Davis, M. A., Walsh, S. D., & Saar, M. O. (2011). Statistically reconstructing continuous isotropic and anisotropic two-phase media while preserving macroscopic material properties. *Physical Review E, 83*(2 Pt 2), 026706. doi:10.1103/PhysRevE.83.026706

de Aguirre, C. M. (2015). The evolution of family law: Changing the rules or changing the game? *BYU Journal of Public Law, 30*(2), 231-250.

Dunphy, P. J. (2006). Location and biosynthesis of monoterpenyl fatty acyl esters in rose petals. *Phytochemistry, 67*(11), 1110-1119. doi:10.1016/j.phytochem.2006.03.023

Durscher, R. J., & Roy, S. (2012). Three-dimensional flow measurements induced from serpentine plasma actuators in quiescent air. *Journal of Physics D: Applied Physics, 45*(3). doi:10.1088/0022-3727/45/3/035202

Ebell, M. H., McKay, B., Dale, A., Guilbault, R., & Ermias, Y. (2019). Accuracy of signs and symptoms for the diagnosis of acute ahinosinusitis and acute bacterial rhinosinusitis. *Annals of Family Medicine, 17*(2), 164-172. doi:10.1370/afm.2354

Elmer, J. W., Palmer, T. A., & Specht, E. D. (2007). Direct observations of sigma phase formation in duplex stainless steels using in-situ synchrotron x-ray diffraction. *Metallurgical and Materials Transactions A, 38*(3), 464-475. doi:10.1007/s11661-006-9076-3

Emery, R. L., Levine, M. D., Cheng, Y., & Marcus, M. D. (2015). Change in body weight does not mediate the relationship between exercise and smoking cessation among weight-concerned women smokers. *Nicotine Tobacco Research, 17*(9), 1142-1148. doi:10.1093/ntr/ntu284

Erlandson, K. M., Lake, J. E., Sim, M., Falutz, J., Prado, C. M., Domingues da Silva, A. R., . . . Guaraldi, G. (2018). Bone mineral density declines twice as quickly among HIV-infected women compared with men. *Journal of Acquired Immune Deficency Syndrome, 77*(3), 288-294. doi:10.1097/QAI.0000000000001591

Farrell, T. S. C., & Ives, J. (2014). Exploring teacher beliefs and classroom practices through reflective practice: A case study. *Language Teaching Research, 19*(5), 594-610. doi:10.1177/1362168814541722

Felton-Dansky, M. (2009). Echoes of Strindberg in contemporary Chile. *PAJ: A Journal of Performance and Art, 31*(3), 83-91.

Ferguson, L., & Goldberg, K. (2019). From the earth: Spirituality, medicine vessels, and consecrated bowls as responses to slavery in the South Carolina Lowcountry. *Journal of African Diaspora Archaeology and Heritage*, 1-29. doi:10.1080/21619441.2019.1690843

Ford, R. T. (2013). Headscarves, accommodation and the problem of joint costs. *Social Identities, 19*(6), 704-722. doi:10.1080/13504630.2013.842670

Francis, I. L., & Newton, J. L. (2005). Determining wine aroma from compositional data. *Australian Journal of Grape and Wine Research, 11*, 114-126.

Frank, R. A. (2000). Medical communication: Non-native English speaking patients and native English speaking professionals *English for Specific Purposes, 19*, 31-62.

Freestone, R. (2015). The exhibition as a lens for planning history. *Planning Perspectives, 30*(3), 433-446. doi:10.1080/02665433.2014.1002105

Frolking, S., Roulet, N. T., Moore, T. R., Richard, P. J. H., Lavoie, M., & Muller, S. D. (2001). Modeling northern peatland decomposition and peat accumulation. *Ecosystems, 4*(5), 479-498. doi:10.1007/s10021-001-0105-1

Geach, M. R., Stokes, M., Telfer, M. W., Mather, A. E., Fyfe, R. M., & Lewin, S. (2014). The application of geospatial interpolation methods in the reconstruction of Quaternary landform records. *Geomorphology, 216*, 234-246. doi:10.1016/j.geomorph.2014.03.036

Gile, K. J. (2011). Improved inference for respondent-driven sampling data with application to HIV prevalence estimation. *Journal of the American Statistical Association, 106*(493), 135-146. doi:10.1198/jasa.2011.ap09475

Gilheaney, O., Zgaga, L., Harpur, I., Sheaf, G., Kiefer, L., Bechet, S., & Walshe, M. (2017). The Prevalence of oropharyngeal dysphagia in adults presenting with temporomandibular disorders associated with rheumatoid arthritis: A systematic review and meta-analysis. *Dysphagia, 32*(5), 587-600. doi:10.1007/s00455-017-9808-0

Gillam, L., Tariq, M., & Ahmad, K. (2005). Terminology and the construction of ontology. *Terminology, 11*(1), 55-81.

Goode, J. R., & Wohl, E. (2010). Substrate controls on the longitudinal profile of bedrock channels: Implications for reach-scale roughness. *Journal of Geophysical Research: Earth Surface, 115*(F3).

Greed, C., & Reeves, D. (2005). Mainstreaming equality into strategic spatial policy making: are town planners losing sight of gender? *Construction Management and Economics, 23*(10), 1059-1070. doi:10.1080/01446190500372353

Greer, N., Fleuriet, K. J., & Cantu, A. G. (2012). Acrylic Rx: A program evaluation of a professionally taught painting class among older Americans. *Arts & Health, 4*(3), 262-273. doi:10.1080/17533015.2012.696072

Griffeth, R. W., Allen, D. G., & Barrett, R. (2006). Integration of family-owned business succession with turnover and life cycle models: Development of a successor retention process model. *Human Resource Management Review, 16*(4), 490-507. doi:10.1016/j.hrmr.2006.08.006

Groom, C. (2012). Non-native attitudes towards teaching English as a lingua franca in Europe. *English Today, 28*(1), 50-57. doi:10.1017/s026607841100068x

Groves, C., & Meiman, J. (2005). Weathering, geomorphic work, and karst landscape evolution in the Cave City groundwater basin, Mammoth Cave, Kentucky. *Geomorphology, 67*(1-2), 115-126. doi:10.1016/j.geomorph.2004.07.008

Hahn, S. L., Borton, K. A., & Sonneville, K. R. (2018). Cross-sectional associations between weight-related health behaviors and weight misperception among U.S. adolescents with overweight/obesity. *BMC Public Health, 18*(1), 514. doi:10.1186/s12889-018-5394-9

Hamilton, R. J., Giningele, M., Aswani, S., & Ecochard, J. L. (2012). Fishing in the dark-local knowledge, night spearfishing and spawning aggregations in the Western Solomon Islands. *Biological Conservation, 145*(1), 246-257. doi:10.1016/j.biocon.2011.11.020

Hart, M. L., Lamper, D., & Johnson, N. F. (2002). An investigation of crash avoidance in a complex system. *Physica A, 316*, 649–661.

Heaton, K., Kelly, S. D., Hoogewerff, J., & Woolfe, M. (2008). Verifying the geographical origin of beef: The application of multi-element isotope and trace element analysis. *Food Chemistry, 107*(1), 506-515. doi:10.1016/j.foodchem.2007.08.010

Herring, A. M. (2006). Inorganic–polymer composite membranes for proton exchange membrane fuel cells. *Journal of Macromolecular Science, Part C: Polymer Reviews, 46*(3), 245-296. doi:10.1080/00222340600796322

Hill, F., Reichart, R., & Korhonen, A. (2015). Simlex-999: Evaluating semantic models with (genuine) similarity estimation. *Computational Linguistics, 41*(4), 665-695. doi:10.1162/COLI

Huddleston, P., Whipple, J., & VanAuken, A. (2003). Food store loyalty: Application of a consumer loyalty framework. *Journal of Targeting, Measurement and Analysis for Marketing, 12*(3), 213-230.

Hughes, J., & Wilson, K. (2004). Playing a part: the impact of youth theatre on young people's personal and social development. *Research in Drama Education: The Journal of Applied Theatre and Performance, 9*(1), 57-72. doi:10.1080/1356978042000185911

Hynes, S., Tinch, D., & Hanley, N. (2013). Valuing improvements to coastal waters using choice experiments: An application to revisions of the EU Bathing Waters Directive. *Marine Policy, 40*, 137-144. doi:10.1016/j.marpol.2012.12.035

Jeffs, C. T., Kennedy, P., Griffith, P., Gripenberg, S., Markesteijn, L., & Lewis, O. T. (2018). Seed predation by insects across a tropical forest precipitation gradient. *Ecological Entomology, 43*(6), 813-822. doi:10.1111/een.12672

Joffe, R. T., MacQueen, G. M., Marriott, M., & Young, L. T. (2004). A prospective, longitudinal study of percentage of time spent ill in patients with bipolar I or bipolar II disorder. *Bipolar Disorders, 6*, 62-66.

Johnstone, N., Haščič, I., Poirier, J., Hemar, M., & Michel, C. (2012). Environmental policy stringency and technological innovation: evidence from survey data and patent counts. *Applied Economics, 44*(17), 2157-2170. doi:10.1080/00036846.2011.560110

Jones, W. P., Lyra, S., & Navarro-Martinez, S. (2012). Numerical investigation of swirling kerosene spray flames using Large Eddy Simulation. *Combustion and Flame, 159*(4), 1539-1561. doi:10.1016/j.combustflame.2011.10.025

Jossen, J., Annunziato, R., Kim, H. S., Chu, J., & Arnon, R. (2017). Liver transplantation for children with primary sclerosing cholangitis and autoimmune hepatitis: UNOS database analysis. *Journal of Pediatric Gastroenterology and Nutrition, 64*(4), e83-e87. doi:10.1097/MPG.0000000000001438

Kagawa, R. M. C., Gary, D. S., Wintemute, G. J., Rudolph, K. E., Pear, V. A., Keyes, K., & Cerda, M. (2019). Psychiatric disorders and gun carrying among adolescents in the United States. *Journal of Pediatrics, 209*, 198-203. doi:10.1016/j.jpeds.2019.01.010

Keck, C. (2006). The use of paraphrase in summary writing: A comparison of L1 and L2 writers. *Journal of Second Language Writing, 15*(4), 261-278. doi:10.1016/j.jslw.2006.09.006

Keller, D. P., & Hood, R. R. (2013). Comparative simulations of dissolved organic matter cycling in idealized oceanic, coastal, and estuarine surface waters. *Journal of Marine Systems, 109-110*, 109-128. doi:10.1016/j.jmarsys.2012.01.002

Kelly, L. B. (2018). Preservice teachers' ceveloping conceptions of teaching English learners. *TESOL Quarterly, 52*(1), 110-136. doi:10.1002/tesq.375

Kessler, G., Bikowski, D., & Boggs, J. (2012). Collaborative writing among second language learners in academic web-based projects. *Language Learning & Technology, 16*(1), 91-109.

LaCour-Little, M. (2007). The home purchase mortgage preferences of low- and moderate-income households. *Real Estate Economics, 35*(3), 265-290.

Lamb, M., & Wedell, M. (2014). Cultural contrasts and commonalities in inspiring language teaching. *Language Teaching Research, 19*(2), 207-224. doi:10.1177/1362168814541716

Lamond, G., Sandbrook, L., Gassner, A., & Sinclair, F. L. (2016). Local knowledge of tree attributes underpins species selection on coffee farms. *Experimental Agriculture, 55*(S1), 35-49. doi:10.1017/s0014479716000168

Langelier, B. C., & Esmaeili, S. (2009). In-situ laser-fabrication and characterization of TiC-containing Ti–Co composite on pure Ti substrate. *Journal of Alloys and Compounds, 482*(1-2), 246-252. doi:10.1016/j.jallcom.2009.03.168

Leibold, M. A., & Loeuille, N. (2015). Species sorting and patch dynamics in harlequin metacommunities affect the relative importance of environment and space. *Ecology, 96*(12), 3227-3233.

Leow, M., & Mues, C. (2012). Predicting loss given default (LGD) for residential mortgage loans: A two-stage model and empirical evidence for UK bank data. *International Journal of Forecasting, 28*(1), 183-195. doi:10.1016/j.ijforecast.2011.01.010

Love, E. W., & Nedorost, S. T. (2009). Fabric preferences of atopic dermatitis patients. *Dermatitis, 20*(1), 29-33. doi:10.2310/6620.2008.08027

Love, J. H., & Mansury, M. A. (2007). External linkages, R&D and innovation performance in US business services. *Industry & Innovation, 14*(5), 477-496. doi:10.1080/13662710701711380

Luke, T. W. (2018). Democracy and imperialism: The United States and three modes of empire. *Telos, 2018*(185), 9-34. doi:10.3817/1218185009

Macdonald, S. (2004). When means become ends: considering the impact of patent strategy on innovation. *Information Economics and Policy, 16*(1), 135-158. doi:10.1016/j.infoecopol.2003.09.008

Mason, W. (2015). Implementing continuous cover forestry in planted forests: Experience with Sitka Spruce (Picea Sitchensis) in the British Isles. *Forests, 6*(12), 879-902. doi:10.3390/f6040879

McCormack, K., Willems, J., van den Bergh, J., Deschoolmeester, D., Willaert, P., Indihar Štemberger, M., . . . Vlahovic, N. (2009). A global investigation of key turning points in business process maturity. *Business Process Management Journal, 15*(5), 792-815. doi:10.1108/14637150910987946

McDonald, D. M. (2010). Russian statecraft after the "imperial turn": The urge to colonize? *Slavic Review, 69*(1), 185-188.

McNeil, L. (2012). Using talk to scaffold referential questions for English language learners. *Teaching and Teacher Education, 28*(3), 396-404. doi:10.1016/j.tate.2011.11.005

Mester, L. J. (2017). The nexus of macroprudential supervision, monetary policy, and financial stability. *Journal of Financial Stability, 30*, 177-180. doi:10.1016/j.jfs.2017.07.003

Miller, B. (2013). Free to manage? A neo-liberal defence of academic freedom in British higher education. *Journal of Higher Education Policy and Management, 36*(2), 143-154. doi:10.1080/1360080x.2013.861055

Mueller, E. R. (2005). Morphologically based model of bed load transport capacity in a headwater stream. *Journal of Geophysical Research, 110*(F2). doi:10.1029/2003jf000117

Nassar-McMillan, S. C. (2014). A framework for cultural competence, advocacy, and social justice: Applications for global multiculturalism and diversity. *International Journal for Educational and Vocational Guidance, 14*(1), 103-118. doi:10.1007/s10775-014-9265-3

Noone, C. J., Torrilhon, M., & Mitsos, A. (2012). Heliostat field optimization: A new computationally efficient model and biomimetic layout. *Solar Energy, 86*(2), 792-803. doi:10.1016/j.solener.2011.12.007

Norcliffe, G. (2011). Neoliberal mobility and its discontents: Working tricycles in China's cities. *City, Culture and Society, 2*(4), 235-242. doi:10.1016/j.ccs.2011.11.006

Orford, S., & Radcliffe, J. (2007). Modelling UK residential dwelling types using OS Mastermap data: A comparison to the 2001 census. *Computers, Environment and Urban Systems, 31*(2), 206-227. doi:10.1016/j.compenvurbsys.2006.08.003

Organ, K., Koenig-Lewis, N., Palmer, A., & Probert, J. (2015). Festivals as agents for behaviour change: A study of food festival engagement and subsequent food choices. *Tourism Management, 48*, 84-99. doi:10.1016/j.tourman.2014.10.021

Owen, A. J., & Goffman, L. (2007). Acoustic correlates of inflectional morphology in the speech of children with specific language impairment and their typically developing peers. *Clinical Linguistics & Phonetics, 21*(7), 501-522. doi:10.1080/02699200701350488

Parker, J. S., & Born, G. H. (2008). Direct lunar halo orbit transfers. *The Journal of the Astronautical Sciences, 56*(4), 441-476.

Parry, L. E., Charman, D. J., & Blake, W. H. (2013). Comparative dating of recent peat deposits using natural and anthropogenic fallout radionuclides and Spheroidal Carbonaceous Particles (SCPs) at a local and landscape scale. *Quaternary Geochronology, 15*, 11-19. doi:10.1016/j.quageo.2013.01.002

Pell, M. D., Jaywant, A., Monetta, L., & Kotz, S. A. (2011). Emotional speech processing: disentangling the effects of prosody and semantic cues. *Cognition & Emotion, 25*(5), 834-853. doi:10.1080/02699931.2010.516915

Pottenger, A. J. (2018). The 'Servant of God': Divine favour and instrumentality under Constantine, 318–25. *Studies in Church History, 54*, 31-45. doi:10.1017/stc.2017.3

Prior, N., & Santoro, M. (2011). Critique and renewal in the sociology of music: Bourdieu and beyond. *Cultural Sociology, 5*(1), 121-138. doi:10.1177/1749975510389723

Purse, B. V., & Thompson, D. J. (2009). Oviposition site selection by Coenagrion mercuriale(Odonata: Coenagrionidae). *International Journal of Odonatology, 12*(2), 257-273. doi:10.1080/13887890.2009.9748344

Raftery, J. P., & Karim, M. N. (2017). Economic viability of consolidated bioprocessing utilizing multiple biomass substrates for commercial-scale cellulosic bioethanol production. *Biomass and Bioenergy, 103*, 35-46. doi:10.1016/j.biombioe.2017.05.012

Reid, S. E. (2016). The Khrushchev kitchen: Domesticating the scientific-technological revolution. *Journal of Contemporary History, 40*(2), 289-316. doi:10.1177/0022009405051554

Rogers, M. (2006). Structuring information in English. *The Translator, 12*(1), 29-64. doi:10.1080/13556509.2006.10799208

Rose, H., & McKinley, J. (2017). Japan's English-medium instruction initiatives and the globalization of higher education. *Higher Education, 75*(1), 111-129. doi:10.1007/s10734-017-0125-1

Rowlett, P. (2013). Do French speakers really have two grammars? *Journal of French Language Studies, 23*(1), 37-57. doi:10.1017/s095926951200035x

Ryan, J. (2011). Teaching and learning for international students: towards a transcultural approach. *Teachers and Teaching, 17*(6), 631-648. doi:10.1080/13540602.2011.625138

Sayer, P., & Ban, R. (2014). Young EFL students' engagements with English outside the classroom. *ELT Journal, 68*(3), 321-329. doi:10.1093/elt/ccu013

Sheaves, B., Porcheret, K., Tsanas, A., Espie, C. A., Foster, R. G., Freeman, D., . . . Goodwin, G. M. (2016). Insomnia, nightmares, and chronotype as markers of risk for severe mental illness: Results from a student population. *Sleep, 39*(1), 173-181. doi:10.5665/sleep.5342

Sheikh, S. I., Pines, D. J., Ray, P. S., Wood, K. S., Lovellette, M. N., & Wolff, M. T. (2006). Spacecraft navigation using X-ray pulsars. *Journal of Guidance, Control, and Dynamics, 29*(1), 49-63.

Sherratt, J. A. (2012). Numerical continuation methods for studying periodic travelling wave (wavetrain) solutions of partial differential equations. *Applied Mathematics and Computation, 218*(9), 4684-4694. doi:10.1016/j.amc.2011.11.005

Short, J. C., Broberg, J. C., Cogliser, C. C., & Brigham, K. H. (2010). Construct validation using computer-aided text analysis (CATA) an illustration using entrepreneurial orientation. *Organizational Research Methods, 13*(2), 320-347.

Slater, W. (2000). A modern-day aint? Metropolitan Ioann and the post-Soviet Russian orthodox church. *Religion, State and Society, 28*(4), 313-325. doi:10.1080/09637490220151425

Smith, G. G., Li, M., Drobisz, J., Park, H.-R., Kim, D., & Smith, S. D. (2013). Play games or study? Computer games in eBooks to learn English vocabulary. *Computers & Education, 69*, 274-286. doi:10.1016/j.compedu.2013.07.015

Smith, K. H., Anderson, C. J., Tao, W., Endo, K., Mumford, K. A., Kentish, S. E., . . . Stevens, G. W. (2012). Pre-combustion capture of CO2—Results from solvent absorption pilot plant trials using 30wt% potassium carbonate and boric acid promoted potassium carbonate solvent. *International Journal of Greenhouse Gas Control, 10*, 64-73. doi:10.1016/j.ijggc.2012.05.018

Smith, N., & Graham, T. (2017). Mapping the anti-vaccination movement on Facebook. *Information, Communication & Society, 22*(9), 1310-1327. doi:10.1080/1369118x.2017.1418406

Stockwell, J., Chen, Z., Niazi, M., Nosib, S., & Cayabyab, F. S. (2016). Protein phosphatase role in adenosine A1 receptor-induced AMPA receptor trafficking and rat hippocampal neuronal damage in hypoxia/reperfusion injury. *Neuropharmacology, 102*, 254-265. doi:10.1016/j.neuropharm.2015.11.018

Story, S., & Brigmon, R. L. (2017). Influence of triethyl phosphate on phosphatase activity in shooting range soil: Isolation of a zinc-resistant bacterium with an acid phosphatase. *Ecotoxicology and Environmental Safety, 137*, 165-171. doi:10.1016/j.ecoenv.2016.12.003

Stott, T. (2013). Review of research in fluvial geomorphology 2010–2011. *Progress in Physical Geography: Earth and Environment, 37*(2), 248-258. doi:10.1177/0309133313477124

Tauskela, J. S., Aylsworth, A., Hewitt, M., Brunette, E., & Blondeau, N. (2016). Failure and rescue of preconditioning-induced neuroprotection in severe stroke-like insults. *Neuropharmacology, 105*, 533-542. doi:10.1016/j.neuropharm.2016.02.007

Tauskela, J. S., Mealing, G., Comas, T., Brunette, E., Monette, R., Small, D. L., & Morley, P. (2003). Protection of cortical neurons against oxygen-glucose deprivation and N-methyl-d-aspartate by DIDS and SITS. *European Journal of Pharmacology, 464*(1), 17-25. doi:10.1016/s0014-2999(03)01371-2

Taylor, S. J., Yadav, P. K., & Zhang, Y. (2010). The information content of implied volatilities and model-free volatility expectations: Evidence from options written on individual stocks. *Journal of Banking & Finance, 34*(4), 871-881. doi:10.1016/j.jbankfin.2009.09.015

Taylor-Gooby, P. (2010). Opportunity and solidarity. *Journal of Social Policy, 40*(3), 453-470. doi:10.1017/s0047279410000681

Thomas, P. R., Bullock, S., Richardson, T. S., & Whidborne, J. F. (2015). Collaborative control in a flying-boom aerial refueling simulation. *Journal of Guidance, Control, and Dynamics, 38*(7), 1274-1289. doi:10.2514/1.G000486

Thong, G., Davies, K., Murphy, E., & Keogh, I. (2017). Significant improvements in quality of life following paediatric tonsillectomy: a prospective cohort study. *Irish Journal of Medical Science, 186*(2), 419-425. doi:10.1007/s11845-016-1398-6

Tierney, S. M., Friedrich, M., Humphreys, W. F., Jones, T. M., Warrant, E. J., & Wcislo, W. T. (2017). Consequences of evolutionary transitions in changing photic environments. *Austral Entomology, 56*(1), 23-46. doi:10.1111/aen.12264

Treleaven, N. C. W., Su, J., Garmory, A., & Page, G. J. (2019). An efficient method to reproduce the effects of acoustic forcing on gas turbine fuel injectors in incompressible simulations. *Flow, Turbulence and Combustion, 103*(2), 417-437. doi:10.1007/s10494-019-00020-4

Tyler, A., & Jan, H. (2016). Be going to and will: talking about the future using embodied experience. *Language and Cognition, 9*(3), 412-445. doi:10.1017/langcog.2016.10

Warde, P. (2018). Trees, trade and textiles: Potash imports and ecological dependency in British industry, c.1550–1770*. *Past & Present, 240*(1), 47-82. doi:10.1093/pastj/gty007

Weatherley-Singh, J., & Gupta, A. (2017). An ecological landscape approach to REDD + in Madagascar: Promise and limitations? *Forest Policy and Economics, 85*, 1-9. doi:10.1016/j.forpol.2017.08.008

Weichel, E. D., Martidis, A., Fineman, M. S., McNamara, J. A., Park, C. H., Vander, J. F., . . . Brown, G. C. (2006). Pars plana vitrectomy versus combined pars plana vitrectomy–scleral buckle for primary repair of pseudophakic retinal detachment. *Ophthalmology, 113*(11), 2033-2040.

Welsh, J. T. (2010). The grammarian C. Iulius Romanus and the "Fabula Togata". *Harvard Studies in Classical Philology, 105*, 255-285.

Wheatley, J. E., Bala, S., Barnes, D. C., Schoolderman, C., Jakab, G., Raynel, G., & Rayner, C. M. (2019). CO2 capture using phenoxide salts; alternatives to amine-based capture agents, and comparative speciation studies as components in solvent blends. *International Journal of Greenhouse Gas Control, 88*, 353-360. doi:10.1016/j.ijggc.2019.06.012

White, F. C. (2008). Beauty of soul and speech in Plato's Symposium. *The Classical Quarterly, 58*(1), 69-81. doi:10.1017/s0009838808000050

Wiggins, B. E., & Bowers, G. B. (2014). Memes as genre: A structurational analysis of the memescape. *New Media & Society, 17*(11), 1886-1906. doi:10.1177/1461444814535194

Williams, C. C., & Nadin, S. (2012). Beyond the commercial versus social entrepreneurship dichotomy: A case study of informal entrepreneurs. *Journal of Developmental Entrepreneurship, 17*(03). doi:10.1142/s1084946712500185

Williams, G. (2008). A multilingual matter: Sinclair and the bilingual dictionary. *International Journal of Lexicography, 21*(3), 255-266. doi:10.1093/ijl/ecn026

Wilson, J., Mantovan, N., & Sauer, R. M. (2020). The economic benefits of volunteering and social class. *Social Science Research, 85*, 102368. doi:10.1016/j.ssresearch.2019.102368

Wilson, P. B., & Madrigal, L. A. (2016). Associations between whole blood and dietary omega-3 polyunsaturated fatty acid levels in collegiate athletes. *International Journal of Sport Nutrition and Exercise Metabolism, 26*(6), 497-505. doi:10.1123/ijsnem.2015-0316

Wilson, S. J., Dickinson, D. K., & Rowe, D. W. (2013). Impact of an early reading first program on the language and literacy achievement of children from diverse language backgrounds. *Early Childhood Research Quarterly, 28*(3), 578-592. doi:10.1016/j.ecresq.2013.03.006

Wren, T. A., Lee, D. C., Hara, R., Rethlefsen, S. A., Kay, R. M., Dorey, F. J., & Gilsanz, V. (2010). Effect of high-frequency, low-magnitude vibration on bone and muscle in children with cerebral palsy. *Journal of Pediatric Orthopaedics, 30*(7), 732-738. doi:10.1097/BPO.0b013e3181efbabc

Zenker, M. J., Borden, R. C., & Barlaz, M. A. (2004). Biodegradation of 1,4-dioxane using trickling filter. *Journal of Environmental Engineering, 130*(9), 926-931. doi:10.1061/(asce)0733-9372(2004)130:9(926)