A Corpus-based Exploration of the Differences between Native and Non-native Texts in Academic English

Pičuljan, Mirna

Master's thesis / Diplomski rad

2016

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Rijeka, Faculty of Humanities and Social Sciences / Sveučilište u Rijeci, Filozofski fakultet u Rijeci**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:186:934036

Rights / Prava: In copyright/Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-05-14**

UNIVERSITY OF RIJEKA

FACULTY OF HUMANITIES AND SOCIAL SCIENCES

DEPARTMENT OF ENGLISH

Mirna Pičuljan

A CORPUS-BASED EXPLORATION OF THE DIFFERENCES BETWEEN NATIVE AND NON-NATIVE TEXTS IN ACADEMIC ENGLISH

Submitted in partial fulfilment of the requirements for the M.A. in English Language and Literature and Informatics at the University of Rijeka

Supervisor:

Dr Giuseppe Palumbo, University of Trieste

Co-supervisor:

Dr Marija Brala-Vukanović

September 2016

Abstract

This thesis explores the differences between native and non-native texts in academic English. The analysis was conducted on a corpus built by gathering scientific articles in the field of linguistics, written by academics coming from six different language backgrounds (English – representing the native speakers; Croatian, German, Italian, Polish and Spanish – representing the non-native speakers). The corpus was examined with the help of corpus analysis software (TagAnt and AntConc). The aim of the research was to determine whether there are differences between the native and non-native academic authors' writings. If such dissimilarities were to be found, the subsequent aim was to discover which forms they take, how salient they are and what they might indicate when it comes to characterizing the writings of the native and non-native speakers of English. The findings point to some differences with regard to some of the aspects analyzed in the two general batches of examined texts (i.e. native speakers vs non-native speakers).

Key words: native, non-native, academic writers, corpus, lexical bundles, parts of speech

Table of contents

List of tables

List of figures

List of abbreviations

AFL	Academic Formula List
CIA	Contrastive interlanguage analysis
CR	Contrastive rhetoric
EFL	English as a foreign language
ESL	English as a second language
L1	First language
L2	Second language
MI	Mutual information
NNS	Non-native speaker
NP	Noun phrase
NS	Native speaker
POS	Part of speech
PP	Prepositional phrase
VP	Verb phrase

1. Introduction

The native versus non-native dichotomy has proved to be a controversial topic in the past couple of decades, whether it is focusing on the role that the existence of these differences (or the lack of them) have for understanding second language acquisition and learning a foreign language, understanding the influence of the mother tongue on the use of English as a second or foreign language, predicting the possible errors non-native writers might make, identifying the native language of an author in forensic linguistics, discussing and understanding attitudes towards native and non-native English teachers or applying all these findings in developing different language learning curricula specifically tailored for different groups of learners and their needs.

This dichotomy is a topic of interest especially in the context of English, the language of academia. Many studies have focused on examining the 'nativeness' factor in the English language, particularly in recent years and with the development of corpus analysis software and various natural language processing tools. However, investigating the similarities and differences of the writings of native and non-native academic professionals has only just begun in the past couple of years.

In this thesis, I will examine a self-built corpus consisting of 1 native and 5 non-native authors' batches of texts with the help of concordance and POS tagging software, juxtapose the individual batches in several aspects and compare the results to those of previous studies in the field. The aim of this study is to explore scientific texts written in English by native and non-native academic professionals to see whether, in which way and to what extent do these writings differ.

I want to find out whether a global difference in used structures and styles in writing exists for different groups of highly educated people according to their linguistic background in order to understand whether it is possible to support the idea of the existence of 'nativeness' as a concept in writing. My hypothesis is that certain differences between the NSs and NNSs will be found when examining the corpora.

In line with what has been said, the paper is organized as follows:

Chapter 2 will provide a review of the existing studies in the field, focusing on several aspects of the native versus non-native writing in general and on the investigation of lexical bundles specifically. This will serve as a basis for what follows: building a corpus, establishing a methodology, analyzing and discussing data.

Chapter 3 will present the corpus built and used, the methodology, taxonomies and analyses applied to the corpus, as well as the results obtained from analyzing the corpus and different NS and NNS batches.

Chapter 4 will discuss and propose possible explanations for the results obtained through the corpus analysis.

Chapter 5, the last chapter of this thesis, will sum up the findings, discuss the original hypothesis in the light of new discoveries, provide a conclusion, state the limitations of the study and a few suggestions for further research.

2. Previous research

2.1. Native versus non-native writing

English has become the lingua franca of academia and research, even though its dominance might be slightly less pronounced in areas such as the social sciences and the humanities (Gnutzmann & Rabe, 2014). The impact that writing and publishing in English has on researchers is evident even from only skimming the list of projects and studies focused on the topic, such as "Publish in English or Perish in German?" by Gnutzmann & Rabe (2012)[1] and "Linguistic Imperialism" by Philipson (1992) or from reading descriptions of English as a "Tyrannosaurus rex" (Swales, 1997, p. 374). This increasing "anglicization" in the academy (Gnutzmann & Rabe, 2014, p. 31) exerts different effects on researchers from both a social perspective (i.e., as regards attitudes, levels of stress and expectations) and a linguistic one (i.e. in terms of lexical preferences, collocational usage, syntactic patterns, stylistic preferences). Hence, apart from the widely recognized advantages of sharing an academic lingua franca, some concerns are being raised as to the potential disadvantages which the use of a (foreign) lingua franca might pose for non-native writers of research publications in English. The question of publishing in English and the native (NS) versus non-native speaker (NNS) dichotomy have become hotly debated topics in the field of applied linguistics in particular (Kuteeva & Mauranen, 2014).

Even expert NNS writers might not always be aware of the influence that cross-linguistic transfer[2] exerts on their writings. This is where contrastive rhetoric (CR), a field of linguistic inquiry which focuses on the author's first language and culture in order to explain their influence on his/her writing in a second language steps in. CR investigates writing, learning and using another language and culture (Atkinson, 2004). As Mauranen (1996) claims, "[t]he 'logical' progression of a text is [...] not a straightforward reflection of a writer's ability to think clearly, but a product of

---

[1] *Publish in English or Perish in German (PPEG)* is an ongoing project investigating the current situation, problems and problem-solving strategies of German scientists publishing in English.

[2] Jarvis & Pavlenko (2007) define cross-linguistic transfer as "the application of linguistic structure of a speaker's native language in the context of a new, foreign language; on various levels" (in Berzak, Reichart & Katz, 2014).

culture and the textual resources of a language" (p. 195). Other authors also emphasize the importance of the effects that a writer's culture and his or her ways of categorizing the world have on linguistic choices. Kramsch & Lam (2009) say that, when writing in another language, an author has to deal with "foreign ways of organizing the world through language" (p. 58). They also emphasize that both native and non-native speakers of a language need to be schooled into academic literacy, given that the conventions used in writing are very different from those used in spoken language. Citing Kachru (1988), who claimed that "different language speaking communities have developed different conventions of writing", Hinkel (1994, p. 353) highlights the impact that a specific culture's concepts have on its underlying characteristics of writing. Her experiment showed that authors from different linguistic backgrounds have different approaches not only to writing but also to the analysis of a piece of text. Based on the native speakers' analyses of given texts, she claims that even "[t]he NNSs with many years of training in L2 writing do not seem to have the NS-like access to this common background knowledge and the contextual assumptions associated with L2 rhetorical notions and conventions and the appropriate rhetorical devices." (Hinkel, 1994, p. 372). In line with all of these claims, Atkinson (2004) argues that culture needs to be given more attention in CR research, as it represents one of the factors influencing authors in such a way that the patterns of coherence in their texts can be classified as either native or non-native-resembling. One of the dichotomies the author puts forward is the "big versus small culture" opposition (Atkinson, 2004, p. 285), emphasizing the common habit of categorizing cultures on the basis of national backgrounds, while, she claims, the concept is much more complex and there are many more layers of cultures to be taken into account. In agreement with Atkinson and taking into account the differences the 'small cultures' might exhibit, the culture I have focused on in this research may be defined as that of academic professionals, or, even more specifically, the culture of bilingual academic linguists who publish in (at least) two languages. While Atkinson (2004) fears that as corpus linguistics tools are becoming more powerful and widespread the notion of culture will be given even less attention than it currently receives, I must disagree – if anything, corpus linguistics and the emergence of new tools might provide more possibilities to engage in contrastive rhetoric research than in the past. The research reported on here is an attempt at backing up this argument: the study aims to provide results and insights that could not have been uncovered without the use of corpus linguistics tools.

Related to the idea of a distinct disciplinary culture is the research conducted by Gnutzmann and Rabe (2014), who focused on the ESL demands posed by the characteristics of a specific discipline. They interviewed German researchers coming from four different disciplines and came to interesting conclusions and results which differ according to the field in question. It seems, then, that both the language choices and language demands for the non-native authors depend, to a significant extent, on the discipline they do research in. This is confirmed by Ebrahimi and Motlagh (2016), whose research into the textual devices used in research abstracts written by NSs and NNSs across four disciplines revealed that the choice and frequency of device selection is influenced by the discipline as well as by the 'nativeness' factor. Also, according to Muresan and Perez-Lllantada's (2014) study and similarly to McGrath's (2014) research, NNSs in the social sciences base their language choices on more factors, including audience and research topic. These "ethnographically oriented research" findings (Gnutzmann & Rabe, 2014, p. 31) are important for analyzing results of any research on native versus non-native academic writing and, the authors argue, more attention should be given to disciplinary cultures as factors influencing writing styles and principles. However, as with cultural aspects, adherence to the conventions of a disciplinary culture does not automatically mean all the writings belonging to a certain category will exhibit the same characteristics, as there will always be a certain degree of individuality which cannot be pigeonholed. Nonetheless, even with the uniqueness of each and every text and its author is taken into account, some clusters of similarities are still expected to emerge, or as Becher puts it: "even between different institutions in the same system, the phenotypical variations can be substantial, but […] one can nonetheless clearly identify genotypical cultures in a particular setting" (Becher, 1994, p. 155, as cited in Gnutzmann & Rabe, 2014, p. 33). As the authors find this true for disciplinary cultures, I hope to find the same outcome in case of different linguistic backgrounds.

In their research, Gnutzmann & Rabe (2014) identified four main concepts of disciplines which affect linguistic choices and demands: rigidity of genre and language, different writing modes, language norms and nature of the data under study. As for the first concept - rigidity of genre and language - they found that writing is considered easier for the NNSs in those fields in which the structure is rigid and vocabulary limited, such as biology. In such fields, the language is much more formulaic, contains fixed phrases and a practice of recycling such constructions, as well as reusing prototypical sentences, templates and blueprints. As far as the field of linguistics is concerned, the interviewees' reports seem to "reflect the position of linguistics between an

empirically experimental science and a more humanities-oriented discipline" (Gnutzmann & Rabe, 2014, p. 34) with some structure provided, but much freedom left to the author. Concerning the third concept - language norms - the results also differ, with linguistics falling somewhere in between reliance on the native speakers' editorial process and the non-native authors', reviewers' and editors' dominance. It is interesting to note that two interviewees reported they ask both natives and non-natives for proofreading – depending on the importance they give to their own articles. Bringing this into connection with my own research, I must note here that there are several articles in my corpora for which the authors stressed the fact they were not edited by native speakers. For other articles no editing information could be found and some were presumably edited by NSs. The corpora of NNS texts that were used for the study are therefore a combination of native- and non-native-edited papers.

Connected to the concept of language norms applied to the NNS texts and the NS editorial process is the study by Yli-Jokipii & Jorgensen (2004), who examined the post-editorial makeup of Danish and Finnish authors' texts, or what they refer to as "academic journalese"[3] published in English. The native speakers had the task of editing these texts in order for them to successfully conform to the standards of UK English. The authors examined the linguistic changes (at the level of cohesion) altering the rhetorical characteristics (strategic aspects) of the texts and looked at the deletions and insertions made by the editors. Their motivation for the research originated from their own previous observations of differences between the relation of explicitness and implicitness in NNS writing and NS editing and from an interest in the differences between Danish and Finnish authors' writings in English.[4] They linked both to the cultural hypothesis stating that differences stem from different native language characteristics. For example, their expectation was that the Finnish writers would have problems with the use of the definitive article "the" as a rhetorical device, as they do not have an equivalent structure in their language. When examining the linguistic

---

[3] Defined by the authors as a "relatively recent type of texts written by researchers or professionals with a background in research that is easily available to large audiences outside the academic world. In particular, (…), texts that are freely available on the Internet." (Yli-Jokipii & Jorgensen, 2004, p. 342)

[4] Explicitness was also one of the focal points of a study by Scarcella (1984) through which she concluded that cultural differences influence the linguistic devices chosen as well as the way in which they are used and that explicitness is not valued to the same extent among NNSs and NSs. However, the author questions to which extent her conclusions are generalizable to NNSs of different linguistic backgrounds.

changes, the authors focused on cohesive devices (ellipsis and deletion, repetition and insertion, reference-derived strategies and conjunction and embedding) as determinants of the degree of explicitness or implicitness of a text. The results of pre- and post-edited texts' analysis showed that the editors employed both increasing and decreasing explicitness. No differences were found in the way that Danish and Finnish texts were treated nor were there differences concerning the problems the two authors' groups encountered in their writing process or the errors exhibited in their final versions. Even the characteristics previously believed to belong to the Finnish writers were found to be shared by the Danish authors.

2.2. Lexical bundles

Lexical bundles, defined as "recurrent lexical sequences" (Biber & Conrad, 1999, p. 182) represent one of the main focuses of my research. Their significance for the present study is best underlined following Biber, Johansson, Leech, Conrad & Finegan (1999), who state that "...producing natural, idiomatic English is not just a matter of constructing well-formed sentences, but of using well-tried lexical expressions in appropriate places" (p. 990). The importance of studying lexical bundles as part of the research on formulaic language is evident not only in corpus and applied linguistics, but in the broader areas of psycholinguistics, cognitive science, first- and second- language acquisition, language instruction and evaluation and testing (O'Donnell, Romer & Ellis, 2013). Although some previous studies focused on examining the use of lexical bundles by student NNSs versus NSs (Adel & Erman, 2012; Chen & Baker, 2010; O'Donnell, Romer & Ellis, 2013) or by student versus expert writers (Chen & Baker, 2010; Cortes, 2004), Pan, Reppen & Biber (2016) were the first to directly address the relation of the use of lexical bundles by L1 and L2[5] academic professionals. In their paper on the comparison of patterns of L1 and L2 English academic professionals, they analyzed lexical bundles found in telecommunications research journals. Their results showed great differences on both structural and functional levels. As was already noted, no previous research on the differences between NS and NNS has been carried out

---

[5] The distinction between the first (L1) and second (L2) language reflects that between the native and non-native speaker. However, in this study, all of the authors (to my knowledge) learned English as a foreign language rather than as an L2.

on a sample of academic writers, thus it is no surprise that researchers disagree over whether there is a difference between academic and other writers (whether NSs or NNSs) concerning the use of lexical bundles and its relation to levels of English-language proficiency. Pan, Reppen & Biber (2016) only refer to the study by Perez Llantada (2014), in which L1 and L2 English professional writers were compared, with the results revealing register as one of the crucial determiners of lexical bundle usage. The other studies mentioned above (i.e. Chen & Baker, 2010 and Cortes, 2004, both based on samples with different proficiency levels) found that L1 English writers used more of both tokens and types[6] of lexical bundles (Adel & Erman, 2012; Chen & Baker, 2010). However, some studies indicate that lower proficiency L2 writers use more lexical bundles than more expert L2 writers (Staples, Egbert, Biber & McClair, 2013), while on the other hand specific bundles tend to be less frequent in student than expert writings, regardless of the L1/L2 distinction. O'Donnell, Romer & Ellis (2013) looked at the frequency-defined and mutual information[7] (MI)-defined formulas used to identify lexical bundles. The results of their study showed that expert NS and both NNS and NS graduate writers used more frequency-defined n-grams than NNS and NS undergraduate authors. The authors attributed these differences to the level of expertise as opposed to the native versus non-native dichotomy. Also, the MI scores showed that the use of specific formulas, or rhetorical devices, marks the difference between novice and expert writers rather than that between the NS and NNS. However, it is important to note that "different definitions of formulaic language produce different patterns of relationship with expertise and L1/L2 status" (O'Donnell, Romer & Ellis, 2013, p. 102). Interestingly, the authors found no differences between NSs and NNSs when it comes to the use of lexical bundles included in the Academic Formula List (AFL)[8], suggesting that "learning these, for natives and non-natives alike, is akin to learning another language", as "it takes a great amount of experience and instruction to become idiomatic in particular specialist genres of English for Academic Purposes" (O'Donnell, Romer & Ellis, 2013, p. 102). To sum up, all of the previous findings on the topic might be interpreted in terms of

---

[6] Tokens represent frequencies, while types represent different categories of lexical bundles.

[7] Mutual information (MI) is a measure of the dependence between two random variables. High MI is an indication of a large reduction in uncertainty about one random variable given knowledge of another; low MI indicates a small reduction; and zero MI indicates the variables are independent. (Latham & Roudi, 2009)

[8] "The AFL includes formulaic sequences identified as (i) frequent recurrent patterns in corpora of written and spoken language, which (ii) occur significantly more often in academic than in non-academic discourse, and (iii) inhabit a wide range of academic genres." (Simpson-Vlach & Ellis, 2010, p. 487)

lower- versus higher-level proficiency in academic writing rather than L1/NS versus L2/NNS proficiency (Romer, 2009).

There are different approaches when it comes to the analysis of the use of lexical bundles: comparing lists categorized by register or authors' proficiency levels/age/gender/field of study; comparing their structural and functional traits; or studying their discourse functions. In their study, Pan, Reppen & Biber (2016) focused on the last two concepts: structural and functional traits, and discourse functions. They controlled for variables that might influence the research outcome: they examined only those texts coming from the same discipline and register – published academic research articles in the field of telecommunications – and made sure that the L2 authors shared the same linguistic background – the Chinese language.[9] Their results revealed that L2 writers use more different lexical bundles than their L1 peers. When it comes to the structural types of the bundles used, it was found that L1 writers use more phrasal (NP and PP-based tokens) lexical bundles, while L2 writers use clausal (VP-based tokens) bundles more frequently. These results were backed up by the study conducted by Güngör & Uysal (2016), who extracted categorized lexical bundles from research articles according to the taxonomies of Biber, Johansson, Leech, Conrad & Finegan (1999) and Hyland (2008). The authors found that Turkish scholars overused clausal bundles, while their native peers relied more heavily on phrasal bundles. Comparing these results with those of previous studies investigating the connection between proficiency level and the use of lexical bundles (which show that novice L1 and L2 writers rely on clausal bundles, while experts rely on phrasal bundles), the authors argue that the developmental pathways of L1 and L2 writers can be seen as similar. In other words, it seems that the switch from clausal to phrasal style of writing happens as the writer's proficiency level reaches a new stage, which is not an easy process for either the native or the non-native writer. Still, their study shows that L1 writers are at an advantage when it comes to the transition, even at the very high level of proficiency which characterizes all academic writers share. When it comes to comparing the functional characteristics of lexical bundles, the results indicate that L1 writers use more referential expressions than the L2 writers, an outcome also indicated by Chen & Baker (2010).

Pan, Reppen & Biber (2016) also found that L1 writers use fewer stance bundles than expected, while L2 writers use more than they would have predicted. The authors claim that in that

[9] A similar design has been used for the present study; see section 3.1 Corpus and methodology.

way the L2 writers "demonstrate better control expressing the degree of doubt and certainty" (p. 69). However, it seems the non-native professionals also tend to use evaluative bundles composed of subjective adjectives, which may make them "vulnerable to violating discourse norms by making their writing appear too personal" (Pan, Reppen & Biber, 2016, p. 69). Looking at this on a more detailed level, that of specific functional subcategories, it is interesting to note that the authors attribute the rare use of section bundles[10] by the L2 writers to a lack of strong audience orientation when compared to the L1 writers, a point already made by Scarcella (1984), who investigated attention-securing devices[11] and found that NNSs have difficulties using them, and Silva (1993), who asserted that NNSs tend to underestimate the reader's knowledge. Wrapping up their paper, Pan, Reppen & Biber (2016) conclude that the differences found cannot be attributed exclusively to either L1 versus L2 writing differences or to the novice versus expert writer dichotomies, but to an "interplay of L1 and expertise" (p. 70).

Another study centered on NNSs' versus NSs' use of lexical bundles is that of the already mentioned Chen & Baker (2010), who compared native expert writing, native student writing and L2 student writing. Their results indicated that L2 student writers used the smallest range of bundles, while at the same time overused certain lexical combinations, which their native peers, both novice and expert writers, rarely used. When looking at the L1 and L2 student writings compared to those of experts, it was found that both student groups underuse certain expressions and that, once again, they prefer using clausal to phrasal bundles, serving as evidence for the hypothesis that both NNS and NS writers share at least some features or stages of development when it comes to academic writing. Overall, Chen and Baker (2010) claim that the range of lexical bundles increases in terms of both types and tokens according to the writer's stage of development or proficiency. However, they warn that this result might be related to corpus size, as larger corpora exhibits fewer lexical bundles, and that this might be one of the reasons their study disagrees with

---

[10] The authors examine three *section* bundles (subcategory of research-oriented bundles), namely: *in the next section, in the previous section, in this section we*

[11] The author looked into cataphoric reference, interrogatives, direct assertion, structural repetition, short, abrupt elements, sentence-initial adverb + verb sequence and historical context. She found that NNSs used 50% less of the listed elements and had a much smaller width of their range, some of the elements never being used at all. Also, she noticed that NNSs relied heavily on only two of these devices - historical context and direct assertion, using them differently from the NSs.

some outcomes of other studies, such as those by De Cock (2000) or Hyland (2008)[12] and conclude that it is not possible to determine if a relationship exists at all between the frequency of bundles and proficiency.

Most of the other studies in the field center on the differences in the use of collocations, discourse markers and connectors as cohesion devices. As regards collocations, Durrant & Schmitt (2009) compared L1 and L2 student writing and found that NS authors use more of the low frequency collocations than do the non-native writers, while part of the non-natives also overuse strong collocations. The authors attributed these differences to the NNSs' "conservatism" – a preference for common and more frequent items, which "creates the feeling that non-native writing lacks 'idiomaticity'" (p. 175).

As for discourse markers, Muller (2004) emphasizes their importance as a feature that might differentiate native speakers' writing from that of non-native speakers. Siepmann (2005) confirms this claim, with his research results indicating NNSs' "unnatural writing as a result of overt errors and unusual frequency of occurrence of particular items" (p. xii). In order to avoid such "unnatural writing", the authors must especially pay attention to cohesion, with connectors representing one of its essential devices. Inappropriate use of connectors might even cause problems in message communication and meaning transfer (Ma & Wang, 2016).

Cohesion is the main focus of the study on connector usage conducted by Granger & Tyson (1996) as part of their contrastive interlanguage analysis (CIA).[13] The authors' hypothesis that there is an over-usage of connectors in the case of French authors (which would be attributable to language transfer) was shown to be invalid. However, they noticed another pattern, i.e. NNSs displaying a tendency to overuse and underuse connectors according to their function (e.g., overusing corroborating connectors, underusing connectors for contrasting and developing an argument). The authors went a step further, presuming that these differences might even cause

---

[12] Other reasons for the disagreement of results include the difference in dispersion requirements, and the inclusion or exclusion of certain types of bundles (e.g., while Hyland included topic-related ones in his research, Chen & Baker did not consider them).

[13] CIA is an area of research "which involves comparing and contrasting what non-native and native speakers of a language do in a comparable situation. (..) the different non-native English varieties are compared with native speaker English and with each other. The results of this interlanguage analysis are then examined in the light of classic contrastive analysis of the native languages (…)" (Granger & Tyson, 1996, p. 18)

different types of argumentation to be attributable to the NSs and NNSs, but leaving it for further research. In another study of connector usage, Goldman & Murray (1989) found that NNSs displayed a greater tendency to erroneous use, while Ma & Wang (2016) in their study of L1 versus L2 student writing demonstrated that some of the connectors are used more frequently by the NNSs and sometimes even misused. It is interesting to note that the authors ask whether the low frequency of the NNSs' use of "because" could be attributed to cultural differences, "since" being perceived as more formal than "because" by the Cantonese students, which once again takes us back to the issues of cross-linguistic transfer and cultural influence.

Bearing in mind all that was found and noted in previous research, a study aimed at exploring the characteristics of a self-built corpus was conducted. In the next chapter, this study's corpus, methodology and the results of data analysis are presented.

3. The present study

3.1. Corpus and methodology

As noted by Blanchard, Tetreault, Higgins, Cahill & Chodorow (2013), even though a substantial body of research has been carried out as the interest in the field grows, there are consistency issues which tend to make it impossible to compare the results obtained. Different authors employ different approaches, datasets and languages. Thus, different authors might obtain different results even when using the same tools. As there is no general rule of thumb to be followed in order to carry out the research, it is important to control for variables which might exert an influence on the final outcomes. Granger & Tyson (1996) emphasize the necessity for the gathered data to be comparable and thus control it for 4 properties: type of learner (e.g. an EFL learner as opposed to an ESL writer), stage of advancement, text type and the comparison with a native corpus of comparable characteristics.

In the present study, the analyzed corpus contains 120 research articles published in English, grouped into 6 different batches of 20, according to the language background of the authors. In particular, 1 batch contains 20 research articles published by authors who are native speakers (NSs) of English, while the remaining 5 batches each include articles by authors that are native speakers of other languages and non-native speakers (NNSs) of English: respectively, Croatian, German, Italian, Polish and Spanish. More specifically, as far as the selection of authors and text to be included in the corpus was concerned, the following criteria were considered:

- For the NNS texts, only texts written by non-native English speaking authors who had also published a minimum of one article in their mother tongue were taken into account and included in the corpus;
- Only texts (written by either NSs or NNSs) presenting or discussing research in the field of linguistics were taken into account and included in the corpus;
- Only texts written by academic professionals and published in academic journals were used. The list of all the texts used can be found in the Appendix.

The corpus characteristics are summed up in Table 1.

Table 1: Corpus characteristics

Corpus components	Authors' mother tongue	Word tokens per batch	Word types per batch	Number of texts
Native speaker (NS) batch	English	135 483	9 701	20
Non-native speaker (NNS) batches	Croatian	130 883	9 537	20
	German	150 564	9 811	20
	Italian	113 103	9 401	20
	Polish	121 696	8 756	20
	Spanish	100 748	7 360	20
Average NNS batch	-	123 398.8	8 973	20
Total corpus	-	752 477	24 953	120

Prior to the analysis, all the texts had to be preprocessed. Each and every one of the texts was converted from the original .pdf or .docx format into .txt format, examined and manually cleaned according to the following principles:

- Each corpus file contains: title, abstract, body of running text, captions and notes separated from the running body of text;
- Elements which had to be deleted include: quotations that are placed outside the body of text; examples that are given outside the body of text; text elements within figures and graphs;
- The sections and headers decided by the author were retained (e.g. *1. Introduction 2. Materials and methods 3. Discussion 4. Conclusion*).

After preparing the corpus for analysis, each individual batch (English, Croatian, German, Italian, Polish, Spanish) and the whole merged NNS corpus were run through the AntConc 3.4.4w

2014 version concordance software and TagAnt 1.2.0w 2015[14] POS tagging software. AntConc was used to investigate word frequencies (with and without stop words – a list of function words self-built by merging several existing, acknowledged and publicly available lists), n-grams and concordances; while the TagAnt tool was used for POS tagging of the whole corpus. The POS tags types were counted by the AntConc's concordance tool when necessary.

The lexical bundles were extracted from the corpora by using AntConc's "N-grams" tool, with N=4 and the range (the number of texts in which a specific N-gram occurs) set to 15% of the total number of texts – 3 texts per each individual batch, 15 texts for the whole NNS batch and 18 texts for the complete corpus containing both NS and NNS batches. Table 2 shows the distribution of 4-grams before and after setting the range, while Figures 1 and 2 show screenshots of the AntConc's analysis of frequency of 4-grams' occurrence and their range after setting it to 15% of the total number of texts for the native and the whole non-native corpora respectively.

Table 2: The distribution of 4-grams before and after setting the range to 15% of the total number of texts per corpus

Corpus components	Authors' mother tongue	Number of 4-gram types (before)	Number of 4-gram tokens (before)	Number of 4-gram types (after)	Number of 4-gram tokens (after)
Native speaker (NS) batch	English	130 453	135 423	261	1 357
Non-native speaker (NNS) batch	Croatian	122 384	130 823	286	1 574
	German	140 737	150 504	303	1 636
	Italian	106 518	113 043	269	1 414
	Polish	114 167	121 636	242	1 307
	Spanish	94 483	100 688	243	1 330
Total NNS batches	-	578 289	616 694	88	3 181
NNS average		115 657.8	123 338.8	268.6	1452.2

---

[14] TagAnt was developed by Lawrence Anthony based on TreeTagger, a tool developed by Helmut Schmid. Both AntConc and TagAnt are publicly available as open-source programs at Laurence Anthony's Website: http://www.laurenceanthony.net/software.html
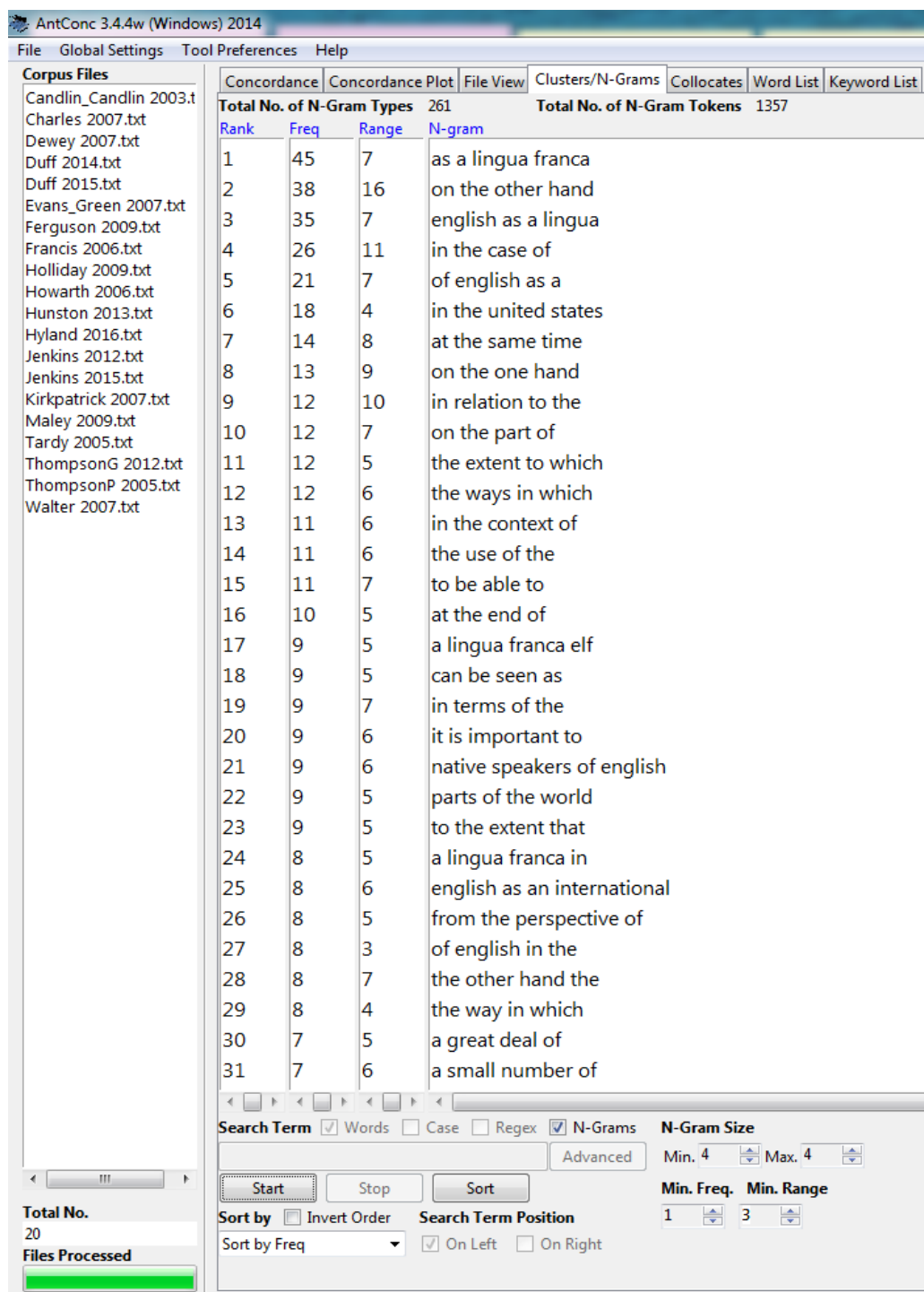
AntConc 3.4.4w (Windows) 2014

File   Global Settings   Tool Preferences   Help

Corpus Files
Candlin_Candlin 2003.t
Charles 2007.txt
Dewey 2007.txt
Duff 2014.txt
Duff 2015.txt
Evans_Green 2007.txt
Ferguson 2009.txt
Francis 2006.txt
Holliday 2009.txt
Howarth 2006.txt
Hunston 2013.txt
Hyland 2016.txt
Jenkins 2012.txt
Jenkins 2015.txt
Kirkpatrick 2007.txt
Maley 2009.txt
Tardy 2005.txt
ThompsonG 2012.txt
ThompsonP 2005.txt
Walter 2007.txt

Concordance | Concordance Plot | File View | Clusters/N-Grams | Collocates | Word List | Keyword List

Total No. of N-Gram Types 261   **Total No. of N-Gram Tokens** 1357

Rank	Freq	Range	N-gram
1	45	7	as a lingua franca
2	38	16	on the other hand
3	35	7	english as a lingua
4	26	11	in the case of
5	21	7	of english as a
6	18	4	in the united states
7	14	8	at the same time
8	13	9	on the one hand
9	12	10	in relation to the
10	12	7	on the part of
11	12	5	the extent to which
12	12	6	the ways in which
13	11	6	in the context of
14	11	6	the use of the
15	11	7	to be able to
16	10	5	at the end of
17	9	5	a lingua franca elf
18	9	5	can be seen as
19	9	7	in terms of the
20	9	6	it is important to
21	9	6	native speakers of english
22	9	5	parts of the world
23	9	5	to the extent that
24	8	5	a lingua franca in
25	8	6	english as an international
26	8	5	from the perspective of
27	8	3	of english in the
28	8	7	the other hand the
29	8	4	the way in which
30	7	5	a great deal of
31	7	6	a small number of

Search Term ☑ Words ☐ Case ☐ Regex ☑ N-Grams   **N-Gram Size**

[   ]   Advanced   Min. 4 ⬍ Max. 4 ⬍

Start   Stop   Sort   **Min. Freq. Min. Range**

Sort by ☐ Invert Order   **Search Term Position**   1 ⬍   3 ⬍

Sort by Freq ▼   ☑ On Left ☐ On Right

Total No.
20
Files Processed

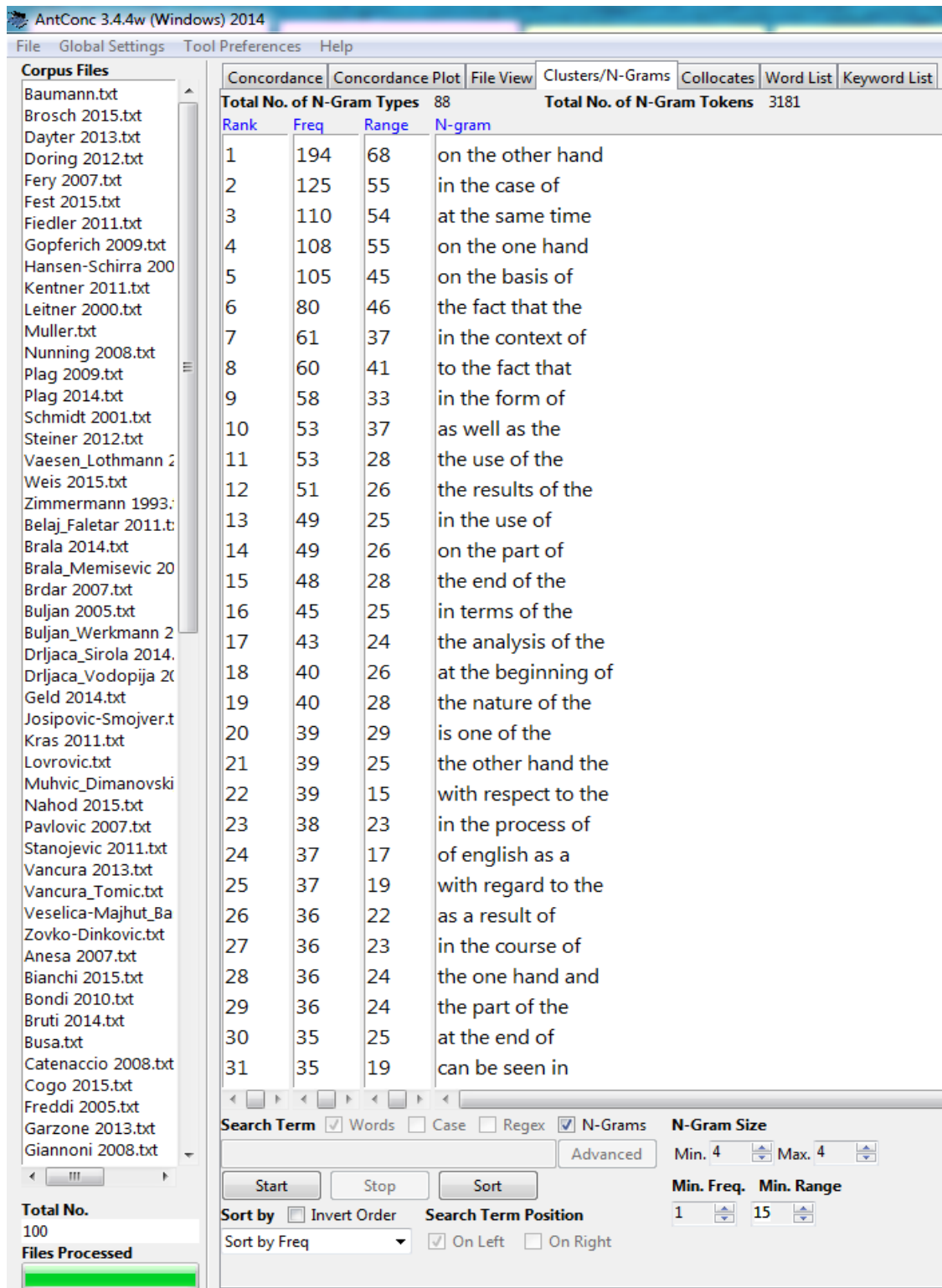**Figure 1: AntConc screenshot - The frequency and range of 4-grams in the NS corpus**

Figure 2: AntConc screenshot - The frequency and range of 4-grams in the NNS corpus (all 5 batches)

As we can see in Table 2, 261 types of 4-grams were identified in the native corpus, while 88 of them were counted in the non-native authors' corpus when analyzed as a whole, comprising all of the 5 individual NNS batches. Thus, it could seem that the native speakers used more and a greater variety of lexical bundle types. However, the number of 4-gram types depends greatly on the size of the batch/corpus in question. In other words, the NS batch (20 texts) is not supposed to be comparable to the complete NNS batch (100 texts). After calculating the NNS average the number of 4-grams is almost exactly the same (261 for the NSs and 268.6 for the NNSs' average). Hence, it can be argued that there seems to be no difference in the amount of bundles used when comparing the native and non-native writers.

Taking a look at the bundles extracted from the NNS corpus as a whole, it is interesting to note that 45 of them are shared between the NSs and NNSs. In other words, 51.1% of the NNS bundles occurred in the NS batch as well. These bundles are presented in Figure 3 (in alphabetical order).

a wide range of	*in terms of the*	*on the part of*
as a foreign language	*in the case of*	*seems to be a*
	in the context of	*that there is a*
as well as the	*in the form of*	*the case of the*
at the beginning of	*in the process of*	*the context of the*
at the end of	*in the use of*	*the end of the*
at the level of	*it is clear that*	*the fact that the*
at the same time	*it is important to*	*the nature of the*
be seen as a	*it is necessary to*	*the one hand and*
can be found in	*it is possible to*	*the other hand the*
can be seen in	*it should be noted*	*the part of the*
for the purpose of	*of English as a*	*the rest of the*
for the purposes of	*on the basis of*	*the use of the*
I would like to	*on the one hand*	*to be able to*
in line with the	*on the other hand*	*to the use of*
in order to be		

Figure 3: NNS lexical bundles occurring in the NS batch

If we disregard the range setting and look at the 100 most frequent 4-grams, we can see that 30% of the bundles are shared among the NNS and NS corpora. These bundles are presented in Figure 4 (in alphabetical order).

applied linguistic and discourse	can be seen as	on the basis of
	English as a lingua	on the one hand
as a foreign language	for the purposes of	on the other hand
as a lingua franca	in terms of the	on the part of
as an international language	in the case of	that there is a
	in the context of	the end of the
as well as the	in the use of	the nature of the
at the beginning of	it is important to	the other hand the
at the end of	it is possible to	the use of the
at the level of	native speakers of English	the way in which
at the same time	of English as a	

Figure 4: Most frequent lexical bundles occurring in both batches

Next, the retrieved bundles were POS tagged by TagAnt and analyzed according to their structural and functional aspects.

3.2. Structural analysis of lexical bundles

The structural analysis of the 4-grams was done by employing the Biber, Conrad & Cortes' (2004) taxonomy and with further reference to the *Longman's Grammar of Spoken and Written English* (Biber, Johansson, Leech, Conrad & Finegan, 1999). Biber, Conrad & Cortes (2004) divided the bundles into 3 types: bundles incorporating noun/prepositional-based or comparative expressions (phrasal bundles), those incorporating verb phrase fragments and those incorporating dependent clause fragments (both of which fall under the category of clausal bundles). In order for

the results of the analysis to be comparable, each batch (5 NNS + 1 NS) was analyzed individually. The top 100 bundles per batch were extracted, POS tagged and analyzed manually. The obtained results, which are presented in Figure 5, are in agreement with those of Biber, Conrad & Cortes (2004), who found that in academic prose phrasal bundles are used to a much greater extent than clausal ones. Considering the NS versus NNS structural distribution of bundles, it was found that the native authors use 9.5% more phrasal bundles and 50.5% less clausal bundles than the average non-native author. These results are in line with those of Pan, Reppen & Biber (2016) and Güngör & Uysal (2016), discussed in section *2.2.*. Examples of each of the types of the identified 4-grams are presented in Table 3.
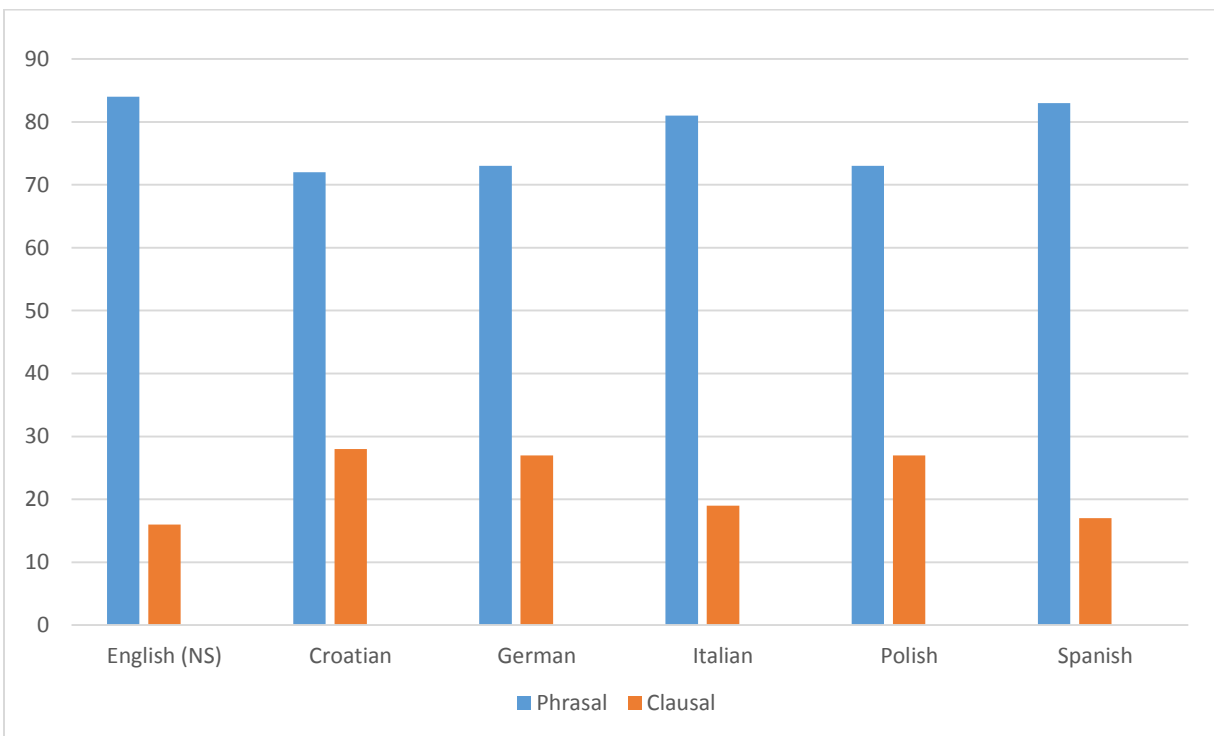

Figure 5: Structural distribution of the first 100 most frequent 4-grams per batch

Table 3: Examples of structural categories and subcategories of the first 100 most frequent 4-grams per batch[15]

Structural type of bundle	Structural subcategory	Example
Phrasal	NP-based	*the fact that the, the meaning of the, the nature of the, the end of the, the results of the, the basis of the, the analysis of the, the use of the, a great deal of, a small number of*
	PP-based	*on the other hand*, in the case of*, at the same time*, on the basis of, in the context of, in terms of the, for the purposes of, at the end of the, one of the most, at the beginning of*
	Comparative expressions	*as well as the*
Clausal	Incorporating verb phrase fragments	*I would like to, it is clear that, it is important to, is based on the, can be regarded as, be used as a, is one of the, used to refer to, is an example of*
	Incorporating dependent clause fragments	*that there is a, as we have seen, if we assume that, to be able to, that seem to be, in order to avoid*

* occurring in all of the NS and NNS batches in the top 10 most frequently used 4-grams

---

[15] For each (sub)category, the most common examples occurring in more than one batch are presented.

3.3 Functional analysis of lexical bundles

Like the structural analysis, the analysis of functional types of bundles was carried out following Biber, Conrad & Cortes' (2004) categorization. Three possible discourse functions of lexical bundles were investigated: stance expressions, discourse organizers and referential expressions. Each of the categories is divided into subcategories. Examples of each of the categories and subcategories found in our corpus are presented in Table 4. In order to be able to classify each bundle as serving one of the three functions, it was necessary to examine their context by looking at their concordances. AntConc's concordance tool was used for this purpose. The distribution of bundles' function types is presented in Table 5 and Figure 6.

In addition and due to the results of the analysis, two subcategories had to be added to the referential type of bundles: purpose bundles and 'subject-specific' bundles (Jablonkai, 2010). The latter have already been defined in Cortes (2004 and 2008) as *subject-bound bundles* and *context-dependent bundles*, and in Hyland (2008) as *research-oriented topic bundles*.

Table 4: Examples of lexical bundles according to their function[16]

Category	Subcategory	Example
1. Stance expressions	a) Epistemic stance	Personal: no occurrences *(e.g. I don't know if)* Impersonal: *the fact that the*
	b) Attitudinal/modality stance	i. Desire Personal: no occurrences *(e.g. I don't want to)* ii. Obligation/directive Personal: no occurrences *(e.g. I want you to)*

---

[16]For each (sub)category, the most common examples occurring in more than one batch are presented. Where no occurrences of a specific type of bundle were found in the corpus, a general example of the category is provided in brackets.

		Impersonal: *it is important to, it is necessary to, it should be noted*
		
		   iii. Intention/prediction
		Personal: no occurrences *(e.g. I'm not going to)*
		Impersonal: *if we assume that, does not have to, it has to be, seems to be*
		
		   iv. Ability
		Personal: no occurrences *(e.g. to be able to)*
		Impersonal: *it is possible to, it is impossible to, to be able to*

2. Discourse organizers	a) Topic introduction/focus	*in this paper I, I would like to, in the following section*
	b) Topic elaboration/classification	*on the other hand, as well as the, is due to the*
3. Referential expressions	a) Identification/focus	*is one of the, one of the main, as an instance of*
	b) Imprecision	no occurrences *(e.g. or something like that)*
	c) Specification of attributes	i. Quantity specification: *a small number of, a large number of* ii. Tangible framing attributes: *in the form of*

		iii. Intangible framing attributes: *in the case of, the nature of the, the result of the*
	d) Time/place/text reference	i. Place reference: *in the United States* ii. Time reference: *at the same time, at the beginning of* iii. Text deixis: *as shown in the, be seen in table* iv. Multi-functional reference: *at the end of, the end of the*
	e) Purpose	*for the purposes of, in order to achieve*
	f) Subject-specific	*English as a lingua, as a lingua franca, native speakers of English, English as an international, as a foreign language, English for academic purposes, of the native speaker, language learning and use, native and non-native, the teaching of English, Oxford international corpus of, Vienna Oxford international corpus, in the target language, as a second language, non-native speakers of*

Table 5: Functional distribution of lexical bundles

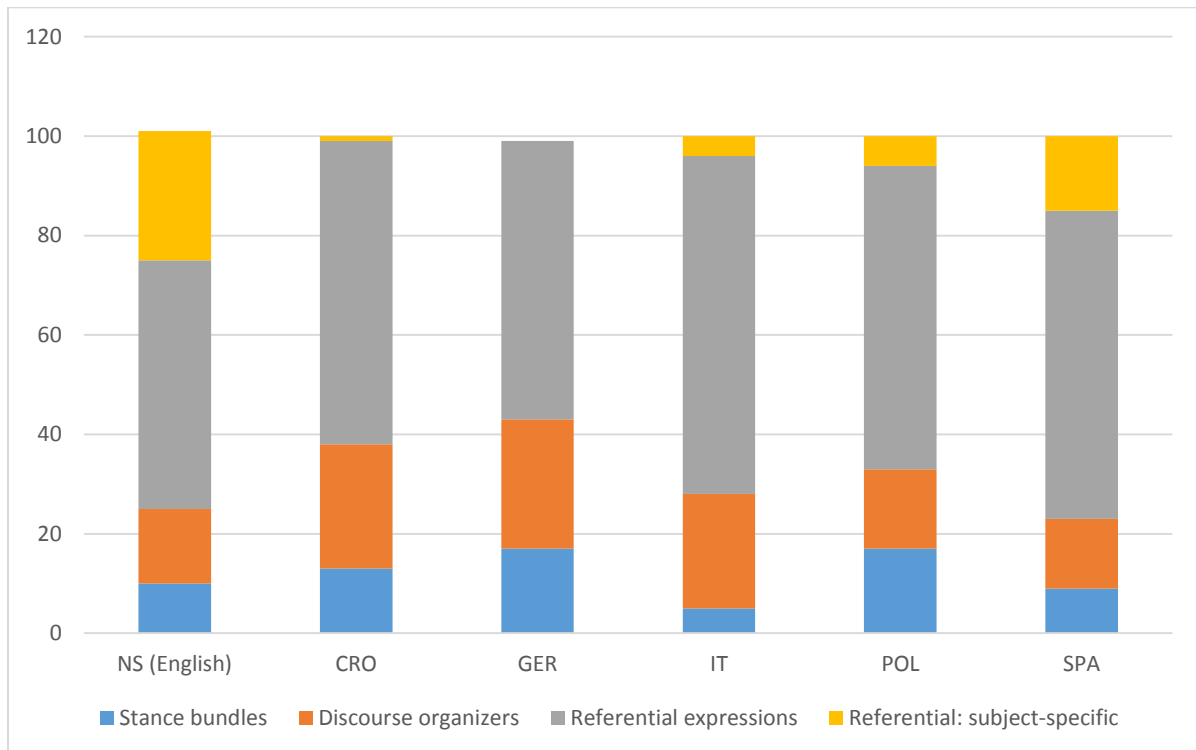| Corpus components	Author's mother tongue	Stance bundles	Discourse organizers	Referential expressions (subject-specific)
Native speaker (NS) batch	English	10	14	76 (26)
Non-native speaker (NNS) batch	Croatian	13	25	62 (1)
	German	17	26	56 (0)
	Italian	5	23	72 (4)
	Polish	17	16	67 (6)
	Spanish	9	14	77 (15)
NNS average	-	12.2	20.8	66.8 (5.2)

Figure 6: Functional distribution of lexical bundles

The results of the functional analysis of lexical bundles generally show a similarity in their usage throughout individual batches, as well as when comparing the native batch to the average non-native one. The greatest difference is in the use of referential bundles, with native authors using an average of 11.6% more bundles of that type than the non-native writers. This finding is in line with those of Pan, Reppen & Biber (2016), Chen & Baker (2010) and Güngör & Uysal (2016). However, an interesting and salient contrast between the NS and NNS batches is revealed if we take a look at the usage of subject-specific bundles, a subcategory of referential bundles. In this case, the native authors used, overall, more than twice as much bundles than the average number of bundles found in the other subcorpora. When looking at individual batches, the difference is especially remarkable, with the Croatian and German batches together containing only 1 subject-specific bundle, as opposed to the 26 bundles of the native one.

3.4. Word frequencies in the corpora

If we take a look at individual word frequencies and compare the top 100 words occurring in our corpus, 74% of them are shared between the NSs and NNSs if stop words are included (Figure 7), while 47% are shared if stop words are excluded (Figure 8).

I	J	K	L	M	N	O	P	Q
a		can		into		or		they
about		case		is		other		this
all		corpus		it		research		to
also		data		its		see		two
an		different		language		some		use
and		discourse		linguistic		students		used
are		english		may		study		was
as		example		more		such		we
at		for		most		text		were
be		from		no		than		what
been		has		not		that		which
between		have		of		the		will
both		however		on		their		with
but		in		one		there		would
by		information		only		these		

Figure 7: 74% of the most frequently used words (including stop words) ahared between NNSs and NSs

D	E	F	G	H	I	J
analysis		important		particular		way
based		information		point		words
case		knowledge		present		
communication		language		research		
context		languages		see		
corpus		learners		speakers		
data		level		students		
different		linguistic		studies		
discourse		model		study		
english		native		terms		
example		new		text		
features		non		texts		
focus		number		time		
general		part		two		
group		participants		use		

Figure 8: 47% of the most frequently used words (excluding stop words) shared between NNSs and NSs

3.5. POS tag analysis of the corpus

The part of speech analysiswas done using the TagAnt's tagging tool for tagging the whole corpus and AntConc's concordance tool for counting the tags. NNS batches' average POS tags were than calculated and used for comparison with the NS batch's figures. POS distribution among the NS and the average NNS batches is presented in Figure 9.
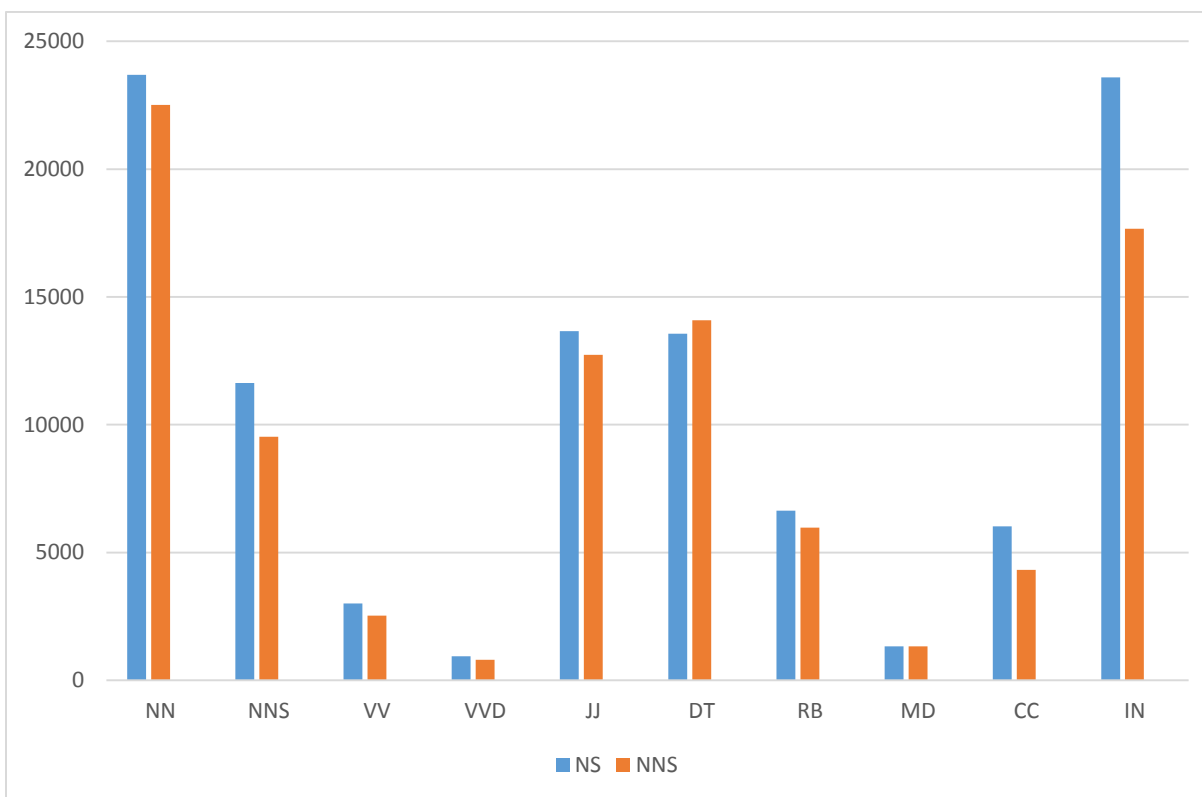
Figure 9: POS distribution in the NS and NNS batches

Legend:

NN – noun (singular/mass)   **DT** – determiner

NNS – noun (plural)   **RB** – adverb

VV – verb (base form)   **MD** – modal

VVD – verb (past tense)   **CC** – coordinating conjunction

JJ – adjective   **IN** – preposition/subordinating conjunction

4. Discussion

4.1. Overall usage of lexical bundles

In general, no prominent differences were found between the NS and NNS batches in terms of the number and frequency of lexical bundles used. This result is not surprising when it comes to academic, professional writers. Previous studies that found differences when analyzing this aspect were conducted on a different sample of participants, comparing authors of different ages and proficiency levels.

4.2. Structural analysis of lexical bundles

The structural analysis of the bundles confirmed previous findings in two ways. Firstly, all the authors, regardless of the NS versus NNS dichotomy, used phrasal bundles to a much greater extent than clausal ones, which is a pattern characteristic for academic prose. This finding is in line with those of Biber, Conrad & Cortes (2004) and Pan, Reppen & Biber (2016). Secondly, when comparing the NS and NNS batches, it was found that native authors use more phrasal and less clausal bundles than the average non-native writer. These findings are in line with those of Pan, Reppen & Biber (2016) and Güngör & Uysal (2016).

When trying to explain these results in the context of L1 expert writers versus L1 novice/L2 learners and professionals, other authors, such as Parkinson & Musgrave (2014), suggest that a higher usage of phrasal as opposed to clausal bundles acts as an indication of an author's higher proficiency level. In line with Pan, Reppen & Biber's (2016) findings, who compared L1 and L2 professionals, we can confirm that these differences are visible at the expert level as well, but in a very subtle manner. Thus, if we consider a greater usage of phrasal as opposed to clausal bundles as an indicator of transition to linguistic expertise, native writers seem to be in advantage over their non-native peers, even in the context of such highly proficient authors.

4.3. Functional analysis of lexical bundles

In line with the results of the structural analysis, the functional analysis' results support previous findings as well. Both the NS and NNS batches show the same pattern of usage identified by Biber, Conrad & Cortes (2004), who discovered that among the three categories (namely discourse organizers, stance bundles and referential expressions), when used in academic prose, referential bundles represent the most commonly used functional type of bundles, the difference between this and the other two categories being quite extreme. When it comes to usage of the other two functional categories, stance bundles and discourse organizers, both batches used more of the discourse than stance bundles, as opposed to the pattern found by Biber, Conrad & Cortes (2004). The difference between these two categories is less pronounced than that between them and the referential expressions. Stance bundles refer to attitudes, modality, intention, prediction, obligation or ability, while discourse organizers are used as means of topic introduction or elaboration. Hence, it seems that regardless of the native language, academic writers focus more on presenting to the reader the interconnectedness of their ideas, the structure, coherence and cohesion of their texts than on providing more subjective, personal judgements and establishing a closer relationship with the reader.

When it comes to comparing the NS and NNS batches, as was found in studies by Pan, Reppen & Biber (2016) and Chen & Baker (2010), native writers used more referential expressions than their average non-native peers. The most interesting result in this category concerns the difference between the usage of subject-specific expressions – 26 for the NS and only 5.2 for the average NNS batch. Biber, Conrad & Cortes (2004) define referential bundles as those serving the purpose of emphasizing a particularly important entity. Thus, it could be argued that native writers are more oriented towards placing the subject of their studies into a wider context of the field in question. In this way, the native authors manage to orient the readers in a broader area, while at the same time keeping them focused on a specific issue.

4.4. Word frequency analysis of the corpora

The results of the word frequency analysis in the NS and NNS batches show that 74% of the most frequent words are shared between the two when including the function words, while 47% are shared when looking only at the content words. This is not a surprising finding, as function words are used as a means of providing cohesion and structure, an aspect in which both the native and non-native academic authors tend to excel at. When focusing on the content words, the ones that are shared between the NSs and NNSs are mainly either subject-specific (e.g. *corpus, communication, English, language, linguistic, native, speakers, texts, words*) or research-specific (e.g. *analysis, example, participants, model, number, point, research, study*). As the subjects of the texts gathered in the corpus differ, this result can be said to be in line with expectations. It can also indicate that both the native and non-native writers think in the same direction when presenting their study in general to the readers. It would be interesting to analyze this aspect in a corpus built of texts belonging to other disciplines and see if the findings can be generalized across them.

4.5. POS tag analysis of the corpora

Comparing the NS and NNS batches in terms of word category usage, it is evident that all parts of speech are used to a comparable extent and in similar degrees, except for the plural noun forms and both coordinating and subordinating conjunctions. These POSs are used by the native authors to a greater degree than by their non-native peers, the difference being more salient when it comes to subordinating conjunctions and subtler when looking at the coordinating conjunctions and plural nouns. When looking further into the differences between batches in the use of conjunctions, it was found that the NSs use not only a greater number of conjunctions but also a wider range of items. While the difference in the usage of plural noun forms might be attributable to purely stylistic preferences, the more extensive and diverse usage of both coordinating and subordinating conjunctions by the native authors might be an indication of a difference in the way they realize cohesion. The NSs might prove to be in a slight advantage over the NNSs in this aspect.

5. Conclusions and limitations

The results of the investigation into the native and non-native authors' corpora are aligned with those of previous research – differences were found in the use of structural and functional lexical bundles, as well as in the distribution of parts of speech. Thus, it can be concluded that the hypothesis posed at the beginning of the research has been proved to be valid – however subtle the contrast, the scientific writings of native academic professionals do differ from those of their non-native peers.

It is necessary to highlight, however, some limitations of this research, which might have exerted an influence on the results. First, the corpus used is relatively small in size; second, information on the editing process of all of the texts comprised in the corpus was not included in the analysis: if considered, such information might have led to different interpretations. Also, the results cannot be generalized to other disciplines.

Taking into account the results and limitations of this research, further research suggestions can be proposed. It would be interesting to repeat the study applying the same methodology on a larger corpus, which would also make the results more significant statistically. Also, in order to make the conclusions more generalizable, the study could be recreated in the context of other disciplines.

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. http://dx.doi.org/10.1016/j.esp.2011.08.004

Atkinson, D. (2004). Contrasting rhetorics/contrasting cultures: why contrastive rhetoric needs a better conceptualization of culture. *Journal of English for Academic Purposes 3,* 277-289.

Becher, T. (1994). The significance of disciplinary differences. *Studies in Higher Education, 19*(2), 151-161.

Berzak, Y., Reichart, R. & Katz, B. (2014). Reconstructing Native Language Typology from Foreign Language Usage. *Proceedings of the Eighteenth Conference on Computational Language Learning*, 21-29.

Biber, D. & Conrad, S. (1999). Lexical Bundles in Conversation and Academic Prose. In H. Hasselggård & S. Johansson (Eds.), *Out of Corpora: Studies in Honour of Stig Johansson* (pp. 181-202). Amsterdam/Atlanta, GA: Rodopi.

Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (1999). *Longman Grammar of Spoken and Written English.* Harlow, England: Longman.

Biber, D., Conrad, S. & Cortes, V. (2004). If you look at…: Lexical Bundles in University Teaching and Textbooks. *Applied Linguistics, 25*(3), 371-405.

Blanchard, D., Tetreault, J., Higgins, D., Cahill, A. and Chodorow, M. (2013). TOEFL11: A corpus of non-native English. *ETS Research Report Series, 2*, i-15. http://dx.doi.org/10.1002/j.2333-8504.2013.tb02331.x

Chen, Y-H. & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology, 14*(2), 30-49.

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. (2008). A comparative analysis of lexical bundles in academic history writing in English and Spanish. *Corpora, 3*(1), 43-57.

De Cock, S. (2000). Repetitive phrasal chunkiness and advanced EFL speech and writing. In C. Mair & M. Hundt (Eds.), Corpus Linguistics and Linguistic Theory (pp. 51–68). Amsterdam: Rodopi.

Durrant, P. & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations?. *IRAL - International Review of Applied Linguistics in Language Teaching*, *47*(2), 157-177. http://dx.doi.org/10.1515/iral.2009.007

Ebrahimi, S. F. & Motlagh, H. S. (2015). Across Disciplinary Study of Textual Devices in Research Article Abstracts Written by Native and Non-Native Writers of English. *Journal of Applied Linguistics and Language Learning, 1*(1), 24-29. http://dx.doi.org/10.5923/j.jalll.20150101.04

Ellis, N.C. & Simpson-Vlach, R. (2010). An Academic Formulas List: New Methods in Phraseology Research. *Applied Linguistics, 31*(4), 487-512. http://dx.doi.org/10.1093/applin/amp058

Gnutzmann, C. & Rabe , F. (2014). 'Theoretical subtleties' or 'text modules'? German researchers' language demands and attitudes across disciplinary cultures. *Journal of English for Academic Purposes, 13,* 31-40.

Goldman, S. R. & Murray, J. D. (1992). Knowledge of Connectors as Cohesion Devices in Text: A Comparative Study of Native English and ESL Speakers. *Journal of Educational Psychology, 84*(4), 504-519. http://dx.doi.org/10.1037/0022-0663.84.4.504

Granger, S. & Tyson, S. (1996). Connector usage in the English essay writing of native and nonnative EFL speakers of English. *World Englishes, 15*(1), 17–21. http://dx.doi.org/10.1111/j.1467971X.1996.tb00089.x

Güngör, F. & Uysal, H. H. (2016). A Comparative Analysis of Lexical Bundles Used by Native and Non-native Scholars. *English Language Teaching, 9*(6), 176-188. http://dx.doi.org/10.5539/elt.v9n6p176

Hinkel, E. (1994). Native and Nonnative Speakers' Pragmatic Interpretations of English Texts. *TESOL Quarterly, 28,* 353–376. http://dx.doi.org/10.2307/3587437

Hyland, K. (2008). As can be seen: lexical bundles and disciplinary variation. *English for Specific Purposes, 27*, 4-21. http://dx.doi.org/10.1016/j.esp.2007.06.001

Jablonkai, R. (2010). English in the context of European integration: A corpus-driven analysis of lexical bundles in English EU documents. *English for Specific Purposes, 29*, 253–267.

Jarvis, S. & Pavlenko, A. (2007). *Crosslinguistic influence in language and cognition*. New York, NY: Routledge.

Kachru, Y. (1988). Writers in Hindi and English. In A. Purves (Ed.), *Writing across languages and cultures: Issues in contrastive rhetoric* (pp. 109–137). Newbury Park, CA: Sage

Kramsch, C. & Lam, W. S. E. (2009). Textual Identities: The Importance of Being Non-Native. In G. Braine (Ed.), *Non-Native Educators in English Language Teaching* (pp. 57-72). New York, NY: Routledge.

Kuteeva, M. & Mauranen, A. (2014). Writing for publication in multilingual contexts: An introduction to the special issue. *Journal of English for Academic Purposes, 13,* 1-4.

Latham, P. E. & Roudi, Y. (2009). Mutual information. *Scholarpedia, 4*(1), 1658. http://www.scholarpedia.org/article/Mutual_information

Ma, Y. & Wang, B. (2016). A Corpus-based Study of Connectors in Student Writing: A Comparison between a Native Speaker (NS) Corpus and a Non-native Speaker (NNS) Learner Corpus. *International Journal of Applied Linguistics & English Literature, 5*(1), 113-118. http://dx.doi.org/10.7575/aiac.ijalel.v.5n.1p.113

Mauranen, A. (1996). Discourse Competence – Evidence from Thematic Development in Native and Non-native Texts. In A. Mauranen & E. Ventola (Eds.), *Academic Writing: Intercultural and textual issues* (pp. 195-230). Amsterdam/Philadelphia: John Benjamins Publishing.

McGrath, L. (2014). Parallel language use in academic and outreach publication: A case study of policy and practice. *Journal of English for Academic Purposes, 13,* 5-16. http://dx.doi.org/10.1016/j.jeap.2013.10.008

Müller, S. (2004). *Discourse Markers in Native and Non-native English discourse.* Amsterdam/Philadelphia: John Benjamins Publishing Company.

Muresan, L.-M. & Perez-Lllantada, C. (2014). English for research publication and dissemination in bi-/multiliterate environments: The case of Romanian academics. *Journal of English for Academic Purposes, 13*, 53–64. http://dx.doi.org/10.1016/j.jeap.2013.10.009

O'Donnell, M. B., Romer, 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.

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.

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.

Perez 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.

Phillipson, R. (1992). *Linguistic imperialism*. Oxford, England: Oxford University Press.

Römer, U. (2009). English in academia: does nativeness matter? *Anglistik: International Journal of English Studies, 20*(2), 89-100.

Scarcella, R. C. (1984). How Writers Orient their Readers in Expositoy Essays: A Comparative Study of Native and Non-Native English Writers. *TESOL Quarterly, 18,* 671–688. http://dx.doi.org/10.2307/3586582

Siepmann, D. (2005). *Discourse Markers Across Languages: A Contrastive Study of Second-Level Discourse Markers in Native and Non-Native Text with Implications for General and Pedagogic Lexicography.* London, England: Routledge.

Silva, T. (1993). Toward an Understanding of the Distinct Nature of L2 Writing: The ESL Research and Its Implications. *TESOL Quarterly, 27*(4), 657-677.

Staples, S., Egbert, J., Biber, D. & McClair, A. (2013). Formulaic sequences and EAP writing development: lexical bundles in the TOEFL iBT writing section. *Journal of English for Academic Purposes, 12*(3), 214-225.

Swales, J. M . (1997). English as Tyrannosaurus rex. *World Englishes, 16*(3), 373-382. http://dx.doi.org/10.1111/1467-971X.00071

Yli-Jokipii, H. & Jorgensen, P. E. F. (2004). Academic journalese for the Internet: a study of native English-speaking editors' changes to texts written by Danish and Finnish professionals. *Journal of English for Academic Purposes 3,* 341-359.

Appendix

List of articles used for building the corpus:

The native-speaker batch (English):

Candlin, C. N. & Candlin, S. (2003). Health care communication: A problematic site for applied linguistics research. *Annual Review of Applied Linguistics, 23*, 134-154.

Charles, M. (2007). Reconciling top-down and bottom- up approaches to graduate writing: using a corpus to teach rhetorical functions. *Journal of English for Academic Purposes, 6*, 289-302.

Dewey, M. (2007). English as a lingua franca and globalization: an interconnected perspective. *International Journal of Applied Linguistics, 17*(3), 332-354.

Duff, P. A. (2014). Case Study Research on language Learning and Use. *Annual Review of Applied Linguistics, 34*, 233-255.

Duff, P. A. (2015). Transnationalism, Multilingualism, and Identity. *Annual Review of Applied Linguistics, 35*, 57-80.

Evans, S. & Green, C. (2007). Why EAP is necessary: A survey of Hong Kong tertiary students. *Journal of English for Academic Purposes, 6*, 3-17.

Ferguson, G. (2009). Issues in researching English as a lingua franca: a conceptual enquiry. *International Journal of Applied Linguistics, 19*(2), 117-135.

Francis, N. (2006). The development of secondary discourse ability and metalinguistic awareness in second language learners. *International Journal of Applied Linguistics, 16*(1), 37-60.

Holliday, A. (2009). The role of culture in English language education: key challenges. *Language and Intercultural Communication, 9*(3), 144-155.

Howarth, P. (2006). The Phraseology of Public International English. *International Journal of English Studies, 6*, 109-129.

Hunston, S. (2013). Systemic functional linguistics, corpus linguistics, and the ideology of science. *Text & Talk, 33*(4-5), 617-640.

Hyland, K. (2016). Methods and methodologies in second language writing research. *System, 59*, 116-125.

Jenkins, J. English as a Lingua Franca from the classroom to the classroom. *ELT Journal 66*(4) Special issue, 486-494.

Jenkins, J. (2015). Repositioning English and multilingualism in English as a Lingua Franca. *Englishes in Practice, 2*(3), 49-85.

Kirkpatrick, A. (2007). Setting attainable and appropriate English language targets in multilingual settings: a case for Hong Kong. *International Journal of Applied Linguistics, 17*(3), 376-391.

Maley, A. (2009). ELF: a teacher's perspective. *Language and Intercultural Communication, 9*(3), 187-200.

Tardy, C. M. (2005). "It's like a story": Rhetorical knowledge development in advanced academic literacy. *Journal of English for Academic Purposes, 4*, 325-338.

Thompson, G. (2012). Intersubjectivity in newspaper editorials. *English Text Construction 5*(1), 77-100.

Thompson, P. (2005). Points of focus and position: Intertextual reference in PhD theses. *Journal of English for Academic Purposes, 4*, 307-323.

Walter, C. (2007). First- to second-language reading comprehension: not transfer, but access. *International Journal of Applied Linguistics, 17*(1), 14-37.

The non-native speaker batch (Croatian):

Belaj, B. & Faletar, G. T. (2011). Cognitive foundations of emotion verbs complementation in Croatian. *Suvremena lingvistika, 37*(72), 153-169.

Brala-Vukanovic, M. (2014). Articles as a Lexical Pointing System. Is Unique Identifiability a Linguistic and Cognitive Universal? *Croatian Journal of Philosophy, 14*(41), 233-252.

Brala-Vukanovic, M. & Memisevic, A. (2012). Moving *od(-)* and *do(-)* in Croatian. An account of sources, goals and dual readings of the dative. *Jezikoslovlje 13*(1), 41-69.

Brdar, M. (2007). Topic-continuity, metonymy and locative adverbials: A cognitive-functional account. *Suvremena lingvistika, 63*(1), 13-29.

Buljan, G. (2005). Interpreting English Verb Conversions: The Role of Metonymy and Metaphor. *Suvremena lingvistika, 57-58*(1-2), 13-30.

Buljan, G. & Werkmann, A. (2013). Knocking on EU's door: on the changing and conflicting metaphorical images of Croatia in selected EU press. *Jezikoslovlje, 14*(2-3), 337-362.

Drljaca Margic, B. & Vodopija-Krstanovic, I. (In press). Introducing EMI at a Croatian university: Can we bridge the gap between global emerging trends and local challenges? In C. Jensen,

S. Dimova & A. K. Hultgren (Eds.), *English as a Scientific and Research Language. English in Europe: Volume 3. Language and Social Processes (LSP) Series* (pp. 43-64). The Hague: De Gruyter Mouton.

Drljaca Margic, B. & Sirola, D. (2014). 'Jamaican and Irish for fun, British to show off': Attitudes of Croatian university students of TEFL to English language varieties. *English Today, 30*, 48-53.

Geld, R. (2014). Investigating meaning construal in the language of the blind: a cognitive linguistic perspective. *Suvremena lingvistika, 40*(77), 27-59.

Josipovic Smojver, V. (In press). Croatian primary school pupils and English pronunciation in light of the emergence of English as a lingua franca. In J. Mihaljević Djigunović & M. Medved Krajnović (Eds.), *Early Learning and Teaching of English: New Dynamics of Primary English* (pp. 37-65). Bristol: Multilingual Matters.

Kras, T. (2011). Acquiring the syntactic constraints on auxiliary change under restructuring in L2 Italian: Implications for the Interface Hypothesis. *Linguistic Approaches to Bilingualism 1*(4), 413–438.

Lovrovic, L. (2014). Developing Learners' Communicative Competence in Task-Based Language Teaching. *Strani jezici 43*(3), 228-236.

Muhvic-Dimanovski, V. & Skelin Horvat, A. (2008). Contests and nominations for new words – why are they interesting and what do they show. *Suvremena lingvistika, 65*(1), 1-26.

Nahod, B. (2015). Domain–specific Cognitive Models in a Multi–Domain Term Base. *Suvremena lingvistika, 41*(80), 105-128.

Pavlovic, N. (2007). Directionality in Translation and Interpreting Practice. Report on a questionnaire survey in Croatia. *Forum 5*(2), 79-99.

Stanojevic, M.-M. (2011). Modeling epistemic distance: the Croatian I-participle. *Suvremena lingvistika, 37*(72), 219-239.

Vancura, A. & Tomic, D. (2013). The Analysis of Insulting Practices - Sticks and Stones in the Croatian Parliament. In G. Kisicek & I. Z. Zagar (Eds.), *What Do We Know about the World? Rhetorical and argumentative perspectives* (pp. 143-158). Charleston, SC: Open Monograph Press.

Vancura, A. (2013). The story of the tortoise and the hare – Speech rate in simultaneous interpretation and its influence on the quality of trainee-interpreters performance. *Jezikoslovlje, 14*(1), 85-99.

Veselica Majhut, S. & Basic, I. (2013). From foot to meter: Rendering of culture-specific items in popular fiction translations from English to Croatian. In A. Peti-Stantic, M.-M. Stanojevic & G. Antunovic (Eds.), *Language Varieties Between Norms and Attitudes* (pp. 197-218). Frankfurt: Peter Lang.

Zovko Dinkovic, I. (2002). Locative Alternation in English and Croatian. *Studia romanica et anglica zagrabiensia*, XLVII-XLVIII, 585-596.

The non-native speaker batch (German):

Baumann, S. (2008). Degrees of Givenness and their Prosodic Marking. In C. M. Riehl & A. Rothe (Eds.). *Was ist linguistische Evidenz? Kolloquium des Zentrums "Sprachenvielfalt und Mehrsprachigkeit* (pp. 35-55). Aachen: Shaker

Brosch, C. (2015). On the Conceptual History of the Term Lingua Franca. *Apples – Journal of Applied Language Studies, 9*(1), 71-85.

Dayter, D. (2014). Self-praise in microblogging. *Journal of Pragmatics, 61*, 91-102.

Doring, S. (2012). The focus sensitivity of sentence adverbs. *Proceedings of ConSOLE XIX*, 201-214.

Fery, C. (2007). Information Structural Notions and the Fallacy of Invariant Correlates. *Interdisciplinary Studies on Information Structure, 6*, 161–184.

Fest, J. (2015). Corpora in the Social Sciences – How corpus-based approaches can support qualitative interview analyses. *Revista de lenguas para fines específicos, 21*(2), 48-69.

Fiedler, S. (2011). English as a lingua franca – a native- culture-free code? Language of communication vs. language of identification. *Apples – Journal of Applied Language Studies, 5*(3), 79-97.

Gopferich, S. (2009). Comprehensibility assessment using the Karlsruhe Comprehensibility Concept. *The Journal of Specialised Translation, 11*, 31-52.

Hansen-Schirra, S. (2003). Linguistic enrichment and exploitation of the Translational English Corpus. *Proceedings of the Corpus Linguistics Conference*, 288-297.

Kentner, G. (2012). Linguistic rhythm guides parsing decisions in written sentence comprehension. *Cognition, 123*(1), 1-20.

Leitner, G. (2000). Lexical frequencies in a 300 million word corpus of Australian newspapers. Analysis and interpretation. *International Journal of Corpus Linguistics 5*(2), 1-32.

Muller, G. (1999). Imperfect Checking. *The Linguistic Review, 16*, 359-404.

Nunning, V. (2004). The Importance of Being English: European Perspectives on Englishness. *European Journal of English Studies, 8*(2), 145-158.

Plag, I. (2010). Compound stress assignment by analogy: the constituent family bias. *Zeitschrift fur Sprachwissenschaft, 29*(2), 243-282.

Plag, I. (2014). Phonological and phonetic variability in complex words: An uncharted territory. *Rivista di Linguistica, 26*(2), 209-228.

Schmid, H.-J. (2001). 'Presupposition can be a bluff': How abstract nouns can be used as presupposition triggers. *Journal of Pragmatics, 33*, 1529-1552.

Steiner, E. (2012). Methodological cross-fertilization: empirical methodologies in (computational) linguistics and translation studies. *Translation: Corpora, Computation, Cognition, 2*(1), 3-21.

Vaeßen, J. & Lothmann, T. (2014). Do You Read Me? Metaphor as a Pathway to the Conceptualisation of Literary Identity. *International Journal of Literary Linguistics, 3*(1), 1-18.

Weiß, H. (2015). When the subject follows the object. On a curiosity in the syntax of personal pronouns in some German dialects. *J Comp German Linguistics, 18*, 65–92.

Zimmermann, T. E. (1993). On the proper treatment of opacity in certain verbs. *Natural Language Semantics, 1*, 149-179.

The non-native speaker batch (Italian):

Bianchi, F. (2015). Subtitling science: An efficient task to learn content and language. *Lingue Linguaggi 15*, 7-25.

Bondi, M. (2010). Metadiscursive Practices in Introductions: Phraseology and Semantic Sequences across Genres. *Nordic Journal of English Studies, 9*(2), 99-123.

Bruti, S. (2015). Compliments in fansubs and in professional subtitles: The case of Lost. *Rivista internazionale di tecnica della traduzione, 16*, 13-34.

Busà, M. G. (2010). Sounding natural: improving oral presentation skills. *Language Value, 2*(1), 51-67.

Catenaccio, P. (2008). Press releases as a hybrid genre: addressing the informative/promotional conundrum. *Pragmatics, 18*(1), 9-31.

Cogo, A. (2015). Complexity, negotiability, and ideologies: a response to Zhu, Pitzl, and Kankaanranta et al. *Journal of English as a Lingua Franca, 4*(1), 149-155.

Freddi, M. (2005). Arguing linguistics: corpus investigation of one functional variety of academic discourse. *Journal of English for Academic Purposes, 4*, 5-26.

Garzone, G. (2013). Variation in the use of modality in legislative texts: Focus on shall. *Journal of Pragmatics, 57*, 68-81.

Giannoni, D. S. (2008). Medical writing at the periphery: The case of Italian journal editorials. *Journal of English for Academic Purposes, 7*, 97-107.

Guido, M. G. (2015). Mediating linguacultural asymmetries through elf in unequal immigration encounters. *Lingue Linguaggi, 15*, 155-175.

Malavasi, D. (2007). Lexical Analysis of Implicit Promotional Devices in Bank Annual Reports. *ILCEA, 9*, 171-184.

Mazzi, D. (2007). The Construction of Argumentation in Judicial Texts: Combining a Genre and a Corpus Perspective. *Argumentation, 21*, 21-38.

Mazzi, D. (2014). The words of history: Phraseology as a key to historical argument. *Lingue Linguaggi, 12*, 147-160.

Milizia, D. (2014). In, out, or half way? The european attitude in the speeches of british leaders. *Lingue Linguaggi, 11*, 157-175.

Morini, M. (2014). From norms to macro-norms? The Translation of Classics seen in a Global, Diachronic Framework. *Lingue Linguaggi, 11*, 177-190.

Musacchio, M. T. (2004).The Distribution of Information in LSP Translation. A Corpus Study of Italian. *Rivista internazionale di tecnica della traduzione, 8*, 89-105.

Pavesi, M. (2009). Pronouns in film dubbing and the dynamics of audiovisual communication. *Vigo International Journal of Applied Linguistics, 6*, 89-107.

Peruzzo, K. (2014). European English terms for Italian legal concepts: the case of the Italian Code of Criminal Procedure. *Rivista internazionale di tecnica della traduzione, 15*, 145-157.

Salvi, R. (2012). Textual equivalence in the translation of specialized languages: discourse adherences and dissimilarities. *Traduzione e riscrittura, 11*, 102-115.

Vettorel, P. (2013). English in Italian advertising. *World Englishes, 32*(2), 261-278.

The non-native speaker batch (Polish):

Adamska-Salaciak, A. (2012). Dictionary definitions: Problems and solutions. *Studia Linguistica Universitatis Iagellonicae Cracoviensis, 129*(4), 7-23.

Biel, L. (2011). Professional Realism in the Legal Translation Classroom: Translation Competence and Translator Competence. *Translators' Journal, 56*(1), 162-178.

Biel, L. (2014). The textual fit of translated EU law: a corpus-based study of deontic modality. *The Translator, 20*(3), 332-355.

Bromberek-Dyzman, K. (2012). Affective Twist in Irony Processing. *Humana. Mente Journal of Philosophical Studies, 23*, 83–111.

Debowska-Kozlowska, K. (2014). Processing Topics from the Beneficial Cognitive Model in Partially and Over-Successful Persuasion Dialogues. *Argumentation, 28*, 325-339.

Duszak, A. & Lewkowicz, J. (2008). Publishing academic texts in English: A Polish perspective. *Journal of English for Academic Purposes, 7*, 108-120.

Jaworska, D. & Jelec, A. (2014). Thoughts on the table: Gesture as a tool for thinking in blind and visually impaired children. *Yearbook of the Poznan Linguistic Meeting, 1*(1), 73–88.

Jopek-Bosiacka, A. (2013). Comparative law and equivalence assessment of system-bound terms in EU legal translation. *Linguistica Antverpiensia, 12*, 110-146.

Kajzer-Wietrzny, M. (2013). Idiosyncratic Features of Interpreting Style. *New Voices in Translation Studies, 9*, 38-52.

Kajzer-Wietrzny, M. & Tymczyńska, M. (2015). Devising a systematic approach to examination marking criteria for audiovisual translation: a case study from Poland. *The Interpreter and Translator Trainer, 9*(3), 342-355.

Krajewska, M. (2010). Cross-Linguistic Evidence for the Strident Stop Hypothesis. *Acta Philologica, 38*, 55-67.

Lew, R. (2012). The role of syntactic class, frequency, and word order in looking up English multi-word expressions. *Lexikos, 22*, 243-260.

Lijewska, A. & Chmiel, A. (2015). Cognate facilitation in sentence context – translation production by interpreting trainees and non-interpreting trilinguals. *International Journal of Multilingualism, 12*(3), 358-375.

Mazur, I. & Chmiel, A. (2012). Towards common European audio description guidelines: results of the Pear Tree Project. *Perspectives, 20*(1), 5-23.

Murawska, M. (2013). Incommensurability of discourses in professional medical communication. *Explorations: A Journal of Language and Literature, 1*, 26-49.

Rataj, K. (2014). Surfing the brainwaves of metaphor comprehension. *Poznań Studies in Contemporary Linguistics 50*(1), 55–73.

Siek-Piskozub, T. & Strugielska, A. (2008). Autonomy, experience and concepts: a study in educational discourses. *Poznań Studies in Contemporary Linguistics 44*(4), 597–616.

Twardzisz, P. (2014). Metonymic Target Identification: In Search of a Balanced Approach. *Nordic Journal of English Studies, 13*(1), 104-29.

Whyatt, B. (2007). Two languages, two cultures, one mind: a study into developmental changes in the students' view of language as a tool in cross-cultural communication. *Poznań Studies in Contemporary Linguistics 43*(2), 133–147.

Wach, A. (2011). Native-speaker and English as a lingua franca pronunciation norms: English majors' views. *Studies in Second Language Learning and Teaching, 2*, 247-266.

The non-native speaker batch (Spanish):

Alonso, P. (2010). The lexico-structural/semantic progression in discourse topic construction: evidence from comment articles. *Círculo de Lingüística Aplicada a la Comunicación (CLAC), 44*, 3-26.

Alonso, R. (2002). Transfer: constraent, process, strategy or inert outcome? *CAUCE, Revista de Filologia y su Didactica, 25*, 85-101.

Barreiro Bilbao, S. C. (2015). Perception of place of articulation in English fricatives by Spanish listeners. *Vigo International Journal of Applied Linguistics, 12*, 37-62.

Carrio-Pastor, M. L. (2014). Cross-cultural Variation in the Use of Modal Verbs in Academic English. *SKY Journal of Linguistics 27*, 153–166.

Catala, R. J. (2005). Using Lex30 to measure the L2 productive vocabulary of Spanish primary learners of EFL. *Vigo International Journal of Applied Linguistics, 2*, 27-44.

Gomez-Jimenez, E. (2015). An introduction to graphology: definition, theoretical background and levels of analysis. *Miscelánea: a journal of English and American studies 51*, 71-85.

Gonzalez-Alvarez, D. (1998). Evaluative Subject Modifiers in Early Modern English. *Sederi, 9*, 23-38.

Fonseca Mora, C. (1999). Foreign language acquisition and melody singing. *ELT Journal, 54*(2), 146-152.

Hidalgo Downing, L. (2006). Text world creation in advertising discourse. *Circulo de Lingüística Aplicada a la Comunicación (CLAC), 13*, 23-44.

Jesus Sanchez, M. (2015). The Effect of Multiple Encounters in Contextualized Learning. *Porta Linguarum, 25*, 49-62.

Lazaro-Ibarrola, A. (2016). Are CLIL learners simply faster or also different? Evidence from L1 use in the repair sequences and discourse markers of CLIL and EFL learners. *Vigo International Journal of Applied Linguistics, 13*, 127-145.

Llinares, A. (2006). A pragmatic analysis of children's interlanguage in EFL preschool contexts. *Intercultural Pragmatics, 3*(2), 171–193.

Llurda, E. (2007). The representation of EFL teachers' views on the role of English as a Lingua Franca. *Vigo International Journal of Applied Linguistics, 4*, 11-23.

Luque Agullo, G. (2001). Oral short stories as a vehicle for communication. *The Journal of English and Foreign Languages, 18*(30), 1-14.

Martinez, R. D. (2005). Covert communication in the promotion of alcohol and tobacco in Spanish press advertisements. *Revista Electronica de Linguistica Aplicada, 5*, 82—102.

Mateo, M. (2014). Exploring Pragmatics and Phonetics for Successful Translation. *Vigo International Journal of Applied Linguistics, 11*, 111-135.

Munoz, C. (2007). Cross-linguistic influence and language switches in L4 oral production. *Vigo International Journal of Applied Linguistics, 4*, 73-94.

Palacios Martinez, I. M. (2011). The Language of British Teenagers. A Preliminary Study of its Main Grammatical Features. *Journal of the Spanish Association of Anglo-American Studies, 33*(1), 105–126.

Perez-Sabater, C. (2012). The Linguistics of Social Networking: A Study of Writing Conventions on Facebook. *Linguistik online, 56*, 81-93.

Perinan-Pascual, C. (2013). A knowledge-engineering approach to the cognitive categorization of lexical meaning. *Vigo International Journal of Applied Linguistics, 10*, 85-104.