Impaginazione
Verena Papagno

EUT Edizioni Università di Trieste, 2017.

ISBN 978-88-8303-912-6 (print)
ISBN 978-88-8303-913-3 (online)

EUT - Edizioni Università di Trieste
Via E. Weiss, 21 – 34128 Trieste
eut@units.it
http://eut.units.it
https://www.facebook.com/EUTEdizioniUniversitaTrieste

Testi, corpora, confronti interlinguistici: approcci qualitativi e quantitativi

a cura di
Giuseppe Palumbo

EUT EDIZIONI UNIVERSITÀ DI TRIESTE

Sommario

DOI: 10.13137/978-88-8303-913-3/20551

6. Notes on investigating the native vs non-native distinction in written academic English

GIUSEPPE PALUMBO
Università di Trieste

Abstract

Texts written in English by non-native speakers can be considered instances of mediated language, where the mediation takes place between a writer's native language and English, seen, respectively, as the "source" and "target" poles. In investigating such texts, the methods of analysis can thus draw on some assumptions and approaches used in translation studies, starting from the idea that in mediated communication the target product always shows traces of interference from features and traits associated with the source material. This chapter reports on an investigation of written academic language in English. The investigation is corpus-based and the texts included in the corpus include research papers in two different academic disciplines written by either native speakers or non-native speakers of English. Initial findings of the investigation are discussed in relation to two specific aspects: part-of-speech distribution and preference for pre- or post-modification in noun groups.

Key words

Academic English, nativeness, language mediation, interference, noun groups.

1. Introduction

Written academic English has been and is still being extensively studied from many different perspectives. The reasons for this continued interest by scholars are manifold. English is the language of choice (some would say of necessity) of today's scientific and scholarly communities, having now established a global dominance that sowed its seeds in the 1950s or perhaps even earlier (Gordin 2015). This global dominance is now being cemented by the decision of many universities worldwide to use English as the medium of instruction in all or some of their degree programmes at both undergraduate and postgraduate level.

As the global language of academia, English is the object of a vast and diverse array of initiatives concerned with language research itself, pedagogy, publishing and language "brokering" (Lillis and Curry 2010). As regards pedagogy, the acquisition of English, whether it be for purposes of comprehension, production or both, is promoted through many instruction programmes, often informed by one or the other approaches or paradigms favoured by scholarly research. In terms of publishing, academic English has become the object of reference works (e.g. OUP 2014) and dedicated publications – some overtly instructional (e.g. Swales and Feak 2013), others with a more popularizing character (e.g. Sword 2012). Such works may be aimed not only at speakers of other languages but also at native English speakers who are still trying to master the specificities of academic expert communication. As regards language brokering (please note that Lillis and Curry (2010) originally used the term "literacy brokering"), this may be seen to include all the activities and practices of those who help authors to shape a manuscript into its final published form, thus including well-established and recognized practices such as copy-editing, translation and proof-reading but also the occasional and more diversified interventions of peers and other actors within an individual researcher's network of contacts.

In this chapter, I start with an overview of recent studies of written academic English, noticing how the focus has over the years increasingly moved on the production of scholarly writers who use English as an additional language to their native language (which in some academic disciplines amounts to say that they *only* use English in their academic publications). I then propose an approach to investigating written academic English using corpus-based methods and employing perspectives that are normally associated with studies in which the dimension of language "contact" or "mediation" is explicitly taken into consideration, in particular translation studies. Finally, I present an exploratory analysis of a corpus comprising research papers in English written by scholars with diverse language backgrounds (including native speakers) and conclude by pointing to ways of refining the analysis and extending it to other relevant aspects.

2. A brief overview of existing research on written academic English

Scholarly research on academic English has seen the emergence of different approaches, most of them sharing a pedagogical preoccupation. These approaches can be grouped according to the research traditions they draw from, the questions they investigate and the methods they employ. A recent critical overview is provided in Tribble (2017), who is essentially concerned with the impact of scholarly research on the effectiveness of academic writing instruction programmes. Tribble (2017: 30-33) distinguishes between genre-informed approaches, approaches based on writing and composition studies, "academic literacies" or "critical" approaches, and approaches based on the notion of English as a Lingua Franca (ELF).[1]

Genre-informed approaches emerge from the essentially UK-based tradition of studies on register and context. Representative, and highly influential, studies following a genre-based approach include Swales (1990) and Hyland (2000). More recently, studies in this tradition have started to apply corpus-based methods, extending their scope to comparisons between spoken and written academic registers (e.g. Biber 2006). The approaches based on composition and writing studies are mainly identified with US scholars and draw on rhetoric-based teaching. An influential study in this tradition (curiously not mentioned in Tribble's overview) is Bazerman (1988). The approaches based on the notion of "academic literacies" (Lea and Street 1998) tend to employ ethnographic methods (as in Lillis and Curry 2010) and are mainly preoccupied with the empowerment of students as writers, adopting at times openly critical stances against established academic conventions (as in Benesch 2001). The last, and more recent, group of approaches identified by Tribble (2017) in his overview is the one incorporating the notion of ELF and referred to by both Tribble and its proponents (e.g. Jenkins 2014) as ELFA, or English as a Lingua Franca Academic. These approaches emphasize the international character of scientific and scholarly communication and focus on how communicative practices and their outcomes are shaped by processes of revision and negotiation between participants.

Tribble's overview of the approaches to research on academic English is not neutral, in the sense that his main aim was to show how ELFA approaches are not as "challenging" or even "paradigm-changing" as some of their proponents (especially Jenkins 2014) would have them be. In particular, Tribble (2017: 33-35) notes the very scarce contribution that ELFA approaches have so far made to pedagogy and warns of the risks of exaggerating the importance and significance of the *native* vs *non-native* dichotomy. In his view, the effectiveness of writing instruction

1 For a definition of English as a Lingua Franca, see Jenkins (2014: 2): "English when it is used as a contact language between people from different first languages (including native English speakers)". Note not all scholars would agree to include native English speakers in this definition, reserving ELF for situations in which communication in English takes place exclusively between non-native speakers of the language.

is enhanced if the focus remains primarily on disciplinary expertise, as already happens in genre-informed approaches.

The relative merits of the approaches identified in Tribble's overview of the research on academic English would deserve a more in-depth discussion than is possible here. As far as the ELFA perspective is concerned, this may well not be – as noted by Tribble – the paradigm-changing approach that some its proponents consider it to be, but it certainly has the merit of having drawn the attention of language scholars and language users alike to the background of linguistic and cultural diversity against which a significant share of today's communication employing English takes place. Higher education in particular can easily be seen as a "prototypical ELF scenario" (Smit 2018: 387; cf. also Palumbo 2015), at least as much as all the other scenarios in which English is the contact language of choice for native speakers of other languages, one example being international institutions such as the UN and the EU.

3. Academic English from a "language mediation" perspective

In consideration of its multilingual background, communication based on ELF could be seen as one instance of "mediated communication"[2] or as a situation of "language contact" (Matras 2009) and thus considered amenable to the research methods employed in language studies that explicitly consider dimensions of mediation or contact. Areas in which these dimensions are explicitly taken into consideration include second language acquisition, pragmatics, linguistic typology, sociolinguistics, contrastive linguistics and translation studies.

On this basis, I would like to propose that texts written in English by non-native speakers can be considered as instances of mediated language, where the mediation takes place between a writer's native language and English, seen, respectively, as the "source" and "target" poles. In investigating such texts, the methods of analysis can thus draw on assumptions and approaches used in translation studies (as also suggested by Cook 2012), starting from the idea that in mediated communication the target product always shows traces of interference from features and traits associated with the source material. This interference, in turn, may also be seen to render translated texts in a given language somewhat different with respect to comparable non-translated texts, i.e. texts in that language that are not the result of translation and belong to the same genre or register.

2 I am using the notion of "mediation" in a more general sense than that sometimes used in translation studies, where it is linked (e.g. in Hatim and Mason 1997: 147) to the ideological significance of the translator's intervention in the transfer process. My idea of translation as "mediated communication" comes very close to Chesterman's (2004: 10-11) characterization of translation as one particular instance of "constrained communication", to be placed alongside reported speech and – especially significant in the present context – communication in a non-native language.

Investigations of translated texts along these lines have tended to focus, alternatively, on the difference between translations and non-translations or on the effects of interference from the source language. Not all distinctive features of translated language are due to interference from the source pole. Some features may be seen to emerge as an effect of the translation process, irrespective of the languages involved (for instance, "explicitation", whose status as a universal trait of translations has however been contested; see Becher 2010). Even interference itself can be observed in relation to two planes: one is that of the individual *texts* to be translated; the other is that of the *systemic* features of the source language. In order to assess the relative influence of these two planes, studies comparing translations and originals should be designed carefully and use appropriate methods and perspectives: textual and corpus-based analysis, in other words, should be complemented with investigations of cognitive processes and social-historical analyses. However, evidence from various studies (see for instance Mauranen 2005 and the other studies mentioned there) suggest that "that it is the source language system that influences the translator, not directly or only elements which the translator is faced with in the text" (Mauranen 2005: 78). Even more importantly for the purposes of the present chapter, Mauranen (2005: 77) warns that "it is unwise to neglect the ubiquity of transfer in bi- or multi-lingual situations".

Interference from the source language is normally considered one of the factors leading to "translationese", or the particular style described as being typical of translated texts. More specifically, interference – as in Toury's (1995) "law of interference" – manifests itself not so much through the presence of errors as through the recurrence of forms and structures that deviate from the norms of the target language or occur with greater than usual frequency in the target language.

Several studies have investigated the particular features of translationese, some of them even questioning its empirical basis. Tirkkonen-Condit (2002), for instance, showed that translationese is not readily recognizable by human subjects on the basis of linguistic features. Other studies have shown that when texts are subject to automatic analysis, there are cues for the distinction between translations and non-translations. Borin and Prütz (2001) investigated the distributional differences of part-of-speech n-grams in translated and non-translated texts and concluded that they "may turn out to be indicative of translation effects in the syntactic domain". Along the same lines, Baroni and Bernardini (2006: 259) found that one of the cues for discriminating between originals and translations was "the distribution of function words and morphosyntactic categories in general".

In the rest of the chapter I'll report on an attempt at framing the investigation of written academic texts in terms of the "translational" perspective briefly sketched above. The key interrelated notions in this perspective are those of transfer and interference. In particular, I will try to establish whether the language background of the writers may be seen to exert any influence on the texts at the level of morpho-syntax, taking a suggestion from some of the studies of

translated language mentioned above (in particular Borin and Prütz 2001) and employing an analytical procedure that was already used in a previous study (Palumbo 2015) with promising results. The following are some initial questions the investigation is intended to provide an answer to:

- Can any differences be observed in the morpho-syntactic profiles of the texts, as observed for instance in terms of part-of-speech (POS) distribution?
- How can the observed differences between texts be related to the author's native languages, or families of languages, in spite of the editorial process the texts have gone through?
- How do the linguistic differences attributable to national identities interact with the writers' disciplinary identities?

In the exploratory case study presented here I will not be able to provide conclusive answers to these questions. The case study is based on a small corpus and only proposes a crude quantitative analysis of some morpho-syntactic and structural features of the texts, based on simple descriptive statistics. The study is essentially intended to set the scene for subsequent, more rigorous analyses, which in turn might require a revision of both the criteria followed for compiling the corpus and the methods of analysis.

4. The corpus

The corpus used for this exploratory study was constructed on the basis of one general aim and according to a set of more detailed criteria following from this aim. The general aim was that of reflecting output in English as produced by two different groups of writers working as academic researchers: native speakers (NSs) of English on the one hand and non-native speakers (NNSs) of English on the other. Defining a "native speaker" for any given language may not be as straightforward as it seems (Davies 2003) and may be next to impossible when specialist writing is concerned (Tribble 2017: 34-35). The general operational criterion taken into account to guarantee that the corpus would reflect the "nativeness vs non-nativeness" distinction was an empirical one, related to a writer's publishing history: writers based in an English-speaking country, affiliated to a research institution there and only publishing articles written in English were taken to be representative of "native" output. Writers based in a non-English speaking country, affiliated to a research institution there and having a history of publishing both in English and in another language (presumably, their "native" language) were taken to be representative of "non-native" writing in English. (A bilingual writer might have ended up being included in this second group, but as defining "bilingualism" may be fraught with as many problems as defining "(non)-nativeness", no attempts were made at identifying possible bilingual authors.)

Not all academic disciplines would easily lend themselves to a search for "non-native" writers as defined above. In several disciplines most, if not all, publishing for research purposes is in English, and therefore authors generally have no history of publishing texts belonging to the same genre (i.e. academic articles) in another language, even when they are based in a non-English speaking country. After some research, two specific disciplines were identified in which authors could be observed to fall within the "native" or the "non-native" categories as described above: linguistics and agricultural economics.

A corpus with two components (or sub-corpora) was then constructed: one for linguistics (LING) and one for agricultural economics (AGRO). Each corpus component includes a total of 120 texts, all of them research articles, distributed as follows (see also Fig. 1):

- 20 texts written by native speakers of English;
- another 100 texts written by non-native speakers of English representing 5 different native-language, or L1, backgrounds, namely Croatian, German, Italian, Polish and Spanish; each of these native language backgrounds is represented with 20 texts.

Each corpus component can thus be seen to be composed of six different L1 sets, English being one of these L1s.

LING	AGRO
120 research articles from linguistics journals, representing the following native languages: English, German, Croatian, Polish, Italian, Spanish (20 texts each). Total size: 581,100 words	120 research articles from agricultural economics, representing the following native languages: English, German, Croatian, Polish, Italian, Spanish (20 texts each). Total size: 571,700 words

Table 1. Corpus composition

All the texts included in the corpus were taken from journals, with a preference for those adopting an open-access policy. All texts were considered in their final, published form. An attempt was also made at selecting both NS and NNS articles from the same journals, but this was not always possible. All texts were cleaned by removing (most) para-textual material. The corpus was compiled and POS-tagged using the Sketch Engine (Kilgarriff et al. 2014).

The six different native-language backgrounds for text authors are the same for the two corpus components. For the non-native English speaking authors, the native languages were chosen for opportunistic reasons (i.e. because it was easier for those languages to find texts written by authors meeting the requirements described above) but also with the aim of representing a variety of language families, so that results from the analysis could also be discussed in language-typological terms. Counting in English itself as one of the native language back-

grounds, both corpus components can thus to be said to represent writers with a Germanic language background (English, German), writers with a Romance language background (Italian, Spanish) and finally writers with a Slavic language background (Croatian, Polish). In the analysis, all the texts written by authors sharing a given native language were treated as separate sets.

5. An exploratory analysis of the corpus

The analysis presented here is a first exploratory attempt at answering the research questions illustrated above. The approach is similar to that already adopted in Palumbo (2015), in which a different corpus was analyzed. The analysis aims to explore differences and similarities among the sets of texts constituting each corpus component, where a set is understood as comprising all the texts written by authors sharing the same native-language, or L1, background. The level of description chosen for the analysis is morpho-syntax, based on the idea that traces of transfer from the writers' native language might emerge at the structural level. Two aspects are investigated in particular: 1) POS distribution in the L1 sets; and 2) the distribution of noun group structures in the L1 sets. The corpus was analyzed using the part-of-speech (POS) tagger available in the Sketch Engine (Kilgarriff et al. 2004). The results given here are no more than basic descriptive statistics; for an informed interpretation of the data, significance tests should be performed, which I reserve for later analyses.

5. 1 Part-of-speech distribution: Native vs non-native speakers

An initial general comparison can be made between POS distribution in the native and non-native texts, i.e. between the set of English L1 writers on the one hand and the other five L1 sets on the other. Figures 1 and 2 present the data for each corpus component (the data for the cumulative NNS set were obtained by averaging out the relative frequencies observed in the individual L1 sets for each POS class).

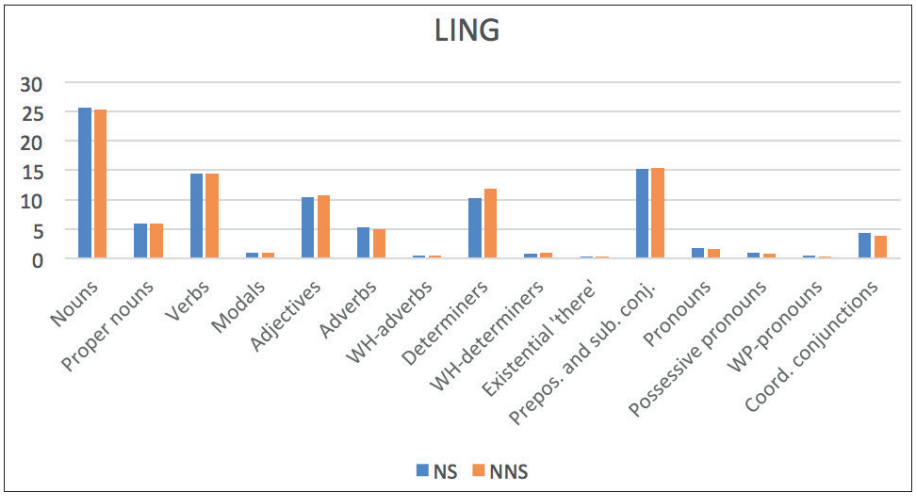

Figure 1. Part-of-speech distribution (in percentage terms) in the LING corpus: NS texts vs NNS texts.

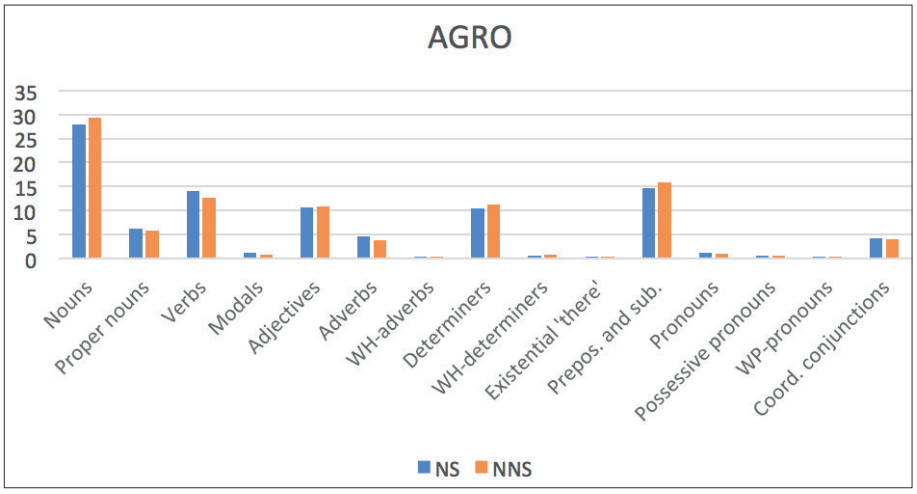

Figure 2. Part-of-speech distribution (in percentage terms) in the AGRO corpus: NS texts vs NNS texts.

What is certainly striking at first glance is the similarity of the distribution across the two disciplines. To take just two examples: the proportion of nouns is roughly double the proportion of verbs in both disciplines; prepositions/subordinating conjunctions and coordinating conjunctions are also distributed very similarly, with the latter amounting to roughly one third of the former in both cases. Consistent adherence to generic conventions of academic writing is likely to be at play here. When the data are considered across language backgrounds,

some differences emerge: determiners, for instance, are used consistently more by NNSs in both disciplines; verbs are used more by the NSs in the AGRO component, whereas in the LING component the percentage for verbs is almost exactly the same (14.45% for NSs and 14.48% for NNSs); finally, adverbs are used slightly more by NSs in both disciplines. Such small differences might of course be due to chance, but it could be worth investigating whether they reflect influences from the L1 of the non-native writers and whether such influences can be distinguished in terms of language families or language-typological differences.

5.2 Noun group structures per L1 sets

Nouns are by far the larger POS category represented in the two corpus components and across all language backgrounds. This marked preference for nominalization seems to be in line with the generic conventions of academic language. The figures so far, however, are only really telling us that all writers in the corpus use far more nouns than any other POS category. It could be worth checking how the nouns are organized, syntactically, into phrases or groups. This could make it possible to observe how pre- and post-modification (e.g. *article selection* vs *the selection of articles*; for a discussion, see Biber et al. 2001: 578-602) are distributed – the assumption being that in some NNS sets the preference for post-modified structures is more marked than in the native sets because of L1 interference. Rough measures in this respect can be obtained by counting occurrences of noun sequences (including proper names) with and without intervening prepositions or apostrophes (as in genitive pre-modified structures). For the purposes of this analysis, the following sequences of elements were searched for and counted in the corpus: 'NOUN + *of/in* (*the*) NOUN'; 'NOUN + NOUN; 'NOUN + NOUN + NOUN'; 'NOUN + *'s* + NOUN'; 'PROPER NOUN + *'s* + NOUN'. To provide comparative measures, the frequency of occurrence of items in each category was turned into a percentage ratio against all nouns present in the set under consideration. The resulting figures are given in Figures 3 and 4, where sequences of two ('NOUN + NOUN') and three nouns ('NOUN + NOUN + NOUN') are conflated into one category.

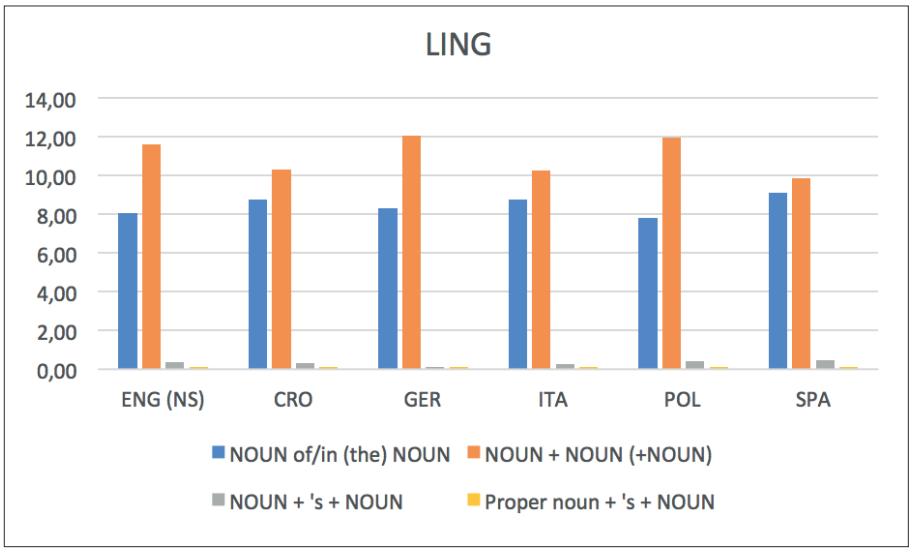

Figure 3. Noun group structures in the LING corpus (per language background of authors).

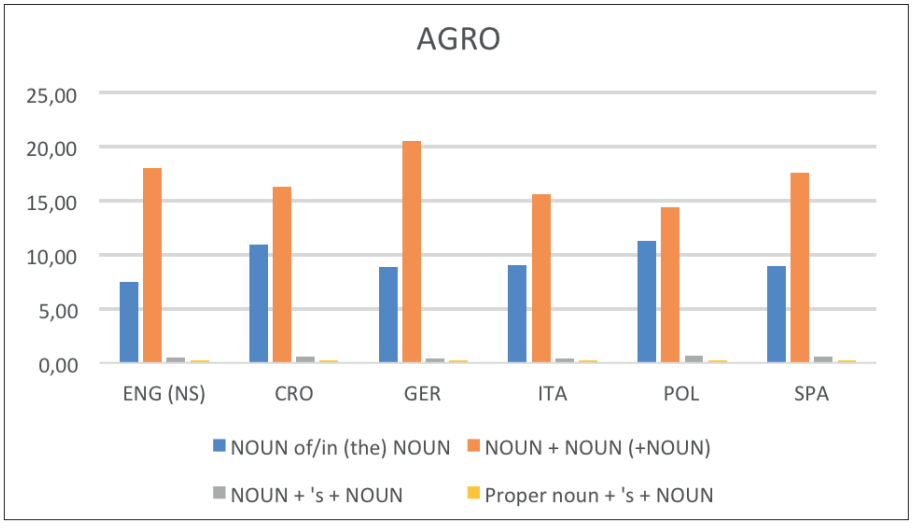

Figure 4. Noun group structures in the AGRO corpus (per language background of authors).

An intuitive expectation might be that writers coming from Romance languages would tend to use pre-modified structures to a lesser degree than writers from other language backgrounds, given that in Romance languages pre-modification is generally not available as a grammatical option for the modification of noun group heads. Conversely, German writers could be expected to make ample use of this syntactic option, which is one that they also have available in their own language. Data from the two corpus components show indeed that German writ-

ers are the ones that use pre-modification the most – even more than native English writers. The data for writers from Romance languages also seem to confirm expectations, at least with respect to writers from Germanic languages: in both AGRO and LING, Italian and Spanish writers use pre-modification less than both English and German writers (in LING Italian and Spanish writers are those that use pre-modification *the least*). For writers with a Slavic language background the data do not exhibit consistent trends. Particularly notable is the diverging behavior of Polish writers in the two corpus components. In the linguistics sub-corpus, Polish academics are among the writers that tend to use pre-modified structures the most, whereas in the agricultural economics sub-corpus Polish writers are the ones that use pre-modification *the least*. Not discounting chance as one possible cause for this difference, other explanations might have to do with the English language competence of writers. As part of their specific disciplinary expertise, Polish linguists might be more alert to the need to consider pre-modification as an option for the noun phrases in their English language writing. At this stage, however, this is no more than speculation.

6. Conclusions

The approach to investigating academic language in English proposed in this chapter tries to build on its nature as a locus of language contact. To reuse a metaphor that has been proposed (by Gellerstam 2006) in relation to translation, the approach is based on the idea that the native language of academics writing in English could leave "fingerprints" in their texts. These fingerprints, in turn, can be identified by comparing non-native writing to native writing. To use still another notion from translation studies, non-native academic writing can be observed in terms of its "textual fit" (Biel 2014) with respect to native writing. The idea of comparing and contrasting native and non-native writing in English in the context of academic language is not new, but to date it has mostly been used for investigating the language of learners (Nesselhauf 2005) or specific aspects of language, such as formulaic sequences (O'Donnel et al. 2013) and lexical bundles (Salazar 2014; Esfandiari and Barbary 2017).

The exploratory investigation proposed in this chapter is mainly intended as a pointer to aspects that appear to be amenable to further study and could take cues from analyses of translated language. Part-of-speech distribution, for instance, would benefit from further analysis based on an approach similar to Borin and Prütz's (2001), where the parts of speech are treated as sequences of n-grams and observed in terms of their positional differences. The texts in the corpus could also be analyzed using authorship attribution methods, such as the various measures of "intertextual distance" that are used to assess the similarity (and dissimilarity) between texts. (For an application of these methods to the analysis of translated texts, see the two chapters by Ondelli and Nadalutti in this volume.)

As far as specific language and textual traits are concerned, future analyses of the corpus might include all the aspects that are traditionally considered in comparisons and contrasts between translated and non-translated language: the distribution verb tenses and passive constructions; the use of encapsulating anaphoric references; the use of the definite article; the use of pronouns, as opposed to equivalent impersonal constructions; the frequency and distribution of connectives and other cohesive devices; the presence and function of questions; the use of lexical repetition, as opposed to the preference for synonyms; variation in sentence length; lexical preferences according to word etymology ('Latinate vs Anglo-Saxon vocabulary'). Any finding on such aspects could also be interpreted in terms of the general attitude displayed by writers with respect to readers, characterized, for instance, in terms of Hinds' (1987) distinction between "writer responsibility" and "reader responsibility" (see also MacKenzie 2015).

Some scholars (e.g., Römer 2009; Tribble 2017) have argued that the distinction between nativeness and non-nativeness in English academic writing is not as relevant as it would appear to be and that disciplinary expertise may play a much more significant role than native language background. Others have pointed out that the "native/non-native distinction remains a useful heuristic" (Gnutzmann and Rabe 2014: 39), while recognizing that the language demands made on researchers may vary significantly across disciplines. Adopting approaches normally associated with the study of translations does not imply, per se, that prominence should be given to the nativeness factor; however; it could contribute a fresh perspective on the study of how, in academic English, language and communicative norms are re-negotiated in an international scenario.

REFERENCES

Baroni M. and S. Bernardini (2006) "A new approach to the study of translationese: Machine-learning the difference between original and translated text", *Literary and Linguistic Computing*, 21(3), pp. 259-274.

Bazerman C. (1988) *Shaping Written Knowledge. The Genre and Activity of the Experimental Article in Science*, Madison, The University of Wisconsin Press.

Becher V. (2010) "Abandoning the notion of 'translation-inherent' explicitation: Against a dogma of translation studies", *Across Languages and Cultures*, 11(1), pp. 1-28.

Benesch S. (2001) *Critical English for Academic Purposes. Theory, Politics and Practice*, London/New York: Routledge.

Biber D. (2006) *University Language. A Corpus-Based Study of Spoken and Written Registers*, Amsterdam/ Philadelphia, John Benjamins.

Biel Ł. (2014) *Lost in the Eurofog: The Textual Fit of Translated Law*, Frankfurt am Main, Peter Lang.

Borin L. and K. Prütz (2001) "Through a glass darkly: part of speech distribution in original and translated text". In *Computational Linguistics in the Netherlands 2000*. Ed. by W. Daelemans, K. Sima'an, J. Veenstra and J. Zavrel, Amsterdam, Rodopi, pp. 30-44.

Chesterman A. (2004) "Hypotheses about translation universals". In *Claims, Changes and Challenges in Translation Studies*. Ed. by G. Hansen, K. Malmkjær and D. Gile, Amsterdam/Philadelphia, John Benjamins, pp. 1–13.

Cook G. (2012) "ELF and translation and interpreting: Common ground, common interest, common cause", in *Journal of English as a Lingua Franca*, 1(2), pp. 241-262.

Davies A. (2003) *The Native Speaker. Myth and Reality*, 2nd ed., Clevedon, Multilingual Matters.

Esfandiari R. and 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, pp. 21–42.

Gellerstam M. (2005) "Fingerprints in Translation", in *In and Out of English: For Better, For Worse?* Ed. by G. Anderman and M. Rogers, Clevedon, Multilingual Matters, pp. 201-2013.

Gordin M.D. (2015) *Scientific Babel. The Language of Science from the Fall of Latin to the Rise of English*, Chicago, The University of Chicago Press.

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

Hatim B. and Mason I. (1997) *The Translator as Communicator*, London/New York: Routledge.

Hinds J. (1987) "Reader versus writer responsibility: A new typology". In *Writing across Languages: Analysis of L2 text*. Ed. by U. Connor and R. B. Kaplan, Reading, Mass., Addison-Wesley, pp. 141-152.

Hyland K. (2000) *Disciplinary Discourses. Social Interactions in Academic Writing*, Harlow, Longman.

Lea M. R. and Street B. V. (1998) "Student writing in higher education: An academic literacies approach", *Studies in Higher Education*, 23(2), pp. 157-172.

Lillis T. and Curry M.J. (2010) *Academic Writing in a Global Context. The Politics and Practices of Publishing in English*, London/New York, Routledge.

Kilgarriff A., Baisa V., Bušta J., Jakubíček M., Kovář V., Michelfeit J., Rychlý P. & Suchomel V. (2014) "The Sketch Engine: Ten years on", *Lexicography*, 1(1), pp. 7-36.

Matras Y. (2009) *Language Contact*, Cambridge, Cambridge University Press.

Mauranen A. (2005) "Contrasting languages and varieties with translational corpora", *Languages in Contrast*, 5(1), pp. 73-92.

MacKenzie I. (2015) "Rethinking Reader and Writer Responsibility in Academic English". *Applied Linguistics Review*, 6(1), pp. 1-21.

Nesselhauf N. (2005) *Collocations in a Learner Corpus*, Amsterdam/Philadelphia, John Benjamins.

Palumbo G. (2015) "Studying ELF Institutional Web-based Communication by Universities: Comparison and Contrast with English Native Texts". In *English for Academic Purposes: Approaches and Implications*. Ed. by P. Thompson and G. Diani, Newcastle upon Tyne, Cambridge Scholars Publishing, pp. 245-264.

O'Donnell M. B., Römer U. and 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), pp. 83-108.

OUP (2014) *Oxford Learner's Dictionary of Academic English*, Oxford, Oxford University Press.

Römer U. (2009) "English in Academia: Does Nativeness Matter?", *Anglistik: International Journal of English Studies*, 20(2), pp. 89-100.

Salazar D. (2014) *Lexical Bundles in Native and Non-Native Scientific Writing: Applying A Corpus-Based Study to Language Teaching*, Amsterdam/Philadelphia, John Benjamins.

Smit U. (2018) "Beyond monolingualism in higher education. A language policy account". In *The Routledge Handbook of English as a Lingua Franca*. Ed. by J. Jenkins, W. Baker and M. Dewey, London/New York, Routledge, pp. 387-399.

Swales J. M. (1990) *Genre Analysis*, Cambridge, Cambridge University Press.

Swales J. M. and Feak. C.B. (2013) *Academic Writing for Graduate Students. Essential Tasks and Skills*, 3rd ed., Ann Arbor, Mich., Michigan University Press.

Sword H. (2012) *Stylish Academic Writing*, Cambridge, Mass., Harvard University Press.

Tirkkonen-Condit S. (2002) "Translationese - a myth or an empirical fact? A study into the linguistic identifiability of translated language", *Target*, 14(2), pp. 207-220.

Toury G. (1995) *Descriptive Translation Studies and Beyond*. Amsterdam/Philadelphia: John Benjamins.

Tribble C. (2017) "ELFA vs. Genre: A new paradigm war in EAP writing instruction?", *Journal of English for Academic Purposes*, 25, pp. 30-44.