

Compiling a Written Corpus of English as an Academic Lingua Franca: Medical Research Articles of Turkish Academics

Neslihan Önder

Abstract

English as lingua franca (ELF) is a “world language whose speakers communicate mainly with other NNSs, often from different L1s than their own” (Jenkins, 2006, p.140). However, given the importance of ELF in the world of publication, surprisingly, empirical research on the linguistic description of ELF has been thin. Available studies are confined to audio-recordings as spoken language and there are not any written ELF corpora. Moreover, there is an urgent need to raise awareness of ELF in mainstream English classrooms. This study aims to build a self-compiled corpus of Turkish academics’ empirical research articles from the field of medicine to investigate the salient features of ELF lexico-grammar. Specifically, the use of ‘the’, ‘a’ and ‘an’ as variants (not errors) were examined. 20 medical research articles were collected containing approximately 59,648 words. Empirical data was analysed manually to explore the potential salient features of ELF lexico-grammar (see Cogo & Dewey, 2006; Seidlhofer, 2004). It is notable that unlike Seidlhofer’s (2004) corpus, which is a computer-based corpus of audio recordings and transcriptions of spoken ELF interactions, this study concentrates on written corpus in the form of medical research articles. The initial findings, with regard to the target grammar-related aspects, revealed that the omission of the definite article ‘the’ had the highest frequency at 2.1% (1,246 occurrences) and inserting the definite article the when redundant amounted to 0.36% (219 occurrences). It was also observed that omissions of indefinite articles ‘a/an’ were 0.26% (153 occurrences). The findings suggest that the usage of ‘the’, ‘a’ and ‘an’ tended to be unproblematic, and did not lead to a breakdown in communication, which could be ELF variants as regular tendencies. Discourse samples are provided from the corpus and pedagogical implications are discussed in light of literature.

Keywords: ELF, English for specific purposes, Corpus, Medical research article.

Introduction

It is well-known that English is one of the most widespread languages in the world and that non-native speakers of English outnumber native speakers (Crystal, 2003; Graddol, 1999, 2006), which leads to variations (differences) in the use of English and paves the way for English as a lingua franca (henceforth ELF). ELF is mainly defined as “a ‘contact language’ between [people] who share neither a common native tongue nor a common (national) culture, and for whom English is the chosen *foreign language of communication*” (Firth, 1996, p.240, emphasis added). ELF should not be regarded as deficient. Rather, it is different in terms of form from English spoken by a native speaker and is a flexible communicative way of interacting with other languages (Hülmbauer, Böhringer & Seidlhofer, 2008). Jenkins (2000) rightly questions the strict patterns of ‘proper’ English in teaching English as a foreign language. As she puts it, indeed, “[t]here is really no justification for doggedly persisting in referring to an item as ‘an error’ if the majority of the world’s L2 English speakers produce and understand it” (p. 160).

Turkish belongs to the Turkic family of languages and spoken mainly in the Republic of Turkey as an official language. Accordingly, Turkish grammar has various commonalities and differences with other languages worldwide (for detailed information on Turkish grammar, see Göksel & Kerslake, 2005). For instance, definite article use is not available in Turkish while it is a salient characteristic of English language. Compatible with the context of the present research, there are empirical studies on ELF in Turkey. Karakaş (2012) conducted an empirical study in which he focused on the use of English as an ELF in academia through identifying the perceived problems of Turkish academics in a state university in Turkey. As a descriptive and qualitative study, the data was collected via a survey questionnaire, which was modified from Coury (2001), with the participation of 27 academics. His findings revealed that English was primarily used for communicative activities i.e., while abroad (90%) whereas only one academic stated that s(h)e used German for communication. There are various studies on English as ELF in Turkish context, including learners’ attitudes and motivation to learn English (Kızıltepe, 2000), English teachers’ attitude towards the present status of ELF or an international language in Turkey (Bayyurt, 2006, 2008) and also the place of English in Turkey as a foreign language context of Turkey (Coşkun, 2010). The present study could provide insightful data that English is dominant as a foreign language in academia in Turkey. However, when ELF literature is examined, there is scant focus on empirical studies for classroom implications. First, few descriptive empirical studies have been conducted on the linguistic description of ELF (but see Breiteneder, 2005; Cogo & Dewey, 2006; Dewey, 2007; House, 1999; Jenkins, 2000; Meierkord, 2002), which has been highlighted by many researchers (e.g., Firth 1996; Jenkins, 2006; Mauranen, 2010; Seidlhofer 2001, to name a few). It is also notable that available empirical research on ELF tends to focus on phonetics and pragmatics. Second, there is a growing interest in compiling corpora from L1 speakers/writers in contrast to non-native speakers. Available established ELF corpora, including ELFA (English as a Lingua Franca in Academic Settings) and VOICE (Vienna-Oxford International Corpus of English), are confined to

audio-recordings of spoken language (see Mauranen, 2003; Seidlhofer, 2001). There are no written corpora of ELF which have been examined for variations. Third, there is an urgent need to realise the relevance and significance of ELF data in mainstream English classrooms. Given the paucity of empirical studies and also due to pedagogical and practical needs, this research aims to build a self-compiled, specialized and synchronic electronic written corpus from Turkish scholars' research articles in order to find out variants regarding Turkish academics' grammatical choices, which tend to be systematic and frequent (Seidlhofer, 2004). This study integrates English for Specific Purposes (ESP) with ELF with the suggestion that there could be some tendencies specific to each discipline, including medicine, with regard to variants in lexico-grammar usage and also the variants might change according to writers' first language such as Turkish. It is notable that English, as a medium of communication, has been used in academia as a means of incentive to promote researchers' studies in the world of publication. Thus, specifically, this study will be based on medical research articles written by academics in the Faculty of Medicine at Uludağ University in Turkey. There is no single description of ELF; therefore, compatible with the specific purpose explained above, in this study, the definition of ELF has been revisited. ELF here refers to the '*written use of academic ELF for medical purposes*' (ELFMP).

Methodology and Procedure

As Smagorinsky (2008) rightly puts it, in order for results to be credible, the methods of collection and analysis need to be highly explicit. To describe how data became results, following Bhatia (1993), two criteria were used to select appropriately constructed texts and build a representative ELF corpus (see Biber, 1993; McEnery, Xiao & Tono, 2006; Onder, 2012, for the representativeness) for the reliability of the findings. The texts that were included in the self-compiled specialized small scale written corpus belong to a specific genre. In terms of their communicative purposes, the texts were distinguishable from other genres (research articles vs. lectures) to investigate "regularities in the highly variable and basically unexplored territory of English" (Mauranen, 2003, p.519). After receiving ethical approval, the actual users of the texts, known as specialist informants, were consulted in the Faculty of Medicine at Uludağ University for their collaboration in building an ELF written corpus. Twenty recent medical research articles were collected, with a total of approximately 59,648 running words to date. It is notable that the articles were written in 2011-2012 academic year and they were about to be submitted to an international journal. The researcher, who is teaching medical English in the Faculty of Medicine, supported the specialists whose expertise are different (e.g. anatomy and biochemistry) in the department to proofread the articles before submission. During the proofreading process, the corpus was built. The empirical data was analysed manually in order to explore the potential salient features of ELF lexico-grammar following Seidlhofer (2004, p.220), who suggests eight types of potential salient features of ELF based on the VOICE as follows: dropping the third person present tense -s; confusing relative pronouns 'who' and 'which'; omission of the definite and indefinite article; incorrect forms in tag questions; unnecessary prepositions; overusing certain verbs of high semantic generality; using that-clauses instead of infinitive constructions; and overdoing explicitness. Seven of these features were confirmed with Cogo and Dewey's (2006) study. However, it is worth noting that unlike Seidlhofer's (2004) corpus, which is a computer corpus of audio recordings and transcriptions of spoken ELF interactions, the present study concentrates on academic writing. In other words, the nature of the texts, i.e., formal or informal text, is highly likely to yield different variants. Thus, there may be other potential variants that are likely to be different in the target corpora.

Results and Discussion

The initial findings with regard to lexico-grammar aspects, including the usage of '*the, a and an*', in Turkish authors' ELF corpus in medicine tend to be unproblematic and do not lead to any breakdown in communication. Definite article '*the*' as an ELF variant is in line with Seidlhofer's (2004) and Cogo and Dewey's (2006) studies. The quantitative analysis of written ELF corpus revealed that the variants include omitting definite and indefinite articles and inserting them when unnecessary (see Table 1).

Table 1. Results of quantitative analysis

<i>RAs</i>	Omission of <i>the</i>	Inserting definite articles when unnecessary	Omission of <i>a/an</i>	No. of words
1	109	9	9	4993
2	32	4	4	1843
3	89	24	9	2788
4	10	3	6	930
5	19	37	3	2698
6	72	10	34	4889
7	16	4	0	2268

8	7	3	1	793
9	101	1	5	2.821
10	430	4	19	5.779
11	46	4	4	2.568
12	72	25	10	3.276
13	10	5	12	2.068
14	9	17	2	2.504
15	29	9	5	3.240
16	11	13	5	3.224
17	47	21	4	2.407
18	34	10	4	3.326
19	8	2	2	1.136
20	95	14	15	6.097
Total	1.246	219	153	5.9648

When the medical articles were analyzed, the initial findings, with regard to the target lexico-grammar aspects, revealed that the omission of the definite article *the* when necessary had the highest frequency at 2.1% (1,246 occurrences) and inserting the definite article *the* when redundant amounted to 0.36% (219 occurrences). Indefinite articles omission as for *a/an* amounted to 0.26% (153 occurrences). In Turkish unlike English, Turkish does not require *a/an* in sentence formation. The total frequencies of variants were shown in Table 2.

Table 2. Total frequencies of target lexico-grammar aspects

Number of Research Articles	Running Words	Omitting definite and indefinite articles when necessary		Inserting definite articles when unnecessary
		<i>the</i>	<i>a / an</i>	<i>the</i>
20	Approximately 59,648	1.246 (2.1%)	153 (0.26%)	219 (0.36%)

The following extracts from the corpus could show omission of the definite article ‘*the*’ is very common in the methods part in Turkish academics’ medical research articles, especially while they are describing the tests they employed and describing the patients:

Extract 1. Omission of the definite article ‘*the*’

All of the MS patients fulfilled the McDonald criteria (McDonald et al., 2001)...

The institutional review board approved this retrospective study and all of the participants gave their informed consent...

The Hotelling T2 test was used to compare the shapes of the cerebellum between the control and relapsing-remitting MS group, ...

For example, hypertension prevalence in 1999 was 58.0 % and 64.3% among the 60 – 69 and 70 – 79 age groups, respectively...

Another regular tendency emerges in the data was use of *the* when it was not needed.

Extract 2. Insertion of ‘*the*’ when unnecessary

Multiple sclerosis can cause atrophy in the different regions of the brain, leading to changes in both the size and shape.

...local shape analysis has recently gained considerable interest because of its potential to extract differences in the shape morphologies.

Although the frequency of the omission of indefinite articles ‘*a* and *an*’ (0.26%) was not as frequent as for *the* (2.1%), the findings revealed that Turkish academics had a tendency to omit indefinite articles as shown in the following extracts:

Extract 3. Omission of indefinite articles ‘*a/an*’

...sepsis-associated encephalopathy (SAE) is a severe complication that is associated with increased morbidity and mortality...

....showed that the diagnostic accuracy of an ED point-of-care blood procalcitonin test was lower ...

..B. hominis infection is an indicator of poor personal hygiene and is a warning sign of intestinal parasitic infection...

A possible explanation for omission of *the*, inserting definite articles when unnecessary and omission of *a/an* could be the difference between Turkish and English grammar. Turkish language does not have such an article as *the*, while English language has, which needs to be confirmed with an interview to reveal academics’ perceived difficulties concerning writing in academic English.

Previous research in academic writing suggests that non-native speakers in periphery countries are at a disadvantage on publishing in English for various reasons, including the influence of first language (Wood, 1997) and its rhetoric (Leki, 2003), inequalities in the world of scholarship (Salager-Meyer, 2008), badly written articles (Coates, Sturgeon, Bohannon & Pasini, 2002). Given the challenges non-native academics face, the variants the findings revealed should not be considered as an error in academia and mainstream English classrooms. It is evident that these variants do not give rise to a breakdown in communication.

Conclusion and Pedagogical Implications

Academic writing in English is comparatively challenging for both native and non-native speakers. However, although English is a common lingua franca, research on ELF mainly focuses on speaking and, surprisingly, writing is ignored. As has been discussed in the present study, empirical studies on Turkish academics’ use of ELF in academic written discourse could provide new insights into variants concerning characteristics of ELF lexico-grammar. Findings reveal specific systematic use of lexico-grammar features, including *the*, *a* and *an*, do not lead to communication problems in academic English use of Turkish scholars. The results are not conclusive; however, it is tentatively suggested that these regular tendencies as variants could be available in Turkish academics’ research articles, which might be considered by journal editors and teachers of English. As a pedagogical finding, while teaching definite and indefinite articles, teachers can create an embracing milieu to discuss the fact that the variations discussed above with regard to the use of articles does not give rise to any communication problems which may foster active use of English a foreign language in speaking and writing. This study has some important implications for future practices. As Cogo (2012) puts it, ELF researchers have started boosting learners, teachers, and ELT practitioners to take part in the debate on English ownership although there is still a lot to do to increase critical awareness of language variation among learners (Sewell, 2012). Moreover, it is worth noting that there is a real dearth of material that reflects authentic use of English as ELF through showing the real world of discourse specific to the discipline for non-native speakers. We might use the extracts from corpus to produce authentic in-house course materials in order to raise awareness on ELF among students and teachers in mainstream classrooms, thereby contributing to pedagogy and teacher training. In line with Cogo and Dewey’s (2006) suggestion “the language used to report these findings is reminiscent of language used in error analysis” (p. 74). This study has several limitations. First, it is a preliminary study with a small corpus. Second, the data is confined to research articles in medicine. Therefore, we must be careful in making generalisations from the findings in this study. Further research could be conducted to ascertain authors’, reviewers’ and editors’ perceptions about lexico-grammar variants of ELF by

speakers of various languages, including Turkish, German, Italian or Spanish, which remains to be investigated in light of empirical data.

References

- Bayyurt, Y. (2008). A lingua franca or an international language: The status of English in Turkey. Retrieved from http://www.academia.edu/1262293/A_Lingua_Franca_or_an_International_Language_The_Status_of_English_in_Turkey
- Bayyurt, Y. (2006). Non-native English language teachers' perspective on culture in English as a foreign language classrooms, *Teacher Development*, 10(2), 233–247.
- Bhatia, V. K. (1993). *Analysing genre: Language use in professional settings*. London: Longman. In D. Biber (1993). Representativeness in corpus design. *Literary and linguistic computing*, 8(4), 243–257.
- Breiteneder, A. (2005). The naturalness of English as a European lingua franca: the case of the 'third person -s', *Vienna English Working Papers*, 14, 3–26.
- Coates, R., Sturgeon, B., Bohannan, J., & Pasini, E. (2002). Language and publication in cardiovascular research articles, *Cardiovascular Research*, 53, 279–285.
- Cogo, A. (2012). English as a Lingua Franca: concepts, use, and implications. *ELT J*, 66(1), 97–105.
- Cogo, A., & Dewey, M. (2006). Efficiency in ELF communication: from pragmatic motives to lexicogrammatical innovation, *Nordic Journal of English Studies*, 5, 59–93.
- Coury, J. G. (2001). English as a lingua franca in the Brazilian academic word. Retrieved from <http://www3.telus.net/linguisticsissues/linguafranca.htm>.
- Coşkun, A. (2010). Whose English should we teach? Reflections from Turkey. *ESP World*, 1(27), Retrieved from http://www.esp-world.info/articles_27/espworld.pdf
- Crystal, D. (2003). *English as a global language*. Cambridge, England: Cambridge University Press.
- Dewey, M. (2007). English as a lingua franca and globalization: an interconnected perspective. *International Journal of Applied Linguistics*, 17, 332–354.
- Firth, A. (1996). The discursive accomplishment of normality. On lingua franca English and conversation analysis, *Journal of Pragmatics*, 26, 237–259.
- Göksel, A., & Kerslake, C. (2005). *Turkish: A comprehensive grammar*. London: Routledge.
- Graddol, D. (1999). The decline of the native speaker, *AILA Review*, 13, 57–68.
- Graddol, D. (2006). *English next?* London: British Council.
- House, J. (1999). Misunderstanding in intercultural communication: Interactions in English as a lingua franca and the myth of mutual intelligibility. In C. Gnutzmann (Ed.), *Teaching and learning English as a global language* (pp. 73–89). Tübingen: Stauffenburg.
- Hülmbauer, C., Böhringer, H., & Seidlhofer, B. (2008). Introducing English as a lingua franca (ELF): Precursor and partner in intercultural communication, *Synergies Europe*, 3, 25–36.
- Jenkins, J. (2006). Points of view and blind spots: ELF and SLA, *International Journal of Applied Linguistics*, 16, 137–62.
- Jenkins, J. (2000). *The phonology of English as an international language. New models, new norms, new goals*. Oxford: Oxford University Press.
- Karakaş, A. (2012). English as a lingua franca: Practices of academics in a Turkish university, *Uşak Üniversitesi Sosyal Bilimler Dergisi*, 5(3), 160–179.
- Kızıltepe, Z. (2000). Attitudes and Motivation of Turkish EFL Students towards second language learning, *ITL*, 129–130.
- Leki, I. (2003). Coda: Publishing L2 writing research, *Journal of Second Language Writing*, 12, 103–105.
- Mauranen, A. (2003). The corpus of English as a lingua franca in academic settings, *TESOL Quarterly*, 37, 513–527.
- Mauranen, A. (2010). Features of English as a lingua franca in academia, *Helsinki English Studies*, 6, 6–28.
- McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies. An advanced resource book*. London: Routledge.
- Meierkord, C. (2002). 'Language stripped bare' or 'linguistic masala'? Culture in lingua franca communication. In Knapp, K., & C. Meierkord, (Eds.), *Lingua Franca Communication*. Peter Lang.: Frankfurt 109–133.
- Önder, N. (2012). An integration of corpus-based and genre-based approaches in EAP: National English examinations in Turkey. *Conference proceeding of 6th LangUE 2011 postgraduate conference*, 31–42, University of Essex, UK.
- Salager-Meyer, F. (2008). Scientific publishing in developing countries: Challenges for future, *English for Academic Purposes*, 7, 121–132.
- Seidlhofer, B. (2001). Closing a conceptual gap: The case for a description of English as a lingua franca, *International Journal of Applied Linguistics*, 11, 133–158.

- Seidlhofer, B. (2004). Research perspectives on teaching English as a lingua franca. *Annual Review of Applied Linguistics*, 24, 209–239.
- Smagorinsky, P. (2008). The method section as conceptual epicenter in constructing social science research reports, *Written Communication*, 25, 389–411.
- Sewell, A. (2012). English as a lingua franca: ontology and ideology, *ELT J*, 66(1), 89–96.
- Wood, A. (1997). International scientific English: Some thoughts on science, language and ownership, *Science Tribune*. Retrieved from: <http://www.tribunes.com/tribune/art97/wooda.htm>.

Anadili İngilizce Olmayan Akademisyenlerin İngilizce Yazılı Metinlerin Derlenmesi: Türk Akademisyenlerin Tıbbi Makaleleri

Özet

Anadili İngilizce olmayan bireylerin farklı ülkeden insanlar ile iletişim kurmak için İngilizceyi kullanmaları literatürde *English as a lingua franca (ELF)* olarak tanımlanmaktadır. Yayın dünyasında ELF'in önemini gözönünde bulundurduğumuzda, ELF'in dilbilimsel tanımı üzerine yapılan bilimsel çalışmalar çok azdır. Literatürdeki çalışmalar İngilizce konuşma dilini yansıtan ses kayıtları ile sınırlıdır ve herhangi bir yazılı ELF bütüncesi yoktur. İngilizce eğitimi verilen sınıflarda ELF ile ilgili bilincin artırılması acil bir ihtiyaçtır. Bu çalışma ile ana dili İngilizce olmayan akademisyenlerin İngilizceyi ELF sözcüğe-dilbilgisi belirgin özellikleri araştırmak için tıp alanında Türk akademisyenlerinin yazdığı İngilizce araştırma makalelerinin araştırmacı tarafından derlenmiş bir bütüncü oluşturmayı hedeflenmektedir. Özellikle 'a' ve 'an' değişkenlerinin kullanımı incelenmiştir. 20 tıbbi İngilizce araştırma makaleleri toplanmış ve derlem yaklaşık 59.648 kelimedenden oluşmuştur. Ampirik veriler hedeflenen potansiyel ELF dilbilgisi yapılarının belirgin özellikleri keşfetmek için elle analiz edildi. Seidlhofer (2004) tarafından oluşturulan ELF etkileşim ses kayıtları ve transkripsiyonları bir bilgisayar tabanlı derlem aksine, bu çalışma tıbbi makalelerden oluşan yazılı bir derleme odaklanarak hazırlanmıştır. Hedeflenen dilbilgisi öğeleri ile ilgili ilk bulgular Türk akademisyenlerin İngilizce tıbbi makalelerinde % 2.1 (1.246 kere) oranıyla en yüksek frekansa sahip belgeli tanımlık olan *the*'nin kullanılması gereken yerlerde kullanılmadığını ve %0.36 (219 kere) oranında *the*'nin kullanılmaması gereken yerlerde kullanıldığını ortaya çıkarmıştır. Belirsiz tanımlık olan *'a/an*'nin kullanılması gereken yerlerde 0.26% (153 kere) oranında kullanılmadığı gözlenmiştir. Bu bulgular *'the'*, *'a'* and *'an'* ile ilgili dilbilgisi yapılarının sorunsuz olduğunu ve iletişimde herhangi bir probleme yol açmadığını göstermiştir. Oluşturulan derlemden söylem örnekleri sunulmuş ve pedagojik sonuçları literatür ışığında tartışılmıştır.

Anahtar sözcükler: ELF, Özel amaçlar için İngilizce, Bütüncü, Tıbbi makale.