

Lexical Bundles in Research Article Acknowledgments: A Corpus Comparison

Araştırma Makalelerinin Teşekkür Bölümlerindeki Sözcüksel Öbekler: Bir Derlem Karşılaştırması

Elif TOKDEMİR DEMİREL¹, Hesamoddin SHAHRİARİ AHMADİ**

ABSTRACT: Benefiting from a corpus based design, this paper attempts to compare acknowledgments across three academic disciplines written by Turkish and Iranian authors in terms of the formulaic language employed. Studies into the academic register have commonly chosen the research article (RA) and its various sections as their object of investigation. Despite the multitude of studies carried out on RAs, there are still many aspects of this register which merit closer analysis. One relatively understudied section of the RA is the acknowledgments. The findings reveal a difference in the rate of multiword expressions used by the two groups of writers. It was found that Iranian authors were inclined to overuse a certain set of formulaic expressions; whereas, Turkish authors generally avoided using these expressions in their texts. Possible explanations for this observed difference are discussed in detail in the paper.

Key Words: lexical bundle, acknowledgments, corpus linguistics

ÖZ: Bu çalışmada derlem yöntemi kullanılarak üç farklı akademik disiplinde, Türk ve İranlı araştırmacılar tarafından yazılmış olan teşekkür bölümleri, kullanılan kalıplaşmış dil açısından karşılaştırılmıştır. Akademik yazı türü üzerine yapılan çalışmalar, araştırma makalelerini ve alt bölümlerini yaygın olarak araştırma konusu yapmışlardır. Araştırma makaleleri üzerine yapılmış olan birçok çalışma bulunmasına rağmen, bu yazı türünün birçok yönü daha yakından araştırılmayı gerektirmektedir. Bilimsel makalelerin yetersiz araştırılmış bir bölümü de teşekkür bölümleridir. Sonuçlar, iki yazar grubunun sözcüksel öbekleri kullanma oranlarında farklılıklar olduğuna işaret etmektedir. İranlı yazarların belli bir grup sözcüksel öbeği aşırı kullanma eğiliminde olduğu, buna karşılık Türk yazarların genellikle bu tür ifadeleri yazılarında kullanmaktan kaçındıkları bulunmuştur. Gözlemlenen bu farkların olası sebepleri detaylı olarak makalede tartışılmaktadır.

Anahtar Sözcükler: sözcüksel öbek, teşekkür bölümü, derlem dilbilimi

1. INTRODUCTION

The scientific community has started to become increasingly globalized with English serving as the lingua franca. Researchers not only from English speaking countries but also from other language backgrounds have been and will be contributing to research in all fields of science. International journals accept publications from countries all around the world and this adds a certain amount of diversity to the use of English in research publications. Although the language of research is commonly thought of as being uniform, it has been shown through research that there is a certain amount of variation across disciplines and writers of different language and cultural backgrounds within academic language. Several language specialists including Brodkey (1987) and Crowley (1991) have even rejected the term 'academic prose' due to the large-scale variability among texts in this register.

Researchers are becoming increasingly interested in variation within academic prose caused by various factors including language background and disciplines. Corpus linguistics is one of the methods commonly used in recent decades for exploring variation in language use owing to the practicality and reliability it offers in handling vast amounts of data. Corpus

¹ Yrd. Doç. Dr., Karadeniz Technical University, Faculty of Letters, Trabzon-Turkey, e-mail: elif@ktu.edu.tr

** Dr., Ferdowsi University of Mashhad, Department of English, Mashhad, Iran, e-mail: hesam.shahriari@gmail.com

linguistics makes possible quantitative analyses of large amounts of language data with the use of computational analytical tools which would not have been possible through traditional methods of research. Studies contrasting specific corpora have proven the existence of variation even within a single genre e.g. among textbooks and research articles in biology and history (Conrad, 2001) among medical research articles (Biber and Finegan, 2001) , among technical texts (Carrio-Pastor,2009). As can be seen, the possibilities corpus linguistics provides for researching variation are endless.

Research on linguistic variation using a corpus approach has revealed differences between disciplines in patterns of language use which are specifically related to the purposes and methods of each discipline, and have consequently deepened our understanding of variation (Biber, 2009; Biber et al., 1998, Aijmer and Stenstrom 2004, Biber, 2006, Biber and Burgess, 2000; Conrad and Biber, 2000). Kachru (2008) states that corpus linguistics methods have direct relevance to the study of subjects such as language variation and lexicography and have potential to provide more reliable results compared to intuitive studies of applied linguistics. The advantages of cross-linguistic studies on phraseology based on corpora over more traditional methods have been acknowledged among others by Colson (2008) who points out that corpus-based studies make possible statistical analyses of the various categories of set phrases as well as offering a very reliable methodology. Biber's (2009) study, for example uses a combination of corpus-based and corpus-driven approaches to investigate the most common multi-word sequences in conversation and academic writing. The study reveals that multi-word patterns common to speech versus academic writing are fundamentally different from each other, and concludes that:

Formulaic language is very important in both conversational and written academic discourse, but it is realized in very different ways linguistically: fixed sequences that represent clause fragments in conversation, versus formulaic frames that consist of noun phrase and prepositional phrase fragments in academic writing. (Biber 2009, p. 301)

Corpus linguistics methods have shed light on our understanding of and have challenged previous assumptions related to the register of academic writing. For example, Biber and Gray (2010) have discovered differences between academic writing and spoken registers which are somewhat in contradiction with common beliefs about academic writing. They have shown that academic writing is not structurally 'elaborated' and that conversation makes use of subordinate clauses much more than academic writing. Instead of using subordination, in academic writing meaning is compressed by means of phrasal modifiers embedded in noun phrases.

The control of multi-word expressions is an important element of fluent language production and presents a challenge specifically to student writers and non-native speakers alike. For example, Romer (2010) who created a model for identifying the phraseological profile of a specific text type for investigating book reviews highlights the importance of knowledge about modification of common multi-word expressions, their functions and positions in a text for non-native writers. Academic clusters as central elements of academic discourse have been proven to provide important insights into the structure of academic discourse. Hyland's (2008) study on academic clusters in research articles, doctoral dissertations and master's theses has shown that research articles contain far fewer clusters which are mostly text-oriented in contrast to master's theses which display different patterns. Thus, academic clusters can function in differentiating genres and reflect the writers' characteristics. For example Hyland (2008) interprets the high

reliance on formulaic expressions by master's theses writers as a strategy employed by less confident or less proficient students in constructing their texts. The study of clusters in academic discourse would benefit instructional practices especially in advanced settings by allowing teachers to focus on the specific ways of creating meaning appropriate to particular kinds of writing rather than relying on the massive literature describing the research article (Hyland 2008, p.60).

Research on the lexical characteristics of academic discourse employing a corpus approach has also started to focus on specific subsets of language and various sections of research articles providing evidence for disciplinary and cross-cultural variation. For example (Martin-Martin & Burgess, 2004) compared research article abstracts in two disciplines of social sciences (psychology and phonetics) written by English and Spanish writers and found cross-linguistic differences in the frequency and style of academic criticism. Öztürk (2007), on the other hand examined introduction sections of research articles in two sub disciplines of applied linguistics. He found substantial differences between the move structures employed in these sub disciplines although they belonged to the same main discipline. In a study on the methods section of research articles, Williams (2010) found that there are cross-cultural differences in the use of first-person verbs in Spanish and English medical research articles.

Although research on sections of research articles has been flourishing, there are relatively fewer studies on research article acknowledgments. Research article acknowledgments nevertheless deserve more attention since they function as one of the shapers of a writers' identity by associating the writer with a group of researchers and a certain circle of thought. Hyland and Tse (2004) state in their article on dissertation acknowledgments that lack of studies on this specific text type affects novice writers' ability to express their capabilities and intellectual autonomy when they are recognizing academic help they receive from others. Acknowledgments are regarded by other researchers as "records of often significant intellectual influence" (Cronin and Overfelt, 1994, p.183), "public gratitude for private gestures" and "expressions of solidarity" (Cronin, Mc.Kenzie & Stiffler, 1992).

Existing studies on research article acknowledgments have revealed that although the format of acknowledgments change from journal to journal and from discipline to discipline (McCain, 1991), the language used in them is quite formulaic following a relatively restricted range of lexico-grammatical patterns (Hyland and Tse, 2004). For example twenty year analysis of acknowledgments by Cronin, Shaw and La Barre (2003) has shown that this established practice in research article publications reflect "symbolic and substantive significance enormously dependent on context, moment and behavioral characteristics of different scholarly tribes." (2003, p.109). Most studies on acknowledgments (Cronin and Franks, 2006; Cronin, Shaw & La Barre, 2003, 2004; Giles and Council, 2004), however focus more on the social and intellectual role of acknowledgments and collaboration in the field of scholarly research but not the lexical characteristics and they do not follow a corpus linguistics approach.

Research on acknowledgment sections of research articles have also been done in various disciplines other than social sciences. For example, Salager-Meyer et al. (2009) have analyzed the acknowledgment paratext of medical research articles and concluded that 'backstage solidarity' as termed by Goffman (1959) shows variability depending on context. They suggest that communicative and sociocultural norms of contributions to academic studies are dependent on

disciplines as well as language and context. Another study which was carried out through a survey with medical writers in the health care industry by Phillips (2009) points to the lack of acknowledgment of medical writers who have been involved in the writing process of medical research articles and concludes that the trend in the industry seems to be towards using more assistance from medical writers and being more transparent in acknowledging them. In the life sciences field Cronin and Franks (2006) have carried out an analysis of 1000 research article acknowledgments from a specific journal in the field: *Cell*. Their study has relevance to the present study as it points to the importance of acknowledgments as indicators of features defining contemporary research in the life sciences such as intense coauthorship and subauthorship.

Drawing from the idea that meaning is expressed differently in different text types, and Romer's (2010) suggestion to focus on subsets of language when studying phraseology, we explore lexical variation in the restricted language of research article acknowledgments across cross-linguistic boundaries. This study compares multi-word sequences in research article acknowledgments written by two groups of non-native speakers: Turkish and Iranian authors using a corpus linguistics approach. Through this study, we intend to shed light on the differences between non-native speaker writers from two different language backgrounds: Turkish and Persian in regards to the most commonly used multi-word sequences in acknowledgments and their functions. The acknowledgments were selected from three disciplines, namely Social Sciences, Medical Sciences and the Hard Sciences. This measure allows for comparisons across disciplines alongside those across L1 background. A parallel sub-corpus of native speaker acknowledgments was used as a reference when interpreting the differences that exist between Turkish and Iranian writers of research.

2. METHOD

2.1. Corpus

The corpus used for the study consisted of 270 Research Article acknowledgments. The acknowledgments varied in length, discipline and authors' first language backgrounds. The selected texts were written by authors from three linguistic backgrounds. These included 90 acknowledgements written by Turkish authors, 90 by Iranian authors and another 90 by authors for whom English was a native language. The 90 acknowledgements in each parallel sub-corpus were equally chosen from three disciplines: the Social Sciences (N=30), the Hard Sciences (N=30) and the Medical Sciences (N=30). The acknowledgments collected for this study varied in length, with the average acknowledgment in the Turkish corpus having 73.1 words, and the average Iranian acknowledgment consisting of almost twice as many words (130.5). The texts were chosen randomly, and therefore the observed difference in length can be seen as being characteristic of all acknowledgments written by authors from the two groups.

Table 1: Details of Sub-corpora Used in the Analysis

Author Background	Social Science	Hard Science	Medical Science	Total
Turkish	30	30	30	90
Iranian	30	30	30	90
Native-English Speaker	30	30	30	90
Total	90	90	90	270

The fact that acknowledgements are not always present in each and every research articles makes them difficult to locate and include in a study of this nature. In addition, the restricted

range of lexico- grammatical patterns used in acknowledgements justifies the cut-off point of 30 texts from each discipline and language background adopted in the present research.

2.2. Procedure

The compiled corpus was analyzed using AntConc 3.2.0. (Anthony, 2011) to identify the existing lexical bundles in each corpus through an empirical analysis. Figure 1 shows a screenshot from AntConc. Our adopted definition of lexical bundles* is borrowed from Biber et al. (1999), who define them as the most recurrent multi-word sequences in a given register. The study focused on units ranging between 3 to 7-word sequences. The frequency cut-off used for the identification of bundles was set at 4 times per million words. In comparison to other studies, this criteria is quite lenient, but given the investigative nature of our study and also, considering the short length of each text, it best serves our purposes and can be more informative. This restriction also helps prevent methodological issues caused by the difference in the number of words in each corpus. In order to avoid idiosyncratic expressions by an individual author, it was agreed that a sequence would have to appear in at least 4 different texts to be included into the analysis. This decision was not restrictive, since the texts in the acknowledgement were too short to include recurrent expressions.

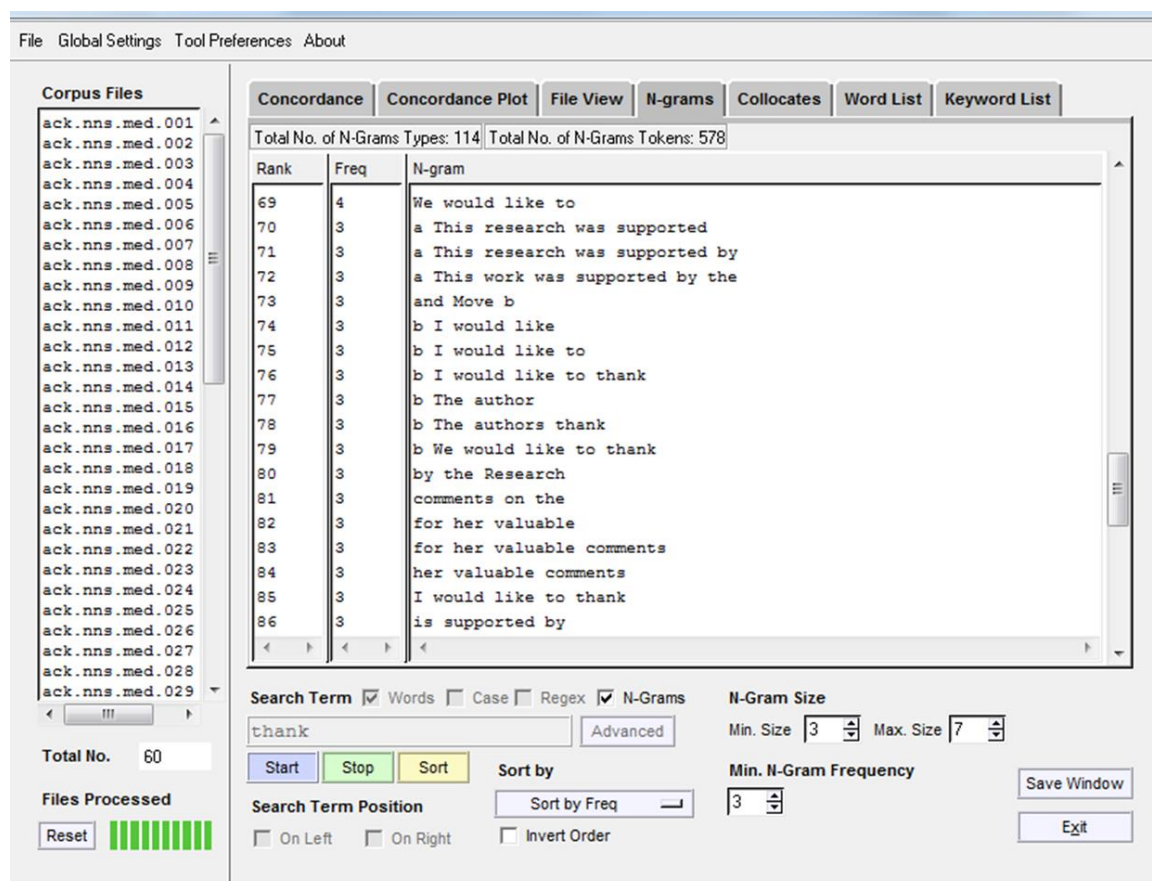


Figure 2. 1. A Screenshot from AntConc 3.2.0 for Lexical Bundles Search of the Acknowledgment Corpus

* The term 'Lexical Bundle' is used in the study to refer to the most recurrent multi-word sequences in a given register as defined by Biber et al. (1999). Throughout the study, similar terms such as 'multi-word patterns/expressions/sequences', 'academic clusters' and 'formulaic expressions' have been used to refer to 'lexical bundles'

The lexical bundles found during the course of this investigation do not represent complete structural units. Rather, they are mostly bridging elements, linking two structural units (i.e., phrases or clauses) to each other such as ‘would like to’ or ‘was supported by’. The relatively high frequency with which these bundles recur in the corpus reveals that they are most likely stored and used as pre-fabricated linguistic patterns. As a result, even though the identified bundles did not always constitute a whole structural unit, they could be used as an index for determining the degree to which authors use the formulaic principle to form their utterances.

Comparisons were made between the sub-corpora in order to identify differences across disciplines, as well as first language background. In the first analysis, Iranian and Turkish authors were compared in terms of the rate of lexical bundles used in their acknowledgements. Following this, comparisons were made between the disciplines (Social Science, Hard Science and Medical Science articles). The corpus of acknowledgments from speakers of English as a native language was only included in this study for referential purposes, and served as a middle-ground standard based on which the two groups of Turkish and Iranian acknowledgments could be compared.

3. FINDINGS

The results of the first round of analysis, comparing Iranian and Turkish acknowledgements, revealed that the Turkish corpus consisted of only 40 identified bundles, while 132 bundles were identified in the Iranian corpus. It should be noted that all bundles containing proper nouns (e.g., university names, funding organizations, etc.) were excluded from the analysis. An analysis of the corpus of native English-speaking researchers, on the other hand, resulted in 117 bundles.

The most frequent lexical bundle in the Iranian corpus of acknowledgements was *would like to* which recurred 42 times per million words. The same bundle ranked third in the list of most frequent bundles in the Turkish corpus, with 11 recurrences per million words. In the corpus of native English-speaking authors, it was the second most frequent bundle with 11 attestations per million words. This bundle followed a subject and came before a verb, forming a complete structure in the acknowledgements corpus. The most common subject for this bundle in the Iranian corpus was *we* (15), followed by *I* (12) and *the authors* (8); and in one instance, an Iranian author referred to himself in the third person by using the subject *he*. *We* was also the most common subject for this bundle in the Turkish corpus, occurring a total of 7 times. The second most subject for Turkish authors was *the authors* (4), and the first person pronoun, *I*, was only seen two times per million words. The ten most frequent lexical bundles found in the two corpora are displayed in Table 2 below: Use of personal pronouns can be interpreted within the frame of academic genre and its cultural implications (Uzun and Huber, 2002).

Table 2: Ten Most Frequent Bundles Found Within the Three Corpora

Iranian corpus		Turkish corpus		Native speaker corpus	
bundles	f	bundles	f	bundles	f
would like to	42	was supported by	22	like to thank	16
like to thank	29	supported by the	14	would like to	16
would like to thank	27	would like to	11	of this article	15
of Medical Sciences	19	This work was	10	would like to thank	13
University of Medical	19	like to thank	9	version of this	11

University of Medical Sciences	18	study was supported	9	the University of	9
to thank the	16	study was supported by	9	This work was	9
like to thank the	14	This study was	9	authors would like	8
was supported by	14	This work was supported	9	authors would like to	8
would like to thank the	13	was supported by the	9	earlier version of	8

The most common lexical bundle found in the Turkish corpus of acknowledgements was *was supported by*, occurring 22 times per million words. The same bundle was the ninth most frequent bundle in the Iranian corpus with 14 attestations. This bundle was preceded by a subject and followed by a single noun or noun phrase. Turkish authors most frequently chose *this study* as their subject (8); after that, they commonly chose *this work* (7) and *this research* (5), respectively as their subjects. Only one Turkish author used the phrase *this project* as the subject of the discussed bundle. Iranian authors, on the other hand, most commonly selected the phrase *this work* (8) followed by *this study* (4) and *this research* (3). A comparison with the corpus of acknowledgements by native English-speaking authors reveals that there was no great difference between the groups in terms of the selected subjects. Similar to Turkish and Iranian writers, authors speaking English as a native language used *this work* (4), *this research* (2) and *this study* (1).

The verbal expressions used in the multiword units were also of great interest. The resulting bundles from the analysis of the Turkish corpus included four main verbal expressions, namely *supported by* (19 bundles), *would like to* (4 bundles), *thank* (4 bundles), and *(to be) grateful* (1 bundle). Going through the Iranian corpus, one cannot fail to notice the greater number of verbal expressions within the identified bundles. The most commonly used verbal expressions by Iranian writers include: *thank* (found in 33 bundles), *would like to* (29 bundles), *(to be) supported by* (9 bundles), *to express* (7 bundles), *wish to* (4 bundles), *(to be) grateful* (4 bundles), *appreciate* (2 bundles) and *(to be) thankful* (1 bundle). From among these verbs, *to express*, *(to be) thankful*, and *wish to thank*, were also spotted in the Turkish corpus, but each with only two instances of occurrence. The verb *appreciate* was not found in the Turkish corpus at all. Comparing these verbs with those found in the corpus of native-speaker acknowledgments also resulted in an interesting observation. Native speakers of English included unique verbal expressions such as *acknowledge* (2 bundles), *(to be) funded* (2 bundles), *participate* (1 bundle) and *to be sponsored* (1 bundle), alongside verbal expressions shared with the other two groups, namely, *would like to* (11 bundles), *thank* (8). The difference in the number of identified verbal expressions in the Turkish and Iranian corpora could be attributed to the difference in the overall number of lexical bundles used by Turkish and Iranian authors. By norming the data, we can see that from every 10 bundles in the Turkish corpus, one contained a verbal expression (10%); and as for the Iranian corpus, 6.8% of bundles included verbal expressions. Of the total number of bundles found in the native-speaker corpus of acknowledgments, 5.9% included verbal expressions, meaning that our two groups of non-native speakers over-used these expressions in their writing compared to their native-speaker counterparts.

4. RESULTS AND DISCUSSION

Interpreting the relatively large gap between the two groups is made possible by looking at the extent to which native speakers use lexical bundles in their corpus of acknowledgements. An analysis of the acknowledgements corpus of native English-speaking authors resulted in 117

bundles. This number is less compared to the Iranian corpus, but still considerably larger than that of the Turkish corpus. The results from this analysis show that Iranian authors tend to rely on pre-fabricated patterns to a great extent, even more than authors who spoke English as a native language. On the other hand, Turkish authors are apparently under-using these expressions and depend more on creatively-formed expressions. Type token ratio (TTR) is a measure used in the study to compare the lexical density of the texts in the two parallel acknowledgments corpora Biber et al. (1998). TTR is calculated by dividing the number of unique word types in a corpus with the number of tokens. Looking at the TTR in each of the two corpora, the finding relating creativity of expressions could be corroborated. For the Turkish and Iranian corpora, the TTR was 0.33 and 0.24, respectively. This basically means that the range of words used by Turkish writers was more varying than that of Iranians, who tended to use the same words repeatedly throughout their acknowledgements.

This difference in the use of varying words can also be traced across the different disciplines in the two languages. The TTR for acknowledgments written by Turkish authors in the field of Social Sciences was 0.52, while the ratio for Iranian acknowledgments in the same discipline was 0.33. In acknowledgments from Hard Sciences, the TTR for Turkish and Iranian texts was 0.38 and 0.26, respectively. Finally, in Medical Science articles, the TTR for Turkish and Iranian acknowledgments was 0.42 and 0.32, respectively. The three disciplines consistently point towards the inclination among Turkish authors to write with varying words and avoid repetition, and the tendency among their Iranian counterparts to use the same words more repeatedly.

Overall, the difference in the rate of lexical bundles used by the two groups of authors seems to be the most significant finding of this study. According to Schmitt (2010), non-native use of formulaic language can be observed along three dimensions: the amount of sequences used; the accuracy with which they are used; and finally, the speed/goodness of the underlying intuitions. The two groups of non-native speakers in this study were different in terms of the amount of sequences they used. In order to determine whether they were also different in their accuracy of use, we went through the identified bundles. We found that one of the bundles frequently recurring in the Iranian corpus (*would like to appreciate*) was incorrect, as judged by native speakers of English. Some mistakes (e.g., *thank to, owe the debt of*) were also found in the Turkish corpus; however, their short length and relatively low frequency prevented them from appearing in the resulting lexical bundles. As a result, it can be said that there is little difference between the two groups with regards to accuracy in the use of lexical bundles.

The infrequent use of bundles by Turkish authors could be attributed to the element of avoidance. Laufer (2000) argues that avoidance is an outcome of cross-linguistic influence and is a strategy used by non-native speakers to overcome difficulties in communication. Avoidance could also be a result of the inherent difficulty of a target form. A study carried out by Laufer and Eliasson (1993) suggests that it is a systematic incongruence between the first and second language which determines avoidance, as opposed to inherent difficulty. As a result, if we were to attribute the difference in the quantity of bundles to Turkish authors' underuse or avoidance, we could claim that the difference between Turkish and English may have brought about this result. Observing the results from a different perspective, one may also argue that the observed difference may be due to Iranian authors' overuse of certain multiword expressions. Various studies have shown that some non-native speakers tend to use certain formulaic sequences frequently and repetitively, because they view them as reliable 'safety nets' which can be confidently used at times of uncertainty (De Cock, 2000; Foster, 2001; Granger, 1998).

This study compared Turkish and Iranian corpora of research article acknowledgments with regards to formulaic language. The lexical bundles found in the course of this descriptive investigation provide a profile of these two groups of non-native English users, and show that

authors from different first language backgrounds, even those coming from higher levels of proficiency, exhibit different patterns of language usage. The most obvious difference between the two groups investigated in this paper was in the frequency of bundles. We feel that this is the outcome of different strategies. While Turkish writers tried to avoid using a range of lexical bundles (either because of cross-linguistic dissimilarities or inherent difficulty), Iranian writers frequently depended on these expressions as a reliable strategy for efficient communication. Future studies can compare other sections of research articles or possibly other registers to see whether these findings are consistent.

Acknowledgments

The authors would like to thank Prof. Dr. Douglas Edward Biber and Shelley Staples from Northern Arizona University who contributed to this study with their valuable feedback and constructive comments.

5. REFERENCES

- Aijmer, K. & Stenström, A. (Eds.) (2004). *Discourse Patterns in Spoken and Written Corpora*. Amsterdam: Benjamins.
- Biber, D. & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Specific Purposes*, 9, 2-20.
- Biber, D. (2006). *University Language: A Corpus-Based Study of Spoken and Written Registers*. Amsterdam: Benjamins.
- Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, 14, 275-311.
- Biber, D. & Finegan, E. (2001). Intra-textual variation within medical research articles. In S. Conrad, and D. Biber (Eds.), *Variation in English: Multidimensional Studies*, pp.108-124, London: Longman.
- Biber, D. & Burges, J. (2000). Historical change in the language use of women and men: gender differences in dramatic dialogue. *Journal of English Linguistics*, 28, 21-37.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *The Longman Grammar of Spoken and Written English*. London: Longman.
- Brodkey, L. (1987). *Academic Writing as Social Practice*. Philadelphia: Temple University Press.
- Carrio-Pastor, M. L. (2009). Contrasting specific English corpora: Language variation. *International Journal of English Studies*, Special Issue, 221-233.
- Colson, J-P. (2008). Cross-linguistic phraseological studies: An overview. In S. Granger & F. Meunier (Eds.), *Phraseology: An Interdisciplinary Perspective* (pp. 191-206), John Benjamins Publishing Company .
- Conrad, S. (2000). Adverbial marking of stance in speech and writing. In S. Hunston & G. Thompson (Eds.), *Evaluation in text: Authorial Stance and the construction of Discourse* (pp. 56-73), Oxford: Oxford University Press.
- Conrad, S. (2001). Variation among disciplinary texts: a comparison of textbooks and journal articles in biology and history. In S. Conrad, and D. Biber (Eds.), *Variation in English: Multidimensional Studies* (pp.94-108), London: Longman.
- Cronin, B. & Franks, S. (2006). Trading Cultures: Resource Mobilization and Service Rendering in the Life Sciences as Revealed in the Journal Article's Paratext. *Journal of the American Society for Information Science and Technology*, 57, 1909-1918.
- Cronin, B. & Overfelt, K. (1994). The scholar's courtesy: A survey of acknowledgment behavior. *Journal of Documentation*, 50, 165-196
- Cronin, B., Mc. Kenzie, G. & Stiffler, K. (1992). Patterns of acknowledgment. *Journal of Documentation*, 48, 491-516

- Cronin, B., Shaw, D., & La Barre, K. (2003). A cast of thousands. Coauthorship and sub-authorship collaboration in the twentieth century as manifested in the scholarly literature of psychology and philosophy. *Journal of the American Society for Information Science and Technology*, 54, 855–871.
- Anthony, L. (2011). AntConc (Version 3.2.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.antlab.sci.waseda.ac.jp/>
- Cronin, B., Shaw, D., & La Barre, K. (2004). Visible, less visible, and invisible work: Patterns of collaboration in twentieth century chemistry. *Journal of the American Society for Information Science and Technology*, 52, 160–168.
- Crowley, S. (1991). A personal essay on freshman English. *Pre/Text*, 12, 155-176.
- De Cock, S. (2000). Repetitive phrasal chunkiness and advanced EFL speech and writing. In Mair, C., and Hundt, M. (Eds.), *Corpus linguistics and linguistic theory* (pp. 51-68), Amsterdam: Rodopi.
- Giles, C.L. & Council, I. G. (2004). Who gets acknowledged: Measuring scientific contribution through automatic acknowledgment indexing. *Proceedings of the National Academy of Sciences of the United States of America*, USA, 101, 17599-17604
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In Cowie, A.P. (ed.), *Phraseology: An interdisciplinary perspective* (pp.145-160), Amsterdam: John Benjamins.
- Hyland, K. & Tse, P. (2004). “I would like to thank my supervisor”: Acknowledgments in graduate dissertations. *International Journal of Applied Linguistics*, 14, 259-275
- Hyland, K. (2008). Academic clusters: text patterning in published and postgraduate writing, *International Journal of Applied Linguistics*, 18, 40-62
- Kachru, Y. (2008). Language variation and corpus linguistics. *World Englishes*, 27, 1–8
- Laufer, B. (2000). Avoidance of idioms in a second language: The effect of L1-L2 degree of similarity. *Studia Linguistica* 54, 186-196.
- Laufer, B., & Eliasson, S. (1993). What causes avoidance in L2 learning: L1-L2 difference, L1-L2 similarity, or L2 complexity? *Studies in Second Language Acquisition*, 15, 35-48.
- Martin-Martin, P., & Burgess, S. (2004). The rhetorical management of academic criticism in research article abstracts. *Text*, 24, 171-195
- McCain, K.W. (1991). Communication, competition, and secrecy: The production and dissemination of research-related information in genetics. *Science, Technology, & Human Values*, 16, 491–516.
- Ozturk, I. (2007). The textual organization of research article introductions in applied linguistics: variability within a single discipline. *English for Specific Purposes*, 26, 25-38
- Phillips, S. G. (2009). Authorship and writing practices in the health care industry. *AMWA Journal*, 24, 4-8
- Romer, U. (2010). Establishing the phraseological profile of a text type: the construction of meaning in academic book reviews. *English Text Construction*, 3, 95-119
- Salager-Meyer, F.; Ariza, M. A. A.; Berbesi, M. P. (2009). “Backstage solidarity” in Spanish- and English-written medical research papers: Publication context and the acknowledgment paratext. *Journal of The American Society for Information Science and Technology*, 60, 307-317
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. New York: Pelgrave Macmillan.
- Uzun, L., & E. Huber (2002) (Eds.). *Türkçede Bilgi Yapısı ve Bilimsel Metinler*. Essen: Die Blaue Eule.
- Williams, I. (2010). Cultural differences in academic discourse: Evidence from first-person verb use in the methods sections of medical research articles. *International Journal of Corpus Linguistics*, 15, 214-239

Genişletilmiş Özet

İngilizcenin akademik bir ortak dil haline gelmesi bilim çevresini küresel bir alan haline getirmiş ve her milletten bilim insanlarının her bilim dalında bu alana katkı yapmasını da olanaklı kılmıştır. Uluslararası bilimsel dergiler dünyanın her tarafından, ana dili İngilizce olmayan araştırmacıların bilimsel yayınlarını kabul etmekte ve bu da yayınlarda kullanılan İngilizce’ye belli bir oranda çeşitlilik katmaktadır. Her ne kadar akademik yazı dilinin değişkenlik içermeyen tek biçimli bir yazı türü olduğu düşünülse

de, arařtırmalar hem farklı alıřma alanları, hem de farklı kltrel ve dilsel altyapılardan gelen yazarlar tarafından kullanılan akademik dilin farklılık gsterdiğine iřaret etmektedir.

Akademik dilde, farklı etmenler tarafından yaratılan bu farklılařma sreci son yıllarda bir arařtırma konusu haline gelmiřtir. Derlem yntemi, byk sayıdaki dil rneklerini kısa zamanda ve pratik olarak arařtırma olanađı sađlaması aısından, dildeki bu tr eřitliliđin arařtırılmasında sıklıkla kullanılan bir yntem haline gelmiřtir. zelleřmiř yazı trlerini karřılařtıran derlem arařtırmaları tek bir yazı trnn bile kendi iinde belli bir oranda eřitlilik gsterdiğini bulmuřtur. rneđin biyoloji ve tarih alanındaki ders kitaplarında (Conrad, 2001), tıp alanındaki arařtırma makalelerinde (Biber and Finegan, 2001) ve teknik metinlerde (Carrio-Pastor, 2009) yapılan arařtırmalar metinlerin kendi iinde eřitlilik ierdiğini gstermiřtir.

Derlem yntemi kullanılan alıřmalar akademik dildeki farklılařma ile ilgili olan bilgilerimizi derinleřirmiřtir. (Biber, 2009; Biber et al., 1998; Aijmer and Stenstrom 2004; Biber, 2006; Biber and Burgess, 2000; Conrad and Biber, 2000). Kachru (2008)'ya gre, derlem dilbilimi yntemleri dilde farklılařma konusunun arařtırılmasına ok yatkındır ve aynı zamanda uygulamalı dilbilimdeki sezgiye dayalı alıřmalara oranla daha gvenilir sonular vermektedir. Ayrıca derlem alıřmaları akademik yazı dili ile ilgili kavrayıřımıza katkıda bulunmuř, hatta varolan bazı yerleřmiř yargıları da rtmřtir. rneđin, Biber ve Gray (2010) derlem yntemini kullanarak, yazı dili ve konuřma dili arasında, varolan inanıřlarla eliřen bazı farklar bulmuřlardır. rneđin, inanılanın aksine, akademik yazı dilinin yapısal olarak ok karmařık olmadıđını ve konuřma dilinin birleřik cmle yapılarını yazı diline gre daha fazla kullandıđını bulmuřlardır.

Bu alıřmada arařtırma makalelerinin teřekkr blmlerinden oluřan bir derlem kullanılarak  farklı akademik disiplinde Trk ve İranlı arařtırmacılar tarafından yazılmıř olan teřekkr blmleri, kullanılan kalıplařmıř dil aısından karřılařtırılmıřtır. Kullanılan derlem toplam 270 teřekkr blm iermekte olup 3 paralel alt-derlemden oluřmaktadır. Alt derlemlerin her birinde İranlı, Trk ve Amerikalı olmak zere farklı yazar grupları tarafından yazılmıř 90 teřekkr blm bulunmaktadır. Ayrıca her alt derlem de, Fen bilimleri, sosyal bilimler ve tıp bilimleri olmak zere,  farklı alıřma alanından eřit sayıda (30) teřekkr blm iermektedir.

Oluřturulan derlem zerinde, AntConc 3.3.0. programı kullanarak szcksel bekler bulunmuřtur. Biber et al. (1999) szcksel bekleri bir yazı trnde en sık olarak tekrarlanan oklu kelime dizileri olarak tanımlamaktadır. Bu alıřmada da bu tanımdan yola ıkılarak szcksel bekler arařtırılmıřtır. alıřmada 3 – 7 kelime uzunluđundaki kelime dizileri gz nne alınmıřtır. Tekrarlanma sıklıđının alt limiti ise her milyon kelimedede 4 defa olarak belirlenmiřtir. Alt limit, derlemi oluřturan yazı rneklerinin kısa oluřu sebebiyle dřk tutulmuřtur. Karřılařtırmalarda hem farklı arařtırma alanları, hem de ana dili farklı olan yazar grupları gz nne alınmıřtır.

Analiz sonucunda, Trk arařtırmacıları tarafından yazılmıř olan teřekkr blmlerinde 40, İranlı yazarlar tarafından yazılan metinlerde 132 ve Amerikalı yazarlar tarafından yazılanlarda ise 117 szcksel bek bulunmuřtur. Bu sonulara gre řu yorum yapılabilir: İranlı yazarlar szcksel bekleri kullanmaya daha eđilimlidirler, fakat bunun karřısında Trk yazarlar bu yapıları daha az tercih etmekte ve birbirinden farklı, zgn ifadeler kullanmaya zen gstermektedirler. Bu yorum derlem ynteminde dil rneklerinin kelime zenginliđini lmekte kullanılan type/token (tr/rneke) oranı ile aıklanabilir. Trk ve İranlı yazarlarının metinleri karřılařtırıldıđında rnek/rneklem oranının Trk yazarlar iin 0.33 ve İranlı yazarlar iin 0.24 olduđu grlmektedir. Bu oranlara gre Trk yazarların metinlerinde kullanılan kelime eřitliliđi, aynı kelimeleri daha ok sıklıkla tekrar kullanan İranlı yazarlara gre daha yksektir.

alıřmanın en gze arpan bulgularından biri, iki yazar grubu arasında szcksel beklerin kullanım oranındaki fark olarak ne ıkmaktadır. Schmitt'e (2010) gre, ikinci bir dilde yazan yazarların kalıplařmıř dili kullanımları kullanılan yapıların oranı, kullanımlarındaki dođruluk oranı ve anlam altyapılarının dođruluđu aılarından deđerlendirilebilir. Bu alıřmada, karřılařtırılan iki grup arasında szcksel beklerin kullanım sıklıđı aısından farklar gzlenmiřtir. Grupların kullanım dođruluđu aısından farklı olup olmadıklarını tespit etmek iin szcksel bekler yakından incelenmiřtir. Bu incelemenin sonucunda hem İranlı hem de Trk yazarların kullanımlarında yanlıřlıklar tespit edilmiřtir: rneđin İran derleminde rastlanan ve kalıplařmıř bir yapı olarak tekrarlanan *would like to appreciate* yapısı anadili İngilizce olan bir deđerlendirici tarafından yanlıř olarak deđerlendirilmiřtir. Trk derleminde de bazı hatalı kullanımlar bulunmuřtur; rneđin *thank to, owe the debt of* gibi. Fakat, Trk derleminde bu yapılar tekrarlanma

sıklıklarının az oluşu (4 defadan az) sebebiyle sözcüksel öbekler arasında yer bulmamıştır. Sonuç olarak, iki grup arasında sözcüksel öbeklerin kullanımın doğruluğu açısından çok az fark bulunmuştur.

Türk yazarlar tarafından sözcüksel öbeklerin az kullanılması kaçınma olarak değerlendirilebilir. Laufer (2000) kaçınmanın dilbilimsel çapraz etkileşim sonucu ortaya çıkabileceğini ve ikinci dil kullanıcılarının iletişim zorluklarını aşmakta kullandıkları bir taktik olduğunu iddia etmektedir. Kaçınma, aynı zamanda hedef ögenin özündeki zorluktan da kaynaklanabilir. Laufer ve Eliason'un bir çalışması da kaçınmanın hedef ögenin zorluğundan çok, birinci ve ikinci diller arasındaki sistematik uyumsuzluğun bir sonucu olduğunu önermektedir. Türk yazarların sözcüksel öbekleri az kullanmalarının veya kullanmaktan kaçınmalarının sebebi, bu durumda, Türkçe ve İngilizce arasındaki uyumsuzluktan kaynaklanıyor olabilir. Sonuçlara farklı bir açıdan bakıldığında ise gözlenen farkın İranlı yazarların belli sözcüksel öbekleri çok fazla kullanmasıyla ilişkili olduğu da söylenebilir. Birçok çalışma, bazı ikinci dil kullanıcılarının sözcüksel öbekleri emin olmadıkları durumlarda bir 'güvenlik ağı' gibi görerek sıklıkla ve tekrar tekrar kullandıklarını göstermiştir (De Cock, 2000; Foster, 2001; Granger, 1998).

Citation Information:

Tokdemir-Demirel, E. & Ahmadi, H.S. (2013). Lexical bundles in research article acknowledgments: a corpus comparison. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi [Hacettepe University Journal of Education]*, 28(2), 457-468.