



Recurrent phrases in Turkish EFL learners' spoken interlanguage: A corpus-driven structural and functional analysis*

Aysel Şahin Kızıl^{a **}, Abdurrahman Kilimci^b

^a*Firat University, Department of English Language and Literature, Elazığ 23119, Turkey*

^b*Çukurova University, Faculty of Education, English Language Teaching Department, Adana 01330, Turkey*

APA Citation:

Şahin Kızıl, A., & Kilimci, A. (2014). Recurrent phrases in Turkish EFL learners' spoken interlanguage: A corpus-driven structural and functional analysis. *Journal of Language and Linguistic Studies*, 10(1), 195-210.

Abstract

Various studies contrasting learner language with native speaker (NS) performance have shown that recurrent multiword expressions that come so naturally to the NSs pose difficulty for the non-native (NNS) speakers and hinder their language production although usually easy to understand. Therefore, the present study aims to explore the structural and functional properties of the recurrent phrases in the spoken English of the Turkish learners of English and the native speakers to find out whether these word combinations also cause difficulty for the learners under investigation. The study adopts "the corpus driven 'recurrent word combination' method" (De Cock, 2004, p. 227) within the framework of the Contrastive Interlanguage Analysis (CIA) (Granger, 1998). The corpora drawn on in the study are the native speaker corpus, the Louvain Corpus of Native English Conversation (LOCNEC), and the subcorpora of the non-native speaker corpus LINDSEI (the Louvain International Database of Spoken English Interlanguage), which contains speech produced by advanced Turkish learners of English. Two taxonomies were used to analyze the recurrent phrases: the structural taxonomy and the functional taxonomy. The study confirms that the recurrent language characterizes both native and nonnative speech despite marked variations in terms of underuse and overuse phenomena in the learner data. The significance of difference as to the structural and functional variations that particular word combinations display in the nonnative corpus as compared to the native speaker corpus is discussed and pedagogical implications are shared.

© 2014 JLLS and the Authors - Published by JLLS.

Keywords: Recurrent word combinations; spoken corpus; interlanguage; corpus-driven method; contrastive interlanguage analysis

1. Introduction

As stated by Gilquin, Granger and Paquot (2007), corpus linguistics and corpus based research has played "a key role in most language-related fields from lexicography to language teaching through natural language processing and literary criticism" (p. 320). The practical and theoretical potential of computer assisted corpus analysis has recently been recognized in the field of Second Language Acquisition (SLA) as well. With the purpose of researching interlanguage through usage-based descriptive and quantitative as well as qualitative analyses, a number of researchers have begun

* This paper is based on some portions of the unpublished doctoral dissertation by Aysel Şahin Kızıl "Recurrent Phrases in Learner English: A Corpus Driven Approach", presented to the Institution of Social Sciences, Cukurova University, 2013.

** Aysel Şahin Kızıl. Tel.: +90-505-253-6321
E-mail address: askizil@firat.edu.tr

developing what are called learner corpora which refer to electronic collections of speech or writing of foreign or second language learners in a variety of language settings.

Although learner corpus compilation is a relatively recent activity, a number of projects have already been started (even some were completed) since 1990s. Among a number of learner corpora, the International Corpus of Learner English (ICLE) is especially noteworthy since, unlike most of the existing learner corpora focusing on one L1 group, ICLE contains data from learners with different L1s. It currently consists of 4.5 million words of argumentative essays written by university students of English with 16 different L1 backgrounds. One strength of ICLE stems from the fact that it has a reference corpus, the Louvain Corpus of Native English Essays (LOCNESS), compiled from the native speakers (NSs) under the same task conditions, which make it an efficient comparable base for learner English. For spoken learner corpora, currently, the biggest one is the Louvain International Database of Spoken English Interlanguage (LINDSEI), which contains interviews with advanced learners of various L1s. To date, it has covered 11 different mother tongue backgrounds. Like ICLE, LINDSEI has also a comparable corpus of NSs, The Louvain Corpus of Native English Conversation (LOCNEC).

Most of the learner corpus projects have been launched in the last decade and the research drawing from them is relatively limited (Tono, 2000). However, there are notable efforts in terms of exploiting the potentials of learner corpora in both written and spoken medium (Ädel & Römer, 2012; Aijmer, 2002, 2004; Altenberg, 2002; De Cock, Granger, Leech, & McEnery, 1998; De Cock, 2004; Ebeling, 2011; Granger & Tyson, 1996; McEnery & Kifle, 2002; Wei, 2009) – though the spoken interlanguage related studies are fairly new.

2. Literature Review

Altenberg (1998) defines the term ‘recurrent word-combinations’ as “any continuous string of words occurring more than once in identical form” (p. 101). The notion of recurrent phrases has been investigated through corpus based methodology to reveal the patterning of language under various terminologies such as recurrent word combinations (Altenberg, 1998), lexical bundles (Biber, Conrad, & Cortes, 2004; Biber, Johansson, Leech, Conrad, & Finegan, 1999; Hyland, 2008a), recurrent sequences (De Cock, 2004), lexical chunks (Ishikawa, 2009) and multi-word constructions (Liu, 2012). De Cock's study (2004) is among the earliest attempts to investigate the recurrent phrases in learner language. Her primary focus is the spoken English produced by French learners. Waibel (2007) conducted an investigation on phrasal verbs in written interlanguage produced by German and Italian learners, which provides insights for the recurrent phrases in learner language. Ping (2009) investigated lexical bundles in written English of Chinese students and compared the findings with native speaker (NS) writing. Likewise, Ishikawa (2009) compared the high-frequent word combinations in English essays written by Japanese learners of English with those used by NSs. Chen and Baker (2010) report a study of recurrent phrases through a written corpus of learner language. More recently, Ädel and Erman (2012) carried out a study on recurrent word combinations in academic written English of Swedish learners in comparison with NSs' written performance.

These studies contrasting learner language with the NS performance have shown that recurrent multiword expressions that come so naturally to NS pose difficulty for non-native (NNS) users (De Cock, 2004; Nesselhauf, 2005). Recurrent phrases are usually easy to understand but they hinder language production for the learners. Learners construct their spontaneous speech by combining individual words, which results in producing unnaturally sounding language although it is grammatically correct. Kjellmer (1991) summarizes this as "their building material is individual blocks rather than prefabricated sections" (p. 124).

As for the Turkish L1 background, the learner corpus research mainly has centred on written interlanguage (Can, 2009, 2012; Kilimci & Can, 2009; Kilimci, 2001, 2003, 2009; Şanal, 2007). Given that, to the best knowledge of the authors, there exist no previous corpus-based studies on the recurrent word combinations in the spoken English of Turkish learners, the study undertaken is the first attempt to explore the recurrent phrases through the contrastive analysis of relatively large native (LOCNEC) and nonnative (LINDSEI-TR) speaker spoken corpora. In this respect, the exploration of this untouched area of research is believed to make a significant contribution to the existing literature of learner corpora.

2.1. *Research questions*

The present study aims to investigate the recurrent phrases in the spoken corpus of informal interviews by the Turkish EFL learners and check the findings against a comparable native speaker spoken corpus (LOCNEC) from structural and functional perspectives within the domain of contrastive interlanguage analysis (CIA) (Granger, 1998). This study is designed to explore the following research questions:

1. What are the structural and functional features of recurrent sequences of two or more word combinations prevalent in the spoken interlanguage of the Turkish EFL learners?
2. To what extent are these recurrent sequences in the Turkish learners' speech similar to and/or different from those in the native speaker speech?

3. Method

3.1. *Design*

The present study adopts “the corpus driven ‘recurrent word combination’ method” (De Cock, 2004, p. 227) and the Contrastive Interlanguage Analysis (CIA) (Granger, 1998) to investigate data from a learner spoken corpus (LINDSEI-TR) and a native speaker spoken corpus (LOCNEC). Two taxonomies were used to analyze the recurrent phrases from structural and functional perspectives. The first is the structural taxonomy modeled by Biber et al. (1999), which was also used in similar studies on recurring word combinations in the field (Cortes, 2002b, 2004; Hyland, 2008a, 2008b). The second is the taxonomy for the functional categorization of the recurrent expressions initially designed by Cortes (2002a, 2002b) but later revised by Biber and his colleagues (Biber et al., 2004; Biber & Barbieri, 2007).

3.2. *Data*

Two types of comparable spoken corpora were used to investigate the recurrent phrases in the spoken performance of the Turkish learners of English and native speakers: the Turkish component of LINDSEI (LINDSEI-TR[†]) (see Kilimci, 2014) and the Louvain Corpus of Native English Conversation (LOCNEC). The LINDSEI corpus with its all sub-corpora and the comparable reference corpus LOCNEC follow the same design structure with at least 50 interviews made up of pre-identified tasks such as telling a story, answering a question or describing a picture. (Gilquin, 2012).

[†]LINDSEI-TR as a research project (Project no: EF2013BAP22) is funded by the Commission of Scientific Research Projects, CukurovaUniversity. Data evaluation and annotation process is still in progress.

3.2.1. Learner corpus: Turkish component of LINDSEI – LINDSEI-TR

LINDSEI is the spoken counterpart to the written learner corpus ICLE, and it is the first large-scale corpus of spoken learner English. The corpus is compiled from informal interviews with higher intermediate to advanced learners of English, employing three type of tasks: a set topic, free discussion and a picture description. The project which started with the compilation of the first component from the French mother tongue learners of English has expanded with the inclusion of other mother tongue backgrounds with the transcripts of 50 interviews of about 1000 words of learner language each. To date, a total of 20 sub-corpora with different L1 backgrounds are in the project, and 11 of them have been completed and made available to public use, yet the others are in progress (Gilquin, 2012).

The Turkish component of LINDSEI (LINDSEI-TR), contains a total of 58 recorded interviews of about fifteen minutes, each transcribed and marked up according to the LINDSEI guidelines. The corpus contains 80.817 words, of which learner turns comprise 63.924 words. The corpus is still being double checked in detail again to ensure that the interviews are transcribed accurately and consistently in terms of the transcription conventions. The present study is based on 58 interviews, but the final corpus is planned to be of 50 interviews like the other components of LINDSEI corpus.

3.2.2. Reference Corpus: LOCNEC

LOCNEC, the Louvain Corpus of Native English Conversation, was compiled as part of the interlanguage research project. It is the mirror image of LINDSEI as the comparable corpus of native speaker English. The interview sessions were recorded non-surreptitiously, and they were not used for any sort of external assessment of the participants who are all university students majoring in English. As to the content of the conversations, the same procedure as LINDSEI was followed. Namely, the interviewees were first introduced general topics identified beforehand, which is proceeded with the follow-up questions depending on what the interviewees had said. The last part of the questionnaire included making-up a story based on the given pictures. The LOCNEC interviews make up a total number of 161,725 words and learner turns only consist of 118.553 words.

3.3. Data collection procedures

The Turkish component of LINDSEI was compiled from the face to face interviews of about 15 minutes by the third and fourth year students studying at the department of English Language Teaching in Çukurova University. The interviews were based on three types of tasks intended to elicit speech samples from the learners. First, interviewees were told that they would be interviewed on topics of their personal interest for about fifteen minutes and their conversation would be recorded. Then, they were asked to choose one of the three topics presented in written form - e.g. an experience you've had which has taught you an important lesson, a country you have visited which has impressed you, a film/play you've seen which you thought was particularly good/bad - and think about what they were going to say for a few minutes. The second part of the interview was an informal conversation based on either the questions related to what learners said or more general topics to keep the conversation going. The final part of the interview was story-telling task. The interviewees were asked to make up a story around a four picture sequence rather than describe it. After the interviews ended, each interviewee was asked to fill out and sign a learner profile, which contained data about learners' social and educational backgrounds and signified the interviewee's consent for the speech recorded to be used for research purposes. Later, the interviews were transcribed verbatim and marked up, paying particular attention to

such speech features as dysfluencies, filled pauses and backchanneling, empty pauses, truncated words, phonetic and prosodic features etc.

3.4. Data analysis

The study focused on recurrent phrases of two-, three-, four-, five- and six-word sequences that occur at least 12, 6, 4, 3 and 3 times respectively to investigate learner language from a broader perspective. Different frequency thresholds were set for each sequence because the length of recurrent word combinations is inversely related to their frequency (Altenberg, 1998; De Cock, 2004). Recurrent phrases and their frequencies were extracted from the native and nonnative speaker corpora using WordSmith Tools v5.0 (Scott, 2010), exclusively focusing on the interviewee turns which were tagged as and in the transcription of the interviews. Since the LOCNEC interviewee turns consist of 118.553 words, which is almost twice that of the LINDSEI-TR corpus, speech samples of 63.900 words were extracted from the LOCNEC corpus in order to make it comparable to the LINDSEI-TR, which is made up of 63.924 words.

4. Results and Discussion

4.1. Structural analysis of recurrent phrases in LINDSEI-TR and LOCNEC

The structural analysis yields three broad types of word sequences: verb phrase fragments, dependent clause fragments and noun/prepositional phrase fragments, which are further divided into sub-categories. Figure 1 displays the overall distribution of main structural types of word combinations in the Turkish learner and the native speaker spoken data.

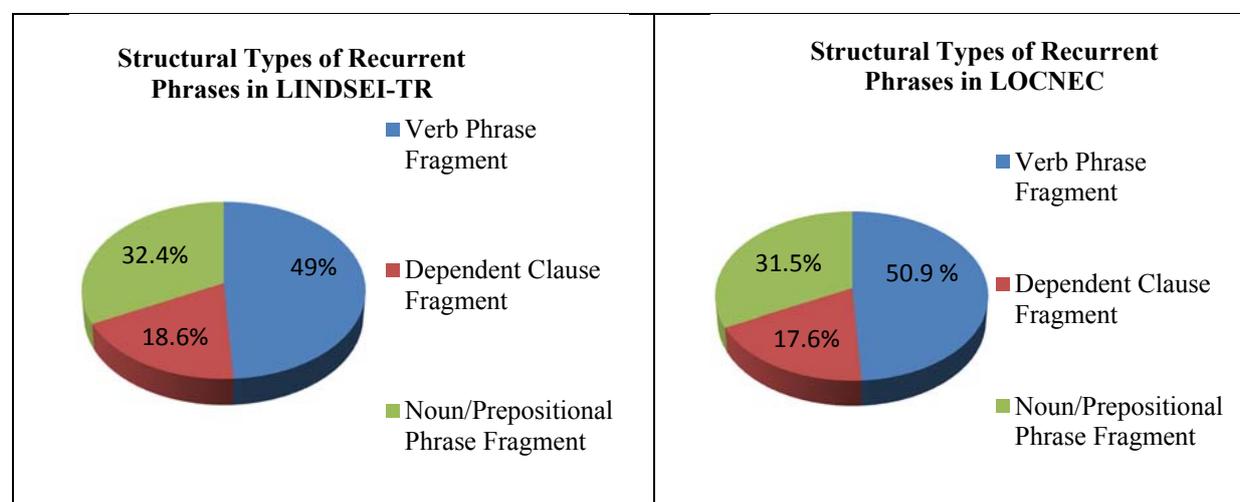


Figure 1. Distribution of major structural types in LINDSEI-TR and LOCNEC

As seen in Figure 1, structural types of word combinations show a very similar distributional pattern in the learner and native language. Verb phrase fragments in both corpora make up the biggest proportion in the overall structural types. This structural profile of data in Turkish learners' and native speakers' speech is, to a great extent, similar to the data distribution across the structural types reported in Hernández (2013). Also, this finding lends support to Biber et al. (2004) who note that in conversation, almost 90% of all the word combinations include verb phrases, and even "50% of them

begin with a personal pronoun + verb phrase” (e.g. I think it is, I don’t know) (p.380). The second largest group of word combinations incorporates noun and prepositional phrases. Dependent clause fragments are the least frequently used structures in the native and nonnative spoken data. These findings suggest that verb, noun and prepositional phrase fragments are the single most important building blocks for on-going discourse, whether in native speaker or in learner spoken communication. Therefore, it is possible to claim that conversation is fundamentally phrasal rather than clausal, which is not surprising when the cognitive load of making full sentences in real-time production is considered. Although Turkish learners show similarities in general distribution of the structural types of word sequences, a detailed analysis dividing these combinations into more specific sub-categories reveals several differences between these two speaker groups. Table 1 displays the major recurrent phrases further categorized according to their structures with examples from both LINDSEI-TR and LOCNEC corpus.

Table 1.

Structural Categories of Recurrent Phrases in LINDSEI-TR and LOCNEC

STRUCTURE	LINDSEI-TR	%	LOCNEC	%
1. Verb Phrase Fragments				
(connector +)1st/2nd person pronoun+VP fragment	<i>I think (and I think) I want (so I want, I want to) you know I don't (I don't know)</i>	61.2	<i>I think (yeah I think) you know (you know you) I mean (yeah I mean)</i>	60.3
(connector +) 3rd person pronoun+ VP fragment	<i>it was (it was very) she doesn't like,</i>	22.5	<i>it was (and it was, it was just) so that was,</i>	28.3
Discourse markers + VP fragment	<i>I think it is, of course I want</i>	2.1	<i>sort of you know, you know it was,</i>	3.7
Verb Phrase (with non-passive verb)	<i>like it,</i>	11.8	<i>paint a picture</i>	3.7
verb phrase with passive verb	-----	-----	<i>was very impressed</i>	1.8
yes/no question fragments	<i>can I say</i>	1.0	-----	
Wh-Question fragments	<i>how can I say</i>	1.4	<i>what else did I</i>	2.2
2. Dependent Clause Fragments				
1 st /2 nd person pronoun+Dependent Clause	<i>she wants him to, I don't know why</i>	16.6	<i>you know it was</i>	15
Wh-clause fragments	<i>when I was a child</i>	11.1	<i>I don't know what</i>	70
if-clause fragments	-----		-----	
verb/adjective+to-clause	<i>I want to do, want to have,</i>	55.5	<i>like to go, like to see, to go to,</i>	15
that-clause fragment	<i>I can say that,</i>	16.6	----	
3. Noun Phrase And Prepositional Phrase Fragments				
Noun phrase with –of phrase fragment	<i>of them, the end of the, a lot of,</i>	22.2	<i>a bit of, a awful lot of</i>	67.8
Noun phrase with other post-modifier	<i>plans for the future</i>	5.6	----	
Other noun phrase expressions	<i>the picture</i>	36.1	<i>a look at</i>	17.9

Prepositional phrase expressions	<i>in high school</i>	36.1	<i>at the moment</i>	14.3
Comparative expressions	-----			
4. Other Expressions	<i>once upon a time</i>		<i>yeah yeah yeah</i>	

The verb phrase fragments category displays a similar frequent use of first person pronoun + verb combination, which could be attributed to text- and task types, where the speaker is urged to talk about himself/herself. In both corpora, this pattern displays a general tendency for the use of know/think/want after the first person pronoun. This corresponds to the observation by Biber et al. (1999), based on their conversational data, that “most of the sequences made up of following elements occur as recurrent [phrases] in conversation: I/you + know/think/want” (p.1001). These phrases are often followed by a complement clause, as in the examples of *I don't know why* from LINDSEI-TR and *you know what I* from LOCNEC.

A closer look at the sub-categories of structural types ascertains several differences between the native speakers' and the Turkish learners' speech as well. The different rates in terms of the use of third person pronoun+VP fragment, especially in the use of pronoun *it* in the NS (28.3%) imply that unlike non-native students, native students do not rely as much on their personal experiences. A similar findings is also reported by Hernández (2013) comparing spoken language of the NS and the NNS students. For the sub-categories of verb phrases with non-passive and passive verbs, it has been observed that there is a tendency towards using active voice verb phrases in both corpora. While passive verbs are not used at all by the Turkish learners, the native speakers use passive voice only at the rate of 1.8%. This finding is in line with the literature since “active voice is the unmarked choice ... all spoken registers use active voice verb phrases over 95% of the time” (Biber, 2006 p. 64). For the comparison of the NS and the NNS in terms of non-passive verbs, Table 1 shows that the learners use non-passive verbs more than the native speakers. That learners generally prefer active discourse frames has been shown in previous studies as well (Granger, 1998; Ishikawa, 2009).

Regarding dependent clause fragments category, while the non-native speakers' speech is dominated by verb + to-clause fragments (55.5%), the native speakers employ wh-clause fragments intensively (70%) in their clausal constructions, which is in line with (Biber et al., 1999) who list wh-clause fragments in the frequent sequences of conversational English and indicate that native speakers often use such fragments as “utterance launchers, presenting a personal stance relative to the information in the following complement clause” (p.1003) as illustrated in the following example (1) taken from LOCNEC.

(1) B> modern Eng= yeah like my major **what I want to do** is .modern English language

On the other hand, the Turkish learners use wh-clause fragments only at the rate of 11.1% in their speech, which displays a striking deviation in the learner speech from the native speaker norm. Comparing the use of clausal fragments in spoken and written texts, Kaltenbock (2004) points to “extra processing effort required by clausal constructions”(p.223) if they are not stored as automatically retrievable sequences. Accordingly, it is possible to claim that the Turkish learners have small repertoire of automatically retrievable wh-clause fragments. They are likely to process such fragments on the basis of grammatical rules, which is relatively difficult in real time production. Another striking difference lies in the use of that-clause fragments between both corpora. Such clauses occur quite frequently in the nonnative speaker data. Biber et al (1999) note that the “retention of *that* is the norm in academic prose, while exceptional in conversation as conversations favor the reduction or omission of the constituents that can easily be reconstructed” (p.680). The tendency of the Turkish learners to retain *that* could be

that the Turkish learners are unaware of register properties. If-clause fragments that do not appear in both corpora might be attributed to the contextual properties of data gathering tasks which don't require the use of such clausal fragments.

Finally, the analysis of the noun phrase and prepositional phrase fragments category shows that the native speaker speech features a higher number of complex noun phrases -noun phrase with -of phrase fragment- (67.8%) and lower number of simple noun phrase -other noun phrase expressions- (17.9%), while the learner speech makes extensive use of simple noun phrases (36.1%) to the neglect of more complex noun phrases (22.2%). This difference could be explained with the fact that "of-phrases is associated with the overall frequency of nouns" (Biber et al., 1999: 302). Although it is typical of written English to make an extensive use of phrasal bundles incorporating noun/prepositional phrases (Biber et al., 2004; Hyland, 2008), the high frequency of noun and prepositional phrases in both corpora seems normal as the interviews that make up the corpus are structured around such communicative tasks as telling a story, answering a question or describing a picture which naturally requires the use of high number of noun and prepositional phrases. In this respect, however, the comparisons reveal that the elicited oral data from learners is obviously less loaded with information and more relies on prepositional phrases and simple noun phrases rather than complex of phrase fragments.

4.2. Functions of recurrent phrases in LINDSEI-TR in comparison with LOCNEC

The functional analysis of the recurrent phrases in the NNS and NS spoken language has pointed to four broad categories: stance expressions, discourse organizers, referential expressions and special conversational expressions. Figure 2 illustrates the distributional pattern of the recurrent phrases in terms of their functions in both NNS and NS corpora.

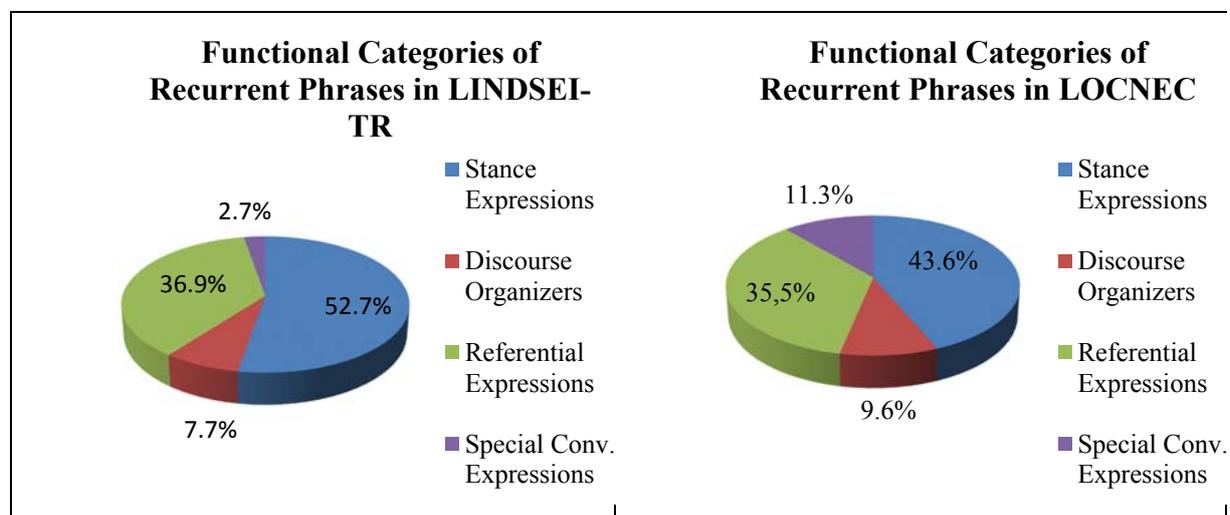


Figure 2. Distribution of major functional categories in LINDSEI-TR and LOCNEC

As seen in Figure 2, stance expressions that cover the word combinations expressing the user's attitudes, judgments and perspective which frame some other propositions have the largest proportion in both corpora. Stance expressions are evaluated in two groups: Epistemic stance and attitudinal stance which have personal and impersonal variations. The data analysis shows that both native and nonnative speakers prefer to use personal stance expressions in conveying their messages. The high proportion of personal stance expressions in spoken language is also observed by Biber and Conrad (2004) who state

that “the most striking aspect of conversation’s use of word combinations is the high proportion of personal stance expressions” (p.67). Similar findings were also reported by Biber and Barbieri (2007), Biber et al. (1999) and Hernández (2013) who note that personal stance expressions make up more than 60% of the typical conversation in English. Referential expressions have the second largest proportion, which is followed by the discourse organizers and special conversational expressions respectively. Special conversational expressions are more widely used by native speakers than the Turkish learners, which points to the Turkish learner’s unfamiliarity with the conversational English.

Biber et al. (1999) further divide these broad categories into sub-classes in accordance with the precise functions the word sequences perform. However, the analysis of LINDSEI-TR and LOCNEC in terms of the functions of recurrent phrases demonstrated that not all sub-categories identified in the original taxonomy are found in the recurrent phrases. For example, phrases expressing impersonal stance or personal/impersonal ability do not appear in both corpora. Therefore, the original taxonomy offered by Biber et al. (1999) has been modified by deleting some categories and by adding a new category of function for the recurrent phrases. It should also be noted that some combinations appear in more than one category as they may perform multiple functions in different context, such as *I don’t know*, which “does not have a single function but is characterized by its broad spectrum of uses”(Aijmer, 2009 p.156). Table 2 further displays the functions of major recurrent phrases in LINDSEI-TR and LOCNEC with examples extracted from transcribed texts.

Table 2.

Functional Categories of Recurrent Phrases in LINDSEI-TR and LOCNEC

Categories	LINDSEI	%	LOCNEC	%	
Sub-Categories					
Stance Expressions	A) Epistemic Stance				
	Personal	<i>I think</i>	16.6	<i>I think</i>	
		<i>I don’t know</i>		<i>you know</i>	29.3
		<i>but</i>			
	B)				
	Attitudinal/Modality Stance				
	B1: Desire	<i>I want</i>	33.3	<i>I would like to</i>	
	Personal	<i>I don’t like,</i>		<i>I wanted to</i>	8.6
		<i>I would like</i>			
	B2: Obligation/Directive				
Personal	<i>I have to</i>	1	<i>you have to</i>	3.1	
B3: Intention/Prediction					
Personal	<i>I will</i>	1.8	<i>I am going to</i>	2.6	
Discourse/Text Organizers	A) Topic Introduction/Focus				
		<i>I want to talk firstly I want to</i>	4.1	<i>I want to talk about</i>	1.7

Referential Expressions	B) Topic Elaboration /Clarification	<i>for example</i>	1.7	<i>I mean I was what I mean</i>	4.8	
	C) Topic Closing/Turn Yielding	<i>I don't know</i>	1.9	<i>I don't know</i>	1.2	
	A) Identification/Focus	<i>in this picture of the film</i>	6.2	<i>that kind of thing that's the only</i>	5.1	
	B) Imprecision / Markers of Vagueness	<i>I don't know</i>	2	<i>and things like that sort of you know</i>	13.7	
	C) Specification of Attributes					
	C1: Quantity specification / Quantifying Sequences	<i>a lot of things there are lots of</i>	8.3	<i>there's a lot of a couple of</i>	12	
	D) Time/Place/Text Reference					
	D1: Markers of Time	<i>at the end of</i>	8.3	<i>all the time yeah</i>	6.1	
	D2: Markers of Place	<i>in high school</i>	4.1			
	D3: Text deixis	<i>as I said before</i>	4.1			
	D4: Multifunctional References	<i>at the end, first of all</i>	3.9			
	Special Conversational Expressions	A) Speech/Thought Reporting		--	<i>I thought it was</i>	2.1
		B) Responses	<i>okay okay</i>	2.7	<i>yeah that's right yeah</i>	9.2

A closer look at Table 2 shows that while epistemic stance occupies a large place in the NSs communication (29.3%), the Turkish learners employ stance expressions regarding personal desire to a great extent (33.3%). As for the intention/prediction expressions in attitudinal stance, both speaker variety choose personal expressions rather than impersonal, which sounds meaningful when the fact that one of the tasks in data gathering was directly about the future plans of the interviewees is considered.

Another difference is observed in the choice of discourse organizers in topic elaboration/clarification. While the NS prefer to use *I mean* and its combinations to clarify the previously stated idea, this discourse item is not so frequent in Turkish learners' data. This confirms the earlier findings by De Cock (2004) and Huang (2011) who also found that *I mean* is underused by non-native speakers of English. What is interesting is that in Turkish learners' speech, *for example* seems to serve the same function as illustrated in the extract (2):

- (2) (eh) so he he argues but (eh) when his father (eh) learned that (eh) he he is ill (er) ..he accepts . his son (eh) . I I am affected (eh) from this film because (eh) it is very similar to my family . (eh) so (er) this was very sad film (eh) ..

<A> (uhu)

for example (eh) my father (eh) have argued (eh) with his father so (eh) he tells all the time (eh) .what he feels about it

As is seen, the learner in his second turn explains why the film he is talking about is very similar to his family by using *for example* just before the clarification. Additionally, the analysis of the NS and the NNS speech has revealed that a new sub-category of discourse organizers which is not proposed in the original taxonomy should be added to the classification of discourse items. When the use of *I don't know* is analyzed in detail, it has been observed that it has an additional function apart from personal epistemic stance. Consider the following utterances (3) from LINDSEI-TR and (4) from LOCNEC.

(3) (eh) well (eh) . I can say that (eh) Turkish people are (eh) more (eh) friendly than (eh) Polish people because (erm) in fact .. for example I stayed there (eh) and no friends (eh) came and (mm) .. we didn't go: (eh) for example to parties so much with friends with classmates they were (em) they weren't so: . smiling

<A> (uhu)

 (eh) maybe because of the climate **I don't know**

<A> (uhu)

(4) yeah I think that might have something to do with it (erm) **I don't know** I've just always felt more comfortable in Ireland and that's maybe where I . I fit in and

As seen in example 3, *I don't know* functions as a topic-closing sequence or as Aijmer (2009) demonstrates it has a floor-yielding function in the conversation. Aijmer (2009) further states that *I don't know* in the potential topic closing function may not always be followed by a new turn since the current speaker may choose to continue as shown in example 4. What is more, Aijmer (2009) notes that this function of *I don't know* is especially common in interviews, which explains the occurrence of this function of *I don't know* in LINDSEI-TR and LOCNEC.

Regarding the referential expressions, what is conspicuous at first glance in Table 2 is the sub-category of imprecision/markers of vagueness. Vagueness tags are the indicator of intersubjectivity and they have a crucial role in informal spoken communications, signaling an assumption of shared experience and social closeness (Aijmer, 2002; De Cock, 2004). It has been observed that the NSs employ vagueness tags (13.7%) in their speech more than the Turkish learners (2%). For instance, one of the commonest vagueness tags –*sort of*– is very frequent in the NS speech while it hasn't occurred in the NNS spoken language at all. This is the case with the learner groups in Aas's (2011) and De Cock's (2004) studies. Therefore, it could be argued that the underuse of vagueness tags or even their not being used in the Turkish learners' speech is an idiosyncratic feature of their spoken interlanguage. Lack of imprecision in an informal conversation is a reason explaining foreign-soundings of the speakers. Thus, it is likely that the Turkish learners' speech sounds non-native as they do not organize their discourse using the characteristics of the informal talk. The last category of the functions of recurrent phrases in the NS and the NNS lend further support to the findings above.

5. Conclusions

In conclusion, this study has set out to explore the use of the recurrent word combinations in terms of structural and functional variation by comparing the native and nonnative spoken corpus. From a pedagogical perspective, the findings of the study could be useful in several ways. To begin with, regarding learners' unfamiliarity with the spoken English, this study suggests course designs exposing learners to the features of spoken English through authentic materials to increase their awareness of linguistic properties of spoken English. Secondly, as underlined by Shirato and Stapleton (2007) word combinations specific to oral communication need to be incorporated into the syllabus of speaking course. Creating a pedagogically useful list of recurrent phrases of spoken English is recommended as a starting point, and the language teachers could arrange the instructional activities based on such a list. Thirdly, explicit instruction on the recurrent phrases used by native speakers in various communicative events (e.g. topic introduction, clarification, turn yielding etc...) would be helpful for learners in terms of foreign-soundingness and in speaking more fluently.

Also, this study has found that the use of vagueness markers and hedging devices (e.g. things like that, sort of, you know, sort of like etc...) is particularly significant considering the striking differences between the learners and the native speakers. Vagueness is central to informal communication and of great use when interlocutors cannot find the right words, and hedging is a characteristic of casual speech softening the tone of conversation (McCarten, 2010). Therefore, in designing speaking courses, these devices should also be incorporated into the syllabus to enhance strategic competence "even before the acquisition of any grammatical competences" (Shirato & Stapleton, 2007, p.408). Last but not the least, this study has important implications about the functions of recurrent phrases regarding the inappropriate use of some word combinations (e.g. for example). Thus, it is necessary that while teaching specific combinations, not only their meaning but also their function in context should be underlined.

There are two noteworthy limitations of this study. First, this study does not provide a complete picture of the spoken interlanguage characteristics of the Turkish learners regarding recurrent word combinations as certain categories such as epistemic tags or markers of vagueness are not considered. Second, since the structural and functional analyses of recurrent phrases were qualitatively conducted manually, it is likely that there might be some possible inconsistencies. Despite the limitations, this study is considered to contribute to the existing knowledge of word combinations in learner English. There is, however, still room for further research to provide better understanding of the use of word sequences in learner language. First, a research project focusing on recurrent phrases across different L1 groups would result in a richer understanding of learner language. Also, comparison with the other sub-corpora of LINDSEI would shed further light on general tendencies across learner populations. Finally, functional patterns of recurrent phrases that emerged from the present study may provide useful onset for further research, leading to practical applications of the recurrent phrases in educational setting.

References

- Aas, H. L. (2011). *Recurrent word-combinations in spoken learner English: A study of corpus data from Swedish and Norwegian advanced learners*. (Unpublished master's thesis). University of Oslo, Norway.
- Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes*, 31(2), 81–92.

- Ädel, A., & Römer, U. (2012). Research on advanced student writing across disciplines and levels: Introducing the Michigan corpus of upper-level student papers. *International Journal of Corpus Linguistics*, 17(1), 3–34.
- Aijmer, K. (2002). *English discourse particles: Evidence from a corpus*. Amsterdam: John Benjamins Publishing Company.
- Aijmer, K. (2004). Pragmatic markers in spoken interlanguage. *Nordic Journal of English Studies*, 3(1), 173-190.
- Aijmer, K. (2009). “So er I just sort I dunno I think it’s just because...”: A corpus study of I don’t know and dunno in learners’ spoken English. In A. H. Jucker, D. Schreier & M. Hundt (Eds.), *Corpora: Pragmatics and discourse* (pp. 151–168). The Netherlands: Rodopi.
- Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word-combinations. In A.P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 101–122). Oxford: Oxford University Press.
- Altenberg, B. (2002). Using bilingual corpus evidence in learner corpus research. In S. Granger, J. Hung & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 37–54). Amsterdam: John Benjamins Publishing Company.
- Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263–286.
- Biber, D., & Conrad, S. (2004). The frequency and use of lexical bundles in conversation and academic prose. *Lexicographica*, 20, 56–71.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at . . . : Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371–405.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow, UK: Pearson.
- Can, C. (2009). İkinci dil edinimi çalışmalarında bilgisayar destekli bir Türk öğrenci İngilizcesi derlemi: ICLE’nin bir altderlemi olarak TICLE. *Dil Dergisi*, 144 (Nisan-Mayıs-Haziran), 16–34.
- Can, C. (2012). Uluslararası Türk öğrenci İngilizcesi derleminde tutum belirteçleri. *Dilbilim Araştırmaları*, 1, 39–53.
- Chen, Y.-H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, 14(2), 30–49.
- Cortes, V. (2002a). Lexical bundles in Freshman composition. In R. Reppen, S. M. Fitzmaurice, & D. Biber (Eds.), *Using corpora to explore linguistic variation* (pp. 131–145). Amsterdam: John Benjamins Publishing Company.
- Cortes, V. (2002b). *Lexical bundles in academic writing in history and biology*. (Unpublished doctoral dissertation). Northern Arizona University, USA.
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23(4), 397–423.

- De Cock, S., Granger, S., Leech, G., & McEnery, T. (1998). An automated approach to the phrasicon of EFL learners. In S. Granger (Ed.), *Learner English on computer* (pp. 68–80). London & New York: Addison Wesley Longman.
- De Cock, S. (2004). Preferred sequences of words in NS and NNS speech. *Belgian Journal of English Language and Literatures (BELL), New Series* (2), 225–246.
- Ebeling, S. O. (2011). Recurrent word-combinations in English student essays. *Nordic Journal of English Studies*, 10(1), 49–76.
- Gilquin, G. (2012, Jan 3). *LINDSEI*. Retrieved from <http://www.uclouvain.be/en-cecl-lindsei.html>
- Gilquin, G., Granger, S., & Paquot, M. (2007). Learner corpora: The missing link in EAP pedagogy. *Journal of English for Academic Purposes*, 6 (4), 319–335.
- Granger, Sylviane, & Tyson, S. (1996). Connector usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes*, 15, 19–29.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A.P.Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 145–160). Oxford: Clarendon Press.
- Hernández, P. S. (2013). Lexical bundles in three oral corpora of university students. *Nordic Journal of English Studies*, 12(1), 187–209.
- Huang, L. (2011). *Discourse markers in spoken English : A corpus study of native speakers and Chinese non-native speakers*. (Unpublished doctoral dissertation). The University of Brimingham, England.
- Hyland, K. (2008a). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4–21.
- Hyland, K. (2008b). Academic clusters : Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18(1), 41–62.
- Ishikawa, S. (2009). Phraseology overused and underused by Japanese learners of English : A contrastive interlanguage analysis. In K. Yagi & T. Kanzaki (Eds.), *Phraseology, corpus linguistics and lexicography: Papers from phraseology* (pp. 87–100). Nishinomiya, Japan: Kwansai Gakuin University Press.
- Kaltenbock, G. (2004). Using non-extraposition in spoken and written texts. In K. Aijmer & A.-B. Stenström (Eds.), *Discourse patterns in spoken and written corpora* (pp. 219–242). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Kilimci, A. (2001). Automatic extraction of the lexical profile of EFL learners through corpus query techniques. *Çukurova Üniversitesi Eğitim Fakültesi Dergisi*, 21(2), 37–47.
- Kilimci, A. (2006). Stance and attitude in advanced Turkish learners' written discourse. In P. Karnowski & I. Szigeti (Eds.), *Proceedings of Language and Language Processing: 38th Linguistics Colloquium* (pp. 347-357), Frankfurt: Peter Lang.
- Kilimci, A. (2009). Negotiation of meaning in L2 academic writing. In G. Socarras (Ed.), *Proceedings of 1st international conference on literature, languages and linguistics*. Athens: ATINER

- Kilimci, A., & Can, C. (2009). Uluslararası Türk öğrenici İngilizcesi derlemi/TICLE: Turkish international corpus of learner English. In M. Sarıca, N. Sarıca, & A. Karaca (Eds.), *XXII. Ulusal Dilbilim Kurultayı Bildirileri* (pp. 1–11). Ankara: Yüzyüncü Yıl Üniversitesi Yayınları.
- Kilimci, A., (2014). LINDSEI-TR: A new spoken corpus of advanced learners of English. *International Journal of Social Sciences and Education*, 4(2), 401-410.
- Kjellmer, G. (1991). A mint of phrases. In K. Aijmer & B. Altenberg (Eds.), *English corpus linguistics* (pp. 111–127). London: Longman.
- Liu, D. (2012). The most frequently-used multi-word constructions in academic written English: A multi-corpus study. *English for Specific Purposes*, 31(1), 25–35.
- McCarten, J. (2010). Corpus-informed course book design. In A. O’Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 413–427). Abingdon, UK: Routledge.
- McEnery, T., & Kifle, N. A. (2002). Epistemic modality in argumentative essays of second-language writers. In J. Flowerdew (Ed.), *Academic discourse* (pp. 182–215). London: Longman.
- Nesselhauf, N. (2005). *Collocations in a learner corpus*. Philadelphia, PA: John Benjamins Publishing Company.
- Ping, P. (2009). A Study on the use of four-word lexical bundles in argumentative essays by Chinese English: A comparative study based on WECCL and LOCNESS. *CELEA Journal*, 32(3), 25–45.
- Scott, M. (2010). WordSmith Tools (Version 5.0), [Lexical Analysis Software]. Available from <http://www.lexically.net/wordsmith/version5/index.html>
- Shirato, J., & Stapleton, P. (2007). Comparing English vocabulary in a spoken learner corpus with a native speaker corpus: Pedagogical implications arising from an empirical study in Japan. *Language Teaching Research*, 11(4), 393–412.
- Şanal, F. (2007). *A learner corpus based study on second language lexicology of Turkish students of English*. (Unpublished doctoral dissertation). Çukurova University, Adana, Turkey.
- Tono, Y. (2000). A Computer Learner Corpus Based Analysis of the Acquisition Order of English Grammatical Morphemes. In L. Burnard & T. McEnery (Eds.), *Rethinking language pedagogy from a corpus perspective: Papers from the third international conference teaching and language corpora* (pp. 123–132). Frankfurt: Peter lang.
- Waibel, B. (2007). *Phrasal verbs in learner English: A corpus-based study of German and Italian students*. Freiburg: Albert-Ludwigs-Universität Freiburg.
- Wei, N. (2009). On the phraseology of Chinese learners spoken English: Evidence of lexical chunks from COLSEC. In A. Jucker, D. Schreier, & M. Hundt (Eds.), *Corpora: Pragmatics and discourse. Papers from the 29th international conference on English language research on computerized corpora (ICAME 29)* (pp. 271–296). Ascona, Switzerland: Rodopi.

İngilizceyi yabancı dil olarak öğrenen Türk öğrencilerinin sözlü aradilinde tekrarlanabilir öbekler: Derleme dayalı yapısal ve işlevsel bir çözümleme

Öz

Anadili konuşuru edimi ile öğrenci dilini kıyaslayan çeşitli çalışmalar anadili konuşurlarına doğal gelen tekrarlanabilir çok sözcüklü öbeklerin anadili konuşuru olmayanlar için zorluk oluşturduğunu ve bu öbeklerin anlaşılmasını kolay olmasına rağmen öğrencilerin dil üretimini engellediğini göstermiştir (De Cock, 2004; Nesselhauf, 2005). Bu nedenle, bu çalışma, tekrarlanabilir sözcük öbeklerinin Türk öğrencileri için zorluk oluşturup oluşturmadığını belirlemek amacıyla Türk öğrencilerinin ve anadili konuşurlarının sözlü İngilizce’inde söz konusu öbeklerin yapısal ve işlevsel özelliklerini araştırmayı hedeflemektedir. Çalışma, tekrarlanabilir öbeklerin incelenmesinde Karşılaştırmalı Aradil Çözümlemesi (Granger, 1998) çerçevesinde, “derleme dayalı ‘tekrarlanabilir kelime öbekleri’ yöntemini” (De Cock, 2004, p.227) kullanmaktadır. Çalışmanın veri kaynağını Louvain Uluslararası Aradil Konuşma İngilizce’si Veritabanı (LINDSEI) oluşturmaktadır. Türk öğrencilerin aradil özelliklerini araştırmak için, bu derlemin, Türk öğrencilerden toplanan verilerle oluşturulan alt derlemi LINDSEI-TR kullanılmıştır. Tekrarlanabilir öbeklerin yapısal ve işlevsel açıdan çözümlemesi iki sınıflama kullanılarak yapılmıştır: yapısal sınıflama ve işlevsel sınıflama. Öğrenci derlemi tekrarlanabilir öbeklerin aşırı kullanımı ve az kullanımı açısından belirgin farklılıklar göstermesine rağmen, tekrarlanabilir öbeklerin hem anadili konuşurlarının hem de anadili konuşuru olmayanların sözlü dillerinin belirgin bir özelliği olduğunu belirlenmiştir. Çalışmanın sonuç bölümünde, anadili konuşuru ve anadili konuşuru olmayanların derlemlerinde bulunan tekrarlanabilir öbeklerin yapısal ve işlevsel farklılıklarının önemi tartışılmış ve öğretimsel sezdirimleri paylaşılmıştır.

Anahtar sözcükler: Tekrarlanabilir sözcükler; sözlü derlem; aradil; derleme dayalı yöntem; karşılaştırmalı aradil çözümlemesi

AUTHORS’ BIODATA

Aysel Şahin Kızıl is an English instructor in the School of Foreign Languages at Firat University, Turkey. She holds an M.A degree on Applied Linguistics and a PhD degree on English Language Teaching. Her current research interests include corpus linguistics, learner language and the use of ICT tools in language education.

Dr. Abdurrahman Kilimci has been working as an assistant professor in the Department of English Language Teaching, Faculty of Education, Çukurova University, where he teaches linguistics, literature and translation courses at the undergraduate level and corpus linguistics and educational technology at the graduate level. He has been involved in corpus compilation projects such as The International Corpus of Learner English (ICLE) and The Louvain International Database of Spoken English Interlanguage (LINDSEI). His main research interests include corpus linguistics, contrastive learner corpus analysis, applied linguistics, discourse analysis and interlanguage pragmatics.