LINDSEI-TR: A New Spoken Corpus of Advanced Learners of English

By

Abdurrahman Kilimci
Cukurova University, Faculty of Education, English Language Teaching Department,
Balcalı, Adana, Turkey

Abstract

The aim of the present study is to describe the LINDSEI-TR, the Turkish component of the LINDSEI (the Louvain International Database of Spoken English), which was initiated to compile a corpus of spoken data produced by learners from varied mother tongues (Gilquin et al., 2010). In this respect, the main objective of the study is to present the aim, development, and the design criteria of the corpus along with its quantitative and qualitative characteristics. The corpus is considered to be of value to researchers in terms of delineating the features of learners' spoken interlanguage and designing teaching materials to improve second language teaching and learning.

Keywords: *Corpus linguistics, spoken corpus, interlanguage, second language teaching and learning*

1. Introduction

Computer learner corpora, which were begun to be compiled in 1990s have since become not only more developed but also varied. Granger (2002:7) defines computer learner corpora as "electronic collections of authentic FL/SL textual data assembled according to explicit design criteria for a particular SLA/FLT purpose. She adds that "they are encoded in a standardized and homogeneous way and documented as to their origin and provenance" (p. 7). Leech (1998) considers learner corpora "a useful resource for anyone wanting to find out how people learn languages and how they can be helped to learn them better" (p xvi). Granger, (2002) points out that the emergence of learner corpus research has brought together the two formerly distinct fields of corpus linguistics and foreign/second language research. She also adds that it has become a theoretical and practical value by availing itself of the main principles, tools and methods from corpus linguistics to provide better descriptions of learner language for a wide range of purposes in foreign/second language acquisition research and the improvement of foreign language teaching.

The ICLE (International Corpus of Learner English) project, launched in 1990, led to the compilation of many sub-corpora of different language backgrounds. In 1995, five years later, a new project, the Louvain International Database of Spoken English Interlanguage (LINDSEI) was started. The project aimed at providing a spoken counterpart to the ICLE, containing oral data produced by advanced learners of English from several mother tongue backgrounds. The compilation of the first component at the CECL with transcripts of 50 interviews of about 100.000 words with French mother tongue learners of English prompted the compilation of other components by the international of the project for different mother tongue backgrounds.

The aim of this article is to present the design criteria and compilation stages of the LINDSEI-TR, the Turkish sub-component of the LINDSEI corpus, providing information on its quantitative and qualitative characteristics. The rest of the work is structured as follows: the next section presents the purpose of the corpus, and then the following respectively report on the design criteria and compilation stages of the LINDSEI-TR from qualitative and quantitative perspectives. A final section highlights the potential of the LINDSEI-TR, and concludes with implications which might help stimulate further avenues for research.

2. Purpose

The LINDSEI-TR is a new spoken English interlanguage corpus compiled as a component of the Louvain International Database of Spoken English Interlanguage (LINDSEI), an international project launched and coordinated at the Centre for English Corpus Linguistics (CECL), the Université Catholique de Louvain in 1995. The corpus is built according to the same principles and conventions specified in LINDSEI guidelines, which are followed by all the international partners compiling other components of the LINDSEI. Therefore, the design of the LINDSEI-TR makes it possible for researchers not only to identify linguistic features of oral interlanguage of Turkish learners but also to compare them with those of other learner varieties. Conducting a cross-linguistic comparison as such can be useful in determining universal and L1-related features of spoken interlanguage of both Turkish learners and other learner varieties. The LINDSEI-TR can also be compared with the LOCNEC (The Louvain Corpus of Native English Conversation), which is a comparable corpus to the LINDSEI in that the same design criteria as in LINDSEI were observed for its compilation. Such a comparison of interlanguage and native language can be helpful to identify nonnative features of the oral interlanguage of Turkish learners. The LINDSEI-TR is the oral counterpart to the TICLE (Turkish International Corpus of Learner English) (Kilimci and Can, 2009; Can, 2009), the Turkish component of the ICLE (International Corpus of Learner English) which contains argumentative essays written by advanced learners of English from a wide array of mother tongue backgrounds such as Bulgarian, Chinese, Czech, Dutch, Finnish, French, German, Italian, Japanese, Norwegian, Polish, Russian, Spanish, Swedish, Tswana, Turkish. Thus, the use of LINDSEI-TR in conjunction with TICLE might prove useful to not only uncover but also compare and contrast the characteristics of spoken and written interlanguage.

3. Design criteria

The LINDSEI-TR component consists of 58 interviews of similar length. The corpus which is orthographically transcribed, but not POS tagged yet totals 80.813 words. The interviewees, native speakers of Turkish, are considered to be advanced learners of English based on the external criteria that they are third and fourth-year students at Faculty of Education, the Department of English Language Teaching, Çukurova University. The LINDSE-TR was collected in line with the LINDSEI format. The consent of the students were sought and documented for the recording of the informal interviews, which last 12 minutes on average. In order to supplement the corpus, such data as the students' mother-tongue, mastery of other foreign languages in decreasing order of proficiency, years of English at school, years of English at university, stays in English-speaking countries etc. were collected so that the effect of learner variables on language proficiency can be studied. The corpus contains information about the interviewer and the interview: for example, the interviewer's gender, mother tongue, mastery of other foreign languages, the introduction topic, and the length of A and B turn both or the learners' turn (B) only etc. The interviews follow the same pattern in that each is made up of three tasks: set topic, free discussion and picture description. Each interview starts with an informal open discussion regarding university life, hobbies, travel or plans for the future. In the second part, the learners are asked to choose one of the three topics and speak on it: (1) an experience that taught them an important lesson, (2) a country which impressed them and (3) a film or a play which they liked or disliked. Each interview ends with a short picture-based story telling.

After the interview, each interviewee was asked to fill in and sign a learner profile, indicating the consent of the learners for their recorded speech to be used for research purposes. The learner profiles provide information about learners' social and language learning variables, such as name, age, sex, nationality, native language, mother tongue, language(s) spoken at home, years of English at school and university, stay in an English-speaking country (Where?, When?, How long?), and other foreign languages in decreasing order of proficiency.

The following Table 1 presents information on the structure of the LINDSEI-TR corpus. As is seen from Table 1, the interviews were conducted by three male instructors, who are native speakers of Turkish. Of the 58 learners who participated in the interview, 19 were males and 39 females. The male and female learners' interviews, on average, last 4.1 and 7.9 hours respectively, which total about 12.12 hours. The shortest interview is 10.41 minutes, while the longest interview is 19.53 minutes. The interviews average 12.50 minutes. The interviewers contribute 20,8%, the male learners 32,8% and the female learners 67,2% of the total 80.813 words in the transcribed corpus. This amounts to 16.845 words for interviewers, 21.826 words for male learners and 42.142 words for female learners. So, the total number of words for <B> turns is 63.968. The mean word count per interviewee is 1148.7 for males and 1080.5 for females.

Table 1 Structure of the LINDSEI-TR corpus

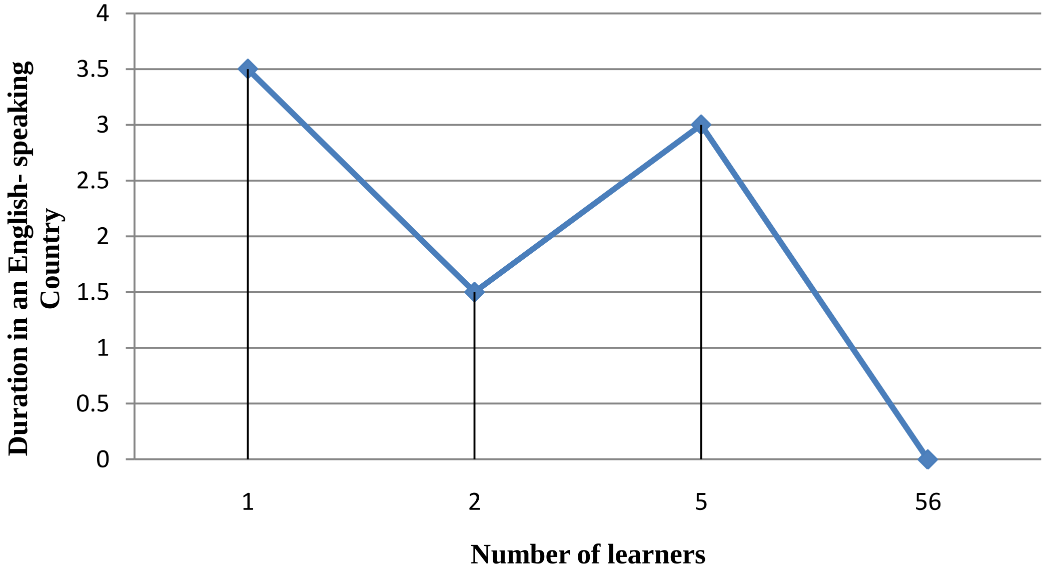
	Male	**%**	**MWC**	**Female**	**MWC**	**%**	**TOTAL**	**%**
A-TURN								
No. of interviewers	3	100	-	-	-	-	3	100
Number of words	16.845	20,8	-	-	-	-	16.845	20,8
B-TURN								
No. of learners	19	32,8	-	39	-	67,2	58	100
Duration/hr	4.1	34,2	-	7.9	-	65,8	12	100
Number of words	21.826	27,0	1148.7	42.142	1080.5	52,2	63.968	79.2
Total no. of words	**38.671**			**42.142**			**80.813**	**100**

MWC: Mean word count per interviewee

Learner Variables

Learner variables were obtained by means of a learner profile questionnaire, which was filled out and signed by learners who volunteered to be interviewed for the compilation of the LINDSEI-TR corpus. The information collected contain a range of variables about learners such as age, nationality, native language, languages spoken at home, education, years of English at school and university, medium of instruction, number of months spent in an English-speaking country, and other foreign languages in decreasing order of proficiency.

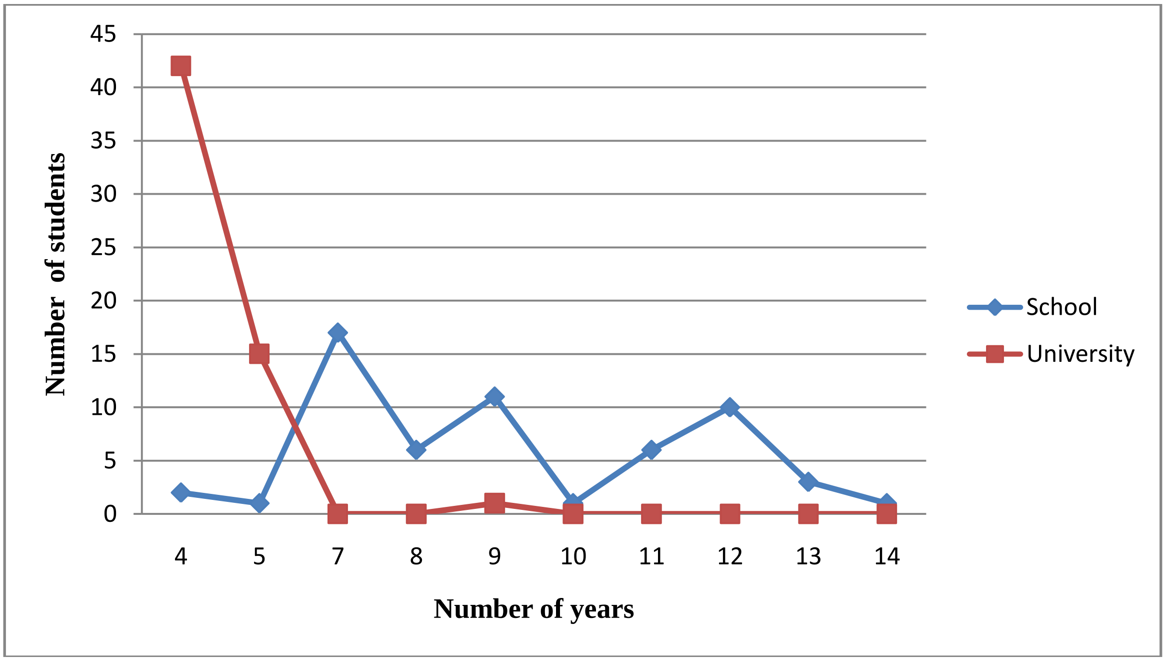
Figure 1 Duration spent in an English-speaking country by number of learners



Along with the LINDSEI guidelines (Gilquin et al., 2010), the proficiency level of the learners is considered to be advanced by means of the institutional status as they are third and fourth year B.A. students at English Language Teaching Department, Education Faculty, Cukurova University. Of the 58 interviewees with an advanced command of English who contributed to the corpus, 39 are female with the average age of 22,1, and 19 are male with the average age of 22,8. The information about how long the students spent in an English-speaking country is presented in Figure 1. As is seen, 56 learners have never been to an English-speaking country. One student has reported to have spent 3.5 months, two students 1.5 months and five 3 months in an English-speaking country.

The data gathered from the learner profiles also provide useful information regarding the students' English background knowledge. Figure 2 shows the learners' duration of English language study at school. As can be seen, 17 students studied English at school for 7 years, 9 students for 11 years, and 10 for 12 years, and 6 for 8 and 11 years. The shortest period of English at school is 4 by 2 students, with the longest being 14 years by 1 student. On average, the learners' exposure to English at school is 9,3 years. As to the study of English at university, Figure 2 illustrates that 42 students studied English for 4 years, which amounts to 72,4 % of the 58 students contributed to the compilation of the corpus. Fifteen students reported that they studied English for 5 years, which indicates that they might have had to attend English Preparatory Class before starting their first year of academic program. Only 1 student stated that he studied for 9 years, which might mean that he had to retake some courses due to failure.

Figure 2 Students' English background knowledge by school and university



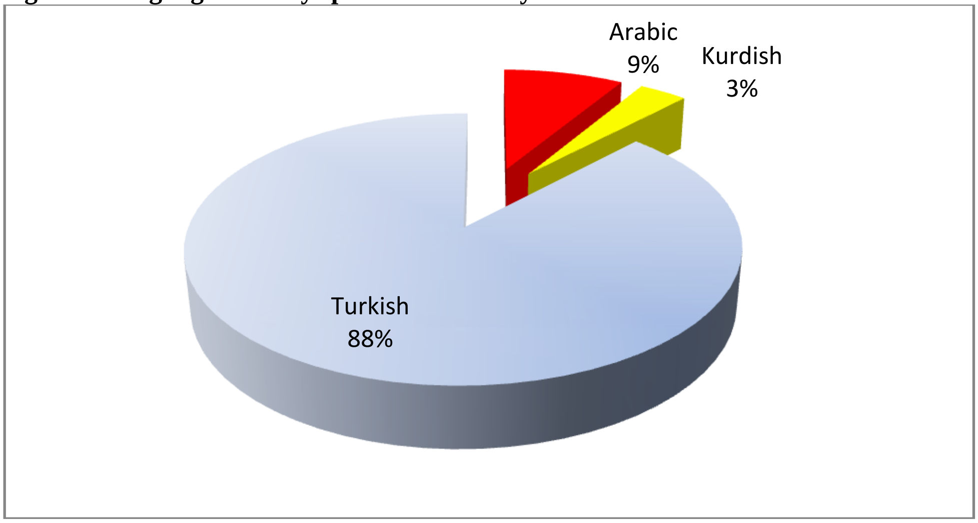
Another learner background variable in the learner profile is the knowledge of other foreign languages. Table 2 illustrates the distribution of other foreign languages apart from English in decreasing order of proficiency by the number of students.

Table 2 Distribution of other foreign languages spoken by learners

	Female		**Male**		*Total*	
	N	%	N	%	N	%
Second Foreign Language						
German	28	48.3	16	27.6	44	75.9
French	10	17.2	2	3.4	12	20.7
Kurdish	-	-	1	1.7	1	1.7
No2Lang	-	-	1	1.7	1	1.7
Third Foreign Language						
No3Lang	30	51.7	17	29.3	47	81.3
French	3	5.2	-	-	3	5.1
German	3	5.2	-	-	3	5.1
Dutch	2	3.4	1	1.7	3	5.1
Arabic	1	1.7	-		1	1.7
Hungarian	-	-	1	1.7	1	1.7
Fourth Foreign Language						
No4 Lang	38	65.5	18	31.0	56	96.6
Spanish	1	1.7	-	-	1	1.7
German	-	-	1	1.7	1	1.7

From the perspective of learners' second other foreign language in order of proficiency, German ranks first, with 75.9 % of the learners who state that they are most proficient in. This is followed by French, with 20.9 % of learners having declared as known. As to the knowledge of a third foreign language, a vast majority of learners (81.3 %) state that they do not know a third foreign language. Those who claim to speak a third language are only about 15 % in total. In this respect, three languages (German, French, and Dutch), are tied with their percentages (5.1 % each) equally distributed. Regarding the fourth foreign language, as is presented in Table 2, almost all of the learners (99.6 %) state that they do not know a fourth foreign language.

Figure 3 Languages usually spoken at home by the learner



Task Variables

Another supplementary piece of information that can be obtained from the learner profiles about the learners involved is the data regarding the languages usually spoken at home. As is seen from Figure 3, Turkish is the most spoken language at home with the highest percentage of learners at 88%, which is followed by Arabic with 9%, and Kurdish 3%.

Three types of speaking tasks were used to elicit speech samples from the learners who volunteered to contribute to the compilation of the corpus.

Task 1 *Set Topic*: Three topics about one of which the learners were asked to choose to talk about for 3-5 minutes.

Topic 1**:** An experience you've had which has taught you an important lesson. You should describe the experience and say what you have learnt from it.

Topic 2**:** A country you have visited which has impressed you. Describe your visit and say why you found the country particularly impressive.

Topic 3**:** A film/play you've seen which you thought was particularly good/bad. Describe the film/play and say why you thought it was good/bad.

Task 2 *Free discussion*: Informal conversation about learners' lives such as likes and dislikes university life, their interests and hobbies etc.

Task 3 Picture *description*: Telling a story based on a 4-picture sequence. (Granger at al, 2002)

Task 1 constitutes the first part of the interviews. Table 3 below illustrates the topic choice in this task in terms of number of learners and number of words and duration of B turns in the interviews which start with that task according to gender. The majority of learners (22 females and 6 males totaling 28 out of 58) opt for topic 3, which is about a film that they have seen. Thus, the B turns in the transcripts beginning with the description of a film amount to 30.627 words and 352.64 minutes (5.8 hours). Topic 1 regarding an experience that has taught them an important lesson is seen to be the second favorite topic among learners (22), with a total number of 23.510 and a total duration of 270.64 minutes (4.50 hours) in the transcripts starting with this topic. Topic 1, a visit to another country which has impressed them, ranks third in the choice of the learners (8), which also means a total of 8 transcripts with only 9.831 words and 1.74 hours recording.

Table 3 Topic choice by gender, turn, and duration of speech

	F	M		F	M		F	M		
	no.	*no.*	*t*	*B/T*	*B/T*	*t*	*Du./min*	*Du./min*	*t/min*	*t/hr*
Country	4	4	8	4.587	5.244	9.831	49.49	54.92	104.41	1.74
Experience	13	9	22	13.167	10.343	23.510	154.83	115.76	270.59	4.50
Film	22	6	28	24.388	6.239	30.627	274.66	77.98	352.64	5.87
Total	39	19	58	42.142	21.826	63.968	478.98	248.66	727.64	12.12
	58			63.968			727:64 min - 12:12 hr			

4. Interview procedure

The speech samples that make up LINDSEI-TR corpus were collected using an interview technique consisting of three parts. Interviews conducted face to face by three instructors were planned to last 15 minutes, but interview length ranged between 10:41 and 19:53 minutes. First, interviewees were told that they would be interviewed on topics of their personal interest for about fifteen minutes and the conversation will be recorded. Then, they were asked to choose one of the three topics presented in written form (see task variables) and think about what they were going to say for a few minutes as they were expected to talk for 3-5 minutes. They were also asked not to take any notes during the thinking period in order for the talk to be spontaneous. This constituted the first part of the interview. After interviewees finished talking about the topic that they chose, the interviewers proceeded to the second part of the interview, which was an informal conversation, by either asking questions related to what learners said or bringing up more general topics to keep the conversation going. All the interviews were concluded with a story-telling task, which made up the third part of the interview. The interviewers were shown a four picture sequence and asked to make up a story around them rather than describe the pictures. After the interviews ended, each interviewee was asked to fill out and sign a learner profile, which helped the collection of data about learners' social and educational backgrounds and signified the interviewee's consent for the speech recorded to be used for research purposes.

5. Transcription

The interviews have been transcribed and marked up according to a standard developed at the Centre for English Corpus Linguistics (CECL), the Université Catholique deLouvain (Granger and Gilquin, 2013). The transcription process begins with assigning specific codes to each interview and goes on with verbatim transcription of everything spoken during the interviews by paying particular attention to such speech features as dysfluencies, filled pauses and backchanneling, empty pauses, truncated words, phonetic and prosodic features etc. These features are identified and accordingly indicated with special markup. The transcripts are later examined and checked in detail to ensure that the interviews are transcribed accurately and consistently in terms of the transcription conventions. The same standard is shared by all international partners who have contributed to the project.

The coding process involves assigning each transcribed interview an identification. For example, the code <h nt="TR" nr="TR001"> signifies that the transcribed speech is the 1st interview with a Turkish mother tongue learner. All interviews end with the </h> tag on a separate line. When these codes are associated with the access file database containing learners' sociolinguistic and educational variables, it is possible to carry out variable specific analysis on the corpus, i.e. comparing females' performance with males'. The following explains the LINDSEI markup conventions employed when transcribing the interviews, presenting examples from the transcribed learner speech.

Speaker and task segmentation

LIDSEI-TR corpus, as in other components constituting the LINSEI database, is composed of transcripts hierarchically segmented into sections and turns. Set of specific section tags are used to identify and mark the three tasks making up the interview. For example, the section tags <S> and </S> label the beginning and the end of the set topic. In the same manner, <F> </F> and <P> </P> section tags mark free discussion and picture description, respectively. The section tags, used before and after the tasks, are placed on a separate line. Such a mark-up is valuable in that it allows topic-based analyses to be carried out on the transcribed data.

(1) <S>
<A> did you manage to choose a topic </A>

Speaker turns are indicated in vertical format with turn boundaries enclosed between angle brackets at the beginning of each turn: "A" for the interviewer and "B" for the interviewee. The end of each turn is labeled with either </A> or </B>.

(2) <A> how long have you been at this university </A>
<B> I am here .. for four years <overlap /> I have been here for four years </B>

Orthography and spelling

Following the British spelling conventions, capital letters on certain specific words such as proper names, I, Mrs, etc. are kept, but words at the beginnings of turns are not capitalized. No punctuation marks are used to indicate the end of a sentence or a clause.

(3) *<B> (eh) and I was a child </B>*

All standard contracted forms are retained as they are typical features of speech (4). Non-standard forms (cos, dunno, gonna, gotta, kinda, wanna and yeah) that appear in the dictionary are transcribed orthographically in their dictionary accepted way (5)

(4) *<B> so it it doesn't change <overlap /> I think it it this is my opinion of course but I think like that (mm) </B>*

(5) *<B> yeah first really I want to be a good person (eh) for everybody in my en= (eh) around and I wanna be a successful teacher .. (eh) and having a good position </B>*

Finally, if acronyms are spelled out as sequences of letters, they are transcribed as a series of upper-case letters separated by spaces. If, on the other hand, acronyms are pronounced as a single word, they are transcribed in capital letters not separated by spaces. Numbers are spelled out as a word instead of using figures in order to avoid the ambiguity that might arise from different ways of saying, for example, "1901".

Disfluency and backchannel signals

Non-word utterances such as filled pauses indicating hesitation or planning what to say (6) and backchanneling used to encourage the speaker to continue (7) are speech events characterizing spontaneous speech. Therefore, these features of speech are transcribed and marked as part of the interview, employing a limited set of fillers, which contain (eh) [brief], (er), (em), (erm), (mm), (uhu) and (mhm).

(6) *<B> for four years yes but <overlap /> this is my last year </B>*
<A> <overlap /> okay so (uhu) can you talk about your life at the university up to now so far </A>

(7) <B> only my I tell I told my mother (laughs) <overlap /> my father don't didn't know anything about it </B>
<A> <overlap /> (uhu) </A>

Silent intervals in speech are indicated with a three-tier system depending on their length: one dot for a "short" pause (< 1 second), two dots for a "medium" pause (1-3 seconds) and three dots for "long" pauses (> 3 seconds).

(8) *<B> collided with (eh) a .. collided with a . big car </B>*

Truncated words are also marked in the transcription. The point at which the word is broken off is identified with an equals sign.

(9) *<B> yes really but in my first year I cried very much ev= nearly every day I cried </B>*

Prosodic features

An additional layer of transcription for the recorded interviews includes the identification of certain phonetic features**.** When the last syllables of such words as *to, so, or* are lengthened, they are marked with a colon attached to the end of the word (10). In addition to marking the lengthened sounds, the pronunciation of the articles *a* or *the* with their strong forms is also indicated. For example, when *a* is pronounced as [ei], it is transcribed as a[ei]. In the same manner, when the article *the* is pronounces as [i:], it is is transcribed as the[i:] (11).

(10) *<B> because (em) he blames (eh) himself from (eh) dying her wife his wife and children so: (em) he (eh) he has lots of pain so </B>*

(11) *<B> well they have to open the gate to the[i:] underworld <overlap /> and she has to give the permission because (eh) it should be her own free will I guess </B>*

Paralinguistic features

Para-linguistic features such as laughing or whispering, which changes the voice quality, are also indicated with particular tags. For example, when a particular stretch of speech is affected by laughing or whispering, it is marked with the insertion of <starts laughing> or <starts whispering> immediately before it and <stops laughing> or <stops whispering> at the end of it (12). In the same way, other non-linguistic sounds such as coughing or sneezing produced by speakers are always transcribed and added between angle brackets (13).

(13) *<B> (eh) actually (eh) lessons are hard <overlap /> <starts laughing> at first <stops laughing> </B>*

(14) <B> *if I can be <coughs> appointed . (eh) I will I will be a teach= I will work as a teacher </B>*

Also, non-linguistic events are added as contextual information between angle brackets only if they seem to have an influence on the course of the interaction and if one of the speakers reacts to it.

Additional markup

Overlap tags are used to indicate the overlap speech regions when two or more utterances occurred simultaneously. The tag <overlap /> (with a space between "overlap" and the slash) marks the beginning of overlapping speech in the first speaker's speech. It is also indicated at the beginning of the second speaker's interruption. The end of overlapping speech is not marked.

(15) *<A> okay (uhu) . and this is your last year (uhu) <overlap /> at the university </A>*
<B> <overlap /> yes my last year I am going to graduate <overlap /> I hope </B>

A three-tier system is used to represent untranscribed regions in speech. Unclear syllables or sounds up to one word are indicated with <X>. Two unclear words or unclear passages more than two words are marked with <XX> and <XXX> respectively (6). Unintelligible words or word endings are indicated with the symbol <?> directly following the word or word ending (7). Unclear names of towns or titles of films for example are replaced by <name of city> or <title of film> (8).

(6) <B> <overlap /> <XX> people but this is what I thought </B>
(7) <B> (eh) . there was a couple <overlap /> (eh) they two love<?>each other very much </B>
(8) *<B> and (eh) it lasted almost one month <overlap /> and (eh) when I was there I had I had a chance to see Washington D C New <overlap /> York and the districts of New York City <overlap /> and <name of city> </B>*

Foreign words are placed between the tags <foreign> and </foreign> (9). As a rule, foreign pronunciation is not noted, except in the case where the foreign word and the English word are identical. If in this case the word is pronounced as a foreign word, this is also marked using the <foreign> tag.

(16) *<B> <foreign> evet </foreign> (eh) yes what <overlap /> I expected </B>*

As a general rule, an appropriate extent of anonymization is maintained during the transcription of interviews, while names of famous people like singers or actors are kept. Whenever speakers involved in the interaction are addressed or referred to, their names are anonimized by using tags like <first name of interviewee>, <first name and full name of interviewer> or <name of professor> to replace names.

(17) *<B> okay hello my name is <first name of interviewee> <overlap /> and <last name of interviewee> I am a student in . . . . . . </B>*

6. Conclusion

The paper has introduced a new spoken learner corpus, the Turkish component of LINDSEI (LINDSEI-TR), particular focusing on the purpose, design criteria and compilation process as well as its qualitative and quantitative features. Although spoken learner corpora are still relatively smaller and fewer than the written learner corpora, their potential benefits has already proven their worth for second language acquisition research. The design criteria of the corpus makes it possible to conduct contrastive interlanguage analyses between LINDSEI-TR and the comparable spoken native speaker corpus LOCNEC, which can be useful to identify various aspects of spoken interlanguage of Turkish learners and accordingly develop approaches to improve pedagogical practices and solve learners' problems. The TR-LINDSEI and its written counterpart TICLE (Kilimci and Can, 2009; Can, 2009) can also be employed together for the purpose of spoken and written contrastive interlanguage analysis as TICLE was compiled considering similar criteria. The TR-LINDSEI is not yet in its final form as it is being double-checked for possible errors or inconsistencies that might have been missed. When the process is complete and it is made available, it is considered to be useful to researchers and practitioners in the fields of both second language acquisition and English language teaching.

References

Granger, S., E. Dagneaux & F. Meunier (Eds.) (2002). *International Corpus of Learner English. Handbook and CDROM*. Louvain-la-Neuve: Presses Universitaires de Louvain.

Granger, S. (2002). A bird's-eye view of learner corpus research. In S. Granger, J. Hung, and S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 3-33). Amsterdam: John Benjamins.

Granger, S., and Gilquin, G. (2013, Sep 15). *The Louvain International Database of Spoken English Interlanguage – LINDSEI.* Retrieved from http:// www.uclouvain.be/en-cecl-lindsei.html

Kilimci, A., Can, C. (2009). TICLE: Uluslararası Turk Oğrenici İngilizcesi Derlemi. M. Sarıca, N. Sarıca and A. Karaca (Eds.). *XXII. Ulusal Dilbilim Kurultayı Bildirileri* (pp. 1-11). Ankara: Yüzüncü Yıl Üniversitesi Yayınları.

Can, Cem (2009). İkinci Dil Edinimi Çalışmalarında Bilgisiyar Destekli Bir Türk Öğrenici İngilizcesi Derlemi: Icle'nin Bir Altderlemi Olarak Ticle. *Dil Dergisi*. (144) , 16-34.

Leech, G. (1998). Preface. In S. Granger (Ed.), *Learner English on Computer* (pp. xiv-xx). New York: Longman.

Gilquin, G., De Cock, S., Granger, S. (Eds.) (2010). *The Louvain International Database of Spoken English Interlanguage. Handbook and CD-ROM*. Louvain-la-Neuve: Presses Universitaires de Louvain.