

Discourse Markers and Spoken English: Nonnative Use in the Turkish EFL Setting

Asuman Aşık¹ & Paşa Tefvik Cephe¹

¹ Department of English Language Teaching, Gazi University, Ankara, Turkey

Correspondence: Asuman Aşık, Department of English Language Teaching, Gazi Faculty of Education, Gazi University, Beşevler, 06500, Ankara, Turkey. E-mail: asuman.asik@gazi.edu.tr

Received: August 22, 2013 Accepted: September 19, 2013 Online Published: November 5, 2013

doi:10.5539/elt.v6n12p144 URL: <http://dx.doi.org/10.5539/elt.v6n12p144>

Abstract

This study investigated the production of discourse markers by non-native speakers of English and their occurrences in their spoken English by comparing them with those used in native speakers' spoken discourse. Because discourse markers (DMs) are significant items in spoken discourse of native speakers, a study about the use of DMs by nonnative speakers is necessary and guiding. Thus, the study was based on two specific corpora. First, a research corpus was composed using the transcriptions of the course presentations of twenty non-native undergraduate students studying at an English Language Teaching (ELT) program in Turkey. To compare the data, transcripts of student presentations of native speakers were attained with the help of MICASE Corpus. The occurrences of the discourse markers in both corpora were determined with frequency analysis. The results indicated that non-native speakers of English use a limited number and less variety of discourse markers in their spoken English. The study therefore highlights the importance of the need for raising non-native speakers' awareness of using discourse markers in their spoken English, and recommends implications for English language teaching.

Keywords: spokendiscourse, non-nativespeakers, discourse markers, corpora

1. Introduction

There are certain invisible rules that govern interactions and are applied by native speakers without noticing (Crozet, 2003). Native speakers of English apply these invisible rules without noticing what kind of elements they should or should not include to their discourse. Particularly, in spoken discourse, native speakers naturally use certain units of talk. Discourse markers (DMs) are among these units of talk uttered by the speakers to make their speech more understandable and rich. Crystal (1988) states that DMs serve as the "oil which helps us perform the complex task of spontaneous speech production and interaction smoothly and efficiently" (p. 48). Therefore, they are also significant in teaching English as a Foreign Language.

According to the recent analyses of corpora of spoken interaction, discourse markers are among the top ten word forms (Allwood, 1996 cited in Fung & Carter, 2007). Thus, there have been numerous studies about discourse markers in English (Svartvik, 1980; Östman, 1981; Schiffrin, 1986; Aijmer, 1987; Schourup, 1985; Erman, 1987). Moreover, the studies of discourse markers in other languages have also been conducted by many authors (Bazzanella, 1990; Gupta, 1995; Chen & He, 2001). However, the studies about the use of discourse markers in English by second or foreign language speakers are limited. Hays (1992), Trillo (2002), Müller (2004) and Fung and Carter (2007) are notable authors within this field of investigation who searched the use of discourse markers by different groups who speak another language as their first language.

Due to the significance of discourse markers in spoken discourse of native speakers of English, there is a necessity to investigate these specific discourse items in spoken discourse of non-native speakers of English as well. Thus, this study is significant in order to identify the discourse markers of Turkish non-native speakers so as to provide essential implications for teaching these units of talk to English as a foreign language (EFL) learner.

1.1 Discourse Markers

Over the last twenty years, the interest towards discourse markers has increased considerably. Quirk, Greenbaum, Leech and Svartvik (1985) emphasise the interactional effect of these words and their importance in developing

an ongoing and intimate relationship with people by explaining that phrases such as *well*, *y'know*, *really* are 'sharing devices' and 'intimacy signals' in everyday conversation. Moreover, Schiffrin (1987) proposed that discourse markers could be looked from a "more theoretical level as members of a functional class of verbal (and non-verbal) devices which provide contextual coordinates for ongoing talk" (p. 41), and pointed out an operational definition by describing "discourse markers as sequentially dependent elements which bracket units of talk" (p. 31). Significantly, discourse markers are seen as "a class of lexical expressions drawn primarily from the syntactic classes of conjunctions, adverbs, and prepositional phrases" (Fraser, 1999, p. 931) by signalling a relationship between the previous utterance and the following one. Moreover, Aijmer (2002) points out that they should be studied pragmatically rather than only grammatically as they are "a class of words with unique formal, functional and pragmatic properties" (p. 2).

There are certain characteristics of discourse markers which are connectivity, multifunctionality, optionality, non-truth conditionality, weak clause association, initiality, orality and multi-categoriality (Schourup, 1999). Connectivity is one of the basic characteristic of DMs as DMs are used to establish a relationship between the current utterance and the previous one. Moreover, DMs are used to fulfil several functions, which make them multi-functional and also multi-categorial. For instance, *well* may function as a hesitation device, denoting thinking process or opening and closing of topics (Fung & Carter, 2007). Another characteristic of DMs is their syntactic and semantic optionality. That is, their removal from the utterance does not change the grammaticality of it. However, this does not mean that they should not be considered as unnecessary elements, they are used to reinforce the statements. Non-truth conditionality is another feature of DMs, which refers that DMs do not contribute anything to truth-conditions of the proposition expressed semantically by an utterance. Furthermore, DMs have weak clause association as they can be treated out of the syntactic structure or not a strong component within sentential structure. Similarly, when analysed syntactically, DMs take place generally in initial positions. In addition, according to Louwerse and Mitchell (2003), DMs occur more often in spoken rather than written discourse, which makes the characteristic of orality significant.

1.2 Discourse Markers and Language Learners

Previous studies on DMs and their functions in the past focused on the issue of how the use of DMs contributes to the pragmatic and communicative competence of speakers or the pedagogical significance of DMs in language teaching (see, e.g. Svartvik, 1980; Östman, 1981; Schiffrin, 1986; Aijmer, 1987). Moreover, in recent years, the number of studies focusing on DMs uttered by not only native speakers but also non-native speakers of English has also been increasing.

Lam (2009) points out that DMs are crucial for learners to communicate successfully at the pragmatic level of interaction. In other words, DMs may help non-native learners of English gain nativeness in the spoken or written discourse of a foreign language. This feeling of nativeness will help learners feel comfortable while learning a foreign language. With the help of DMs in spoken discourse, the naturalness of talk can be attained and similarly in written discourse, the text gains a higher level of coherence (Halliday & Hasan, 1976).

By discussing DMs in terms of 'pragmatic fossilization' which is 'the phenomenon by which a non-native speaker systematically uses certain forms inappropriately at the pragmatic level of communication' (p. 770), Trillo's (2002) study analyses the development of certain DMs (*listen*, *well* and *you know*) in spoken discourse of Spanish non-native speakers of English by comparing them with the use of native speakers. This study displays that in terms of quantity and diversity, non-native adults' usage of DMs is more limited than that of native children and there are some pragmatic failures committed by non-native speakers. The main reasons for these failures are listed as the non-natural teaching environment and inadequate pragmatic resources in the learning process.

Hellermann and Vergun (2007) has studied on 17 beginning adult learners of English who has no previous formal English language instruction to find out the frequency of use and certain functions which are not explicitly taught. They emphasize that "many language learners have 'grammatical' language as the primary goal of their language learning experiences" (p. 158). However, this 'grammatical' target proficiency is often defined as what native speakers of the language consider accurate usage of syntax, phonology, morphology, and semantics so that the propositional content of an utterance is made clear. Thus, as DMs are words or phrases that function within the linguistic system to establish relationships between topics or grammatical units in discourse, the learners will have a better proficiency.

Müller (2004, 2005) analysed the use of DMs by German EFL speakers as compared to its use by American native speakers in detail both quantitatively and qualitatively. The results show that non-native data consists different types of pragmatic functions when compared with the ones of native speakers. This study also

acknowledges that if language learners are capable of using DMs effectively and adequately in spoken discourse, their utterances will be much more understandable for the hearer or the listener.

Recently, Fung and Carter (2007) have also focused on the production of discourse markers in pedagogic settings by using data of Hong Kong learners of English and British native speakers. Their study also displays that non-native speakers tend to use less frequently the kind of DMs British speakers usually use and the diversity of functions in non-native corpus is limited. Thus, Fung and Carter (2007) propose that language learners should learn discourse markers “in order to facilitate more successful overall language use and at the very least for reception purposes” (p. 434).

1.3 Research Questions

Although the number of the studies on DMs of non-native speakers of English has increased recently, there is still need for further research of these pragmatic items in spoken discourse of non-native speakers of English from different L1 background by comparing them with the ones of native speakers. The current study, therefore, will provide a corpus-driven investigation of DMs in terms of cross-cultural research so that specific pedagogical implications can be offered for EFL learners. To this aim, the following research questions have been formulated:

- 1) Which DMs are used by Turkish non-native speakers of English in spoken discourse?
- 2) What is the frequency level of the DMs used by Turkish non-native speakers of English in spoken discourse?
- 3) Are there any differences between the DMs used by Turkish non-native speakers and native speakers of English in spoken discourse?

2. Method

To reach the objectives of the study, the data were collected from two specific corpora: a research corpus constituted particularly for the current study for non-native data and a sub-corpus from the Michigan Corpus of Academic Spoken English (hereafter, MICASE) for native data. MICASE is the outcome of a research conducted by the English Language Institute (ELI) at the University of Michigan, and was taken as a standard for comparison with the corpus gleaned in the current study to be used as baseline data for the interpretation of the non-native language production. In terms of comparability of the two dataset, specific criteria have been set within the current study such as discourse context, discourse mode, number of students, course type and language context, which are all explained in the following sections.

2.1 Non-Native Data in the Turkish EFL Context

2.1.1 Participants

The participants of the research corpus are 20 senior undergraduate students who were studying in the Department of English Language Teaching (ELT), a pre-service teacher education program at a large state-run university, Gazi University, in Ankara, Turkey. The ELT program at this university is one of the leading and populous departments within the field in Turkey with more than 1,250 undergraduate and nearly 40 graduate students. Instruction on the ELT program is conducted in the medium of English.

2.1.2 Transcription

The data were collected along with the presentations done by the participants about the certain topics assigned to them by the instructor of two different courses of the fourth grade curriculum of 2010-2011 Academic Year Spring Semester. These courses are specifically optional courses named as “Sociolinguistics and Language Teaching” and “Pragmatics and Language Teaching.”

For the purpose of collecting the data within spoken discourse of non-native speakers of English, the student presentations of Turkish undergraduate students were first audio-recorded and then saved as sound files on a computer. The recordings were selected according to their audio-quality and representation of the items searched.

Because the research corpus is constituted by taking MICASE as the basis for the study, the transcription process was conducted in a similar way and manner with MICASE. The transcription of the research corpus was done according to MICASE orthographic transcription conventions and mark-up system (see in Appendix) which are organized to allow for ease of readability, while including enough details to ensure adequate comprehension from the text of the transcript alone. The selected recordings were transcribed by using the conventions stated above and directly into a computer file using a computer program that was originally developed for the MICASE Project, called Sound Scribe. During transcription, speech errors made by the participants were not corrected and were transcribed as how they had actually occurred.

2.1.3 Data Selection

Transcripts of the student presentations provide a context that involves EFL students' use of foreign language in a range of certain abilities such as conveying and exchanging information in an academic setting. In this respect, this context may also offer several ways for using DMs in their spoken discourse for different functions such as "cognitive, structural, referential and interpersonal functions" (Fung & Carter, 2007, p. 418). Moreover, student presentations were chosen as the academic speech that can be transcribed more effectively and properly and analysed as the samples of spoken discourse both of native and non-native speakers of English.

In order to develop the small-scale research corpus of non-native speakers of English, thirty presentations were recorded. All of the presentations were listened and a selected fragment of totally 315 minutes by twenty speakers was taken as the main focus for analysis and interpretation as it constitutes the most representative and richest section in terms of oral interaction among the participants. Thus the research corpus consists of students presenting their assigned topic and their interaction with their class-mates during the presentation. Moreover, the research corpus contains data totalling 34,420 words.

The research accepts that the research corpus containing Turkish non-native speakers' data is to be treated as suggestive rather than conclusive about Turkish language learner use. When compared to MICASE which is a large, systematically collected corpus, the research corpus is rather a small-scale one. However, the research corpus is mainly treated as to identify DMs and their occurrences in spoken discourse of Turkish non-native speakers of English while native data provides a kind of comparison criteria about how native speakers approach DMs in their spoken discourse.

2.2 Native Data

Native data was attained from a sub-corpus of Michigan Corpus of Academic Spoken English (MICASE). The MICASE Corpus is a spoken language corpus which is available on-line by consisting of approximately 1.8 million words (nearly 200 hours of academic speech) including contemporary university speech within the microcosm of University of Michigan, Michigan, USA. Within the corpus, there are 152 speech events available; small and large lectures (62), public interdisciplinary or departmental colloquia (13), discussion sections (9), student presentations (11), seminars (8), undergraduate lab sessions (8), lab group and other meetings (6), one-on-one tutorials (3), office hours (8), advising consultations (5), dissertation defences (4), study groups (8), interviews (3), campus/museum tours (2), and service encounters (2).

For comparability of the two specific corpora, in MICASE, "student presentations" of native speakers of English who are senior-undergraduates were chosen as they would better serve for the objective of the study. Moreover, the permission for using these transcripts of MICASE was taken from its authors. Transcripts selected for the study include 19 students as speakers presenting a specific topic consecutively. The presentations given by the students in the fields of social sciences, education and humanities were chosen in order to have a comparable context in transcripts with the research corpus. Similarly, the native data consists of students' spoken discourse while presenting their topic and classroom interaction. Moreover, total duration of the selected fragment is 300 minutes with total words counted as 41,173.

2.3 Data Analysis

The research is a corpus-based study through mainly quantitative analysis which was conducted by the use of descriptive statistics to display the occurrences and distribution of DMs in the discourse through lexical size and frequency counts. However, in order to support the analysis of the research, instances of DMs are also examined for providing concrete examples. There were several procedural steps performed within the quantitative analysis of the research. First, all transcripts were analysed in detail regarding which words or phrases were qualified as DMs. During this part of analysis, the functions of DMs proposed by Schiffrin (1987), Brinton (1996), Fraser (1999), Müller (2005) and Fung and Carter (2007) were taken as the basis to search for DMs. Thus, some instances of the discourse item to be examined were excluded from the research. For example, *well* is an item which should be treated carefully during analysis. *Well* is itemized when it is used to fulfil the function of denoting thinking process, as a hesitation marker or used to open and close the topics. However, in cases when *well* collocates with "very" as an adverb qualifying an adjective, it cannot be identified as DM as it does not fulfil any DM function. For example, in the following extract *well* is used as an adverb qualifying an adjective rather a sentence so these kinds of instances were excluded.

...In other words, if you close uhh if you feel close to someone because that person is related to you, or you know him or her well uhh or he or she is similar to you in terms of your age, social class, occupation etc. you feel -ness uhh you feel less need to employ indirectness... (Research Corpus)

A similar analysis was then conducted for each instance of the DMs in the transcripts, and the list of DMs was composed. Then, in order to display and process results of a corpus of language in an effective way and reach to the analysis for descriptive statistics, a concordance program, *AntConc* was used. Thus, with the help of *AntConc*, each DM was displayed within the concordance lines through which each one of the instances in which DMs occur was analysed. It should be emphasized at this point that the significance of this research is that all the items that serve the function of a DM were identified. In other words, not only a limited number of DMs identified *a priori* were analysed within the corpus. Afterall, a comparative analysis was done for the purpose of displaying the contrastive frequency.

3. Results

3.1 Descriptive Analysis

For the quantitative analysis of the research, the word counts of each transcript were used to display the frequency of the items identified in transcripts. DMs are displayed in the tables for every 100 words in the corpus. Analysis of the data has been conducted upon the frequency of DMs in total word count in each corpus. The occurrences and frequencies of the DMs identified in MICASE and the research corpus are presented throughout the tables (See Table 1 and 2).

Total word count of transcripts within the research corpus is 34,420, which was taken as the basis for calculating the frequency of each DM. Within 34,420 words uttered by non-native speakers, 79 different DMs were determined. However, the number of occurrences of each DM differs particularly. Moreover, total number of occurrences of all DMs is 3,839 which compose 11.15 % of the total word count.

Table 1. Occurrences and frequencies of the DMs in the research corpus

Total Word Count : 34,420					
DM	Occurrence	Percent	DM	Occurrence	Percent
uhh	1,401	4.07	just	26	0.08
and	637	1.85	now, yeah	25	0.07
so	186	0.54	however	23	0.07
yes	180	0.52	Firstly, first, first of all	21	0.06
but	153	0.44	really	19	0.06
umm	119	0.35	of course	14	0.04
for example	114	0.33	such as, you know	13	0.04
let's...	63	0.18	clearly	11	0.03
as	96	0.28	i mean, you see	10	0.03
because	91	0.26	alright	8	0.02
or	89	0.26	by the way, even, exactly, for this reason, I think, only, right	7	0.02
okay	99	0.29	as a brief, generally	6	0.02
also	65	0.19	i guess, obviously, probably	5	0.01
like	47	0.14	in other words, well	4	0.01
then	37	0.11	as an example, especially, maybe, moreover	3	0.01
hmm	34	0.10	although, as a result of, in fact, in this way, indeed, just like, lastly, otherwise, simply, specifically, this is to say	2	0.01
as you see	32	0.09	Other 18 items	1	0.00
huh, hi-huh	27	0.08			

Table 1 summarizes the occurrences of each DM and their representation within the corpus. DMs which have the same number of hits in the corpus are displayed together. Their occurrences and percentages are stated as per each DM. Table 1 clearly displays that the most frequent item identified within the research corpus is *uhh*, (with 4.07 % out of a total of 11.15 %) which significantly composes the majority of the DMs. Furthermore, *and*, *so*, *yes* and *but* compose a great deal of DM representation within the research corpus. Through the list, it is

noticeable that there are several DMs used with *as* such as *as you see, as a brief, as I mean, as you know, as a result of* and *an example*. *As* was counted separately from these markers and these phrases were identified as DMs and their occurrences were counted individually. Among these items, *as you see* has the largest number of occurrences with 32 hits by having 0.09 %.

Moreover, within the list, some markers are displayed together; for example, the group of *huh, hi-huh, uh-huh*; the group of *firstly, first, first of all* and the groups of *for this/that reason, for these reasons* are the same DMs with slight changes in their word forms. Moreover, after the first 17 items in the list, the frequencies of the DMs display similarities group by group such as the group *huh, hi-huh, uh-huh, just* and the group *now, yeah* and *however*. Other 18 DMs (e.g. *above all, absolutely, anyway, I mean*) are not displayed in the table as which they have only hit in the corpus and don't represent any particular contribution to the frequency analysis.

The results of the analysis in MICASE Corpus were obtained by following a similar process with the one conducted in displaying the results of the analysis in the research corpus in order to arrange tools for a comparative analysis (see Table 2). In particular, the number of total words uttered by the native speakers of English is 41,173 and the occurrences of DMs are totally 21.48 % with 8,844 hits. There are 104 different DMs identified with several frequency rates. Specifically, *and* is the most recurrent DM used by native speakers of English with 1,519 hits by representing 3.69 % out of 21.48 %. The DM *umm* follows *and* with 1,232 hits by representing 2.99 %. Other recurrent DMs are *like, uh* and *so* with considerable hits.

Table 2. Occurrences and frequencies of the DMs in MICASE corpus

Total Word Count : 41,173					
DM	Occurrence	Percent	DM	Occurrence	Percent
and	1,519	3.69	alright	53	0.13
umm	1,232	2.99	pretty, stuff	49	0.12
so	541	1.31	still	48	0.12
but	383	0.93	let's	43	0.10
just	367	0.89	first, firstly	40	0.10
okay	211	0.51	yes	39	0.09
you know	205	0.50	i guess	31	0.08
really	188	0.46	probably	30	0.07
because	179	0.43	later	29	0.07
yeah	147	0.36	definitely	27	0.07
or	131	0.32	though	24	0.06
also	126	0.31	i know	21	0.05
actually, i mean	116	0.28	sort of/sorta	19	0.05
as	115	0.28	especially, exactly, just like	15	0.04
right	110	0.27	obviously	14	0.03
I think	101	0.25	mainly	13	0.03
kind of/kinda	122	0.24	secondly, you see	12	0.03
basically	94	0.23	although, even though, for example, instead	11	0.03
now	82	0.20	of course, sure	10	0.02
well	81	0.20	anyway, great, however, totally	9	0.02
cuz	73	0.18	as well, for instance, in conclusion	8	0.02
even	69	0.17	such as	7	0.02
only	69	0.17	in fact, so far, yet	6	0.01
maybe	68	0.17	as you (can) see, at the same time, extremely, i believe, somehow,	5	0.01
oh	67	0.16	eventually, in that sense, next, recently, whereas	4	0.01
mhm, uhuh	64	0.16	finally, frequently, in general, initially, quite, right now, simply	3	0.01
			Other 19 items	1 or 2	0.00

Similarly, some items were gathered within the table; such as *mhm* and *uhuh*, *first* and *first one*, *sort of* and *sorta* and *kind of* and *kinda*, as they are the same DMs with slight changes in their word forms. Furthermore, several items represent the same percentage. Other 19 DMs (e.g. *as a result*, *briefly*, *generally*, *by the way*) are not displayed in the table as which they have only hit in the corpus and don't represent any particular contribution to the frequency analysis.

3.2 Comparative Analysis

As an overall evaluation (see Table 3), there were 3,839 occurrences of DMs identified in the research corpus, which consists of 34,420 words. On the other hand, within 41,173 words, there were 8,837 occurrences of DMs identified in MICASE. These results indicate that the frequency of DMs in the research corpus, 11.15 %, is lower than the one in MICASE with 21.46 %.

Table 3. Comparative presentation of the DMs in the research corpus and MICASE

Corpus	Occurrences	Percent	Word count
Research Corpus	3,839	11.15	34,420
MICASE	8,837	21.40	41,173

For the purpose of analysing the DMs particularly in each corpus, the *median* was used. The median is the numerical value separating the higher half of a sample from the lower half, and is preferred to use to differentiate more frequent items from less frequent items in each particular corpus. Thus, medians of both corpora were determined and a comparative analysis was conducted within more frequent items of each corpus. Since some of the DMs had just one occurrence within the corpora, they have no contribution to the frequency analysis. Therefore, these items were excluded within the lists to be used to find median in each corpus. In accordance with the results of the median, the items that are more frequent in both corpora and their frequencies are presented in Table 4.

Table 4. Comparative results of the frequencies in the research corpus and MICASE corpus

DM	MICASE (%)	Research Corpus (%)	Representation of DMs in research corpus as compared with MICASE
and	3.69	1.85	Less frequent
umm	2.99	0.35	Less frequent
like	1.49	0.14	Less frequent
uhh	1.44	4.07	More frequent
so	1.31	0.54	Less frequent
but	0.93	0.44	Less frequent
just	0.89	0.08	Less frequent
then	0.52	0.11	Less frequent
okay	0.51	0.29	Less frequent
you know	0.50	0.04	Less frequent
really	0.46	0.06	Less frequent
because	0.43	0.26	Less frequent
yeah	0.36	0.07	Less frequent
or	0.32	0.26	Less frequent
also	0.31	0.19	Less frequent
actually	0.28	0.04	Less frequent
i mean	0.28	0.03	Less frequent
as	0.28	0.28	Comparable
now	0.20	0.07	Less frequent
Let's...	0.10	0.18	More frequent
first, first one, firstly, first of all	0.10	0.06	Less frequent
yes	0.09	0.52	More frequent

The more frequent items in each corpus are matched, and the result indicates that there are 22 items identified as the same in each corpus. However, their frequencies in each particular corpus indicate discrepancies. Table 3 also presents the representation of DMs in the research corpus when compared with MICASE. In other words, it indicates whether the DMs of non-native speakers of English are used more frequently or less frequently by comparing with the same DMs used by native speakers of English.

As a consequence, 18 DMs (for instance *and, umm, you know, okay, just, actually* etc.) out of 22 are used more frequently by native speakers of English while only three DMs (*uhh, let's... and yes*) are used more frequently by non-native speakers. Besides, *as* is the only one which has the same frequency in both corpus.

Moreover, the other DMs in the lists should be taken into account as they also reflect significant discrepancies. For example, the DMs in MICASE such as *kind of/kinda, right, I think, basically, well, cuz, even, only, maybe, oh, alright, pretty, I guess, probably, later, definitely, though* and *I know* do not exist in the more frequency table of the research corpus. On the other hand, there are also some items of the research corpus that do not exist in more frequent table of MICASE, which are *for example, as you see, however, of course, such as, clearly, you see/see*.

4. Discussion

Through the quantitative analysis, the research reaches its objectives which are identifying the DMs and their frequencies in the Turkish non-native speakers of English and native speakers' spoken corpus and then conducting a comparative analysis through the results. Thus, the findings are analysed to make certain comments about the DMs used by both non-native speakers and native speakers of English.

As an overall evaluation of the occurrences of the DMs between the two groups of speakers, the findings of the study show that native speakers use DMs more frequently than non-native speakers in terms of occurrences within the total word count of the transcripts and also native speakers use much more different DMs with several functions, that is, their spoken discourse has a variety of DMs. This overall finding is similar to the outcomes of the previous studies by Weinert (1998), Trillo (2002) and Hellerman and Vergun (2007).

However, it cannot be stated that non-native speakers within the study have used DMs in limited number. In fact, it can be argued that there is a tendency of non-native speakers to use DMs in their presentations. This result also supports the previous studies on the DMs by Hays (1992), Lee (1999) and Hellermann & Vergun (2007), who claim that students with a higher proficiency in the learned language are more likely to use DMs. Although the participants of the current research corpus are not acculturated to the foreign language environment, they are upper level students of English and their use of DMs is significant, which supports Hellerman and Vergun's (2007) statement that students of higher proficiency levels use more focal DMs. Tendency of using discourse markers in spoken discourse should be supported to make Turkish non-native speakers fluent in their spoken discourse.

Moreover, it can be noted that the most frequent DMs used by native speakers of English such as *and, like, so, but, just, then, okay, you know, really, yeah, I mean* exist in the table of the most recurrent items used by Turkish non-native speakers. However, when compared with native speakers, these tokens are less frequently used by non-native speakers. Moreover, particular DMs such as *kind of/kinda, right, i think, basically, well* and *cuz* do not take place in the more frequent items of non-native speakers although they are considerably used by native speakers. These DMs are significant in spoken discourse of native speakers. However, the results may suggest that Turkish non-native speakers of English have not been exposed to these pragmatic items while learning English as a foreign language through instructional materials or their language teachers.

One of the fundamental points to be discussed is that the most frequent item of Turkish non-native speakers is *uhh*. In particular, it represents 4.07 percent of the whole corpus. This frequency is considerably significant when compared to the frequencies of other DMs in the research corpus. Moreover, Turkish non-native speakers of English use *uhh* for several purposes such as hesitation, denoting thinking process, searching for the right word and fillers such as in the following extract:

And she is risky, the message is not sense correctly uhh the indirectness uhh doesn't work. uhh and this is uhh causative because the girl uhh is not mature enough to process...and this is the example from the book. uhh you can follow uhh on your book. (Research Corpus)

Therefore, it can be concluded that *uhh* is a 'savior' for Turkish non-native speakers and instead of using any kind of DMs they utter *uhh* most of the time. The reasons behind overusing *uhh* might be that non-native speakers feel insecure of expression in the target language or they are using it during thinking process as a filler. Thus, this result clearly supports the argument of the research and also the main reason of overusing and

fossilization of *uhh* is that Turkish non-native speakers are not aware of and competent in using particular DMs for particular contexts and with several functions. Therefore, there is a need to learn more types of DMs and their functions. For example, in the example above, instead of using *uhh* most of the time, Turkish non-native speakers may prefer specific functional DMs according to the context such as *you know* for sharing knowledge, *I mean* for reformulation or self-correction and *actually* for indicating attitudes. This kind of use of the DMs in their utterances will be helpful for language learners to feel more secure in their explanations and to be more competent in the foreign language.

Another significant observation is the use of *yes* and *yeah* by the speakers. *Yes*, formal form of *yeah*, is one of the few DMs that is more frequent in non-native corpus. On the other hand, *yeah* is a more frequent item used by native speakers of English such as in the following extract:

yeah *um she she said that she would wait she would say she would go up until the last day, and then she would email and say some plans came up <SS: LAUGH>yeah and so that would be avoidance, that was one of the people who avoided responding or avoided rejecting someone, so, yeah that that was interesting. anything else? yeah. (MICASE)*

This finding also supports the result of Fung and Carter's (2007) research in which they are comparing the frequencies of native speakers and Hong Kong language learners. Fung and Carter (2007) also conclude that "there is an over reliance on *yes* rather than *yeah* among the Hong-Kong subjects and they did not use the range of possibilities available with *yeah* that native speakers do as a way to exhibit understanding or acknowledgement (interpersonal category), or as a continuer of the progress of the primary speaker's turn (structural category)" (p. 431). Moreover, native speakers use different types of DMs to show responses like *yeah* such as *sure, right*. On the other hand, again non-native speakers do not have this kind of variety.

Similarly, non-native speakers use *because* for referential purposes while native speakers again use both *because* and *cuz*. However, non-native speakers do not use *cuz* at any instances, but *cuz* functions in a more overtly discourse-marking role in spoken discourse rather than *because* as it serves continuation of topics as well as causal relationships between utterances.

um that I could get thrown into that classroom, um... and... i don't know i was kinda, it kinda intimidates me cuzi don't know what i would do, you know if i was in her place, um cuz i don't know karate. <SS: LAUGH>i asked myself you know why i'm so scared of this classroom, um. (MICASE)

Furthermore, the findings show that non-native speakers have used *well* with a very limited frequency (% 0.01) while native speakers have used more frequently (% 0.20) for several purposes. *Well* is the most frequently analysed DM by different authors (e.g. Cuenca, 2008; Blakemore, 2002; Aijmer and Simon-Vandenberg, 2003) due to its significance in spoken discourse by being fully pragmatic in terms of structural, cognitive and interpersonal functions. The extract below illustrates only two instances of *well* those functions as opening new topic or topic shifting in native corpus. However, in research corpus, it has been noted that non-native speakers do not tend to use this particular DM in their spoken discourse. The reason may be that they do not have necessary knowledge about how *well* may function while speaking English.

S4: um well thank you for being present for our presentation and um, um our project is about, is a combination of cross-sectional, studies and pragmatics, and well through this class we all know that there are, (MICASE)

To summarize, the findings display that DMs are not totally excluded in the non-native data but they are used less frequently. That is, Turkish non-native speakers tend to less frequently use the kind of DMs that native speakers use mainly for interpersonal and cognitive purposes such as *I mean, you know, like, etc.* Turkish non-native speakers prefer to use more textual and structural DMs in their spoken discourse such as *and, so, but*, which may be due to their transfer of knowledge of coordinates from their written discourse or lack of awareness about the range of possibilities of these items. Therefore, the variety and the range of DMs used by Turkish non-native speakers are limited in and confined to particular items, and thus there is an overreliance on certain DMs which may lead to pragmatic fossilization. This outcome is also in parallel with Qun's (2009) study, which highlights overused and fossilized items avoid the use of other types of DMs.

5. Conclusion

This study aimed at identifying the DMs used by Turkish non-native speakers of English through a corpus-based analysis. The findings of the study suggest that in language teaching environment, knowledge of DMs can be made more significant in terms of variety and their functions in spoken discourse, in correspondence with Trillo (2002), Müller (2005) and Fung and Carter (2007). Limited use of DMs in spoken discourse reflects the unnatural context that EFL speakers are exposed to. DMs are one of the significant items which help language

learners use language in culturally, socially and situationally appropriate to maintain cohesion and effectiveness in their discourse and interpersonal interaction (Wierzbicka, 1991). Lack of or misused DMs in spoken discourse may make the utterances of non-native speakers vague, misunderstood, incoherent and inappropriate.

Because the study displays that Turkish non-native speakers of English are not using DMs effectively and in sufficient variety in their spoken discourse, their awareness should be raised towards the variety and functions of DMs. DMs can be taught by both explicit and implicit teaching (Rose & Kasper, 2001; McCarthy & Carter, 1998). In particular, integration of several activities like language observation, problem-solving, and cross-language comparisons (Fung & Carter, 2007) and language samples from daily conversations of native speakers (Hellermann and Vergun, 2007) can be suggested to increase language learners' awareness towards the variety and the use of DMs. Moreover, use of DMs by non-native teachers in classroom interaction and encouraging students to use DMs in interactions may also help.

Finally, further research can be conducted in this area. Informed sessions about the use of DMs in language classrooms can be done to observe the progress in functional use of DMs by language learners, or use of DMs in different discourse contexts can be searched as well. Additionally, use of DMs by non-native speakers with different origins might contribute to comparative analysis among different groups of non-native speakers of English and might be significant in the field of intercultural pragmatics.

References

- Aijmer, K. (1987). Oh and ah in English conversation. In W. Meijs (Ed.), *Corpus Linguistics and Beyond* (pp. 61-86). Amsterdam: Rodopi.
- Aijmer, K. (2002). *English Discourse Particles*. Amsterdam: John Benjamins.
- Aijmer, K., & Simon-Vanderbergen, A. (2003). Well in English, Swedish and Dutch. *Linguistics*, 41(6), 1123-1161. <http://dx.doi.org/10.1515/ling.2003.036>
- Bazzanella, C. (1990). Phatic connectives as interactional cues in contemporary spoken Italian. *Journal of Pragmatics*, 14, 629-647. [http://dx.doi.org/10.1016/0378-2166\(90\)90034-B](http://dx.doi.org/10.1016/0378-2166(90)90034-B)
- Blakemore, D. (2002). *Relevance and linguistic meaning: The semantics and pragmatics of discourse markers*. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511486456>
- Brinton, L. J. (1996). *Pragmatic markers in English: Grammaticalization and discourse functions*. Berlin/New York: Mouton de Gruyter. <http://dx.doi.org/10.1515/9783110907582>
- Chen Y., & He, A. W. (2001). Dui bu dui as a pragmatic marker: Evidence from Chinese classroom discourse. *Journal of Pragmatics*, 33, 1441-1465. [http://dx.doi.org/10.1016/S0378-2166\(00\)00084-9](http://dx.doi.org/10.1016/S0378-2166(00)00084-9)
- Crozet, C. (2003). A conceptual framework to help teachers identify where culture is located in language use. In J. Lo Bianco, & C. Crozet (Eds.), *Teaching invisible culture: Classroom practice and theory* (pp. 34-39). Melbourne: Language Australia.
- Crystal, D. (1988). Another look at, well, you know. *English Today*, 13, 47-59.
- Cuenca, M. (2008). Pragmatic markers in contrast: The case of well. *Journal of Pragmatics*, 40(8), 1373-1391. <http://dx.doi.org/10.1016/j.pragma.2008.02.013>
- Erman, B. (1987). *Pragmatic expressions in English: A study of you know, you see and I mean in face-to-face conversation*. Stockholm: Almqvist & Wiksell.
- Fraser, B. (1999). What are discourse markers? *Journal of Pragmatics*, 31(7), 931-932. [http://dx.doi.org/10.1016/S0378-2166\(98\)00101-5](http://dx.doi.org/10.1016/S0378-2166(98)00101-5)
- Fung, L., & Carter, R. (2007). Discourse markers and spoken English: Native and learner use in pedagogic settings. *Applied Linguistics*, 28(3), 410-439. <http://dx.doi.org/10.1093/applin/amm030>
- Gupta, A. F. (1995). Singaporean English: Pragmatic particles. *Speech and Hearing Association (Singapore) Newsletter*. Retrieved June 1, 2010, from <http://www.leeds.ac.uk/english/staff/afg/shastxt.html/#Three>
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Hays, P. R. (1992). Discourse markers and L2 acquisition. In D. Staub, & D. Cheryl (Eds.), *The Proceedings of the Twelfth Second Language Research Forum* (pp. 24-34). Michigan: Papers in Applied Linguistics.
- Hellermann, J., & Vergun, A. (2007). Language which is not taught: The discourse marker use of beginning adult learners of English. *Journal of Pragmatics*, 39(1), 157-179. <http://dx.doi.org/10.1016/j.pragma.2006.04.008>

- Lam, P. W. Y. (2009). Discourse particles in corpus data and textbooks: The case of well. *Applied Linguistics*, 31(2), 260-281. <http://dx.doi.org/10.1093/applin/amp026>
- Lee, H. (1999). The acquisition of colloquial features by Korean Americans. *Journal of the Pan-Pacific Association of Applied Linguistics*, 3, 71-87.
- Louwerse, M., & Mitchell, H. H. (2003). Towards a taxonomy of a set of discourse markers in dialog: A theoretical and computational linguistic account. *Discourse Processes*, 35, 199-239. http://dx.doi.org/10.1207/S15326950DP3503_1
- McCarthy, M., & Carter, R. (1998). Spoken grammar: What is it and how can we teach it? *English Language Teaching Journal*, 49(3), 207-218. <http://dx.doi.org/10.1093/elt/49.3.207>
- Müller, S. (2004). Well you know that type of person: Functions of well in the speech of American and German students. *Journal of Pragmatics*, 36(1), 1157-1182. <http://dx.doi.org/10.1016/j.pragma.2004.01.008>
- Müller, S. (2005). *Discourse markers in native and non-native English discourse*. Amsterdam/Philadelphia: John Benjamins. <http://dx.doi.org/10.1177/1461445607081332>
- Östman, J. (1981). *You know: A discourse functional approach*. Amsterdam: John Benjamins.
- Quirk, R., Svartvik, J., Leech, G., & Greenbaum, S. (1985). *A comprehensive grammar of the English language*. London: Longman.
- Qun, Z. (2009). Fossilization of “so” as a discourse marker in cross-cultural communication. *US-China Foreign Language*, 7(8), 1-5.
- Trillo, J. R. (2002). The pragmatic fossilization of discourse markers in non-native speakers of English. *Journal of Pragmatics*, 34(6), 769-784. [http://dx.doi.org/10.1016/S0378-2166\(02\)00022-X](http://dx.doi.org/10.1016/S0378-2166(02)00022-X)
- Rose, K. R., & Kasper, G. (Eds.) (2001). *Pragmatics and Language Teaching*. Cambridge, NY: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9781139524797>
- Schiffirin, D. (1986). Functions of and in discourse. *Journal of Pragmatics*, 10(1), 41-66. [http://dx.doi.org/10.1016/0378-2166\(86\)90099-8](http://dx.doi.org/10.1016/0378-2166(86)90099-8)
- Schiffirin, D. (1987). *Discourse Markers*. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511611841>
- Schourup, L. (1985). *Common discourse particles in English Conversation: Like, well, y'know*. New York: Garland.
- Schourup, L. (1999). Discourse Markers. *Lingua*, 107(3), 227-265. [http://dx.doi.org/10.1016/S0024-3841\(96\)90026-1](http://dx.doi.org/10.1016/S0024-3841(96)90026-1)
- Simpson, R. C., Briggs, S. L., Ovens, J., & Swales, J. M. (2002). *The Michigan Corpus of Academic Spoken English*. Ann Arbor, MI: The Regents of the University of Michigan. Retrieved from <http://micase.elicorpora.info/>
- Svartvik, J. (1980). Well in conversation. In Sidney Greenbaum, Geoffrey Leech, & Jan Svartvik (Eds.), *Studies in English Linguistics for Randolph Quirk* (pp. 167-177). London: Longman.
- Weinert, R. (1998). Discourse organisation in the spoken language of L2 learners of German. *Linguistische Berichte*, 176, 459-488.
- Wierzbicka, A. (1991). *Cross-cultural Pragmatics: The Semantics of Human Interaction*. Berlin: Mouton de Gruyter.

Appendix

Micase Transcription and Mark-Up Conventions

Meaning/Description	Appearance Inon-Line Transcripts (Html Version)
Speaker IDs, assigned in the order they first speak.	S1: at the beginning of each turn or interruption/backchannel.
Unknown speaker, without and with gender identified	SU: SU-f, SU-m
Probable but not definite identity of speaker	SU-1:
Two or more speakers, in unison (used mostly for laughter)	SS:

Pauses of 4 seconds or longer are timed to the nearest second.	<P: 05>
Comma indicates a brief (1-2 second) mid-utterance pause with non-phrase-final intonation contour.	,
Period indicates a brief pause accompanied by an utterance final (falling) intonation contour; not used in a syntactic sense to indicate complete sentences.	.
Ellipses indicate a pause of 2-3 seconds	...
This tag encloses speech that is spoken simultaneously, either at the ends and beginnings of turns, or as interruptions or backchannel cues in the middle of one speaker's turn. All overlaps are approximate and shown to the nearest word; a word is generally not split by an overlap tag.	Text of overlapping speech is in blue.
Backchannel cues from a speaker who doesn't hold the floor and unsuccessful attempts to take the floor are embedded within the current speaker's turn, and not shown as a separate line/paragraph.	[S3: Text of embedded speech is in orange and surrounded by orange square brackets.]
Backchannel cues or unsuccessful interruptions that overlap with the main speaker's speech.	[S3: Text of embedded speech that is overlapped is in blue and surrounded by orange speaker ID and square brackets.]
All laughter is marked. Speaker ID not marked if current speaker laughs.	<LAUGH>, <S8 LAUGH>, <SS LAUGH>, etc.
Various contextual (non-speech) events are noted, usually only when they affect comprehension of the surrounding discourse.	<WRITING ON BOARD>
Used when part of an utterance is read verbatim.	<READING>xxx</READING>
Used for non-English words or phrases.	Italics e.g.: the mother says c'est quoi? and Annika says to parceque eh and then,...
Used when an unexpected pronunciation is used that would affect comprehension of the surrounding discourse. Dialect or other phonological variations are generally not represented.	Pronunciation guide follows the word e.g.: ...they asked the librarian for pictures of old Celtic <PRON: /seltik/> uniforms the basketball team, and it turns out that the project was he was supposed to find Celtic <PRON: /keltik/> costumes.
Used when a speaker makes a mistake without self-correcting, and the error might otherwise appear to be a transcribing error.	(Sic) follows the word. e.g.: despite the fact that that was the era of Women's Liberation like I say on the cover of Newsweek, and Gloria Steinman (sic) and uh Betty Friedan...
Two x's in parentheses indicate one or more words that are completely unintelligible. Words surrounded by parentheses indicate the transcription is uncertain.	I don't (xx) whole (xx) analysis it just struck me... lemme not write it that way (lest it be confused) with C syntax...

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).