Journal of Language Education and Research

7(1), 2021

Journal of Language Education and Research

Research Article

Discourse Markers in Learner Speech: A Corpus Based Comparative Study*

Aysel ŞAHİN KIZIL **

ARTICLE INFO

Received 14.07.2020
Revised form 30.11.2020
Accepted 13.12.2020
Doi: 10.31464/jlere.769613

Keywords:
Discourse markers
Learner corpus
Spoken corpora
Relevance Theory

ABSTRACT

Having a vital role in organizing conversational interaction, discourse markers (DMs) are considered to be a significant part of communicative competence. Combining insights from corpus linguistics, second language acquisition research and Relevance Theory, this study is an attempt to explore how non-native speakers of English (NNS) use the DMs in their spoken language. Following the framework of Contrastive Interlanguage Analysis (CIA), this study is based on two specific corpora: Turkish sub-corpus of The Louvain International Database of Spoken English Interlanguage (LINDSEI-TR) for NNS speech and The Louvain Corpus of Native English Conversation (LOCNEC) for native speech. The occurrences of DMs in both corpora were determined using inferential and descriptive statistics. The findings of the study are discussed in connection with implications for language practitioners and syllabus designers in the field of language teaching.

Acknowledgments The author thanks the journal reviewers and editors for their guidance in improving this paper.

Statement of Publication Ethics This study used existing data collected for the compilation of LINDSEI and LOCNEC. All the procedures performed in this study were in accordance with the research ethical standards, and no unlawful method or materials were used.

Conflict of Interest The author declares that there is no conflict of interest.

ISSN: 2149-5602

Introduction

Discourse markers are crucial interactional manoeuvres to organize and structure communication on various levels. As a linguistic category, the term, discourse marker, broadly refers to a syntactically heterogeneous group of expressions that possess distinctive pragmatic and semantic functions in discourse (Blakemore, 2004). Depending on the fact that studying discourse markers is at the interjection of syntax, pragmatics and semantics, it is admittedly a difficult task to create definitive lists of DMs in a language, - even little agreement has been reached on its terminology. Various researchers examined the concept under various labels such as "discourse particles" (Aijmer & Stenström, 2002), "pragmatic markers" (Aijmer, 2004; Downing, 2006; Wei, 2011), "discourse connectives" (Blakemore, 1992), among them "discourse markers" (Bu, 2013; Buysse, 2012; Fraser, 2006; Gilquin, 2016; Müller, 2005; Polat, 2011; Romero-Trillo, 2002) being the most widely used term in the literature.

Though its terminology and definition are still open to debate (Huang, 2019), discourse markers, as one of the distinct characteristics of spoken English, have informed a quantity of studies from a variety of perspectives ranging from descriptive-functional approaches (Buysse, 2012; Fraser, 2006), taxonomic discussions (Wang & Zhu, 2005; Yong, Jingli, & Zhou, 2010) to investigation of gender differences (Erman, 2001). One common conclusion reached by these studies is that discourse markers add greatly to the communication altering pragmatic meaning of utterances; thus, are a central part of pragmatic competence of language users facilitating the complex task of speech production and interaction (Andersen, 2011; Müller, 2005). Most of the studies in the relevant literature, however, have primarily focused on the language produced by native speakers (NS) of English, leaving the area of discourse markers used by non-native speakers (NNS) largely unexplored (Fung & Carter, 2007).

Underlining the importance of discourse markers by NNSs, Aijmer (2002) states that misuse or underuse of discourse markers in non-native speech may cause misunderstanding, which could impair the whole communication. In the same vein, Hellermann and Vergun (2007) note that lack of discourse markers in a speech "mark a speaker as disfluent in the target language in subtle ways", which obliquely put the speaker at the risk of "being marked as separate from the target speech community, which may then inhibit their chances for continued meaningful interaction in that speech community" (p. 161). Terraschke (2007) puts it differently by stating that discourse markers are necessary "to create an informal and friendly conversational atmosphere and to relate [non-native speakers] better to their native interlocutors and ultimately to fit better with their native peers" (p.158).

Drawing from the literature sketched out above demonstrating the significance of discourse markers in native and non-native speech and implying the need for further exploration of the topic, this study investigates the use of discourse markers by Turkish learners of English in comparison with that of native speakers of English. From a Relevance Theoretic perspective and combining insights from corpus linguistics and SLA research, this study aims to contribute to growing amount of research into learner speech

with a special focus on Turkish learners' spoken interlanguage, which makes it one of the first attempts in its scope.

Literature Review

Theoretical Framework: Relevance Theory and Discourse Markers

The framework of Relevance Theory developed by Sperber and Wilson (1986, 1995) has underpinned much of the analytical work on discourse markers as it provides a comprehensive pragmatic model for their analysis.

Building upon Grice's idea that communication is more than a process of codification and decodification of utterances as speakers make use of maxims in conversation involving an inference labor, Relevance Theory is a pragmatic approach to language based on cognitive principles. It postulates that the impact of new information is worked out by the receiver of the communicative act against the background of extant assumptions. In a cognitive environment shared by all members of a speech community, the hearer is supposed to choose a context for an utterance in order to make the correct inferences about the speakers' meaning. Choice of the context is constrained by the principle of "optimal relevance" which refers to assumption that every act of communication is optimally relevant if it is worth the audience's effort to process it (Blakemore, 2004; Sperber & Wilson, 1995). In other words, an utterance is relevant enough if it achieves adequate cognitive effects. Less cognitive effort means greater relevance (Urgelles-Coll, 2010). Accordingly, relevancy of the information in an utterance could be realized in three ways. First, new information can combine with existing assumptions to provide a contextual implication. Second, existing assumptions of the hearer can be strengthened by the new information. Third, new information can introduce a conflict to an existing assumption (Blakemore, 2002).

Within the framework of Relevance Theory, discourse markers are defined as "expressions that constrain the interpretation of the utterances that contain them by virtue of the inferential connections they express" (Blakemore, 1987 in Huang, 2011). Discourse markers are considered to be the signals used by the speaker to guide the hearer's interpretative process. As they have procedural meaning "encoding a constraint on pragmatic inferences" (Blakemore, 2002, p. 4), discourse markers function to narrow down the number of potential interpretations through the specification of certain properties of the context. More specifically, discourse markers constrain the relevant context for the utterance by reinforcing some inferences or relegating some other potential ones. Thus, they help process the information.

Three main characteristics attributed to discourse markers from a relevance-theoretic perspective are non-truth conditionality, optionality and connectivity which guide the analysis and interpretation process of the present study as well (Blakemore, 2004; Huang, 2011). Non-truth conditionality refers to the idea that discourse markers do not generally influence the truth-conditions of the proposition articulated by an utterance (Blakemore, 2002). This property of discourse markers differentiates them from the content words encompassing, for instance, manner adverbial uses of words (e.g., sadly).

The second characteristics of discourse markers is that they are syntactically optional, which means that the presence or absence of discourse markers does not affect the syntactic structure of the sentence. However, it should be noted that optionality does not reduce discourse markers to an unnecessary status. While in the absence of a discourse marker, the relationship it marks is still available to the hearer, it is no longer explicitly cued or reinforced as intended by the speaker. Turning to the "optimal relevance" principle in relevance theory, inexplicitly signaled relation would increase the cognitive effort for the processing, which will reduce the relevancy of the utterance though it is grammatically correct. The last property of discourse markers is the connectivity, which states that discourse markers link the utterance to the context; thus, they facilitate process of interpretation and interaction in the discourse (Fung & Carter, 2007).

Overall, the most important contribution of Relevance Theory to the study of discourse markers is the semantic-pragmatic characterization of these units, referring them as significant aids for communication through facilitating inferences (Müller, 2005). It is obvious that discourse markers have a pivotal role to play in communication, which leads to the assumption that they are significant constituents of interaction that cannot be glossed over by non-native speakers as well. Following section elaborates on discourse markers in non-native language use.

Discourse markers and non-native speakers

As aforementioned, a considerable amount of previous research has concentrated on the use of discourse markers in NS language (Aijmer, 2002; Biber, Johansson, Leech, Conrad, & Finegan, 2007; Fraser, 1999; Jucker, 1993). Recent years, however, have witnessed a shift of interest from NS language to comparative uses of discourse markers by NS and NNS speakers. To Gilquin (2016), compilation of interlanguage corpora is the main drive behind the growing number of research investigating DMs by language learners. It is because recent developments within the field of learner corpus research have enabled researchers to extract and analyze DMs in spoken interlanguage effectively.

Using corpus tools and techniques, a number of researchers conducted studies investigating frequency and functions of DM in learner English. Studies vary in accordance with the L1 backgrounds as well as their scope regarding the range of DMs included. The L1 backgrounds investigated so far are Spanish (Romero-Trillo, 2002), German (Müller, 2005), Chinese (Fung & Carter, 2007; Huang, 2013; Zhao, 2013), Swedish (Aas, 2011; Aijmer, 2011), French (Gilquin, 2008), Japanese (Shimada, 2014) and Turkish (Aşik & Cephe, 2013; Şahin Kızıl & Kilimci, 2014).

As for the scope, the studies conducted thus far focused on various linguistic units in learner English as DMs. Being one of the earliest examples of research on discourse markers in NNS language, Romero-Trillo (2002) analyzed the use of *listen, well* and *you know* in spoken interlanguage of Spanish speakers of English by comparing them with the use in NS speech. The focus of Müller (2005) was *so, well, you know* and *like.* Working on a relatively comprehensive list of DMs, Fung and Carter (2007) studied the appearance of *and, but, because, okay, so, yeah, really, say, sort of, I see, you see, well, right, actually, cos, you know* in the speech of Chinese EFL learners. Apart from this list, major discourse

markers reported in the studies targeting Chinese learners of English are *like, oh, well, you know, I mean, you see, I think* and *now* (Huang, 2013; Huang, 2019). Also investigating the same DMs as in the other studies, Shimada (2014) researched the use some DMs that are not frequently studied in the literature including *basically, right, alright, actually, then* etc. for the Japanese learners. The DMs studied on the side of the Turkish EFL learners are a list of 59 linguistic units including *yeah, I mean, you know, well* etc (Aşik & Cephe, 2013).

The conclusion reached by most of these studies is a complex picture of overuse, underuse and misuse of discourse markers by learners, which justifies the need for further investigation of the topic. Findings of the study by Fung and Carter (2007), for example, suggest that discourse markers such as *so, you know* and *like* are used by NNSs less than by NSs. Similarly, Huang (2019) reports the underuse of *well* as a DM by Chinese EFL learners. This is in contradiction with Müller (2005) who reported that German EFL learners underused all the DMs under investigation with the exception of *well*. *Well* is also overused by Swedish EFL learners and French EFL learners (Aijmer, 2011; Gilquin, 2008). Apart from *well*, findings regarding the overuse of DMs by learners are published by House (2009) for *you know*, by Polat (2011) for *well* and *you know* and by Buysse (2012) for *so* as a DM. In addition to this pattern regarding frequent and infrequent use, extant literature also remarks on the misuse of certain DMs by EFL learners (Granger, 2015; Romero-Trillo, 2002). Qualitatively analyzing the learner speech, Aijmer (2011) demonstrates that Swedish learners use some DMs inappropriately neglecting their primary functions. Similarly, Yong et al., (2010) found that Chinese EFL learners use certain DMs out of their functions.

To explain the reasons behind this idiosyncratic pattern of DMs in learner speech, what is generally referred is the lack of explicit instruction of DMs in language classes and limited exposure to DMs out of language classes. The first one, lack of explicit instruction is worded by Mukherjee and Rohrbach (2006) as " discourse markers are notoriously underrepresented even in modern materials" (p.216). Although spoken input is used in many language teaching contexts, it generally contains inauthentic speech or scripted talk (e.g. broadcast news), which makes the input poor in the typical spoken features such as hesitation markers and DMs (Gilquin, 2016). If EFL learners are not taught DMs explicitly through pedagogical materials, the only way they can learn these markers is the exposure to DMs out of classroom. However, this is not likely as many EFL learners have contact with English only in the classroom. When the fact that most of the EFL teachers are non-native is considered, learners' exposure to naturally occurring DMs in speech gets more complicated as their speech may "exhibit the common interlanguage feature of not including many discourse markers" (Gilquin, 2016 p.216). Even if non-native teachers use DMs, they generally use relational ones (e.g. alright, now, okay) to manage the pedagogical activities (Hellermann & Vergun, 2007).

Taken together, the literature on the use of DMs by EFL learners has shown that DMs as a significant part of communication is worth further investigation. Although there are some emerging patterns and connected causes regarding the use of DMs by EFL learners, data from different learner populations doubtlessly adds to current knowledge and

deepens our understanding regarding the topic. This accumulation of knowledge, in turn, has the potential to contribute to finding pedagogical solutions in terms of making EFL learners better communicators. With this purpose in mind, this study focuses on Turkish EFL learners and investigates the use of three DMs (*you know, I mean, I think,*) in their spoken interlanguage. Following section presents the methodology of the study.

Methodology

In all phases of this study, Research and Publication Ethics are complied with. The methodology followed in this study is Granger's (2009) Contrastive Interlanguage Analysis (CIA), a corpus-based approach that employs two types of comparisons: "between native language and learner language (L1 vs L2) and between different varieties of interlanguage (L2 vs L2)" (p.18). This study made use of both types by comparing the data from LINDSEI-TR for NNS speech and LOCNEC for NS speech. Also included is the comparison of DMs in the interlanguage of learners from different L1 backgrounds (see Fig. 2), which represents the second type of comparison in CIA.

The data for this study comes from two comparable corpora as displayed in Table 1: Louvain International Database of Spoken English Interlanguage (LINDSEI) for learner speech and The Louvain Corpus of Native English Conversation (LOCNEC) for native speech.

Table 1. Corpora under investigation

	Corpora	Participants	Size
Learner	LINDSEI-TR (B Turns)	50 Turkish University Students	54,419
Reference	LOCNEC (B Turns)	50 Native University Students	118.553

LINDSEI is a spoken corpus consisting of interviews by university undergraduates from different L1 backgrounds. To find out the use of DMs by Turkish learners of English, LINDSEI-TR –the sub-corpus of LINDSEI- which contains 50 interviews by Turkish learners considered to be advanced learners of English based on external criteria was employed. Each interview in LINDSEI-TR has an average length of 12 minutes, and followed the same pattern made up of three tasks on a set topic, free discussion and picture description. In transcribing the spoken data, a set of guidelines provided by LINDSEI team were employed. For a fuller description of LINDSEI-TR with more details on sampling, representativeness, see Kilimci (2014). The instances of DMs in native speech were taken from LOCNEC which is a comparable corpus of LINDSEI as the same design criteria and transcription guidelines we observed for its compilation. In line with the focus of this study, only B turns representing learners' turns in the interviews in both corpora were taken into consideration for analysis.

Three discourse markers were selected to analyze for both practical and theoretical reasons: *You know, I mean* and *well*. In practical terms, these three linguistic items that occurred in sufficient numbers in both corpora as tested by a preliminary frequency search are considered to serve for meaningful quantitative analysis. Additionally, these markers

are among the most widely studied items in previous research both in NS and NNS language, thus providing a fruitful comparative base for uncovering idiosyncratic properties of said DMs in various L1 groups. By comparing the findings of this study to these published analyses, the study will further enrich the relevant literature. In theoretical terms, these three markers might potentially have a number of functions performing in a speech and analysing them within a relevance theoretic framework could provide a plausible explanation for a wide range of occurrences. That is, all examples of these markers could be handled with one unique explanation through relevance theory (Jucker, 1993).

Data Analysis

To analyse the frequency of the selected DMs in both NS and NNS corpora, WordSmith Tools 5.0, a standard corpus investigation software (Scott, 2010), was employed. The *Concord* Tool in *WordSmith 5* was also used to differentiate phrases under investigation used as DMs from phrases used to fill in some grammatical functions. In deciding whether the linguistic items are DMs or not, the characteristics provided in Relevance Theory which are explained in section 2.1. were referred. An example of different uses of the same phrase is given below:

Using well as a DM:

(1) <B> travelling most of the time *well* .. it'd be difficult to like come to the socials at night <\B> (LOCNEC, nt="A" nr="E002")

(2) <B> (eh) *well* I don't know I guess it how it works </B> (LINDSEI-TR nt="TR" nr="TR040")

Using well as a non-DM:

(3) <B> ... and I applied to a few others like .. and erm .. I applied to: .. Bangor as *well* . in Wales <\B> (LOCNEC, nt="A" nr="E002")

(4) <B> ... in my opinion he is *well* educated and he . he she taught (eh) us (eh) making .. (eh) (eh) especially while preparing some material (eh) what (eh) can we focus on ... </B> (LINDSEI-TR nt="TR" nr="TR028").

As shown in (1) and (2), the word *well* carries the properties of a DM from a relevance-theoretic perspective as a) it does not add to the truth-condition of the proposition, b) omitting it from the utterance does not change the syntactic pattern of the utterance and c) it connects the former utterance to the following one especially in (1). In the utterances (3) and (4), on the contrary, *well* is used as a content word. Therefore, its omission from the utterance will affect the expression syntactically, and its presence is necessary for the truth condition of the utterance as it helps to hold a positive ground on the part of the speaker.

Both corpora were manually examined to distinguish DM use of the selected items from their non-DM use. This step was followed by the calculation of the frequency of *well*, *you know*, and *I mean* as DMs in both corpora. After identifying the frequency, the raw frequency of each DM was standardized as a frequency per 10,000 words. This procedure results in normalized frequencies (NF) which is considered to be necessary when two

corpora that are of not equal size are compared (Biber et al., 2007). To find out if a certain difference of frequency between NS and NNS speech is statistically significant or not, the chi-square test that determines the level of certainty in the observed difference being either statistically marked or a merely product of chance (McEnery, Xiao, & Tono, 2006) was employed.

These quantitative analyses were followed by qualitative observations about discourse functions of the selected DMs based on the relevant literature. These observations could provide crucial details on the functions of DMs in actual learner speech.

Results

Frequency and Statistical Analysis

To investigate the use of discourse markers in both NS and NNS speech, *well*, *you know*, and *I mean* as DMs were compared in terms of frequency in each corpora. Table 2 displays the type/token ratio, raw frequencies, log likelihood and chi-square values of each selected DM in both corpora.

Table 2. Comparison of the use of DMs in Turkish EFL learners' speech (LINDSEI-TR) and Native Speakers of English speech (LOCNEC)

Discourse Markers	LINDSEI-TR	LOCNEC	Log Likelihood	Chi-Square Value
***Well* as a DM**	19	560	-299.84	212.631***
***You Know* as a DM**	120	632	-92.15	81.296***
***I mean* as a DM**	28	444	-187.73	141.863***
Type/Token Ratio	5,40	4.73		

*p< .001

An overall look at Table 2 clearly demonstrates that compared with NSs, Turkish EFL learners underuse DMs in their speech. The most striking difference between two corpora is observed in the use of *well*, which is reported to be one of the commonest DMs in American and British conversation (Biber et al., 2007). While NSs use *well* as a DM 560 times, it appears only 19 times in Turkish learners' speech. One point to be clarified here is that the overall frequency of *well* in LINDSEI-TR was found to be 45 in the raw data. However, the manual analysis of the corpus to distinguish between *well* as a DM use and non-DM use reduced the frequency to 19. That corresponds to the fact that Turkish learners have the word *well* in their repository of vocabulary; however, they are not aware of its use as a DM.

As stated above, the corpora under investigation is of different size, which requires to calculate the normalized frequencies for each DM. Normalizing the frequencies makes the findings comparable with the other results provided in the relevant literature (O'Keeffe & Farr, 2003). The frequencies in this study have been normalized to tokens per 10.000 word. Figure 1 shows the results.

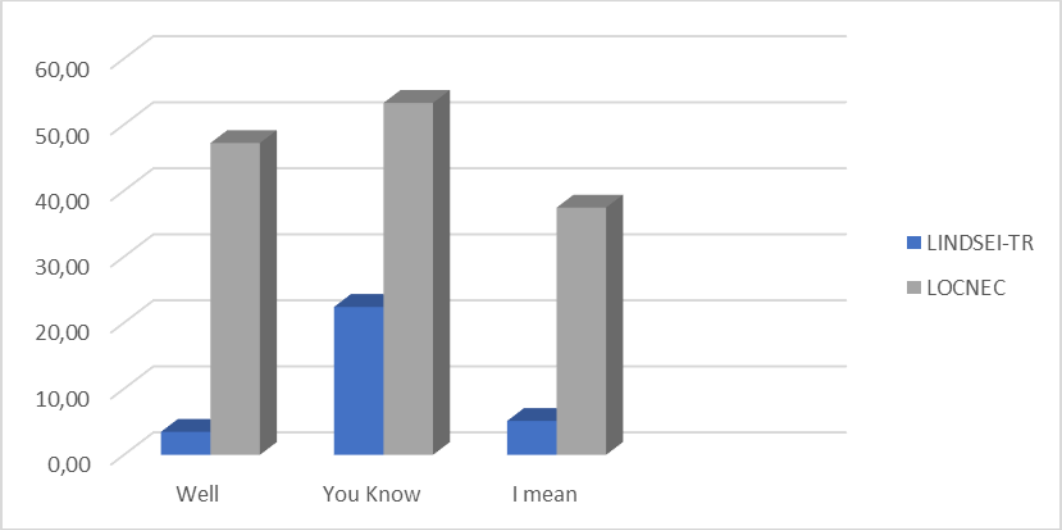

Figure 1. Normalized frequencies of selected DMs across NS and NNS speech.

From the normalized frequencies, it is self-evident that *you know* is the most frequent DM followed by *well* and *I mean* in native speakers' speech. This finding slightly contradicts with the results by Aijmer (2011) who employed the same comparable corpus (i.e. LOCNEC) to investigate the use of *well* in NNS speech. In her research, Aijmer (2011) reports the frequency of *well* as 54,4 per 10.000 word, concluding that *well* is the most frequent marker in LOCNEC followed by *you know* and *I mean* respectively. However, it has not been specified in Aijmer (2011) whether the uses of *well* as a DM and as a non-DM have been differentiated or not. As the present study distinguishes between the DM and non-DM uses of *well* using the framework provided by Relevance Theory, the frequency of *well* as a purely DM has been found to be different from the previous research, ranking it after *you know*.

The native speaker corpus has been used as a norm for interpreting the overuse/underuse phenomena on the part of NNSs. Figure 1, when interpreted in combination with the chi-square values in Table 2, explicitly shows that all the DMs of the focus of this study are underused by Turkish learners of English, and the difference between NS and Turkish learners in terms of the use of all DMs are statistically significant (i.e. well=212.631, p<.001; you know=81.296 p<.001; I mean=141.863 p<.001).

The pattern of overuse/underuse of *well*, *you know* and *I mean* has also been shown in the speech of learners of English from different L1 background: for Swedish learners' use of *I mean* and *you know* in Aas (2011); Swedish learners' use of *well* in Aijmer (2011) LINDSEI-SW; for French learners' use of *well, I mean* and *you know* in Gilquin (2008) LINDSEI-FR and for Japanese learners of English by Shimada[1] (2014)1 LINDSEI-JP. An inclusive look at all these studies is considered to provide interesting insights on DMs in NNS speech. Figure 2 illustrates the normal frequencies of *well, you know* and *I mean* in aforementioned learner groups.

[1] All these cited studies used the relevant components of LINDSEI and LOCNEC, which makes it applicable to discuss their findings in comparison with the findings of the present study.

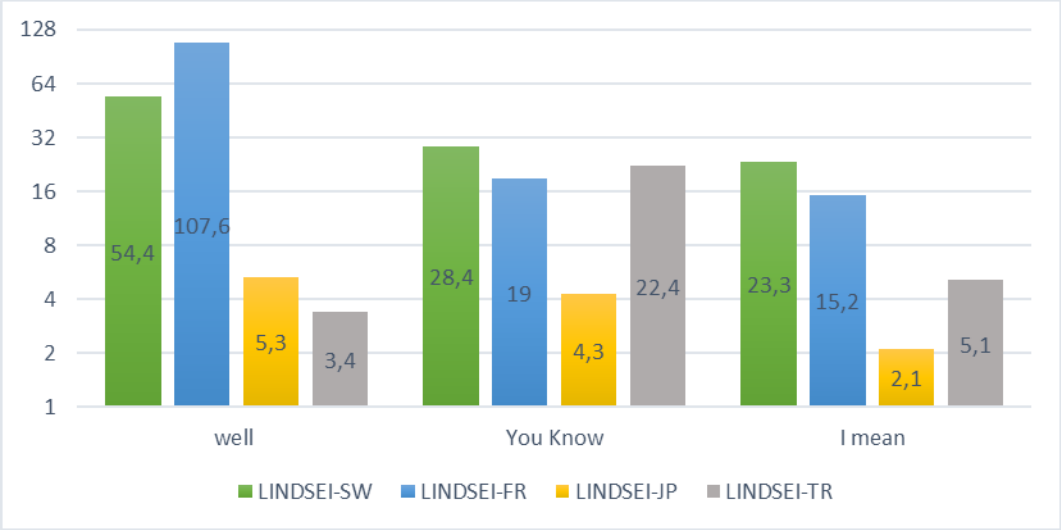

Figure 2. Frequency of *well*, *you know*, and *I mean* per 10,000 words in other sub-corpora of LINDSEI

As is seen in Figure 2, Turkish learners is the group that uses *well* the least in their conversational English. For the use of *you know* as a DM, Figure 2 shows that Turkish learners perform a similar tendency with the other learner groups, which could make a base for considering underuse of *you know* in conversation as a property of non-native speech.

Functional Analysis

Functions of DMs have been analyzed from many perspectives in previous research (see Müller, 2005 for a comprehensive overview); however, there is little agreement on the theoretical orientation or the type of categories explaining the multifunctional uses of DMs (Aijmer, 2011). This study dwells on the Relevance Theory for the analysis of the functions of DMs as it is believed to provide a general framework of communication based on cognitive principles covering all the functions through a combined explanation (Jucker, 1993).

For lack of space, the functional analysis of the DMs in this study is confined to *you know*, which is found to be the most frequent DM in Turkish learner speech. From a Relevance-Theoretic perspective, *you know* is categorized as an "addressee-centered presentation marker which relates the information to the presumed knowledge state of the addressee" (Jucker & Smith, 1998 p.174). Figure 3 presents the specific functions *you know* and their distribution in both NS and NNS corpora.

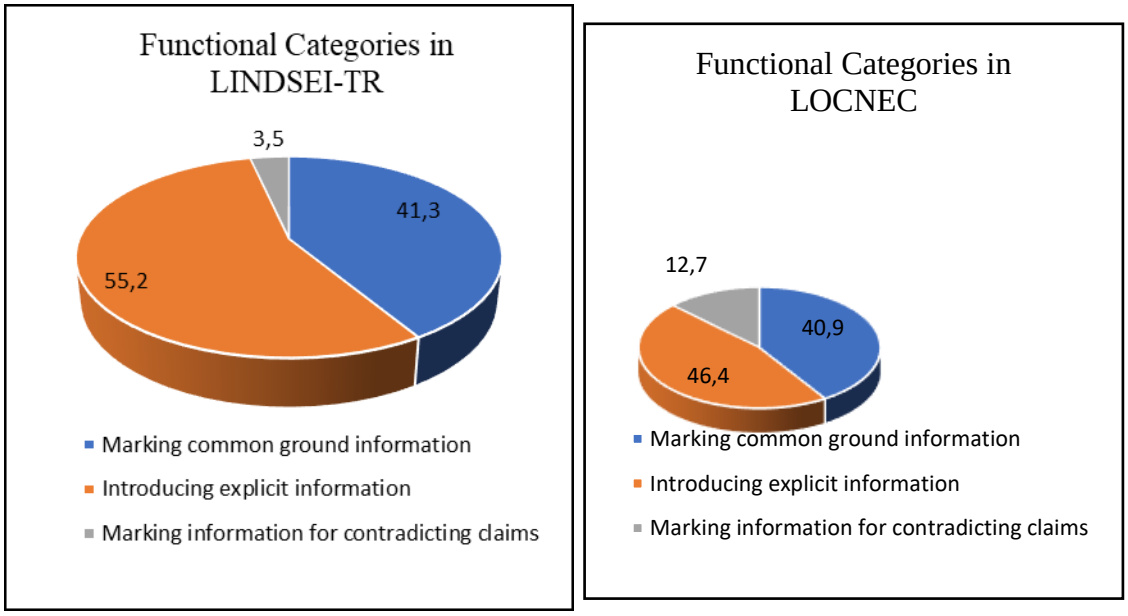

Figure 3. Functional categories of *you know* and their distribution across corpora

As seen in Figure 3, there is a similar pattern of distribution regarding the functions of *you know* in both corpora. Accordingly, using *you know* to mark the information that contradicts earlier claims by the addressee is the least frequent function used by both NSs and NNSs. One explanation for this occurrence could be related to the data collection method used in both corpora as the interviews "were not produced for real communicative purposes, but for classroom (corpus collection) purposes" (Gilquin & De Cock, 2011 p.157).

A similar distributional pattern between NNS and NS speech has been observed in the functional category of *you know* marking common ground information. This function of *you know* has been reported in previous research as well (Bu, 2013; House, 2009; Huang, 2011; Müller, 2005). An example of this assumed common ground is seen in (5) and (6).

(5) <B> and also the most depressing because everyone dies in the[i:] end so I thought oh this is gonna be fun <X> *you know what I mean* . so we went into Manchester from college <\B> (LOCNEC, nt="A" nr="E010").

(6) <B> (eh) generally (mm) this year I can't very much because we have a big exam *you know* for teacher </B> (LINDSEI, nt="TR" nr="TR018").

Finally, *you know* functioning as a DM to introduce explicit information occupies the largest place in both native speaker and Turkish learners' speech as exemplified in (7) and (8) respectively.

<B> .... I expected like you know . a big building with a great big sign saying Nuffield Theatre <\B> (LOCNEC, nt="A" nr="E010").

<B> .... university its it's not just like a high school it's of course we learned English we learned how to teach English but (em) . it's not an army school *you know* it's (eh) it's like we came here (eh) and then we should leave here in different person but in positive way </B> (LINDSEI, nt="TR" nr="TR018"). This finding is in line with the results by Hellermann and Vergun (2007) who found that *you know* is frequently used to mark more specific information being introduced into the conversation.

To summarize, findings show that Turkish learners rarely employ DMs in their speech while native speakers use them quite frequently. This underuse of DMs by Turkish learners leads them to sound non-native and to be less assertive in their communication, which in turn affect the hearer-speaker interaction (House, 2009). Comparison with the other learner groups imply that *well* and *I mean* show a specific use idiosyncratic to Turkish learners in terms of frequency whereas frequency of *you know* as a DM by Turkish learners forms the only common basis with other non-native speakers of English. Functional analysis which is confined to *you know* being the most frequent DM in Turkish learners' speech reveal that they use it appropriately in most instances with a similar distributional pattern to the native speakers' use.

Discussion

Based on the quantitative analysis of the data reflecting spoken English of NNSs and NSs, findings of this study have led to several claims regarding the use of DMs by EFL learners. First, findings have shown that *well* is not primarily used as a DM in the NNS speech, which is in line with the previous research (Huang, 2011). In other words, Turkish EFL learners have *well* in their repertoire, yet they are not aware of its function as a DM. L1 effect and instructional properties of the target setting are the two commonly referred factors in the literature to account for the underuse of DMs. Regarding the case of *well* as observed in this study, L1 effect does not seem to be a valid account. In most of the studies on discourse markers in Turkish, the particle *şey* is considered to be correspondence of *well* in English, and *şey* is reported to be among the most frequent DMs in Turkish (Yılmaz, 2004). Therefore, it would be unreasonable to refer to L1 effect to account for the underuse of *well* by Turkish learners. One explanation for this occurrence, then, could be the underrepresentation of *well* as a DM in instructional materials in EFL instructional setting for Turkish learners. This explanation related to properties of instructional setting could also be applicable for the underuse of *I mean* (its Turkish correspondence is considered to be "*yani"*, another commonest DM in Turkish language (Yılmaz, 2004) by Turkish learners compared with other learner groups. While Turkish learners very frequently employ "yani" in their conversation, they are seemingly not aware of its English correspondence as a DM. Second, underuse of *well* as a DM could also give interesting hints about the English proficiency level in LINDESI-TR. Previous research reported positive correlation of the use of *well* as a DM with the proficiency level of the learners (Fung & Carter, 2007; Huang, 2019). In other words, it is stated that the more proficient the learner is, the more frequent the DMs are in their speech. In LINDSEI-TR, the learners providing information to the corpus are considered to be "advanced" learners based on the external criteria (i.e. their being third and fourth graders at university).

However, the fact that they are not able to use DMs in their speech hints that either their attributed proficiency level should be re-considered for Turkish EFL context or pedagogic intervention is necessary even in advanced level.

Additionally, the findings of this study obtained through the functional analysis of Turkish EFL learners' speech bear notable insights about properties of their English. The underuse of *you know*, for example, by Turkish learners could be interpreted as a trait signalling their being less assertive in target language. Being a highly frequent DM in native speech, *you know* is found to function as a "shared knowledge indicator" pointing to the speaker's confidence in the existence of common information (House, 2009). Although Turkish learners use *you know* to mark common ground information, -as shown by the qualitative analysis- its being infrequent in the speech makes them sound non-native and unassertive in English. Similarly, the underuse of DM *I mean* by Turkish learners contributes to their foreign soundingness, suggesting that *I mean* is not conventionalized in Turkish learners' English to the extent in native speech.

Conclusion

Within the framework of CIA and combining insights from Relevance Theory and learner corpus research, this study aimed to uncover the use of DMs (*well, I mean, You know*) by Turkish EFL learners in comparison with that of native speakers. The findings summarized above bear significant implications.

As it is quite obvious from the findings of this study that DMs are highly frequently used in native speaker interaction, "especially *well, you know* and *I mean*, [t]hey should also be used by learners (Aijmer, 2011). Therefore, as noted by Fung and Carter (2007), Müller, (2005) and Romero-Trillo, (2002) knowledge of DMs in terms of variety and their functions in spoken English should be underlined in language teaching environment as lack of or inappropriate use of DMs could make the utterances vague, incoherent and misunderstood. DMs could be taught explicitly through integration of language activities such as problem-solving, cross-language comparison (Fung & Carter, 2007) or through the analysis of authentic materials including daily conversation of native speakers (Hellermann & Vergun, 2007). Language instruction based on the principles of Data Driven Learning could also make an efficient solution for teaching DMs by exemplifying actual language use since previous studies have shown effectiveness of this method (Şahin Kızıl & Kilimci, 2017; Şahin Kızıl & Savran, 2018; Vyatkina & Boulton, 2017). The awareness about DMs could also be raised implicitly through the use of DMs by teachers in classroom interaction (Hellermann & Vergun, 2007).

The findings of this study which is limited in scope suggest that there is more to explore on the use of DMs by NNSs. Future research could focus on DMs other than the ones analyzed here to have a full understating of DMs by NNSs. Additionally, comparing learners at different acquisition stages would also yield interesting results about how DMs are learnt. Finally, investigations of DMs in various communicative contexts by NNSs other than structured interviews may provide further insights about contextual use of DMs.

References

Aas, H. L. (2011). *Recurrent Word-Combinations in Spoken Learner English*. [Unpublished Master's Thesis] University of Oslo.

Aijmer, K. (2002). *English Discourse Particles: Evidence from a corpus*. Amsterdam & Philadelphia: John Benjamins Publishing Company.

Aijmer, K. (2004). Pragmatic Markers in Spoken Interlanguage. *Nordic Journal of English Studies*, *3*(1), 173–190.

Aijmer, K. (2011). Well I'm not sure I think…The use of well by non-native speakers. *International Journal of Corpus Linguistics*, *16*(2), 231–254.

Aijmer, K., & Stenström, A.-B. (Eds.). (2002). *Discourse patterns in spoken and written corpora*. Amsterdam & Philadelphia: John Benjamins Publishing Company.

Andersen, G. (2011). Corpus based pragmatics I: Qualitative studies. In W. Bublitz & N. R. Norrick (Eds.), *Foundations of Pragmatics* (pp. 587–627). Berlin: Walter de Gruyter.

Aşik, A., & Cephe, P. T. (2013). Discourse markers and spoken English: Nonnative use in the Turkish EFL setting. *English Language Teaching*, *6*(12), 144–155.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (2007). *Longman Grammar of Spoken and Written English*. England: Pearson Education Limited.

Blakemore, D. (1992). *Understanding Utterances: An introduction to pragmatics*. Oxford: Blackwell Publishing.

Blakemore, D. (2002). *Relevance and linguistic meaning: The semantics and pragmatics of discourse markers*. New York: Cambridge University Press.

Blakemore, D. (2004). Discourse markers. In L. R. Horn & G. Ward (Eds.), *The Handbook of Pragmatics* (pp. 221–240). Oxford: Blackwell Publishing.

Bu, J. (2013). A study of the acquisition of discourse markers by Chinese learners of English. *International Journal of English Studies*, *13*(1), 29–50.

Buysse, L. (2012). So as a multifunctional discourse marker in native and learner speech. *Journal of Pragmatics*, *44*(13), 1764–1782.

Downing, A. (2006). The English Pragmatic Marker surely and its Functional Counterparts in Spanish. In K. Aijmer & A. M. S. Vandenbergen (Eds.), *Pragmatic Markers in Contrast* (pp. 39–58). Oxford: Elsevier Inc.

Erman, B. (2001). Pragmatic markers revisited with a focus on you know in adult and adolescent talk. *Journal of Pragmatics*, *33*, 1337–1359.

Fraser, B. (1999). What are discourse markers ? *Journal of Pragmatics*, *31*, 931–952.

Fraser, B. (2006). On the Universality of Discourse Markers. In K. Aijmer & A. M. S. Vandenbergen (Eds.), *Pragmatic Markers in Contrast* (pp. 79–92). Oxford: Elsevier Inc.

Fung, L., & Carter, R. (2007). Discourse markers and spoken English: Native and learner use in pedagogic settings. *Applied Linguistics*, *28*(3), 410–439.

Gilquin, G. (2008). Hesitation markers among EFL learners: Pragmatic deficiency or difference? In J. Romero-Trillo (Ed.), *Pragmatics and Corpus Linguistics: A Mutualistic Entente* (pp. 119–149). Berlin, Heidelberg & New York: Mouton de Gruyter.

Gilquin, G. (2016). Discourse markers in L2 English: From classroom to naturalistic input. In O. Timofeeva, A.-C. Gardner, A. Honkapohja, & S. Chevalier (Eds.), *New Approaches to*

English Linguistics: Building Bridges (pp. 213–249). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Gilquin, G., & De Cock, S. (2011). Errors and disfluencies in spoken corpora: Setting the scene. *International Journal of Corpus Linguistics*, *16*(2), 141–172. *Discourse Markers in Learner Speech*

Granger, S. (2009). The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation. In K. Aijmer (Ed.), *Corpora and Language Teaching* (pp. 13–32). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Granger, S. (2015). The contribution of learner corpora to reference and instructional materials design. In G. Gilquin, S. Granger, & F. Meunier (Eds.), *The Cambridge Handbook of Learner Corpus Research* (pp. 485–510).

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.

House, J. (2009). Subjectivity in English as Lingua Franca discourse: The case of you know. *Intercultural Pragmatics*, *6*(2), 171–193.

Huang, L. F. (2011). *Discourse markers in spoken English: A corpus study of native speakers and Chinese non-native speakers*. The University of Brimingham.

Huang, L. F. (2013). Pedagogical implications of the corpus-based investigation of discourse markers. *Learner Corpus Studies in Asia and the World*, *1*, 227–254.

Huang, L. F. (2019). A Corpus-Based Exploration of the Discourse Marker Well in Spoken Interlanguage. *Language and Speech*, *62*(3), 570–593.

Jucker, A. H. (1993). The discourse marker well: A relevance-theoretical account. *Journal of Pragmatics*, *19*, 435–452.

Jucker, A. H., & Smith, S. W. (1998). And people just you know like 'wow': Discourse markers as negotiating strategies. In A. H. Jucker & Y. Ziv (Eds.), *Discourse Markers: Description and Theory* (pp. 171–204). Amsterdam: John Benjamins.

Kilimci, A. (2014). LINDSEI-TR: A New Spoken Corpus of Advanced Learners of English. *International Journal of Social Science and Education*, *4*(2), 401–410. Retrieved from

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-Based Language Studies An Advanced Resource Book*. London: Routledge.

Mukherjee, J., & Rohrbach, J.-M. (2006). Rethinking applied corpus linguistics from a language-pedagogical perspective: New departures in learner corpus research. In B. Ketteman & G. Marko (Eds.), *Planning, Glueing and Painting Corpora: Inside the Applied Corpus Linguist's Workshop* (pp. 205–232). Frankfurt: Peter Lang.

Müller, S. (2005). *Discourse markers in native and non-native English Discourse*. Amsterdam & Philadelphia: John Benjamins Publishing Company.

O'Keeffe, A., & Farr, F. (2003). Using language corpora in initial teacher education: Pedagogic issues and practical applications. *TESOL Quarterly*, *37*(3), 389–418.

Polat, B. (2011). Investigating acquisition of discourse markers through a developmental learner corpus. *Journal of Pragmatics*, *43*(15), 3745–3756.

Romero-Trillo, J. (2002). The pragmatic fossilization of discourse markers in non-native speakers of English. *Journal of Pragmatics*, *34*, 769–784.

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.

Şahin Kızıl, A., & Kilimci, A. (2017). Teaching Collocations through Web-Based Concordancing: A Corpus Informed Quasi-Experimental Study. In D. Tafazoli & M. Romero (Eds.), *Multiculturalism and Technology-Enhanced Language Learning* (pp. 142–156) Hershey: IGI Global.

Şahin Kızıl, A., & Savran, Z. (2018). The integration of corpus into EFL speaking instruction: A study of learner perceptions. *International Online Journal of Education and Teaching*, *5*(2), 376–389. *Aysel ŞAHİN KIZIL*

Scott, M. (2010). *WordSmith Tools Version 5.0.* (Vol. 5). Liverpool: Lexical Analysis Software.

Shimada, K. (2014). Contrastive interlanguage analysis of discourse markers Used by non-native and native English speakers. *JALT Journal*, *36*(1), 47–68.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition*. Oxford: Blackwell Publishing.

Terraschke, A. (2007). Use of general extenders by German non-native speakers of English. *IRAL*, *45*(2), 141–160.

Urgelles-Coll, M. (2010). *The Syntax and Semantics of Discourse Markers*. London & New York: Continuum.

Vyatkina, N., & Boulton, A. (2017). Corpora in language learning and teaching. *Language Learning & Technology*, *21*(3), 1–8.

Wang, L. F., & Zhu, W. H. (2005). A corpus-based study of the use of discourse markers in Chinese EFL learners' spoken English. *Foreign Language Research, 3*, 40–48.

Wei, M. (2011). Investigating the oral proficiency of English learners in China: A comparative study of the use of pragmatic markers. *Journal of Pragmatics*, *43*(14), 3455–3472.

Yılmaz, E. (2004). *A pragmatic analysis of Turkish discourse particles: Yani, işte and şey*. (Unpublished Doctoral Dissertation) Middle East Technical University.

Yong, W., Jingli, W., & Zhou, C. (2010). The Use of I think by Chinese EFL Learners: A Study Revisited. *Chinese Journal of Applied Linguistis*, *33*(1), 3–23.

Zhao, H. (2013). A study on the pragmatic fossilization of discourse markers among Chinese English learners. *Journal of Language Teaching and Research*, *4*(4), 707–714.