International Journal of Language Academy
ISSN:2342-0251
DOI Number: http://dx.doi.org/10.18033/ijla.4050
Research Article

Volume 6/5 December 2018
p. 210/229

ArticleHistory:
Received
21/10/2018
Received in revised form
21/10/2018
Accepted
07/12/2018
Available online
15/12/2018

A CORPUS-DRIVEN APPROACH TO I THINK AS A PRAGMATIC MARKER IN NATIVE AND NON-NATIVE DISCOURSE

Bir Edimbilimsel Belirteç Olarak Anadil ve Yabancı Dil Söyleminde 'I Think' Kullanımına Derlem Çıkışlı Yaklaşım[1]

Cem CAN[2]
İlyas YAKUT[3]

Abstract
Loosely attached to other syntactic elements in an utterance, pragmatic markers (PM) are one of the most frequented pragmatic elements that fulfill various functions. Among a great number of PMs encountered in daily conversations, I think is one of the most prevalent PMs in native speaker and non-native speaker discourse. Thus, the purpose of this study was to unravel and compare the overall uses and functions of I think among Turkish learners of English and native speakers of British English. Adopting corpus-driven approach to explore the use of I think in both groups, the uses of the marker were classified depending on their functions that were designated through integrated reading of the data by using Wordsmith Tools 6.0. According to the results, the functions of I think were categorized under the domains of shield, booster, structural, evaluative and interactional. Log likelihood statistics was administrated to figure out whether the difference across the groups was statistical in their use of I think with regard to frequency and functional variety. The results revealed that there was not a statistical difference between the groups regarding the overall uses of the marker. However, differences in the functional distributions of I think across the groups were observed.
Key Words: Corpus-driven approach, functional analysis, pragmatic marker, I think.

Öz
Sözce içerisindeki diğer sözdizimsel unsurlarla zayıfça ilişkilendirilen edimbilimsel belirteçler birçok işlevi yerine getiren edimbilimsel unsurlardan bir tanesidir. Günlük konuşmalarda karşılaşılan birçok edimbilimsel belirteçler arasında I think anadil ve yabancı dil söyleminde en yaygın olan belirteçlerden birisidir. Bu yüzden, bu çalışmanın amacı bir edimbilimsel belirteç olarak I think'in kullanımlarını ve işlevlerini ortaya çıkarmak ve İngilizce öğrenen Türk bireylerle anadili İngiltere

[1] This study is a part of the 2nd author's PhD thesis presented at Erciyes University Institute of Social Sciences
[2] Assoc. Prof. Cem CAN, Çukurova University, cemcan01@gmail.com
[3] Res. Asst. İlyas YAKUT, Zonguldak Bülent Ecevit University, yakutilyas@gmail.com

İngilizce'si olan bireyleri bu bağlamda karşılaştırmak olmuştur. I think'in her iki grupta da kullanımını araştırmak için derlem çıkışlı bir yaklaşım benimsenmiş ve Wordsmith Tools 6.0 yazılımı vasıtasıyla bütüncül okuma yöntemiyle belirlenen işlevler üzerinden bir sınıflandırmaya gidilmiştir. Elde edilen sonuçlar ışığında, I think'in işlevleri kaçınma, vurgulama, yapısal, değerlendirme ve etkileşimsel alanlarda sınıflandırılmıştır. Log likelihood testi kullanılarak gruplar arasında I think'in bir belirteç olarak kullanım sıklıklarında ve işlevsel çeşitliliklerinde bir fark olup olmadığı incelenmiştir. İstatistiksel test sonuçları belirtecin kullanım sıklıkları açısından gruplar arasında bir fark oluşturmadığını göstermiştir. Ancak, işlevsel dağılımlarda gruplar arasında farklılıklar gözlemlenmiştir.
Anahtar Kelimler: Derlem çıkışlı yaklaşım, işlevsel çözümleme, edimbilimsel belirteç, I think.

1. Introduction

Communication, either spoken or written, is a process which enables the interlocutors to transmit intended meaning with the help of structural, interactional and/or interpersonal properties of human language (Fetzer, 2011). Along with propositional content of the interaction, pragmatic properties of language systems serve to expedite the intended meaning to the addressee(s) even though they do not add to propositional meaning. For this reason, Östman (1981) remarks that "[l]anguage is a communicative, context-sensitive system, a socio-psychological instrument which is being used to communicate our 'underlying' intentions" (p. 3). To underscore the prominence of functional and pragmatic aspects of language systems in an interaction, Van Bogaert (2010) asserts that interlocutors could not decipher the intended meaning by only focusing on grammatical and structural domains of the message. For this reason, linguistic elements tied to functional and pragmatic facets of the discourse play important roles in the process of comprehending the intended meaning as these features carry the properties of roles and attitudes of the interlocutors related to conceptual meaning of the message.

Contrary to the importance of pragmatic resources in the course of encoding and decoding the explicitly and/or implicitly conveyed meaning, pragmatic and functional sides of communication were undervalued in language related studies as the grammatical properties such as syntax and morphology were more eminent in traditional grammar. However, as functional analysis of language has gained importance after the introduction of communicative competence by Hymes (1972), uses of pragmatically encoded resources with procedural features have gained importance. Thus, with the emergence of functional analysis, pragmatic elements in a language have been explored by considering the language as a whole unit in the light of co-textual (e.g. syntactic and morphologic elements) and contextual factors. Thanks to the shift from traditional approaches to functional analysis of language, the field of pragmatics has gained popularity since the field deals with linguistic possibilities affecting the quality of the interaction (Lakoff, 1993).

Interlocutors make use of pragmatic resources appropriately to enhance the quality of the interaction as these elements provide a smoother conversation between the addresser and addressee. Through benefiting from the functional properties of pragmatic elements, the intended meaning could be transmitted to the hearer(s) with ease as these elements convey the beliefs, feelings and attitudes of the interlocutors along with the propositional content conveyed thanks to the appropriate uses of grammatical resources. Among various pragmatic resources, pragmatic markers (PMs) play paramount roles in the course of spoken interactions. Thus, along with other pragmatic resources, PMs have vital roles in the process of establishing a camaraderie between the interlocutors and constructing a successful and smooth interaction.

PMs are ubiquitous in spoken communication settings and since their numbers vary from a dozen to five hundreds (e.g. *well, okay, yes/yeah, I mean, y'know* etc.) (Brinton, 1996), various definitions have been offered depending on the items taken into account, linguistic theories or approaches adopted, functions assigned to the markers and the beliefs of the scholars. Among various definitions suggested in the related literature (Östman, 1981; Levinson, 1983; Erman, 1986; Fraser, 1996), Bazzenella's (2006) definition seems to encompass a majority of features of a PM as she states that PMs are linguistic elements "which are useful in locating the utterance in an interpersonal and interactive dimension, in connecting and structuring phrasal, inter-phrasal, and extra-phrasal elements in discourse, and in marking some on-going cognitive processes and attitudes." (p. 456). Even though Bazzanella's definition is a comprehensive one, Rühlemann and Aijmer (2015) remark that research in PM is an evolving area for the last 40 years. Thus, new expressions and items are continuously included into the category of PMs resulting in reforming and expanding the definitions of PMs so as to cover the characteristics and functions of new PMs.

PMs used to be associated with disfluency as they were considered to be gambits (Keller, 1979) and fumbles (Edmonson, 1981). However, as functional analysis of language came into prominence, it was understood that these small items are loaded with various functions that are shaped according to the situational and social contexts. Hence, PMs are also considered to be sociolinguistic elements that are framed by the identities of the interlocutors (Aijmer, 2013). Furthermore, these elements are generally exploited in informal or semi-formal spoken interactions. Thus, PM in spoken interactions has long been a focus of interest in the literature since 1980s. As the uses of PMs and their functions are context-bound, PM related studies, especially the recent ones, recruit corpora of spoken interactions since they are "produced in natural communicative setting" (Gilquin & Gries, 2009: p. 7) and they represent "a particular linguistic variety/ register/ genre" (p. 6).

Regarding the PM related research field, some studies focused on a group of PMs (Schiffrin, 1987; Aijmer, 2002; Fung & Carter, 2007) whereas some others were specifically conducted to explore uses and functions of a small number of PMs (Östman, 1981; Erman, 2001; Buyyse, 2012; Mei, 2012; Aijmer, 2015). While the studies investigating the uses of all PMs in a discourse unit are inclined to explain frequencies and main features of PMs, research related to the uses of one or two PMs tries to probe the functions of specific markers through making in-depth analysis of the markers in concern.

As there is a wide range of PMs that are exploited in spoken interactions, a through exploration of all PMs seems improbable. Among various PMs, *I think* as a pragmatic parenthetical (Brinton, 2008) is one of the most frequented pragmatic expressions in spoken communication settings (Karkkainen, 2003). Even though this I+cognitive verb phrase does not display the characteristics of a typical PM in all occasions since it could be employed as a main clause of an utterance (Aijmer, 2015), a majority of the uses are considered to be a member of PM category (Brinton, 2008; Kaltenböck, 2009; 2010; Dehe & Wichmann, 2010; Karkkainen, 2010).

Since *I think* is accepted as a kind of PM, a large and growing body of literature has dealt with the marker in question. Some studies were specifically conducted on the possible functions of *I think* among native speakers of English and the classifications of the possible functions were presented (Aijmer, 1997; Karkkainen, 2003; 2010; Kaltenböck, 2009; 2010; Zhang, 2014). There is also substantial research on the use of *I think* in

different genres (Simon-Vandenbergen, 2000; Fetzer, 2011; 2014; Aijmer, 2015). *I think* as a PM has also been explored to compare uses of it in native speaker (NS) and non-native speaker (NNS) corpora (Aijmer, 2004; Fung & Carter, 2007; Huang, 2011; Liu, 2013). In addition, overall uses of the marker in concern among NSs and Turkish learners of English have also been investigated (Aşık, 2012; Aşık & Cephe, 2013; Zorluer Özer & Okan, 2018). However, previous studies conducted on the use of *I think* among Turkish learners of English did not focus on the functional analyses of it across the two groups since their purpose was to find out frequencies of all the markers appearing in Turkish and NS discourse. Hence, to the best of our knowledge, there is not a study investigating the functions of *I think* among Turkish learners of English and NSs of British English. For this reason, this present study concentrates on functional analysis of *I think* in NS and NNS corpora with regard to its usage rates and functional distributions. Thereof, this current study adopts a corpus-driven contrastive interlanguage analysis approach to reach its aforementioned purpose. Therefore, we aim to answer the following research questions:

1. Do native speakers of English and Turkish learners of English statistically differ regarding the overall frequencies of *I think*?
2. What pragmatic functions do native speakers of English and Turkish learners of English attribute to *I think* in spoken interaction?
3. Do native speakers of English and Turkish learners of English statistically differ with regard to the distributions of pragmatic functions of *I think*?

2. Literature Review

PMs "are grammatically optional but at the same time serve important pragmatic functions (and are, in a sense, pragmatically non-optional)" (Brinton, 2017: p. 8). Since PMs are indispensable part of the spoken conversations, omitting these items will make the speaker "dogmatic, impolite, boring [and] awkward to talk to" (Svartvik, 1980: p. 171) even though the other interlocutors could not pinpoint an error when PMs in spoken discourse are eliminated. Similar to other PMs, *I think* also fulfils various functions in spoken interactions. For this reason, the uses of this I+cognitive verb combination have been investigated in relation to its employment rates, functional dispersions, proposition positions of occurrences and uses in different genres. In addition, uses in NS and NNS discourse have been explored.

The focal point of some researchers was to unravel the possible functions of *I think* in specific discourse types. Aijmer (1997) explored the uses of *I think* in London Lund Corpus of Spoken English (LLC) and she found that the core meaning of the marker was 'cogitation'. According to her, the pragmatic functions of the marker in question emerge out of the core meaning of the marker. The functions of *I think* were divided into 'deliberative' and 'tentative' ones. While the deliberative functions indicated the certainty and reassurance of the speaker, the tentative ones were related to uncertainty and attenuation. Despite taking Aijmer (1997) as a baseline, Karkkainen (2003) revealed that the functions of *I think* could not be confined to Aijmer's binary classification of the functions of *I think*. Thus, she remarked that *I think* could have various new functions depending on the situational and social contextual needs. As she believed that new functions of *I think* could always emerge, in a later study, Karkkainen (2010) investigated the uses of *I think* in Santa Barbara Corpus of Spoken American English (SBCSAE). The results indicated that *I think* was used by NSs with various functions such as self-repair,

online-planning, commitment qualification (p. 214), other-repair (p. 218), indicating disagreement through softening the degree of imposition (p. 219).

Kaltenböck (2010) explored the uses of *I think* in IGE-GB by designating the core meaning of it as speaker cogitation. Depending on the contextual and co-textual variables, such as position, prosody and elements encompassing *I think* (p. 237), he suggested a functional taxonomy with four main functions which were shield, approximator, structural and booster functions and each function had various sub-functions. Even though he offered a functional taxonomy of *I think*, he also stated that "[t]he four categories have to be understood as a basic frame of reference which has to allow for a certain amount of overlap and fuzzy boundaries between the categories, thus representing a functional cline" (p. 260). Additionally, Zhang (2014) stated that the functions of *I think* are "fluid, overlapping, complementary and therefore co-exist" (p. 225). Despite this, she offered a functional taxonomy of *I think* through analyzing the data collected from institutional conversations. According to her, the functions of *I think* could be divided into five main categories: emphatic, evaluative, mitigating, tentative and discursive. Furthermore, she also indicated that the evaluative function could be labeled as the prototype function of *I think*.

In addition to functional categorization of *I think* as a PM, there are also studies aiming to explore uses of *I think* in different genres with different speech events. Simon-Vandenbergen (2000) compared the uses of *I think* in political and casual discourse types. According to her results, the initial position was the most frequently observed position in both groups. However, while politicians used *I think* as a booster, the participants in casual conversations used *I think* with belief or probability based opinions. Furthermore, *I think* as a PM was followed by utterances including obligation and necessity in political discourse. As for the uses of *I think* in monologues and interviews in political discourse, Fetzer (2014) found that the marker was more frequented in interactive communication settings when compared to uses in monologue based speeches. In addition to these studies, Aijmer (2015) compared the uses of *I think* in the genre of broadcast talks and discussions. The results indicated that impromptu speech types such as interviews and discussions included more *I think* when compared to monologues and pre-rehearsed speeches. Naturally, the functional dispersion in interactional communication settings was more diverse. Thus, she concluded that uses and functions of *I think* were greatly affected by the speech event.

The overall uses and/or functions of *I think* as a PM in NS and NNS data have also been examined in the related literature. Even though most of the PMs are found to be less frequent in NNS data, among various studies regarding the uses of *I think*, some researchers (Aijmer, 2004; Gablasova, Brezina, McEnery & Boyd, 2015) found that *I think* was employed in NS and NNS data with nearly equal frequencies. Furthermore, some others (Fung & Carter, 2007; Baumgarten & House, 2010; Huang, 2011; Liu, 2013) indicated that *I think* was outnumbered among NNSs when compared to the uses in NS data. However, some others revealed that NNSs confined their uses of *I think* to a limited number (Jabeen & Rai, 2011; Aşık & Cephe, 2013; Zorluer Özer & Okan, 2018). Among various studies related to uses of *I think* among NSs and NNSs, Aşık and Cephe (2013) and Zorluer Özer and Okan (2018) explored the uses of the marker among Turkish learners/speakers of English and NSs of English. Even though they analyzed the density of overall *I think* uses in both groups, they did not focus on the functions activated by NNSs and NSs of English.

3. Methodology

This current study aims to explore the overall uses of *I think* as a PM among Turkish learners of English and native speakers of British English. In addition, this study tries to explore the functions of *I think* emerging in spoken interactions of NSs and NNSs of English. Since the functions of PMs are formed and reformed depending on the context (Karkkainen, 2003), a corpus driven functional approach to the marker in question was adopted in this study.

Regarding the comparison of the results, Contrastive Interlanguage Analysis (CIA) composes the methodological framework of this study as it is predominantly used in corpus-driven studies. As postulated by Granger (1996), there are two approaches to CIA. The first approach enables the researchers to compare the similarities and differences between interlanguage systems of two groups coming from different first language backgrounds. The second approach, however, sanctions the comparison of the NSs and NNSs of a specific language. As stated by Gilquin (2001), "by comparing learner corpora with a control corpus of the TL, it is possible to pinpoint the features of nonnativeness of the interlanguage" (p. 99) and this could be realized through CIA.

3.1. Data

According to Granger (2008), authentic data plays paramount roles in second language acquisition research and it could be attained through corpora. Despite this fact, she also reveals that "learner language is influenced by a wide variety of linguistic, situational and psycholinguistic factors" (p. 262). Thus, corpora need to be compiled by setting pre-determined clear criteria. For this reason, the uses of *I think* in Turkish component of Louvain International Database of Spoken English Interlanguage (LINDSEI-TR) (Kilimci, 2014) and Louvain Corpus of Native English Conversation (LOCNEC) (De Cock, 2004). While LINDSEI-TR was collected from advanced level Turkish learners of English, LOCNEC is used as a reference corpus of LINDSEI compiled from young native speakers of British English. LINDESI-TR consists of 80813 words, while the learner turns constitute 63924 words of it. LOCNEC, on the other hand, accommodates 161725 words and 118553 were produced by interviewees.

3.2. Instruments

Wordsmith Tools 6.0 developed by Scott (2001) was used to calculate the occurrences of *I think* in both corpora. As indicated by Ghadessy, Henry and Roseberry (2001), "Wordsmith Tools provides almost instantaneous display of word frequency lists; concordances, which allow all the uses of a given word in its contexts; and lists of keywords, words that appear more often in a corpus than chance alone would dictate" (p. xix). Since Wordsmith enables the scholars to conduct vertical and horizontal reading, the overall uses of *I think* were found through vertical reading and the uses of *I think* as a PM and their functions in the context were designated through horizontal reading as the tool permits the researcher to see the occurrences in the context. For the referential statistics, the uses and functions of *I think* across the two corpora were compared through Log likelihood (LL) statistics which is known as a reliable alternative to Chi-Square test (Dunning, 1993).

3.3. Data Analysis

Corpus linguistics deals with the conventionalized language whereas pragmatics focus on non-conventionalized aspects of language uses. Thus, while corpus linguistics allows scholars to annotate the data automatically, pragmatics requires semi-automatic or

manual annotation as the context designates the meanings or functions of pragmatic elements (Rühlemann & Aijmer, 2015).

Initially, each set of corpus was uploaded to WordSmith 6.0 and the occurrences of *I think* were found through vertical reading which is available in concord interface. Then, the uses of *I think* by the interviewers and the cases that did not show the features of a PM were eliminated through horizontal reading of the data.

Following those steps, 10% of the occurrences of *I think* in LOCNEC were annotated with regard to their functions and proposition positions in an utterance. While designating the functions of *I think*, we benefited from the previous works of Aijmer (1997), Simon-Vandenbergen (2000), Kaltenböck (2009; 2010) and Zhang (2014). However, since this study adopted a corpus-driven approach to *I think* as a PM, the functional categorization of the data was not confined to the abovementioned scholars' classifications. After deciding on the functions of *I think*, they were grouped under the main categories of shield, booster, structural, evaluative and interactional. Thus, through a hypothesis-finding process, a new functional classification system emerged. As for the proposition position of *I think* in an utterance, the uses were labeled as initial, medial and final positions depending on the place of occurrence in an utterance. As indicated by Halliday (1994), each part of an utterance having a role in the transitivity of a clause is accepted as a proposition.

As the functional annotation of the data was a rather subjective process, 10% of the uses in LOCNEC were annotated by another rater depending on the functional categories we offered. According to Loewen & Plonsky (2016), inter-rater reliability test is "[a] method of dual coding of data to ensure that the coding categories or scores are being used in a consistent manner" (p. 93). The results of reliability test revealed that the raters were consistent in their pragmatic annotation of *I think* since the Cronbach's Alpha level was above .80 (p=.89).

Based on the process mentioned above, the overall frequencies of *I think* as a PM in both corpora were measured and the results were normalized to uses in every 10000 words since the word counts in each corpus were not equal. To measure whether usage rates of *I think* in both corpora were statistically different with regard to overall uses and functional dispersions, LL test was administrated.

4. Results and Discussion

4.1. An overview of I think in two corpora

Table 1 exhibits the overall distributions of *I think* in LINDSEI-TR and LOCNEC. As shown in raw frequency counts, *I think* occurred 193 times in NNS data while NS participants employed the marker 393 times in their spoken interactions. Even though raw frequency counts indicated that the uses in LOCNEC overrode the occurrences in LINDSEI-TR, this might be a misleading result since the corpus sizes were not equal. Thus, the raw frequencies were normalized to uses per 10000 words and the relative frequency counts indicated that the marker in concern was employed 30.19 and 33.15 times in every 10000 words in LINDSEI-TR and LOCNEC respectively.

Table 1. Overall distribution of *I think* in two corpora

	LINDSEI-TR	LOCNEC
Corpus size in words	63924	118553
I think (n)	193	393
n /10.000	30.19	33.15

According to the normalized frequency counts, it is evident that the NNSs of English underused *I think* in their informal conversations even though the discrepancy between the groups was rather small. In order to measure whether the discrepancy across the groups in relation to overall uses of *I think* was statistically significant, LL ratio test was administrated. Table 2 illustrates the LL ratio test result of overall uses of *I think* in LINDSEI-TR and LOCNEC.

Table 2. LL ratio of *I think* in two corpora

	LINDSEI-TR		**LOCNEC**		**LL Ratio (p<0.05)**	**ELL**
	(O1)	%1	(O2)	%2		
I think	193	0.30	393	0.33	-1.14	0.00000

O1 is observed frequency in Corpus 1
O2 is observed frequency in Corpus 2
%1 and %2 values show relative frequencies in the texts
+ indicates overuse in O1 relative to O2
- indicates underuse in O1 relative to O2

As shown in Table 2, *I think* was employed 0.30 and 0.33 times in every 100 words in LINDSEI-TR and LOCNEC successively. The LL ratio test result revealed that even though the marker was underused by NNSs against NSs, the difference between the groups was not statistical at $p<0.05$ level (LL=-1.14). This result is in line with Aijmer (2004) who compared the uses of various PMs as well as *I think* among Swedish learners of English and NSs of British English. Additionally, Baumgarten and House (2010) stated that *I think* as a PM was the most frequented I+cognitive verb combination among NNSs. Furthermore, Gablasova, Brezina, McEnery and Boyd (2015) revealed that *I think* constituted 42.5% of all PM uses in NNS data.

As for the uses of *I think* among Turkish learners of English, our result contradicts with the findings of Aşık and Cephe (2013) and Zorluer Özer and Okan (2018). Aşık and Cephe (2013) found that prospective teachers of English did not resort to that marker as frequently as the NSs did in their classroom presentations. Thus, they voiced that "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" (p. 151). In addition, Zorluer Özer and Okan (2018) found that *I think* was underused by Turkish speakers of English against NSs and the difference across the groups was statistical according to LL ratio test result.

4.2. Functions of I think in two corpora

The functions of *I think* are tied to its core meaning. For this reason, the functions of *I think* are polysemic in nature (Kaltenböck, 2010; Zhang, 2014). Even though the functions of the marker in question were designated through a bottom-up approach in this study, functional categorizations offered by various scholars (Aijmer, 1997; Simon-Vandenbergen, 2000, Kaltenböck, 2009; 2010 and Zhang, 2014) were used as the baseline for establishing functional categorization system of *I think* in this study.

The functions of *I think* found in this current research were divided into the categories of shield, booster, structural, evaluative and interactional. The parent categories of shield, booster and structural were previously introduced by Kaltenböck (2009) while *I think* as an evaluative device was postulated by Zhang (2014). However, the category of interactional was added by the researchers as there were cases in which

interpersonal/interactive functions of *I think* were more prominent. In Table 3, the functional categories of *I think* with various sub-functions are presented.

Table 3. Pragmatic functions of *I think*

Shield	**Booster**	**Structural**	**Evaluative**	**Interactional**
Approximation Afterthought Expressing uncertainty	Elaborating/expanding by adding more information Expressing contradiction Exemplification Explicitness Expressing agreement Indicating facts Necessity Reason Reiterating previous claim Expressing authority of knowledge Expressing certainty Rejecting request	Topic shift End of turn Online planning Repair Result Quotative	Draw conclusion Indicating personal attitude Sharing experience	Appeal for understanding

The parent category of shield refers to uses of *I think* as a tentative and uncertainty device. The booster category, however, encompasses the functions of *I think* revealing the deliberative uses of the marker. Thus, booster functions strengthen the illocutionary force of the message. While shield refers to lack of speaker's commitment, full-commitment of the speaker is highlighted through boosting function. Therefore, Preisler (1986) remarks that *I think* as a booster exhibits "careful deliberation, objectivity and so, perhaps, authority" (p. 107) of the speaker. Structural functions of *I think* indicate the uses of the marker with textual functions. Kaltenböck (2010) states that *I think* is used with structural functions "to segment text, structure information flow, signal and alleviate disfluencies, or for the sequential organization of talk" (p. 254). The category of evaluative encompasses the uses of *I think* with "a propositional attitude, not emphasizing nor mitigating, but simply assessing or judging the truth and aptness of what is said" (Zhang, 2014: p. 227). *I think* as an evaluative device is used to express mere opinion of the speaker, and the marker generally clusters with evaluative adjectives (Aijmer, 2015). Finally, the category of interactional refers to the employment of the marker which signals the reciprocal contribution of the interlocutors to the establishment of a smooth interaction.

4.3. Comparison of functions of I think in two corpora

Table 4 below presents an overview of the category-based functional distribution of *I think* in LINDSEI-TR and LOCNEC. The raw frequency counts and relative frequencies in every 10000 words are exhibited in Table 4.

Table 4. Cross-tabulation of main categories of functions of *I think* in two corpora

	LINDSEI-TR		**LOCNEC**	
	n	n/10000	n	n/10000
Shield	25	3.9	85	7.2
Booster	50	7.8	121	10.2
Structural	57	8.9	104	8.8
Evaluative	54	8.4	76	6.4
Interactional	7	1.1	7	0.6
TOTAL	193	30.2	393	33.1

n: raw frequency of *I think*
n/10.000: frequency of *I think* per 10.000 words

As shown in Table 4, the functions within shield and booster categories were employed more frequently by NSs against NNSs. According to relative frequency counts, *I think* as a shield was employed 3.9 times in LINDSEI-TR and 7.2 times in LOCNEC per 10000 words. *I think* as a booster was employed 7.8 and 10.2 times among NNSs and NSs respectively. Regarding the uses of *I think* within the categories of structural, evaluative and interactional, Turkish learners of English overused the marker with varying differences. The category of structural was the most frequented one in LINDSEI-TR while *I think* as a booster had the highest number of occurrences in LOCNEC. The overall results obtained from the data contradict with Aijmer (1997) and Liu's (2013) findings. According to Aijmer (1997), 85% of *I think* occurrences had the function of shield while only 15% represented deliberative uses of *I think*. Similarly, Liu (2013) reported that shield function was more favored in her data when compared to other functional domains. However, as stated by Aijmer (2015), functions of a PM, *I think* in this case, are bound to context of speech including genre and formality level of the conversation. In addition, the attitudes, identities and relationship between the interlocutors could affect the functional preferences of PMs.

According to the overall results obtained from the data, the prototype proposition position of *I think* as a shield was the final position in LINDSEI-TR while it was initial position in LOCNEC. Contrary to that difference in the typical position of shield function, majority of the booster functions were observed in initial position of the proposition in both corpora. Similarly, initial position was the default position of *I think* within the domain of structural category. The prototypical proposition position of evaluative functions was also initial position in the two groups. However, *I think* as an evaluative device was observed in final position with a close frequency to the uses in initial position in LINDSEI-TR. Regarding the typical position of *I think* within interactional domain, it was found that the most favored position in LINDSEI-TR was final position while initial position was the most frequented one in LOCNEC.

Since the usage rates in the two corpora show differences in relation to parent category based functional distributions of *I think*, LL ratio test was administrated. Table 5 illustrates the LL ratio test results of main functions of *I think* in two corpora.

Table 5. LL ratio of main categories of functions of *I think* in two corpora

	LINDSEI-TR		**LOCNEC**		**LL Ratio (p< 0.05)**	**ELL**
	(O1)	%1	(O2)	%2		
Shield	25	0.04	85	0.07	-7.85	0.00001
Booster	50	0.08	121	0.10	-2.59	0.00000
Structural	57	0.09	104	0.09	+0.01	0.00000
Evaluative	54	0.08	76	0.06	+2.36	0.00000
Interactional	7	0.01	7	0.01	+1.31	0.00000

O1 is observed frequency in Corpus 1
O2 is observed frequency in Corpus 2
%1 and %2 values show relative frequencies in the texts
+ indicates overuse in O1 relative to O2
- indicates underuse in O1 relative to O2

According to the results obtained from LL analyses, *I think* as a shield was underused by NNSs against NSs and the difference across the groups was statistical since the LL value was -7.85 with 0.00001 effect size. As for the uses of *I think* within booster category, even though it was evident that employment rates in NNS data were lower when compared to NS data, the difference across the groups was not statistical at $p<0.05$ level (LL=-2.59)

since the discrepancy level was smaller than 3.84. Regarding the comparison of overall uses of *I think* within the structural category, the difference across the two groups did not yield a significant difference since the LL ratio revealed an almost equal usage rates between the two corpora (LL=+0.01). *I think* in evaluative and interactional categories were overused by NNS, yet the differences across the groups did not indicate statistical results since the LL values were +2.36 and +1.31 in the uses of evaluative and interactional functions respectively.

Each functional parent category of *I think* was represented by various sub-functions in the data. Thus, the employment rates of each sub-function in each corpus were calculated and the results were compared through LL statistics. Table 6 exhibits the distributions of individual functions of *I think* in LINDSEI-TR and LOCNEC.

Table 6. Functional distribution of *I think* in two corpora

	FUNCTION	LINDSEI-TR		LOCNEC	
		N	n/10000	n	n/10000
Shield	1. Approximation	8	1.25	16	1.35
	2. Afterthought	17	2.66	30	2.53
	3. Expressing uncertainty	0	0.00	39	3.29
Booster	4. Expressing contradiction	3	0.47	19	1.60
	5. Elaborating/expanding by adding more information	5	0.78	19	1.60
	6. Exemplification	1	0.16	2	0.17
	7. Explicitness	4	0.63	8	0.67
	8. Expressing Agreement	2	0.31	5	0.42
	9. Indicating facts	10	1.56	4	0.34
	10. Necessity	3	0.47	4	0.34
	11. Reason	5	0.78	14	1.18
	12. Reiterating previous claim	3	0.47	2	0.17
	13. Expressing authority of knowledge	4	0.63	22	1.86
	14. Expressing certainty	9	1.41	22	1.86
	15. Rejecting request	1	0.16	0	0.00
Structural	16. Topic shift	0	0.00	1	0.08
	17. End of turn	6	0.94	1	0.08
	18. Online planning	39	6.10	92	7.76
	19. Repair	8	1.25	8	0.67
	20. Result	0	0.00	2	0.17
	21. Quotative	4	0.63	0	0.00
Evaluative	22. Draw conclusion	11	1.72	9	0.76
	23. Indicating personal attitude	37	5.79	57	4.81
	24. Sharing Experience	6	0.94	10	0.84
Interactional	25. Appeal for understanding	7	1.10	7	0.59
TOTAL		193	30.19	393	33.15

n: raw frequency of *I think*
n /10.000: frequency of *I think* per 10.000 words

As shown in Table 6, afterthought (n=17, n/10000=2.66), which is used in final position of a proposition to lower the assertiveness of the preceding message, was the most

frequented sub-function of shield category among NNSs while expressing uncertainty (n=39, n/10000=3.29) had the highest frequency counts among NS participants. Interestingly, the function of expressing uncertainty, mainly used in initial and medial positions, was never employed by NNSs. Examples 1 and 2 extracted from LINDSEI-TR embody the uses of *I think* in afterthought function. Scripts in example 3 and 4 exemplifies the uses of *I think* with expressing uncertainty function.

Example 1

<A> so are you happy with it with this <overlap /> experience </A>

<B> <overlap /> sometimes sometimes *I think* about this because the world (eh) has been has changed a lot and (em) people aren't good <laughs> for most time . and sometimes it's bad for me but I say the truth (eh) every time ***I think*** </B>

Extracted from LINDSEI-TR 19

Example 2

<B> (eh) they have (eh) Turk= (eh) they have (eh) mother tongue mother (eh) la= tongue language but (eh) we should . (eh) teach English as a second language but we should learn the same <overlap /> </B>

<A> <overlap /> (uhu) </A>

<B> (eh) a= (eh) as a similar way like <overlap /> Turkish </B>

<A> <overlap /> (uhu) </A>

<B> ***I think*** </B>

Extracted from LINDSEI-TR 2

Example 3

<B> so it was it was quite relevant to us .. er I'm not sure ***I think*** it's a recent translation as well it's not a kind of a dated translation <\B>

Extracted from LOCNEC 9

Example 4

<A> yeah well it depends perhaps it's going to be the same in Italy <\A>

<B> well possibly I . ***I think*** I'll probably try and teach .. adults or <\B>

Extracted from LOCNEC 14

Regarding the most frequently employed individual function within booster category, indicating facts (n=10, n/10000=1.56) was the most preferred one in LINDSEI-TR while the functions of expressing authority of knowledge and expressing certainty (n=22, n/10000=1.86) had the highest frequencies in LOCNEC. Example 5 illustrates the use of *I think* with indicating facts function in picture narration. Example 6 represents the expressing authority of knowledge function and example 7 embodies the function of expressing certainty. *I think* in example 6 was employed to explain what 'yellow things' were used for. In example 7, *I think* was accompanied with *actually* known as a certainty marker.

Example 5

<B> she has a painter to paint herself .. and (eh) in in the[i:] second . picture she looks at this (eh) painting and ***I think*** she doesn't like it . (eh) that much and she ask<?> the[i:] she ask<?> the[i:] painter to paint herself (eh) more . as more beautiful than she really is and: the painter does what she wants . and she in the last photo she probably shows off showing the beautiful (erm) impressive painting of . herself to: her friends </B>

Extracted from LINDSEI-TR 12

Example 6

<A> like yesterday I saw that sometimes they put some yellow things to block [the the wheels <\A>

<B> [yeah . clamps .. clamps I've I've not had that done thank god but ***I think*** that's mainly for people who erm who aren't .. supposed to park on campus at all because there are some parking spaces for visitors <\B>

Extracted from LOCNEC 18

Example 7

<B> ***I think*** that's a pretty interesting <begin_laughter> country actually <end_laughter> <\B>

Extracted from LOCNEC 26

The most frequented function of *I think* within structural category was online planning in both corpora as it emerged 39 (n/10000=6.10) and 92 (n/10000= 7.76) times in LINDSEI-TR and LOCNEC respectively. Furthermore, that function was the most frequented one among the whole functional list suggested in this study. This finding is in agreement with Huang (2011), as she indicated that *I think* with online planning function was the most prevalent one among Chinese learners of English.

The comparison of relative frequency counts revealed that the online planning function was preferred more frequently by NSs. As shown in the scripts below, *I think* with online planning function was employed when the speaker was not mentally ready to resume the talk. For this reason, *I think* with that function was accompanied by pauses and verbal repetitions.

Example 8

<B> then Al= Alex thought Alex thought that ***I think*** he (erm) hurt her friend's brea= (eh) hu= hurt (eh) break .. her friend's heart </B>

Extracted from LINDSEI-TR 41

Example 9

<A> cos it's cheap do you think it's cheaper than in town <\A>

<B> erm well ***I think*** the[i:] advantage of living on campus is that er .. cos you save on travel expenses <\B>

Extracted from LOCNEC 47

Regarding the most frequented sub-function of *I think* under evaluative category, indicating personal attitude had the second highest hit in both corpora since that function was activated 37 (n/10000= 5.79)times by NNSs while NSs employed it 57 (n/10000=4.81) times. According to the relative frequency counts, *I think* with indicating personal attitude was overused by NNSs since a 0.98 difference in every 10000 words was found across the groups. *I think* with that function generally clustered with evaluative adjectives such as *good, wonderful, bad* etc. Examples 10 and 11 extracted from both corpora show the uses of *I think* with that function in LINDSEI-TR and LOCNEC

Example 10

<A> <overlap /> what did you do that made you feel happy (uhu) </A>

<B> to be in my (eh) with my friends it was very very good thing *I think* I like my friends and like (eh) to good time with my friends . (eh) for example festivals ***I think*** it was very good </B>

Extracted from LINDSEI-TR 10

Example 11

<A> [well Clint Eastwood <\A>

<B> Clint ***I think*** Clint Eastwood is very lovely in it I mean [he's sort of quite <\B>

Extracted from LOCNEC 14

As for the uses of *I think* within interactional category, the individual function of appeal for understanding represented that category in both groups. The function was activated 7 times in each corpora, yet their relative frequency counts were 1.10 and 0.59 in LINDSEI-TR and LOCNEC consecutively as the word counts in each corpus were not equal. Examples 12 and 13 disclose the actual uses of *I think* with appeal for understanding function in NNS and NS corpora.

Example 148

<B> (em) *I think* she wants to be a beautiful girl </B>

<A> (mhm) </A>

<B> ***I think*** that thing should be </B>

Extracted from LINDSEI-TR 27

Example 149

<B> and people who have it .. <X> [*I think* ***I think*** <\B>

<A> [they have to develop it <\A>

<B> yeah er <X> *I think* they should be glorified actually <\B>

Extracted from LONCEC 4

As exhibited in the scripts above, the speakers were not able to resume the talk due to not being able to recall the information, they asked for help from the other interlocutors to be able to establish a smooth interaction.

The descriptive statistics and the examples shared above have given some insights about the functional preferences of Turkish learners of English and NSs of British English. However, the LL ratio test was applied to unravel whether the usage rates of each individual function across the groups were statistically different. The LL ratio test results of functions of *I think* in the two corpora are presented in Table 7.

Table 7. LL ratio of functions of *I think* in two corpora

Function	**LINDSEI- TR**		**LOCNEC**			
	(O1)	%1	(O2)	%2	**LL Ratio**	**ELL**
Approximation	8	0.01	16	0.01	-0.03	0.00000
Afterthought	17	0.03	30	0.03	+0.03	0.00000
Expressing contradiction	3	0.00	19	0.02	-5.16	0.00001
Elaborating/expanding by adding more info.	5	0.01	19	0.02	-2.31	0.00001
Exemplification	1	0.00	2	0.00	-0.00	0.00000
Explicitness	4	0.01	8	0.01	-0.02	0.00000
Expressing agreement	2	0.00	5	0.00	-0.13	0.00000
Indicating facts	10	0.02	4	0.00	+7.68	0.00003
Necessity	3	0.00	4	0.00	+0.18	0.00000
Reason	5	0.01	14	0.01	-0.66	0.00000
Reiterating previous claim	3	0.00	2	0.00	+1.29	0.00001
Expressing authority of knowledge	4	0.01	22	0.02	-5.04	0.00001
Expressing certainty	9	0.01	22	0.02	-0.51	0.00000
End of turn	6	0.01	1	0.00	+7.71	0.00005
Online planning	39	0.06	92	0.08	-1.63	0.00000
Repair	8	0.01	8	0.01	+1.50	0.00000
Draw conclusion	11	0.02	9	0.01	+3.31	0.00001
Indicating personal attitude	37	0.06	57	0.05	+0.76	0.00000
Sharing experience	6	0.01	10	0.01	+0.04	0.00000
Appeal for understanding	7	0.01	7	0.01	+1.31	0.00000

O1 is observed frequency in Corpus 1
O2 is observed frequency in Corpus 2
%1 and %2 values show relative frequencies in the texts
+ indicates overuse in O1 relative to O2
- indicates underuse in O1 relative to O2

As shown in Table 7, the largest discrepancy across the groups was found in the function of end of turn since the LL value was +7.71 with 0.00005 revealing an overuse of the function by NNSs. Additionally, the function of indicating facts was also overused in LINDSEI-TR as the LL ratio test result revealed a significant difference between the groups at $p<0.05$ level (LL=+7.68) with 0.00003 effect size. Despite the statistical differences due to the overuse of the marker by NNSs, the function of expressing contradiction was underused by NNSs and the LL value indicated a significant difference between the groups at $p<0.05$ (LL=-5.16) with 0.00001 effect size. Furthermore, the difference across the groups in relation to uses of *I think* with expressing authority of knowledge function yielded a statistical result at $p<0.05$ level (LL=-5.04). The function of drawing conclusion did not yield a significant difference between the groups. However,

participants in LINDSEI-TR inclined to overuse *I think* with that function as the LL value was +3.31 indicating a close value to the critical value of 3.84.

Other functions were employed more frequently by either NNSs or NSs. However, the differences across LINDSEI-TR and LOCNEC were not significant since the LL ratios did not reach the critical value of 3.84. Interestingly, the function of exemplification was the only function that was used almost equally by each group since the LL value was -0.00 indicating that the groups used *I think* with exemplification function with similar frequencies.

5. Conclusion

This study aimed at comparing the uses of *I think* as a PM in spoken interactions of Turkish learners of English and NSs of British English. Being corpus-driven in nature, the paper explored the occurrences of *I think* with regard to overall uses and their functions formed in the social and situational contexts. The first research question was concerned about whether the overall usage rates of *I think* among NNSs and NSs differed significantly. According to the results, NNSs underused *I think* as a PM against NSs, yet the difference across the groups was not statistical. The second research question was formed to find out the functions of *I think* emerging in the context at stake. The results regarding the second research question revealed that *I think* was employed with 25 different functions. While 22 types were recruited by Turkish learners of English, 23 types of functions were exploited by NSs of British English. This study did not adopt any of the previously suggested functional taxonomies of *I think* since new functions of PMs, *I think* in this case, could always emerge depending on the discourse type. However, we benefited from the findings and suggestions of various previous studies related to functions of *I think* as pragmatics based studies are cumulative. Since this study was set out to find out the functions of *I think* through a bottom-up approach, a new functional categorization of the marker at hand was suggested. Depending on the features of each function, they were grouped under the categories of shield, booster, structural, evaluative and interactional. As for the third research question, the results indicated that structural category was the most prominent functional domain among NNSs while the category of booster was the most favored functional domain among NSs. The categories of shield and booster were underused by NNSs against NSs while *I think* within structural, evaluative and interactional domains were overused by NNSs when compared to the uses in NS data. However, the level of difference across the groups regarding the usage rates of *I think* within each functional domain was significant only in the use of shield category. Usage rates of sub-functions of *I think* in both corpora revealed that online planning was the most prevalent function in both corpora. Indicating personal attitude was the second most frequented function in NS and NNS data. Contrarily, afterthought was the third most frequently preferred function among NNSs whereas expressing uncertainty was in the third place in NS data. The largest discrepancy across the groups was observed in the usage rates of end of turn function and the function of indicating facts followed it. Interestingly, these two functions were overused by Turkish learners of English against NSs.

Broadly speaking, the findings of this study emphasize the vital role of *I think* as a PM in spoken discourse. Even though PMs are expected to be used less frequently by NNSs against NSs (Müller, 2005; Huang, 2011; Mei, 2012), the PM of *I think* seems to be an exception since its usages rates in NNS data are close to the frequency counts in NS data. This result is in line with a number of previous studies (Aijmer, 2004; Baumgarten & House, 2010). Even though overall frequency counts of *I think* in the two corpora are close

to each other, the functional distributions of the marker indicate that the preferences of NSs and NNSs do not resemble each other. For this reason, the overuse of *I think* in NNS data is explained by Wu, Wang, and Cai (2010) as "delay, habit, inadequate language proficiency, pragmatic overgeneralization, and probably situational anxiety" (p. 20).

Since functional distributions of *I think* in both corpora show differences, learners need to be trained regarding the pragmatics aspect of language and language use. As revealed by Blum-Kulka and Sheffer (1993), development of pragmatic competence in language acquisition/ learning is the most challenging phase of language acquisition process. Thus, the teachers of English need to provide activities aiming to improve pragmatic competence of the learners (Kasper & Schmidt, 1996). For this reason, learners need to receive pragmatic input from their language instructors which could be implemented into curricula through implicit or explicit teaching of pragmatic elements.

In addition to teaching pragmatic elements, PMs in particular, uses of these elements in different discourse types should be tracked and instructors need to provide feedback to the learners to diminish the possible effects of fossilizations in terms of quantity and functional distributions (Romero-Trillo, 2002). When learners acquire the appropriate uses of PMs in different discourse types, they could interpret or re-interpret the polysemic nature of these elements and they could use them in different communication settings with ease. House (2013) remarks that when learners have the pragmatic fluency known as "the ability to master smooth continuity in ongoing talk" (p. 65), markers could be employed to steer the talk and to shape or re-shape the content depending on their own attitudes and beliefs.

References

Aijmer, K. (1997). *I think* – An English modal particle. In T. Swan & O. J. Westvik (Eds.), *Modality in Germanic languages: Historical and comparative perspectives* (pp. 1-47). Berlin: Mouton de Gruyter.

Aijmer, K. (2002). *English discourse particles: Evidence from a corpus.* Amsterdam: John Benjamins.

Aijmer, K. (2004). Pragmatic markers in spoken interlanguage. *Nordic Journal of English Studies, 3*(1), 173-190.

Aijmer, K. (2013). *Understanding pragmatic markers: A variational pragmatic approach.* Edinburgh: Edinburgh University Press.

Aijmer, K. (2015). Pragmatic markers. In K. Aijmer & C. Rühlemann (Eds.), *Corpus pragmatics: A handbook* (pp. 195-218). Cambridge: Cambridge University Press.

Aşık, A. (2012). *Discourse markers and spoken English: Nonnative use in the Turkish setting.* (Unpublished doctoral dissertation). Gazi University, Ankara.

Aşık, A., & Cephe, T. P. (2013). Discourse markers and spoken English: Nonnative use in the Turkish EFL setting. *English Language Teaching, 6*(12), 144-155. http://dx.doi.org/10.5539/elt.v6n12p144

Baumgarten, N., & House, J. (2010). *I think* and I don't know in English as lingua franca and native English discourse. *Journal of Pragmatics 42*(5), 1184-1200. https://doi.org/10.1016/j.pragma.2009.09.018

Bazzanella, C. (2006). Discourse markers in Italian: Towards a 'compositional' meaning. In K. Fischer (Ed.), *Approaches to discourse particles* (pp. 449-464). Amsterdam: Elsevier.

Blum-Kulka, S., & Sheffer, H., (1993). The metapragmatic discourse of American-Israel families at dinner. In G. Kasper, & S. Blum-Kulka (Eds.), *Interlanguage pragmatics* (pp.196-223). Oxford: Oxford University Press.

Brinton, L. J. (1996). *Pragmatic markers in English.* New York, NY: Mouton de Gruyter.
Brinton, L. J. (2008). *The comment clause in English: Syntactic origins and pragmatic development.* New York, NY: Cambridge University Press.
Brinton, L. J. (2017). *The evolution of pragmatic markers in English: Pathways of change.* Cambridge: Cambridge University Press.
Buyyse, L. (2012). So as a multifunctional discourse marker in native and learner speech. *Journal of Pragmatics, 44*(13), 1764-1782. https://doi.org/10.1016/j.pragma.2012.08.012
De Cock, S. (2004). Preferred sequences of words in NS and NNS speech. *Belgian Journal of English Language and Literatures (BELL)*, 225–246.
Dehe, N., & Wichmann, A. (2010). The multifunctionality of epistemic parentheticals in discourse: Prosodic cues to the semantic-pragmatic boundary. *Function of Language, 17*(1), 1-28. https://dx.doi.org/10.1075/fol.17.1.01deh
Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics, 19*(1): 61-74. Retrieved from http://aclweb.org/anthology/J93-1003
Edmonson, W. (1981). *Spoken discourse: A model for analysis.* New York, NY: Longman.
Erman, B. (1986). Some pragmatic expressions in English conversation. In G. Tottie & I. Bäcklund (Eds.), *English in speech and writing. A symposium* (pp. 131-147). Stockholm: Almqvist & Wiksell.
Erman, B. (2001). Pragmatic markers revisited with a focus on you know in adult and adolescent talk. *Journal of Pragmatics, 33*(9), 1337-1359. https://doi.org/10.1016/S0378-2166(00)00066-7
Fetzer, A. (2011). *I think* this is I mean perhaps this is too erm too tough a view of the world but I often think... Redundancy as a contextualization device. *Language Sciences, 33*(2), 255-267. https://doi.org/10.1016/j.langsci.2010.10.003
Fetzer, A. (2014). *I think*, I mean and I believe in political discourse: Collocates, functions and distribution. *Functions of Language, 21*(1), 67-91. https://doi.org/10.1075/fol.21.1.05fet
Fraser, B. (1996). Pragmatic markers. *International Pragmatics Association, 6*(2), 167-190. Retrived from https://journals.linguisticsociety.org
Fung, L., & Carter, R. (2007a). Discourse markers and spoken English native and learner use in pedagogic settings. *Applied Linguistics, 28*(3), 410-439. https://doi.org/10.1093/applin/amm030
Gablasova D., Brezina, V., McEnery, T., & Boyd, E. (2015). Epistemic stance in spoken L2 English: The effect of task and speaker style. *Applied Linguistics, 38*(5), 1-26. doi: 10.1093/applin/amv055
Ghadessy, M., Henry, A., & Roseberry, R. L. (2001). Introduction. In M. Ghadessy, A. Henry & R. L. Roseberry (Eds.), *Small corpus studies and ELT* (p. xvii-xxii). Philadelphia: John Benjamin Publishing.
Gilquin G., & Gries, S. Th. (2009). Corpora and experimental methods: A state-of-the-art review. *Corpus Linguistics and Linguistic Theory, 5*(1), 1-26. doi: 10.1515/CLLT.2009.001
Gilquin, G. (2001). The integrated contrastive model: Spicing up your data. *International Journal for Contrastive Linguistics, 3*(1), 95-123. doi: 10.1075/lic.3.1.05gil
Granger, S. (1996). From CA to CIA and back: An integrated contrastive approach to computerized bilingual and learner corpora. In K. Aijmer, B. Altenberg & M. Johansson (Eds.), *Languages in contrast. Text-based cross-linguistic studies* (pp. 37-51). Lund: Lund University Press.

Granger, S. (2008). Learner corpora. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics: An international handbook* (pp. 259–275). Berlin: Walter de Gruyter.

House, J. (2013). Developing pragmatic competence in English as a lingua franca: Using discourse markers to express (inter)subjectivity and connectivity. *Journal of Pragmatics, 59*(A), 57-67. https://doi.org/10.1016/j.pragma.2013.03.001

Huang, L. F. (2011). *Discourse markers in spoken English a corpus study of native speakers and Chinese non-native speakers* (Unpublished doctoral dissertation). The University of Birmingham, Birmingham.

Hymes, D. H. (1972). On Communicative Competence. In Pride, J. B., & Holmes, J. (Eds.), *Sociolinguistics: Selected readings* (pp. 269-293). Baltimore, MD: Penguin.

Jabeen, F., & Rai, A. (2011). A corpus based study of discourse markers in British and Pakistani speech. *International Journal of Language Studies, 5*(4), 69-86. Retrieved from https://www.researchgate.net/publication/260042226

Kaltenböck, G. (2009). Initial *I think* : Main or comment clause? *Discourse and Interaction, (2)*1, 49-70. Retrieved from https://journals.muni.cz

Kaltenböck, G. (2010). Pragmatic functions of parenthetical *I think*. In G. Kaltenböck, W. Mihatsch & S. Schneider (Eds.), *New approaches to hedging* (pp. 237-266). Bingley: Emerald Group.

Karkkainen, E. (2003). Stance-taking as an interactive activity: The case of *I think*. In E. Karkkainen (Ed.), *Epistemic stance in English conversation: A description of its interactional functions, with a focus on I think* (pp. 105-182). Amsterdam: John Benjamins.

Karkkainen, E. (2010). Position and scope of epistemic phrases in planned and unplanned American English. In G. Kaltenböck, W. Mihatsch & S. Schneider (Eds.), *New approaches to hedging* (pp. 203-236). Bingley: Emerald Group.

Kasper, G., & Schmidt, R. (1996). Developmental issues in interlanguage pragmatics. *Studies in Second Language Acquisition, 18*(2), 149-169. https://doi.org/10.1017/S0272263100014868

Keller, E. (1979). Gambits: Conversational strategy signals. *Journal of Pragmatics, 3*(3-4), 219-238. https://doi.org/10.1016/0378-2166(79)90032-8

Kilimci, A. (2014). LINDSEI-TR: A new spoken corpus of advanced learners of English. *International J. Soc. Sci. & Education, 4*(2), 401-410. Retrieved from http://ijsse.com/sites/default/files/issues/2013/v4i2/Paper-10.pdf

Lakoff, R. T. (1993). Lewis Caroll: Subversive pragmatics. *Pragmatics 3*(4), 367-385. doi: 10.1075/prag.3.4.01lak

Levinson, S. (1983). *Pragmatics*. Cambridge: Cambridge University Press.

Liu, B. (2013). Effect of first language on the use of English discourse markers by L1 Chinese speakers of English. *Journal of Pragmatics, 45*(1), 149-172. https://doi.org/10.1016/j.pragma.2012.11.002

Loewen, S., & Plonsky, L. (2016). *A an A-Z of applied linguistics research methods*. New York, NY: Palgrave Macmillan.

Mei, W. (2012). *A corpus-based comparative study of pragmatic markers: I mean and you know in native and non-native conversation* (Unpublished doctoral dissertation). University of Liverpool, Liverpool.

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

Östman, J. O. (1981). *You know: A discourse-functional study.* Amsterdam: John Benjamins.

Preisler, B. (1986). *Linguistic sex roles in conversation*: *Social variation in the expression of tentativeness in English*. Berlin: Mouton de Gruyter.

Romero-Trillo, J. (2002). The pragmatic fossilization of discourse markers in non-native speakers of English. *Journal of Pragmatics, 34*(6), 769-784. https://doi.org/10.1016/S0378-2166(02)00022-X

Rühlemann, C., & Aijmer, K. (2015). Corpus pragmatics: Laying the foundations. In K. Aijmer & C. Rühlemann (Eds.), *Corpus pragmatics: A handbook* (pp. 1-26). Cambridge: Cambridge University Press.

Schiffrin, D. (1987). *Discourse markers.* Cambridge: Cambridge University Press.

Scott, M. (2001). Comparing corpora and identifying key words, collocations, and frequency distribution through the Wordsmith tools suite of computer programs. In M. Ghadessy, A. Henry & R. L. Roseberry (Eds.), *Small corpus studies and ELT* (pp. 47-67). Philadelphia, PA: John Benjamin Publishing.

Simon-Vandenbergen, A. M. (2000). The functions of *I think* in political discourse. *International Journal of Applied Linguistics, 10*(1), 41-63. https://doi.org/10.1111/j.1473-4192.2000.tb00139.x

Svartvik, J. (1980). Well in conversation. In S. Greenbaum, G. Leech & J. Svartvik (Eds.), *Studies in English linguistics for Randolph Quirk* (pp. 167-177). London: Longman.

Van Bogaert, J. (2010). A constructional taxonomy of *I think* and related expressions: Accounting for the variability of complement-taking mental predicates. *English Language and Linguistics, 14*(3), 399-427. doi:10.1017/S1360674310000134

Wu, Y., Wang, J., & Cai, Z. (2010). The use of '*I think*' by Chinese EFL learners: A study revisited. *Chinese Journal of Applied Linguistics 33*(1), 3–24. Retrieved from http://www.celea.org.cn/teic/89/10042201.pdf

Zhang, G. (2014). The elasticity of *I think*: Stretching its pragmatic functions. *Intercultural Pragmatics, 11*(2), 225-257. doi: 10.1515/ip-2014-0010

Zorluer Özer, H., & Okan, Z. (2018). Discourse markers in EFL classrooms: A corpus-driven research. *Journal of Language and Linguistics Studies, 14*(1), 20-66. Retrieved from http://www.jlls.org/index.php/jlls/article/view/759