Contents

Overuse of Discourse Markers in Turkish English as a Foreign Language (EFL) Learners' Writings:
The Case of *'I Think'* and *'in My Opinion'*

Ihsan Unaldi

University of Gaziantep, Faculty of Education, Gaziantep, 27310, Turkey
E-mail: ihsanunaldi@gmail.com

ABSTRACT This study tries to dwell on overuse of two discourse markers *I think* and *in my opinion* in Turkish EFL learners' written productions. The data was collected from 161 Turkish EFL learners and a corpus of 58.046 words was compiled and it was compared with a native English corpus of 54.285. The focus of the comparison was the frequency of the phrases *I think* and *in my opinion*. A raw frequency calculation of the phrases revealed that the Turkish EFL learners actually used a considerable amount of them when compared with the written productions of native speakers of English. In the inferential analysis process, the variances of these phrases in the two corpora were calculated and since the calculations yielded statistically significant differences, a non-parametric test, Mann Whitney U-test, was employed. The results validated the obvious difference of the phrases in terms of frequency, which means that there is a plethora of the phrases of *I think* and *in my opinion* in Turkish EFL learners' written productions.

INTRODUCTION

Written or spoken productions of learners of English as a foreign language (EFL) have been analyzed from different aspects with different concerns. Throughout years, data gleaned from EFL learners from many different L1 backgrounds have been subject to numerous quantitative analyses. These analyses focused on issues such as perspectives on grammar (Biber and Reppen 1998; Meunier 2002), error analysis (Dagneaux et al. 1998; Flowerdew 1998; Granger 1999; Flowerdew 2000; Abe and Tono 2005), chunks and phraseology (Granger 1998b; De Cock 2000), pragmatic developments (Flowerdew 1998; Belz and Vyatkina 2005; Callies 2009;) discourse (Aarts and Granger 1998; Mulak 2000; Aijmer 2001; Pulcini and Furiassi 2004; Gilquin 2008) and even on very specific concerns like punctuation marks (Celik and Elkatmis 2013). When the registers are taken into account, it is not surprising that written collections of L2 productions outnumber those of spoken ones (O'Keeffe et al. 2007).

Address for correspondence:
Dr. Ihsan Unaldi
University of Gaziantep,
Faculty of Education (27310),
Gaziantep, Turkey
Telephone: +90 (342) 360 12 00 / 3795
E-mail: ihsanunaldi@gmail.com

Common Features of EFL Written Productions

In a well-known meta-analysis, Silva (1993) compared L1 essays with L2 written productions collected from EFL learners coming from different language backgrounds such as Arabic, Chinese, Japanese and Spanish. L2 writing appeared to be distinct from and less effective than L1 writing. Moreover, L2 writings appeared to have certain organization issues. Similarly, Hinkel (2001) compared essays of native speakers of English with writings collected from speakers of Chinese, Japanese, Korean, Vietnamese, and Indonesian. Frequency rates of overt exemplification markers in the texts such as *(as) an example, for example, for instance, in (my/our/his/her/their) example, like, mainly, namely, such as ..., that is (to say)* were calculated and analyzed via non-parametric statistical techniques. The results showed that the non-native group employed far more example markers (conjunctions), first person pronouns, and past tense verbs in their academic texts. Again, in another study Hinkel (2002) analyzed 68 lexical, syntactic and rhetorical features of L2 text. The related corpus included texts written by advanced learners of English from six different L1 backgrounds: Arabic, Chinese, Indonesian, Japanese, Korean and Vietnamese. The results of the study indicate that L2 writers have a severely limited lexical and syntactic repertoire. This led the learners to produce simplistic texts which are rooted

in conversational discourse in English language. The results revealed a big gap between L1 and L2 texts in terms of basic academic writing.

Features of EFL written productions have been discussed extensively, and the differences between L1 and L2 writing was summarized in terms of micro and macro features (Hinkel 2005). Macro features refer to the global aspects of texts such as cohesion, coherence and organization of ideas. Textual features that have the function of marking discourse organization fall under the category of micro features. In terms of micro features, compared to texts written by native speakers of English, L2 texts naturally exhibit less lexical variety, fewer idioms, shorter sentences, more repetitions, fewer passive constructions and more personal pronouns (Hinkel 2011).

Discourse Markers

It could be argued that, since there has been a paradigm shift from form-focused language instruction to a communicative-focused one, second language learning studies have challenged the *building block* method by making use of the developments in discourse analysis (Celce-Murcia and Olshtain 2005). In linguistics terms, the term discourse refers to the macro-level aspects of language. Macro-level is actually what happens beyond sentence level, which means that in the process of communication, either spoken or written, language is not a set of small separate fragments but rather a whole controlling these small fragments. This whole is generally referred to as discourse. Below is a more elaborate definition.

... an instance of spoken or written language that has describable internal relationships of form and meaning that relate coherently to an external communicative function or purpose and a given audience/interlocutor. Furthermore, the external function or purpose can only be determined if one takes into account the context and participants (i.e., all the relevant situational, social, and cultural factors) in which the piece of discourse occurs. (Celce-Murcia and Olshtain 2000: 4)

A discourse marker, on the other hand, is a word or phrase that does not change the meaning of the sentence, and has an almost empty meaning (Moder and Martinovic-Zic 2004). Some examples of discourse markers are *firstly, how-ever, so, in other words, in summary, actually* and *I mean* (Parrott 2000). Some of these discourse markers are mostly used in managing conversations. These include *actually, anyway, by the way, I mean, OK, now, right, so well, yes, you know* and *you see*. This feature of spoken register is maintained naturally during the course of communication and, the language used in online social networks set aside, it is unusual to see these markers in written productions of native speakers of English. The interesting issue at this point is that the transfer of spoken discourse units into written discourse seems to be one of the universal micro features of EFL writing (Hinkel 2011).

In a corpus-driven study, Trillo (2002) focuses on the discourse markers in native and non-native spoken registers. In the study, he dwells on *pragmatic fossilization* which he defines as the phenomenon by which a non-native speaker systematically uses certain forms inappropriately at the pragmatic level of communication. After comparing four groups including children and adults (two native, tow non-native), he suggests that the development for the grammatical and the pragmatic aspects of language in L2 show different rates. Furthermore, since non-native speakers lack the pragmatic resources that the native speakers enjoy, pragmatic markers go through a process of fossilization both in quality and diversity and this situation is likely to cause, in terms of pragmatics, unfitting linguistic elements used in communication.

Spoken Register of English Language

By making use of the data gathered from the Cambridge and Nottingham Corpus of Discourse in English (CANCODE), O'Keeffe et al. (2007) analyzed common two-word chunks such as *you know*, *really* and *I think* and came to the conclusion that these chunks actually "occur with greater frequency than some everyday single words". The following figure, which makes a comparison among everyday single words and two-word chunks, makes the point clear (O'Keeffe et al. 2007: 69).

As is illustrated in the Figure 1, the two-word chunk *I think* is the third most common phrase in CANCODE preceded by *you know* and *really*. The insight readily available in this situation is that chunks are actually as crucial as single most frequent words in communication. In addi-
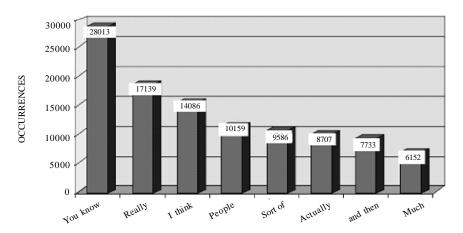
Fig. 1. Two-word chunks and common single words

tion, we can assume that EFL learners are exposed to these chunks as much as single words. Among these high-frequency chunks and single words in spoken register, hedging could be a topic of concern. Hedging, a term coined by Lakoff (1972), generally refers to an effort to mitigate the directness of an utterance and actually is an important part of spoken discourse. By some, it is regarded as a pragmatic marker (Carter and McCarthy 2006) and helps the speakers in avoiding to sound blunt and assertive. For example, if we consider the phrase *I think* in the above figure, there seems to be a considerable amount of difference in terms of pragmatics between the utterances '*I think you should stop smoking.*' and '*Stop smoking.*', which is actually in line with the *less direct is politer* rule (Sadock 2006).

The Overuse Issue in EFL Written Production

Discourse markers in EFL production have been studied from various aspects. Some words or phrases have been reported to be overused by EFL learners. For example**,** Tanko (2004), in order to analyze Hungarian EFL learners' writings, built an original corpus consisting of 93 argumentative essays written in an examination environment. The texts consisted of 500 words in average. The essays were compared with a reference native corpus, and the results of the analysis showed a comparative overuse of adverbial connectors in their essays. The explanation for this was that the Hungarian language does not require the overt marking of relations between linguistic units of the text. Another attempt to analyze connectives in EFL learners' was Altenberg and Tapper's (1998) study. They investigated how advanced Swedish EFL learners use connectives in argumentative essays in comparison with American university students' usages. The data was collected from the International Corpus of Learner English (ICLE): the Swedish sub-corpus and the control corpus of American university student essays. The aim of their study was to examine the use of three types of connectives: adverbial conjuncts (for example, *therefore, in particular*); certain style and content disjuncts (for example, *actually, indeed*); and some lexical discourse markers (for example, *result, compare*). The results implied that advanced Swedish EFL learners tended to overuse adverbial connectives and more types of connectives compared to the native group. So, there was not only a quantitative difference between the two groups in terms of connectives but also a difference in variance was present. Similarly, Schleppegrell (1996) analyzed ESL writers use strategies for the conjunction *because*. The participants in this study were students who were mainly Asian immigrants and had been living in US for different periods. The analysis of their essays revealed that there was a significant overuse of the conjunction *because* in the learners' essays; non-native essays included twice as many instances of *because* compared to the native essays. Another important point is that there was an obvious parallelism between uses of *because* clauses in spoken English and ESL essays, which was interpreted as an indica-

tion of how ESL writers draw on spoken registers inappropriately in constructing their academic essays.

The Overuse of 'I think'

As was mentioned before, native speakers of English tend to use the phrase *I think* in spoken register mostly. After analyzing London Lund Corpus of Spoken English (LLC) Aijmer (1997) proposed that the recurrent phrase *I think* has evolved into a discourse marker or a modal particle. Such discourse markers, she goes on to explain, are pragmaticalized as they tend to involve the attitude of the speaker to the hearer. It's usage without the clause marker *that* is actually much more frequent among the native speakers. She also mentions that *I think* has a structural flexibility which provides it with a variety of positions in an utterance, which means that it could occur in the front, mid or end position. Again, this interpretation is based on the spoken register of English. In addition to this, native speakers tend to use *I think* with *that* with a ratio of 7 % and the rest of the occurrences (97 %) appear without a *that clause* (Thompson and Mulac 1991). The reason for this could be related with the issue of omission of *that*, which makes it possible for this phrase to be used without a clause in the end position. However, in order to make deductions about the functions of *I think* in a context, we need to know its most frequent collocations. To serve this purpose, the spoken version of the British National Corpus (BNC), which is a corpus involving data collected from native speakers of English, was used. Table 1 provides relevant information.

Potential collocations of *I think* in spoken section of the BNC are presented in Table 1. These collocations of *I think* are calculated by taking into account its 2287 incidences in 1 million words, which comes to a percentage of 0.23 %. It is clear from the table that the insight attributing a pragmatic functions to this phrase (Aijmer 1997; De Cock et al. 1998) seems reasonable as its strongest collocations are spoken discourse markers or fillers (er, erm, well, mm etc…).

In previous studies concerning EFL learners from different backgrounds, an overuse of *I think (that)* as a universal phraseology have been reported (Aijmer 2001; Ishikawa 2009; and Yong 2010), and according to Granger (1998b) there is a tendency among EFL learners from

Table 1: Potential collocations of 'I think' in the BNC spoken corpus

N	Word	*f*
1)	think	241
2)	Yeah	238
3)	er	233
4)	erm	208
5)	know	137
6)	Well	137
7)	Yes	112
8)	Oh	108
9)	got	105
10)	all	98
11)	Mm	92
12)	well	89
13)	just	82
14)	mean	77
15)	very	77
16)	Erm	75
17)	really	73
18)	can	71
19)	yeah	69
20)	like	68
21)	get	64
22)	go	60
23)	people	60
24)	probably	58
25)	quite	57
26)	time	56
27)	right	54
28)	gonna	53
29)	now	52
30)	said	52

different L1 backgrounds to make use of active discourse frames more than passive ones. The following examples of active frames used by EFL learners illustrate this point (Granger 1998b).

- we/one/you can/cannot/may/could/might say that: *75 occurrences (vs 4 in NS corpus)*
- I think that: *72 occurrences (vs 3 in NS corpus)*
- we/one can/could/should/may/must notice that: *16 occurrences (vs no occurrences in NS corpus)*
- we/one may/should/must not forget that: *13 occurrences (vs no occurrences in NS corpus)*

Among the active discourse frames mentioned above, the phrase *I think* appears to have been used by the non-native speakers of English with a ratio of 72 to 3, which is obviously excessive. The reason for this overuse is claimed to be the indication of learners' lack of the appropriate repertoire to introduce arguments.

By taking into account the related literature mentioned so far, this study tries to deal with the following research questions.

1. Is there a quantitative difference of the phrase *I think* between the texts written by native speakers of English and Turkish EFL learners?
2. What are the potential collocations of the phrase *I think* in texts written by Turkish EFL learners?
3. Is there a quantitative difference of the phrase *in my opinion* between the texts written by native speakers of English and Turkish EFL learners?

METHODOLOGY

The Participants

The learner data was collected from university freshman engineering students at a state university in Turkey. In order to meet the requirements for a learner corpus (Granger 2003), proficiency levels of these EFL learners were determined using a valid and reliable placement test (Allen 1992) and their English proficiency levels varied from A2 (elementary) to B2 (upper-intermediate). To keep homogeneity, elementary level learners were removed from the study and a total of 161 EFL intermediate and upper-intermediate learners composed the learner data. As for the native data, the native corpus created by Granger's (1998a) LOCNESS (Louvain Corpus of Native English Essays) was used. This dataset was gathered from the essays of native speakers of English and includes 300,000 tokens. In order to obtain a quantitative balance between the learner and the native corpora, one of the subsets of LOCNESS was selected. The rationale for this selection can be seen in Table 2.

Table 2: Comparison of L1 and L2 corpora

	N (Subjects)	N (Words)	Essay type	Prompt
Learner	161	58.046	Argumentative	Exam/Timed
Native	46	54.285	Argumentative	Untimed

It's clear from the table that although the total numbers of the subjects in the groups are very different, the total numbers of the words produced by the two groups are quite similar. What we have is actually 58,046 words produced by 161 Turkish EFL learners and 54,285 words produced by 46 native speakers of English. The essays are all argumentative on topics like euthanasia, controversy in the classroom, capital punishment, money and school systems. On the other hand, the essay prompts are different; the learners produced their essays in a strictly controlled exam environment, whereas the native essays are untimed and written outside the classroom. The exams whereby the texts were collected lasted 50 minutes; dictionaries weren't allowed and the learners were asked to write essays of about 250 words.

Software and Statistics

The data gathered as explained above was analyzed by using the software package Word-Smith tools, version 6.0 (Scott 2012). As the first step, the essays in the corresponding datasets were extracted into separate files so as to make statistical computations possible. In other words, since each piece of writing needs to be assigned a value in terms of the target words and phrases, each essay was processed individually. With the rationale explained before, queries concerning two phrases, *I think* and *in my opinion*, were carried out. While doing so, in order to include every incidence of the phrase *I think,* the related query was performed as *I*think*. Consequently, incidences such as *I don't think*, *I never think* or *I sometimes think* were also taken into account in the analysis process. The results were transferred into SPSS (version 21), another software package for statistical computations.

Normally, when the sets of corpora to be compared are of different sizes, raw and normalized frequencies are analyzed. That is to say, if the total number of words in a corpus is so different as to distort the statistical calculations, a normalization process is required. For example, it would be mistake to compare a set of corpus made up of 40.000 words with a corpus containing 100.000 words because the raw frequency calculations would not reflect the real proportions of lexical items, in which case relevant interpretations would not be valid. However, as can be clearly seen in Table 1, the native and learner corpus used in the current study are so similar in terms of the total number of words used that normalization is not required.

RESULTS

The results of the statistical analyses mentioned before and their interpretations are presented in this section. First of all, descriptive

statistics and the related Mann U-test results of the phrase *I think* are given. Its different positions in learner texts and potential collocations with their word classes are analyzed next. In addition, the phrase *in my opinion* is compared with the native corpora in terms raw frequency. To begin with, descriptive statistics concerning the frequency of *I think* is presented in Table 3.

Table 3: Descriptive results for 'I think'

Group	N	f	Means	sd
Learner	161	104	.65	.97
Native	46	9	.20	.58

As is clear from Table 3, in the learner corpus 161 EFL learners used the phrase *I think* 104 times with an average of .65 and a standard deviation of .97, whereas their counterparts used it 9 times with an average of .20 and a standard deviation of .58. From these results, one can conclude that the learner group outnumbered the native one about the use of this phrase. In order to validate this insight, an inferential statistical calculation is to be applied. Since there are two independent groups a t-test is required, but the results of Levene's test of variance indicate that the group variances are significantly different (p<.01). Therefore, a non-parametric equivalent of t-test, Mann Whitney U-test, was employed (Field 2009:540). The results are presented in Table 4.

Table 4: Mann Whitney U-test Results for 'I think'

Group	N	Mean rank	Sum of ranks	U	p
Learner	161	110.93	17859.50	2587.50	.000
Native	46	79.75	3668.50		

Table 4 exhibits the results of the Mann Whitney U-test results for the phrase *I think*. Learner groups mean rank appears to be 110.93 and the native group scores 79.75 in the same calculation. The results indicate that there is a statistically significant difference between the two groups (U=2587, p<.01). It means that Turkish EFL learners tend to make use of the phrase *I think* significantly more than the native speakers of English. As was mentioned previously, the place of *I think* is quite flexible in spoken register. Table 5 provides a descriptive compari-

Table 5: Positional distribution of 'I think' in learner corpus

Position	Learner f	Learner %	Native f	Native %
Front	45	43.27	3	33.33
Mid	59	56.73	6	66.66
End	0	0	0	0
Total	104		9	

son of the two groups about the different positions of *I think*.

It is clear from the above table that both learners and the native speakers tend to use *I think* in the mid position more (Learner= 56.73 %; Native= 66.66 %). However, the end use of it appears to be nonexistent in both groups. This situation might be related to the fact that the end use of *I think* is a peculiarity of spoken register of English language (Petch–Tyson 1998) with a very low percentage of 3.2 (Mullan 2010).

The next concern of the current study is related to the potential collocations of *I think*. In order to be able to make deductions about a phrase or a single word, one needs to be familiar with the words and structures surrounding it, which means frequent constructions that relates to the level of language between the lexicon and grammar (Turan 2010). Table 6 provides the potential collocations for *I think* along with their word classes, frequencies (*f*) and mutual information scores (*MI*).

Table 6 presents the potential collocations of *I think* in the learner corpus with their word classes, frequencies and mutual information scores. Word class refers to the lexical category to which an item belongs. Lexical items such as nouns, verbs and adverbs all belong to content words category while words with no or ambiguous meanings and serve to express grammatical relationships fall under the category of function words. In the third column, raw frequencies of the related items are given. In the last column, mutual information scores of the items, which show the probability that the two items occur together and they belong together McEnery and Wilson (2001), are exhibited. Items with MI scores equal to or greater than 3 are taken into consideration because typically, scores of about 3 or above show a semantic bonding between the two words (Davies 2008). From the data presented in the above table, it is clear that in the learner corpus the phrase *I think* has a tendency to co-occur with function words like *is,*

Table 6: Collocation candidates of 'I think'

N	Word	Class	f	MI
1)	is	function	19	3.25
2)	that	function	17	3.62
3)	but	function	14	4.32
4)	it	function	14	3.35
5)	should	function	13	5.19
6)	are	function	13	3.18
7)	we	function	12	3.26
8)	be	function	11	3.48
9)	media	content	9	3.39
10)	important	content	8	4.48
11)	have	function	8	3.36
12)	more	function	6	3.99
13)	so	function	6	3.61
14)	must	function	5	4.74
15)	technology	content	5	4.06
16)	there	function	5	3.13
17)	mass	content	4	3.84
18)	don't	function	4	3.64
19)	use	content	4	3.15
20)	consequently	content	3	7.11
21)	practical	content	3	5.69
22)	two	content	3	5.30
23)	harmful	content	3	5.23
24)	benefits	content	3	5.02
25)	reason	content	3	5.02
26)	dreaming	content	3	4.71
27)	enough	content	3	4.60
28)	science	content	3	4.20
29)	radio	content	3	3.50
30)	shouldn't	function	2	5.46
31)	develop	content	2	5.41
32)	really	content	2	5.41
33)	works	content	2	5.35
34)	those	function	2	5.30
35)	practice	content	2	5.20
36)	reasons	content	2	4.94
37)	role	content	2	4.82
38)	right	content	2	4.79
39)	sometimes	content	2	4.52
40)	being	function	2	4.52
41)	conclusion	content	2	4.43
42)	then	function	2	4.41
43)	industrialization	content	2	4.33
44)	affect	content	2	4.18
45)	education	content	2	3.68
46)	first	content	2	3.66
47)	no	function	2	3.48
48)	bad	content	2	3.27
Total			245	

that and *but*. Table 7 presents the related frequencies and percentages.

Table 7: Frequency and percentage of collocation candidates for 'I think'

Word class	f	%
Function	144	58.78
Content	101	41.22
Total	245	

A quick analysis of the above table will make it clear that the phrase *I think* in learner corpus have a greater tendency to co-occur with function words ($f_{function words}$=144, %=58.78). When this outcome is compared with the one presented previously in Table 1, there seems to be a remarkable difference in the use of *I think* between Turkish EFL learners and native speakers of English. No statistical comparison can be carried out between these two data as the former was collected from learner texts while the latter comes from spoken register and the learner corpus consists of 58.046 words and the reference corpus, the BNC, consists of 1 million words. However, when we consider that *I think* mainly belongs to spoken register of English (Aijmer 1997) there seems to be a twofold problem: Turkish EFL learners transfer the phrase *I think* from spoken register to the written one, and they seem to be using it with a very different orientation from the native speakers of English who mainly use it in their conversations for pragmatic concerns mostly (Aijmer 1997). The same procedure couldn't be carried out for the native essays as there were only a small number of incidences of *I think* in the corpus (see Table 3), which made statistical computations meaningless.

Another problem concerning texts produced by Turkish EFL learners is the seemingly overuse of the phrase *in my opinion*. From observations, it was noticed that Turkish EFL learners make use of this phrase abundantly both in written and spoken productions; however, its incidences seemed to be rarer in native productions of English. In order to validate this observation, statistical analyses were carried out and descriptive statistics about this phrase is presented in Table 8.

Table 8: Descriptive results for 'in my opinion'

Group	N	f	Means	sd
Learner	161	39	.26	.48
Native	46	3	.07	.25

Table 8 presents the frequency, mean and standard deviation for *in my opinion*. As is clear from the table, the learner group used *in my opinion* 39 times in their essays with a mean of .26 and a standard deviation of .48. The native group, on the other hand, used it only 3 times which comes to a mean of .07 and a standard deviation of .25. These figures alone denote that there is an overuse of the phrase in learner corpus. In order to make statistical inferences about

it, statistical comparison of the two groups is required. However, as was the case with the phrase *I think*, the results of Levene's test of variance indicated that the group variances are significantly different (p<.01). That's why, a non-parametric test, Mann Whitney U-test, was conducted and the results are revealed in Table 9.

Table 9: Mann Whitney U-test results for 'in my opinion'

Group	N	Mean rank	Sum of ranks	U	p
Learner	161	108.10	17404	3043	.008
Native	46	89.95	4121		

Table 9 reveals the mean and sums of ranks of the two groups. It is obvious from the table that there is statistically significant difference between the two groups in terms of the use of *in my opinion* (U= 3043, p< .01). This means that the obvious difference between the two groups is statistically validated. The interesting point here is that, although the related literature abounds with discourse marker studies (*see the introduction*), there seems to be no study concerning the use/overuse of this phrase by EFL learners.

DISCUSSION

The findings of the current study provide a clear answer to the first research question related to the quantitative aspect of *I think* in Turkish EFL learners' texts. There actually is a statistically significant difference between the learner and native texts. Turkish EFL learners tend to use this phrase more than necessary and with different concerns from the native speakers of English. This finding is in line with the related literature (Aijmer 2001; Ishikawa 2009; Yong 2010). The addition of Turkish EFL learners to the related discussions could be counted as the new perspective that this study brings.

As for the second research question, *I think* in Turkish EFL texts tend to co-occur with function words mostly. Among these, function words mainly used to introduce a new clause in a sentence like *but, that, so* and *then* attract attention. This result might indicate that Turkish EFL learners might be using this phrase in their writings for reasons other than pragmatic ones, which is noteworthy because unlike EFL learners, the native speakers of English use its pragmaticalized version as it involves the attitude of

the speaker to the hearer (Aijmer 2001). The results of the current study also suggest a reference to Trillo's (2002) pragmatic fossilization concept. As was mentioned before, since non-native speakers lack the pragmatic resources, pragmatic markers might go through a process of fossilization leading to unfitting linguistic elements. Also, Kjellrner's hypothesis (1991: 124) that learners' foreign-soundingness may be due to the fact that '[their] building material is individual bricks rather than prefabricated sections'. This means that, Turkish EFL learners have the phrase *I think* as a brick, but they seem to be having problems when it comes to creating coherent contexts with them.

As was indicated previously O'Keeffe et al. (2007) since the phrase *I think* is statistically among the most frequent phrases used in English, EFL learners are very likely to be exposed to these chunks as well, which might be one of the reasons behind its overuse. However, one could raise the following question: Why don't EFL learners use the most frequent phrase *you know* in their texts?

The results related to the third research question trying to determine whether there is a significant quantitative difference between the texts written by native speakers of English and Turkish EFL students concerning the phrase *in my opinion* seem to validate the observation that these learners use this phrase more than the native speakers. Since the related literature lacks studies about this issue, no interpretations or comparisons can be made at this point.

One interesting study points out that a non-traditional writing practice in which the learners write to their peers rather than to the teacher yields better results than traditional ones (Kingir 2013). From this perspective, the current study could be repeated after collecting written data from EFL learners that are written to their peers, but not to the teacher. It is a strong assumption that there should be significant differences in terms of discourse markers along with other points. Such a study, for higher impact, could be backed up with a qualitative approach as well (Hos and Topal 2013) by adapting it into a multi-cultural framework, which is a globally increasing trend these days (Aydin 2013; Tok and Karakus 2013; Demirli 2013).

CONCLUSION

The main question to be asked here is whether the overuse of *I think* and *in my opinion* ful-

fills a pragmatic function as suggests for the native speakers of English or a transfer of spoken register into written by EFL learners which means that they try to write like they speak. Although the current study provides some insights about this issue to an extent, it is apparent that further studies, both quantitative and qualitative, are needed for further and clearer answers to this question.

All in all, the researcher holds the opinion that, although this study reveals how non-native Turkish EFL learners' texts look like, the main concern should not be training learners to write like native speakers but to promote the notion of 'successful language learner' because language is deeply interwoven with people's native cultures and it wouldn't be appropriate for language teachers to expect their students to suppress what is in them while producing in a target language, because after all, there is a great possibility that EFL learners might actually be writing in their own languages just by using the English lexicon.

RECOMMENDATIONS

Now that it has been statistically proven that Turkish EFL learners use some spoken discourse markers significantly more than the native speakers of English, it would be interesting to see Turkish EFL learners' situation with discourse markers in spoken register. In order to make valid interpretations about this issue along with many others, a spoken corpus of Turkish EFL learners is of immediate need.

ACKNOWLEDGEMENTS

REFERENCES

Aarts J, Granger S 1998. Tag sequences in learner corpora: A key to interlanguage grammar and discourse. In: Sylviane Granger (Ed.): *Learner English on Computer*. London and New York: Addison Wesley Longman, pp. 132-141.
Abe M, Tono Y 2005. Variations in L2 Spoken and Written English. *Proceedings from The Corpus Linguistics Conference Series* in University of Birmingham, Birmingham, July 14 to 17, 2005.
Aijmer K 1997. *I think* – an English modal particle. In: Toril Swan, Olaf Westvik (Eds.): *Modality in Germanic Languages*. Berlin: Mouton de Gruyter, pp. 1-47.
Aijmer K 2001. I think as a marker of discourse style in argumentative Swedish student writing. In: Karin Aijmer (Ed.): *A Wealth of English*. Goteborg: Acta Universitatis Gothoburgensis, pp. 247-257.
Allen D 1992. *Oxford Placement Test 2*. Oxford: Oxford University Press.
Altenberg B, Tapper M 1998. The use of adverbial connectors in advanced Swedish learners' written English. In: Sylviane Granger (Ed.): *Learner English on Computer*. London: Addison Wesley Longman, pp. 80-93.
Aydin H 2013. A literature-based approaches on multicultural education. *Anthropologist*, 16(1-2): 31-44.
Belz J, Vyatkina N 2005. Learner corpus research and the development of l2 pragmatic competence in networked intercultural language study. *Canadian Modern Language Review,* 62(1): 17-48.
Biber G, Reppen R 1998. Comparing native and learner perspectives on English grammar. In: Sylviane Granger (Ed.): *Learner English on Computer*. London and New York: Addison Wesley Longman, pp. 145-158.
Callies M 2009. Information Highlighting in Advanced Learner English: *Pragmatics and Beyond*. Volume 186. Amsterdam: Benjamins.
Carter R, McCarthy M 2006. *Cambridge Grammar of English*. Cambridge: Cambridge University Press.
Celce-Murcia M, Olshtain E 2000. *Discourse and Context in Language Teaching*. New York: Cambridge University Press.
Celce-Murcia M, Olshtain E 2005. Discourse-based approaches. In: Eli Hinkel (Ed.): *Handbook of Research in Second Language Teaching and Learning*. Mahwah, NJ: Lawrence Erlbaum, pp. 729-741.
Celik S, Elkatmis M 2013. The effect of corpus assisted language teaching on the learners' proper use of punctuation marks, *Educational Sciences: Theory and Practice*, 13(2): 1090-1094.
Dagneaux E, Denness S, Granger S 1998. Computer-aided error analysis. *System,* 26(2): 163-174.
Davies M 2008. The Corpus of Contemporary American English. From <http://corpus.byu.edu/coca> (Retrieved on March 15, 2013).
De Cock S, Granger S, Leech G, McEney T 1998. An automated approach to the phrasicon of EFL learners. In: Sylvaine Granger (Ed.): *Learner English on Computer*. London: Longman, pp. 67-79.
De Cock S 2000. Repetitive phrasal chunkiness and advanced EFL speech and writing. In: Christian Mair, Marianne Hundt (Eds.): *Corpus Linguistics and Linguistic Theory*. Amsterdam: Rodopi, pp. 51-68.
Demirli C 2013. ICT usage of pre-service teachers: Cultural comparison for Turkey and Bosnia and Herzegovina, *Educational Sciences: Theory and Practice,* 13(2): 095-1105.
Field A 2009. *Discovering Statistics Using SPSS*. London: Sage Publications.
Flowerdew L 1998. A Corpus-based Analysis of Referential and Pragmatic Errors in Student Writing.

TACL98 Proceedings, in Keble College, Oxford, July 24 to 27, 1998.

Flowerdew L 2000. Investigating referential and progmatic errors in a learner corpus. In: Lou Burnard, Tony McEnery (Eds.): *Rethinking Language Pedagogy from a Corpus Perspective.* Frankfurt am Main: Peter Lang, pp. 145-154.

Gilquin G 2008. Discourse markers in native and non-native English discourse. *Linguistics,* 46(3): 667-677.

Granger S 1998a. *Learner English on Computer.* London and New York: Addison Wesley Longman.

Granger S 1998b. Prefabricated patterns in advanced EFL writing: Collocations and formulae. In: Anthony Paul Cowie (Ed.): *Phraseology.* Oxford: Oxford University Press, pp. 145-160.

Granger S 1999. Use of tenses by advanced efl learners: Evidence from an error-tagged computer corpus. In: Oksefjell Signe, Hilde Hasselgard (Eds.): *Out of Corpora.* Amsterdam and Atlanta: Rodopi, pp. 191-202.

Granger S 2003. The international corpus of learner English: A new resource for foreign language learning and teaching and second language acquisition research. *TESOL Quarterly,* 37(3): 538–545.

Hinkel E 2001. Giving examples and telling stories in academic essays. *Issues in Applied Linguistics*, 12: 149–170.

Hinkel E 2002. *Second Language Writers' Text.* Mawah, NJ: Lawrence Erlbaum Associates.

Hinkel E 2005. Analyses of second language texts and what can be learned from them. In: Eli Hinkel (Ed.): *Handbook of Research in Second Language Teaching and Learning.* Mahwah, NJ: Lawrence Erlbaum, pp. 615-628.

Hinkel E 2011. What research on second language writing tells us and what it doesn't. In: Eli Hinkel (Ed.): *Handbook of Research in Second Language Teaching and Learning.* Volume 2. New York: Routledge, pp. 523-538.

Hos R, Topal H 2013. The current status of English as a Foreign Language (EFL) teachers' professional development in Turkey: A systematic review of literature. *Anthropologist*, 16(1-2): 293-305.

Ishikawa S 2009. Phraseology overused and underused by Japanese learners of English: A contrastive interlanguage analysis. In: Katsumasa Yagi, Takaaki Kanzaki (Eds.): *Phraseology.* Nishinomiya: Kwansei Gakuin University Press, pp. 87-100.

Kjellmer G 1991. A mint of phrases. In: Karin Aijmer, Bengt Altenberg (Eds.): *English Corpus Linguistics.* London and New York: Longman, pp. 111-127.

Kingir S 2013. Using non-traditional writing as a tool in learning chemistry. *Eurasia Journal of Mathematics. Science and Technology Education*, 9(2): 101-114.

Lakoff G 1972 . Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic,* 2: 458-508.

McEnery T, Wilson A 2001. *Corpus Linguistics: An Introduction.* Edinburgh: Edinburgh University Press.

Meunier F 2002. The role of learner and native corpora in grammar teaching. In: Sylviane Granger, Joseph Hung, Stephanie Petch-Tyson (Eds.): *Computer Learner Corpora, Second Language Acqui-*

sition and Foreign Language Teaching. Amsterdam and Philadelphia: Benjamins, pp. 119-142.

Moder C, Martinovic-Zic A 2004. *Discourse Across Languages and Cultures.* Amsterdam: John Benjamins.

Mulak J 2000. The discourse functions of because clauses in l1 and l2 argumentative writing including a quantitative survey of reason/cause, concession and condition clauses. In: Tuija Virtanen, Jens Agerstrom (Eds.): *Three Studies of Learner Discourse.* Vaxjo: Vaxjo Universitet, pp. 81-139.

Mullan K 2010. *Expressing Opinions in French and Australian English.* Amsterdam: John Benjamins.

O'Keeffe A, McCarthy M, Carter R 2007. *From Corpus to Classroom.* Cambridge: Cambridge University Press.

Parrott M 2000. *Grammar for English Language Teachers.* Cambridge: Cambridge University Press.

Petch–Tyson S 1998. Writer/reader visibility in EFL written discourse. In: Sylvaine Granger (Ed.): *Learner English on Computer.* London: Longman, pp. 107-118.

Pulcini V, Furiassi C 2004. Spoken interaction and discourse markers in a corpus of learner English. In: Alan Partington, John Morley, Louann Haarman (Eds.): *Corpora and Discourse.* Bern: Peter Lang, pp. 107-123.

Sadock J 2006. Speech acts. In: Laurence Horn, Gregory Ward (Eds.): *Handbook of Pragmatics.* Oxford: Blackwell, pp. 53-73.

Schiffrin D 1986. *Discourse Markers.* Cambridge: Cambridgeshire.

Scott M 2012. *WordSmith Tools Version 6.* Liverpool: Lexical Analysis Software.

Silva T 1993. Toward an understanding of the distinct nature of L2 writing: The ESL research and its implications. *TESOL Quarterly*, 27(4): 657–677.

Schleppegrell M 1996. Conjunction in spoken English and ESL writing. *Applied Linguistics*, 17(3): 271–285.

Tanko G 2004. The use of adverbial connectors in hungarian university students' argumentative essays. In: John Sinclair (Ed.): *How to Use Corpora in Language Teaching.* Amsterdam: John Benjamins, pp. 157-181.

Thompson S, Mulac A 1991. The discourse conditions for the use of the complementizer 'that' in conversational English. *Journal of Pragmatics,* 15: 237-251.

Tok H, Karakus M 2013. Cross-cultural reliability and validity of the Attitude towards College Instructor Authority (ACIA) survey, *Anthropologist*, 16(1-2): 285-291.

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

Turan U D 2012. Collocations with *mind* in corpus and implications for language teaching, *Egitim Arastirmalari - Eurasian Journal of Educational Research*, 49/A: 331- 348.

Yong W 2010. The use of 'I think' by Chinese EFL learners: A study revisited. *Chinese Journal of Applied Linguistics*, 33(1): 3-23.