

***By the way* as a Multifunctional Discourse Marker**

Yejin Kim

(Seoul National University)

Kim, Yejin. 2018. *By the way* as a Multifunctional Discourse Marker. *SNU Working Papers in English Linguistics and Language* 16, 57-82. This study explored the diverse pragmatic aspects of the discourse marker ‘by the way’ using three spoken corpora which were the MICASE corpus, Corpus of Contemporary American English, and British National Corpus (BYU-BNC). Despite ‘by the way’s most widely perceived usage as a digression marker signaling sudden topic change, there have been some noteworthy instances in which it was found to serve other discourse functions as well. Based on that observation, this paper first set out to identify other generally overlooked discourse functions that ‘by the way’ serves. The study then proceeded to take a closer look into how those various functions were differentially realized in each spoken corpus representative of distinct spoken registers. The results rendered suggested that ‘by the way’ does not have one single, static function and instead, its predominant function changes depending on the various discourse settings in which it occurs. From this variable nature of ‘by the way’, it was concluded that further discourse considerations would have to be incorporated if we are to arrive at a more enriched characterization of ‘by the way’.

(Seoul National University)

Keywords: by the way, discourse markers, spoken corpus, conversation analysis, Pragmatics

1. Introduction

In pragmatics, discourse makers are lexical items also classified as “conjuncts”. These conjuncts serve to “express the speaker’s assessment of the relation between two linguistic units” (Quirk, 1985). There have been many attempts to classify these discourse markers in the pragmatics field. However, there always existed a certain amount of variety in different classification schemes. It was argued that this is mostly because “there is no single, uniquely correct inventory of the types of conjunctive relation” since each type of classification would highlight different aspects even within a single discourse marker (Halliday & Hasan, 1976).

Consequently, several attempts to classify discourse markers have engendered the view that discourse markers are usually multi-functional and polysemous (Bell, 1988). As regards the meaning of these discourse markers, it was suggested that discourse markers, in effect, do have their inherent, semantical core meaning. Schiffrin (1987) claims that the core meaning of these markers do not “fluctuate from use to use”. If that is the case, what changes is only the discourse slot in which they appear. Fraser (1988) further maintains that even a single discourse marker may allow a wide range of interpretations, all of which presumably emerge from a ‘core’ sense. On this view, different interpretations of a discourse marker are only one of many manifestations of its semantically enriched meanings.

Even if we adhere to the view that discourse markers have their inherent semantic meanings, they could only be identified on a post-hoc analysis after those meanings have been ‘fleshed out’ throughout through their instantiations. Also, it would be hard to differentiate between their supposedly true semantic meanings and the specific discourse functions they serve. The same can be said when we attempt to analyze common discourse markers such as *by the way*. Since it is hard to define the definitive meaning of *by the way* apart from its widely perceived discourse functions, it would seem more reasonable to investigate their diverse usage patterns in order to arrive at a more enriched characterization.

Although there already exists a large body of research on discourse markers, not much research has been carried out on *by the way* alone. In some of the previous literature on discourse markers, *by the way* was, for the most part, explained in terms of the discourse function it serves. For example, *by the way* was often treated as signaling a following ‘side statement’ that often diverges from the previous topic. (Dailey & Palomares, 2004). Fraser (2009) classified *by the way* as one of many “topic orientation markers”, by which the speaker’s intentions concerning the immediate future topic of the discourse can be conveyed.

In another research (Croucher, 2004) that investigated four formal textual functions of discourse markers, *by the way* was categorized as a discourse marker that introduces digression from the topic under discussion, often in conjunction with another discourse marker ‘oh’. Note that in many cases, *by the way* was explained in terms of the most primary discourse function it serves that is ‘digression’. In the following section, several dictionaries have been consulted in order to take a closer look into the prescribed definitions of *by the way*.

2. Research Question

The aim of this study is two-fold. First, it attempts to identify diverse features of *by the way* through close analysis of corpora concordance lines in which it was found. A pilot study using the MICASE corpus has been carried out in advance to derive some of the most salient features of *by the way*, with the main focus on eliciting and categorizing its pragmatic functions. Since there has been no individual case study done on *by the way* alone, I had to come up with an original list of its possible functions based on this pilot study. The main purpose was to identify linguistic features of *by the way* as well as the diverse pragmatic functions it serves in spoken discourse. After the manual categorization of *by the way*’s characteristics based on my originally devised categorization table, it would then be possible to see whether *by the way* only functions as an ‘empty conversational filler’ or has other various functions that have been unrecognized so far. It would also be possible to see whether *by the way* is predominantly used to introduce a context-irrelevant topic or whether it still shares a strong connection to the overall context in which it is found.

After identification of *by the way*’s linguistic characteristics as well as discourse functions it performs, the study will proceed to compare its occurrence patterns across three different spoken corpora in order to

investigate how it differentially features in each corpus. The main categorization features for comparison will be restricted to *by the way*'s pragmatic function and its context relevance. This analysis based on each corpus will allow us to see which pragmatic function of *by the way* appears most predominantly in each corpus. It will also be able to provide us with answers as to whether *by the way* is recognizable through its link to the whole context or whether it is loosely linked to the extent that it almost exclusively functions as a 'digression marker'.

3. Data Collection

3.1 Dictionary Definitions of *by the way*

The basic assumption for this study was that *by the way* would serve more pragmatically rich functions than its most widely recognized 'digression' function. In order to see if *by the way* actually carries any default meaning, it was necessary to look into some of its generally accepted definitions established by authoritative dictionaries. I selected 5 dictionaries as the most authoritative dictionaries currently used and searched the keyword *by the way* on each online dictionary. The definitions of *by the way* provided by these dictionaries are as follows. First, the Merriam-Webster dictionary defines *by the way*'s meaning through its function "interjection or digression" and presents its synonym as "incidentally". Based on this definition, we can derive *by the way*'s main function as a digression marker that serves to introduce a new topic in the conversation. The next dictionary I consulted was the Cambridge English Dictionary. This dictionary listed two different meanings of *by the way*, with one being the English (British) definition based on British English and the other being the American definition based on American English. On the British English definition, *by the way* is normally used to "introduce a new subject to be considered or to give further information". From this definition, we can elicit two subtly different

functions that is *by the way* as an ‘elaboration marker’ and *by the way* as a ‘digression marker’. It only gives examples such as “I think we’ve discussed everything we need to- *by the way*, what time is it?” to instantiate *by the way*’s function as a digression marker. On the American English definition, *by the way* is used for “introducing a statement or subject that may not be directly related to the subject being discussed”. This definition again enables us to elicit *by the way*’s main function as a ‘digression’ marker that signals sudden subject change. The third dictionary I consulted was the Longman Dictionary of Contemporary English which states that *by the way* is used “when saying something that is not related to the main subject you were talking about before”. This definition also emphasizes *by the way*’s function as a digression marker that signals how the direction of the conversation is going to change. The fourth dictionary I consulted was the Collins CoBuild English Dictionary. The dictionary states that *by the way* is used when you “add something to what you are saying, especially something that you have just thought of”. This definition further fleshes out the bare meaning of *by the way* as it implies how *by the way* signals temporary memory failure and the following attempt at memory retrieval. It does not restrict *by the way*’s function as a mere ‘digression marker’ but instead accentuates its usage as an ‘elaboration marker’ introducing an afterthought. It also explicitly states *by the way*’s grammatical function as a ‘sentence modifier’. The last fifth dictionary I consulted was the Macmillan Dictionary which states that *by the way* is used for “saying a remark not relevant to the main subject of your conversation”. This definition again highlights *by the way*’s discourse function as a ‘digression marker’ signaling that the following remark going to depart from the previous one and will be irrelevant to the preceding context. What is most noticeable throughout these dictionary definitions of *by the way* is that they do not provide a clear-cut definition of *by the way* and are instead giving only a description of how its functions are carried out in discourse settings. The primary functions of *by the way* these dictionaries identified so far are

only ‘digression’ and ‘elaboration’. If we take these two functions as the default functions (or essentially the meaning) of *by the way*, we can take a closer look at how these functions are realized in spoken discourse and at the same time, consider the possibility of there being other functions as well that are yet to have been identified.

3.2 Explanation of the Corpora

In total, three corpora have been used in this study which are the COCA Spoken corpus, MICASE corpus, and the BYU-BNC Spoken corpus further limited to its conversation subcorpus. The COCA corpus is composed of more than 560 million words in 220,225 texts, including 20 million words each year from 1990-2017. The corpus is evenly divided between the five genres of spoken, fiction, popular magazines, newspapers, and academic journals. As for the spoken subcorpus, there is a total of 118 million words. Transcripts come from conversations in more than 150 different TV and radio programs. The next corpus used was the MICASE corpus which has a collection of transcripts from academic lectures recorded at the University of Michigan. There are currently 152 scripts with a total of 184 million words available in this corpus. The last corpus used was the BYU-BNC corpus originally created by Oxford University press in the 1980s to early 1990s. It contains 100 million words with texts from a wide range of genres (e.g. spoken, fiction, magazines, newspapers, and academic). The spoken part (10%) consists of transcriptions of unscripted informal conversations recorded by volunteers from different age, region and social classes. The spoken data were collected in different contexts, ranging from formal business and government meetings to radio shows and phone-ins. In this spoken BYU-BNC corpus, I controlled the search feature to look up concordance lines only from the ‘conversation’ subsection that would allow me to see how *by the way* occurs in natural and ordinary conversations.

The search of the keyword *by the way* was then carried out on each corpus. The abbreviated forms BTW and btw were disregarded due to their primary usage in written text messages and emails. As for the MICASE corpus, the search rendered a total of 118 concordance lines without any delimitation process done. As for the COCA spoken corpus, there was a total of 12668 concordance lines from which 200 samples were randomly selected for this study. As for the last BYU-BNC spoken corpus, only the 'conversation' subcorpus was selected for this study, which rendered a total of 176 raw concordance lines. These three corpora were selected to investigate diverse instances of *by the way* in three different spoken registers. That is because even though the data were all from these spoken corpora, the nature of the data is considered to be different in terms of their degree of formality, level of interaction between speakers, and the naturalness of the spoken language itself.

First, as regards the degree of formality in discourse context, it was posited that data from the BYU-BNC spoken corpus would demonstrate the least degree of formality, with the data from COCA representing average formality, and the data from MICASE corpus displaying the greatest degree of formality by comparison. Since MICASE is a corpus of academic spoken English, the academic setting itself would contribute to the strict formality of the language used in such settings. As for the interactivity level between speakers, BYU-BNC spoken subcorpus showed the highest degree of interactivity with a lot of turn-takings taking place throughout discourse. These turn-takings were found several times even within a single transcript. Most of the data from MICASE corpus revealed that conversations were often carried out unidirectionally, with the academic faculty member holding the conversation turn for most of the time. Data from the COCA corpus displayed a moderate degree of interactivity since transcripts were from either scripted or unscripted radio broadcasts, news, and interviews. The interaction between speakers frequently showed structured and expected turn-takings which in a sense reflect the characteristic of the corpus. Last,

as for the naturalness of speech, data from BYU-BNC corpus showed aspects of everyday speech with many false starts, conversational breakdowns, and incorrect language use. On the other hand, data from the MICASE displayed aspects of highly structured speech with the speech content heavily loaded with academic domain-specific words. In addition, the main focus of language use in MICASE was on clear delivery and minimum errors. COCA was placed in the middle of these two corpora as the data showed some aspects of unnaturalness despite the preponderance of scripted language. This is due to the fact that there could have been some amount of spontaneous speech as well in live broadcasts and radio news.

3.3 Delimitation Process

Repetition of the exact same phrase immediately preceding and following *by the way* showed that these tokens of *by the way* were just dummy fillers. Also, concordance lines with unclear phrases immediately preceding and following *by the way* that made its interpretation impossible were disregarded in this study. In addition, concordance lines with ‘independently used’ *by the way* such that the phrase itself stood alone in the whole context were disregarded as well. There were other concordance lines in which it was not easy to determine whether *by the way* was modifying the immediately preceding sentence or the immediately following sentence. They were altogether disregarded as it was hard to categorize *by the way*’s function. Last, concordance lines in which *by the way* was used to denote ‘by means of’ and ‘through’ were disregarded as they were judged irrelevant to the purpose of this study.

- (1) ...what? um yeah. PAUSE duration :05 **by the way** no smaller than, what? tw- eleven point? twelve point? this is you guys. in terms of your uh font. -> preceding sentence unclear
- (2)... of, this system so... expected single-command cycle, is what we're trying to get. this is **by the way** uh the__ before i put the

question mark maybe i i'll define it. this is the expected... single-comman -> *by the way* standing alone as a separate sentence

(3) stripes. i would say yes ***by the way*** they're flying. swallow or -> ambiguousness of the sentence *by the way* modifies

(4)... revival architecture. Manet, very aware of making contemporary art but showing its contemporaneity ***by the way*** that it both engages the past and differentiates itself from art of the past. Manet for one thing an -> *by the way* denoting 'by means of'

3.4 Categorization Features for Manual Categorization

There were four primary categorization features of *by the way* that I derived from the pilot study on MICASE. The first one was the sentence position of *by the way*. If *by the way* appeared at the beginning of the sentence, it was categorized as belonging to the 'sentence-initial' group. If it appeared in the middle of a sentence or around somewhere near the middle, it was categorized as belonging to the 'mid-sentence' group. If it appeared at the end of a sentence, it was categorized as belonging to the 'sentence-final' group. The second categorization feature was the type of the sentence in which *by the way* appeared. Following this, all the sentences containing *by the way* were divided into declarative, interrogative, exclamatory, and imperative sentence types. The third categorization feature was the pragmatic function that *by the way* performed in the sentence in which it was found. The discourse function of the sentence that *by the way* modified was classified as additional elaboration, qualifying elaboration, personal opinion, digression, question, order, or suggestion. Additional elaboration refers to addition of 'factual' information while qualifying elaboration refers to provision of factual information phrased in negation form. For example, if *by the way* served to introduce a partial (ex. not necessarily, not always) or total negation of the preceding sentence, the discourse function of that sentence that *by the way* modified was classified as 'qualifying

elaboration'. The personal opinion function had to do with *by the way* being used to introduce a value judgment, strong assertion such as claiming that A is B, or a personal disclaimer. For any sentence in which *by the way* introduced a personal opinion (any statement other than pure information), the discourse function of that sentence was classified as 'personal opinion'. The 'question' function was limited to genuine questions demanding a definite answer, not rhetorical questions or any tag questions. 'Digression' function was defined in the strictest sense since this function encompassed any instances of *by the way* being used for topic shift. One may raise the possibility of some overlap among additional elaboration, question, and digression. However, additional elaboration function mainly had to do with supply of relevant information and the question function also had to do with questions still pertaining to the context. Therefore, the discourse function of any sentence in which *by the way* suggested even a hint of digression was classified as 'digression'. The last categorization feature was the context relevance. If a sentence that *by the way* modified was directly relevant to the preceding sentence, that sentence was classified as belonging to 'directly relevant' group. If there was an identifiable referent, demonstrative, or a repetition of the same word close to *by the way*, it was taken as the basis on which to classify the sentence that that *by the way* modified as 'directly relevant'. If a sentence that *by the way* modified was indirectly relevant to the preceding sentence, that sentence was classified as 'indirectly relevant'. In this classification of 'indirectly relevant' sentences containing *by the way*, the relevance had to be determined based on contextual cues and the actual discourse content. Sentences classified as 'irrelevant' included ones in which *by the way* was used to lead to a sudden topic change. In general, if the 'sentence position' and the 'sentence type' had to do with classification based on the surface linguistic feature of *by the way*, other features such as 'pragmatic function' and 'context relevance' features had to do with classification based on pragmatic inference from the discourse content.

4. Results

4.1 Results from MICASE

Figure 1.

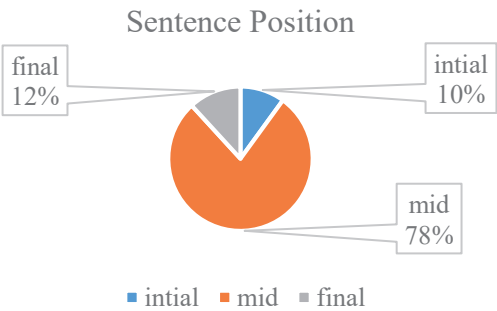


Figure 2.

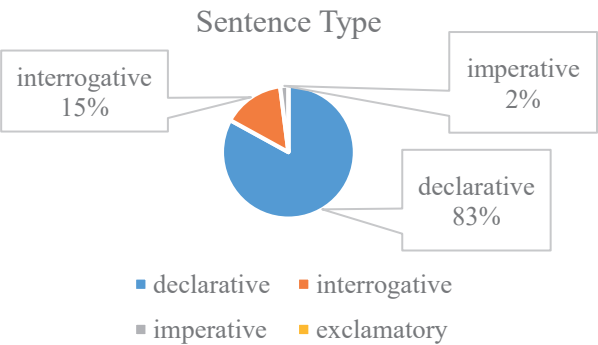


Figure 3.

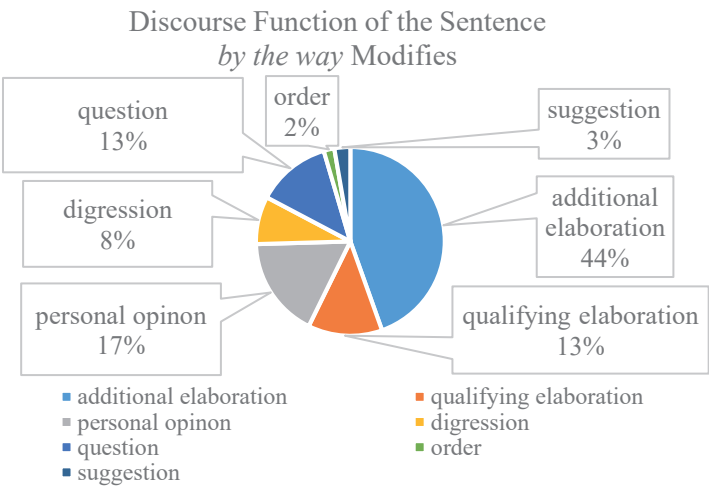


Figure 4.

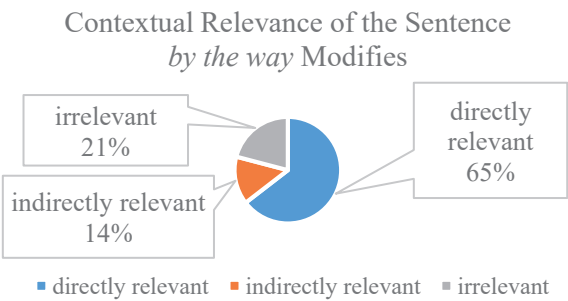


Figure 5. AntConc Collocation Search (Search Window 1L-1R)

#Total No. of Collocate Types: 10					
#Total No. of Collocate Tokens: 51					
1	17	17	0	2.54700	and
2	6	6	0	3.65921	which
3	5	5	0	1.46544	this
4	4	4	0	1.29289	uh
5	4	4	0	5.07425	cake
6	3	3	0	3.85185	who
7	3	3	0	-0.02729	that
8	3	3	0	2.85185	okay
9	3	3	0	3.85185	oh
10	3	3	0	0.30166	is

Figure 1 to Figure 4 all focus on one categorization feature and the classification results based on that one specific feature. Figure 5 shows the collocational pattern of *by the way* through its co-occurrence with other words in the same sentence. From Figure 1 to Figure 5, it was found that in MICASE, *by the way* mainly appeared in the middle of a sentence (78%), also mainly in declarative sentences. Furthermore, it was found that most of the time, *by the way* served to introduce a phrase that either ‘additionally elaborates’ or ‘qualifies’ the immediately preceding sentence, which characterizes *by the way* as an ‘elaboration’ discourse marker. This result is in correspondence with the result from the AntConc search I have carried out. Using the AntConc collocation search feature, I narrowed down the search window range from 1L(left) to 1R(right) in order to find out which words would co-occur most frequently with *by the way*. The result rendered showed that ‘and’ was the most frequently co-occurring word (frequency 17), which further reinforces the previous finding in which *by the way* was often identified as ‘elaboration’ discourse markers. Elaboration function of the sentence *by the way* modified accounted for as much as 57% of all the functions, which was then followed by personal opinion (17%) and question (13%) function. Because most of the speech was recorded in academical settings in which some of the most sensitive and controversial academic issues were treated as well, strong assertion and personal disclaimers were sporadically found throughout the transcripts. Below are some of the example concordance lines in which the sentences *by the way* modified were either an elaboration, strong assertion, or personal disclaimer.

As we can see from the figure 4, the context relevance of the sentence modified by *by the way* was for the most part, directly relevant (65%). If we add all the indirectly relevant cases into the calculation, the final percentage of contextual relevance would amount to as much as 79%. This outcome is presumably a reflection of the nature of MICASE corpus in which academic speech is focused on the coherence of the academic lecture’s content. Any irrelevant sentence introduced by *by the way*

would have disrupted the overall flow of the academic speech, which in turn would have led to some undesirable consequences such as misunderstanding between the lecturer and the students.

4.2 Results from COCA

Figure 6.

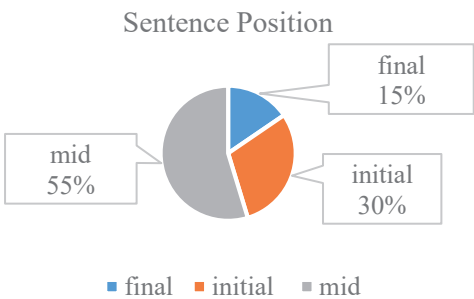


Figure 7.

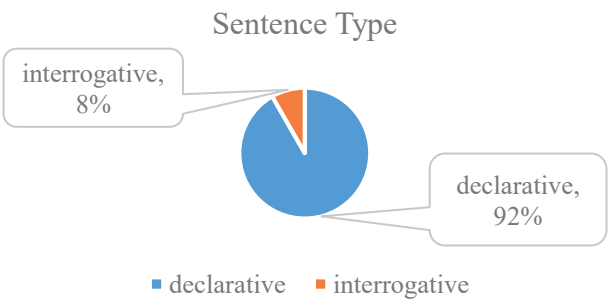


Figure 8.

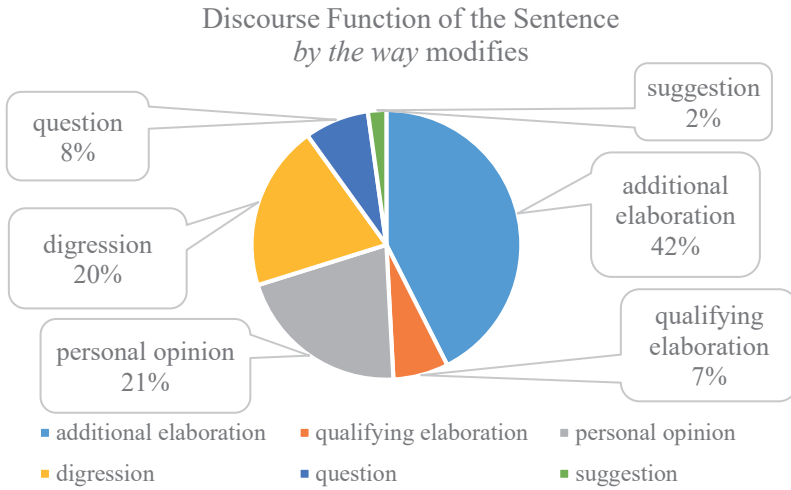


Figure 9.

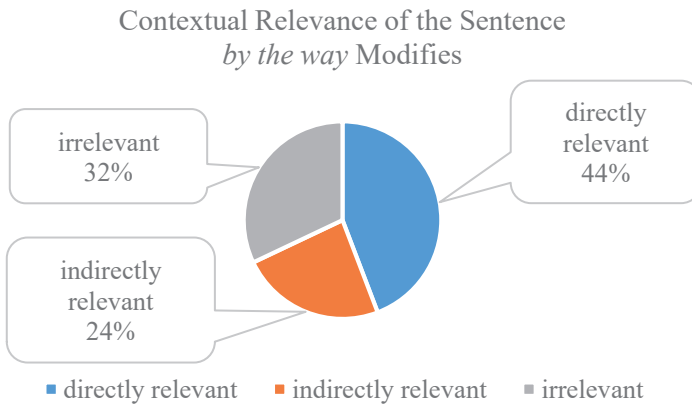


Figure 10. AntConc Collocation Search (Search Window 1L-1R)

#Total No. of Collocate Types: 139					
#Total No. of Collocate Tokens: 209					
1	40	40	0	2.81406	and
2	7	7	0	4.33901	which
3	4	4	0	2.22045	kotb
4	3	3	0	-0.70592	you
5	3	3	0	0.60472	this
6	3	3	0	-0.69074	that
7	3	3	0	2.91731	oh
8	2	2	0	3.11662	years
9	2	2	0	3.96462	williams
10	2	2	0	1.45365	who
11	2	2	0	2.28654	us
12	2	2	0	0.84360	there
13	2	2	0	0.70158	so
14	2	2	0	5.70158	scary
15	2	2	0	1.89423	say

As opposed to the result from MICASE corpus analysis, *by the way* moved more freely within a sentence as is shown by its sentence-initial and sentence-final position accounting for as much as 45% in the Figure 6. This is mainly due to the fact that there was a considerable amount of turn takings taking place even within one transcript, which reflects the highly interactive nature of recorded speech in the COCA spoken corpus. Because neither of the speakers were ‘monopolizing’ the conversation for a long time, each speaker’s speech was either cut short at the utterance of *by the way* or initiated with the utterance of *by the way*.

Figure 7 demonstrates the prevalent usage of *by the way* in declarative sentences, which again corresponds to the result from the MICASE corpus analysis. Figure 8 highlights the elaboration (additional and qualifying) function of the sentence *by the way* modified, which accounted for as much as half (49%) along with ‘digression’ and ‘personal opinion’ function almost equally taking the place of the second most frequent function. There was a considerable increase observed in the percentage of ‘digression’ function (20%) compared to the percentage from the MICASE analysis (8%). This can be ascribed to the nature of the spoken COCA corpus which consists of radio shows, news, and interviews that are susceptible to sudden topic change given the short amount of air time.

Figure 9 shows that the contextual relevance of sentences *by the way* modified decreased in general, with the irrelevant cases accounting for

nearly one third (32%) of all the cases. This result may be a reflection of the increase in digression functions indicated in the previous Figure 8. If *by the way* is used to signal conversational digression, the following sentence would consequently lose its connection to the preceding sentence and be considered irrelevant to the topic discussed so far. After search of *by the way*'s collocations on AntConc, it was discovered that 'and' was again the most frequently co-occurring word with *by the way*. This result further supports the result from the Figure 8 in that 'elaboration' was still found to be the most predominant function of the sentence *by the way* modified.

4.3 Results from BYU-BNC

Figure 11.

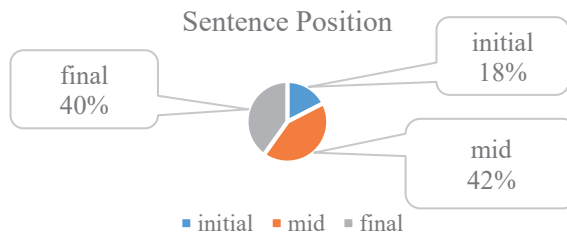


Figure 12.

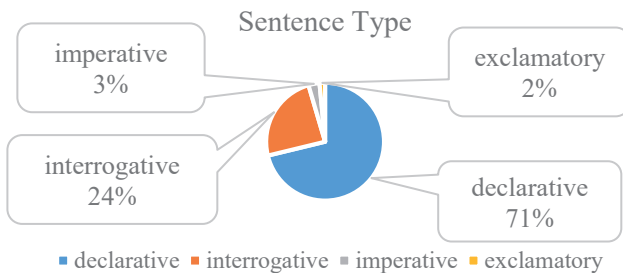


Figure 13.

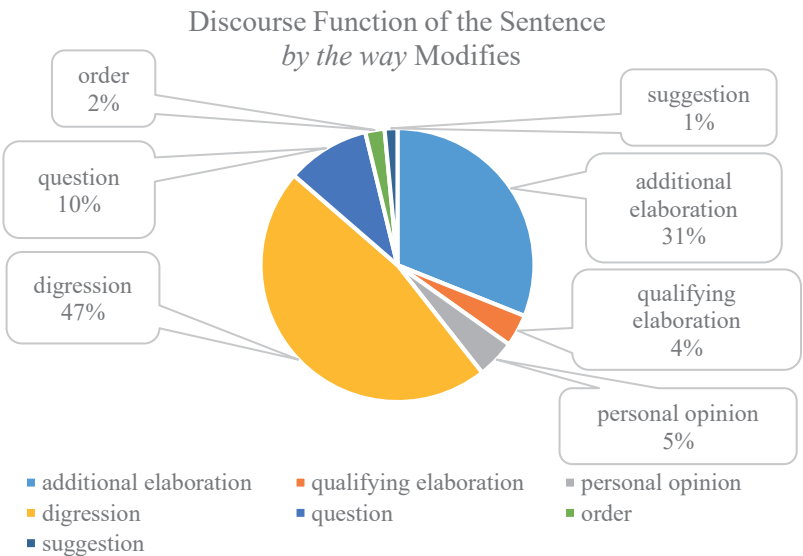


Figure 14.

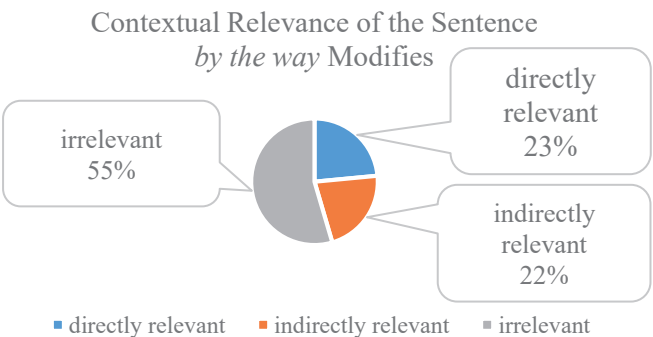


Figure 15. AntConc Collocation Search (Search Window 1L-1R)

#Total No. of Collocate Types: 112					
#Total No. of Collocate Tokens: 179					
1	25	25	0	3.15140	oh
2	8	8	0	1.06394	it
3	5	5	0	0.73636	and
4	4	4	0	1.18178	yeah
5	4	4	0	4.41444	runs
6	3	3	0	-0.95480	you
7	3	3	0	0.73636	unclear
8	3	3	0	1.36713	this
9	3	3	0	1.15140	said
10	3	3	0	-0.52416	pause
11	3	3	0	2.41444	back
12	2	2	0	1.15140	well
13	2	2	0	-1.79502	way
14	2	2	0	1.56644	there
15	2	2	0	4.15140	shit

Figure 11 shows that in the ‘conversation’ subcorpus of the BYU-BNC spoken corpus, *by the way* appeared almost equally in the sentence-final position and the mid-sentence position. Figure 12 shows that the sentence that *by the way* modified was mostly declarative, but a noticeable difference in sentence type is that exclamatory sentences were found for the first time and interrogative sentences also displayed a slight increase. Nevertheless, this result could be a mere reflection of the nature of the BYU-BNC spoken corpus as there would normally be much more diverse sentence types such as these exclamatory sentences in ordinary conversations. In the Figure 13, what first draws attention is the substantial increase in the percentage of ‘digression’ function (47%) in contrast to the percentage from the MICASE (8%) and COCA (20%) corpus analysis. In other words, this implies that about half of the sentences that *by the way* modified digressed from the immediately preceding sentence. If we take the speech data from the spoken BYU-BNC corpus to be the most representative of natural and ordinary conversations among these three corpora, this means that *by the way* would most likely act as a ‘digression’ marker in everyday discourse settings. Result from the figure 14 largely corresponds to the previous results. If a conversation takes a directional change to the extent that it virtually digresses from the original topic, the following conversation is highly likely to lose its relevance to the previous one and be perceived as irrelevant. Additionally, an interesting result was found from the

AntConc search. It was found that ‘oh’ was the most frequently co-occurring word with *by the way*. Speakers showed a strong tendency to combine ‘oh’ and ‘*by the way*’ together instead of using *by the way* on its own. ‘Oh’ is a well-known discourse marker that shows how a speaker has just thought of something different and intends to shift the conversation’s direction. Therefore, the fact that ‘oh’ and ‘*by the way*’ were often used in conjunction is evidence for the claim that *by the way*’s most primary function in casual spoken discourse is to signal digression.

5. Discussion

Table 2. Sentence Position in Each Corpus

	MICASE	COCA	BYU-BNC
Initial	10%	30%	18%
Mid	78%	55%	42%
Final	12%	15%	40%

Table 3. Sentence Type in Each Corpus

	MICASE	COCA	BYU-BNC
Declarative	83%	92%	71%
Interrogative	15%	8%	24%
Imperative	2%	0%	3%
Exclamatory	0%	0%	2%

Table 4. Discourse Function of the Sentence *by the way* Modifies in Each Corpus

	MICASE	COCA	BYU-BNC
Additional elaboration	44%	42%	31%
Qualifying elaboration	13%	7%	4%

Personal opinion	17%	21%	5%
Digression	8%	20%	47%
Question	13%	8%	10%
Order	2%	0%	2%
Suggestion	3%	2%	1%

Table 5. Contextual Relevance of the Sentence *by the way* Modifies in Each Corpus

	MICASE	COCA	BYU-BNC
Directly relevant	65%	44%	23%
Indirectly relevant	14%	24%	22%
Irrelevant	21%	32%	55%

Figure 16.

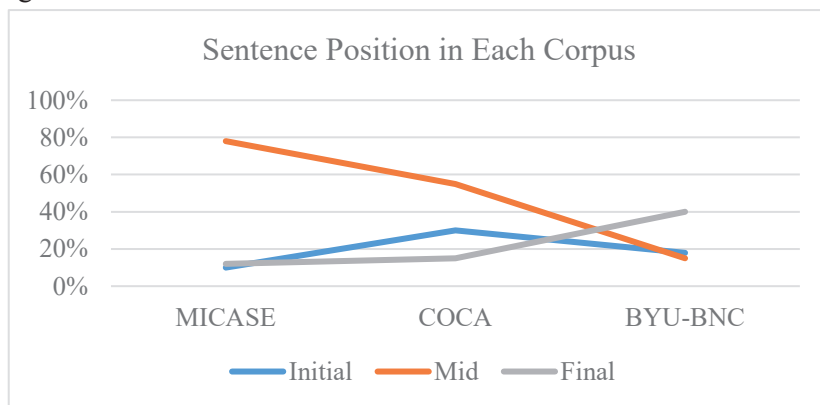


Figure17.

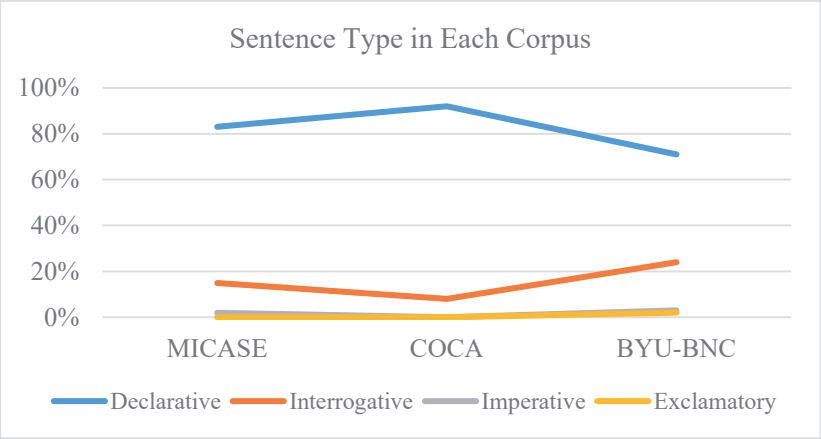


Figure 18.

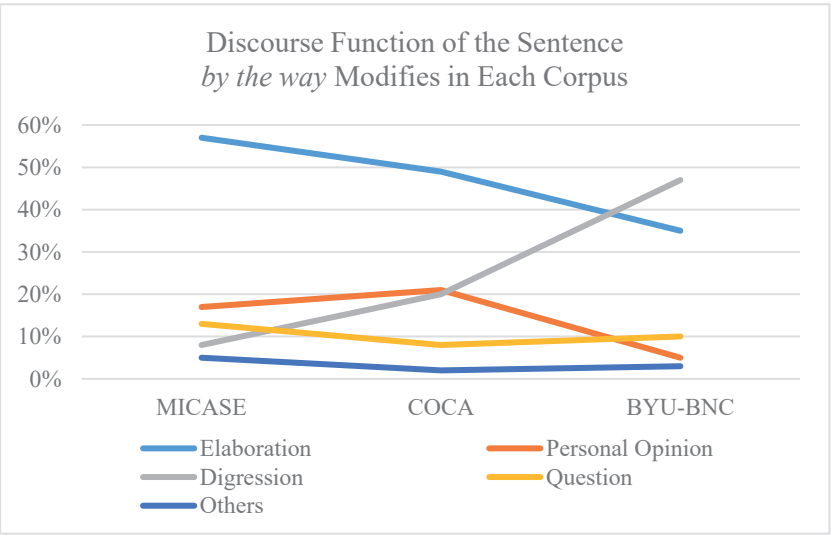


Figure 19.

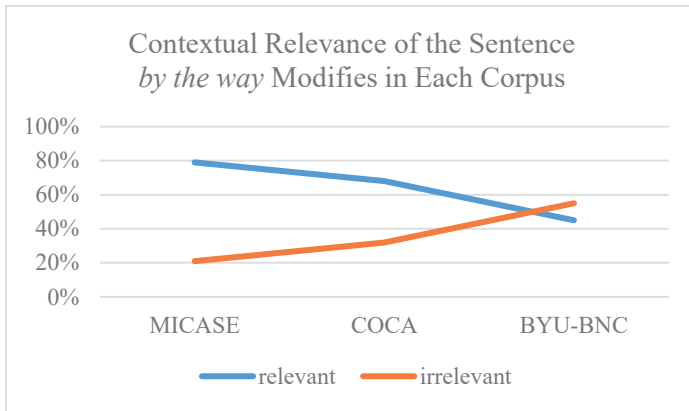


Table 2 to Table 5 combined all the figures from the previous result section in order to better compare and illustrate how *by the way*'s characteristics(features) differentially manifested themselves in each corpus. Figures 16 to 19 were created based on the same figures from these tables. Note that in Figure 18, figures from the 'additional elaboration' and 'qualifying elaboration' were added to calculate one final figure for 'elaboration'. Also, note that in the Figure 19, figures from the 'directly relevant' and the 'indirectly relevant' were combined to calculate one final figure for 'relevant'.

As is evident from the Figure 16, the position of *by the way* in a sentence significantly varied depending on the corpus in which it was found. If we take the difference among these corpora to be the difference in the discourse type, then we can logically infer that there arises a certain correlation between the position of *by the way* and the type of discourse in which it occurs. For example, in MICASE, *by the way* appeared predominantly in the mid-sentence position. However, in the COCA, the sentence position of *by the way* changed to become more flexible, as evidenced by the decrease in the percentage of mid-sentence position and an increase in the percentage of sentence-initial and sentence-final positions. In the spoken BYU-BNC corpus, the percentage of mid-

sentence *by the way* decreased substantially in contrast to the noticeable increase in the percentage of sentence-final *by the way*. This means that more flexibility was observed in *by the way*'s sentence position as discourse settings changed from formal to informal, unnatural to natural, and monologic to interactive. In the Figure 17, it was found that regardless of discourse type, declarative sentence turned out to be the most predominant type of sentence in which *by the way* was found. Exclamatory sentences were only observed in the spoken BYU-BNC corpus. In the Figure 18, it was found that discourse functions of the sentence *by the way* modified differed across the three corpora. Among the five representative functions, the 'elaboration' function and the 'digression' function displayed a significant difference in their percentage as the corpus of interest changed. That is, the 'elaboration' function showed a gradual decrease as the corpus of interest moved from MICASE towards the more informal BYU-BNC spoken corpus. In contrast, the 'digression' function demonstrated an incremental increase, eventually taking the place of the most predominant function in the BYU-BNC corpus. This reverse in pattern is once again repeated in the result from Figure 19 in which the contextual relevance of the sentence *by the way* modified showed a slow decrease as the corpus of interest changed to BYU-BNC corpus.

To sum up, the general trend seems that as conversations shift toward the more informal, interactional, and ordinary conversations with the reference corpus changing from MICASE to BYU-BNC, there is more flexibility in the sentence position at which *by the way* appears. The sentence type, however, remains unchanged with the declarative sentence type always being the most predominant sentence type observed across all three corpora. The only noticeable difference is the increase in the interrogative sentence types in the BYU-BNC corpus, which presumably correlates to the increase in the sentence-final position of *by the way* in the Figure 16. However, further inter-feature analysis is required to confirm this correlation, which is not within the scope of this

study. The discourse function of the sentence *by the way* modified also changed as a function of the discourse type. It was shown in Figure 18 that *by the way* tended to serve more frequently as a ‘digression marker’ in informal, natural, and interactive discourse settings. This observation was repeated in the Figure 19 that showed how contextual relevance of the sentence *by the way* modified dropped sharply in the BYU-BNC corpus. These observations mirror the variable nature of *by the way* in that they demonstrate how each function manifests itself differentially depending on different discourse settings. In academic discourse settings, *by the way* was predominantly used to introduce an ‘elaboration’ of the preceding sentence. On the other hand, in more casual speech settings, *by the way* was predominantly used to signal ‘digression’ in the form of topic change. Apart from these two dominant functions, there were additional functions identified upon close analysis of *by the way*’s usages, albeit with a small percentage. Put differently, this implies that *by the way* does not necessarily ‘predict’ which conversational direction the following sentence is going to take.

6. Conclusion

In conclusion, it appears that *by the way* does not have a single fixed, static discourse function and instead carries a set of different functions that are selectively constrained and realized in various discourse situations. Therefore, these discourse considerations also need to be taken into account to allow for its more precise characterization. Further inter-feature analysis (ex. sentence position- discourse function) is necessary to determine whether any other specific feature of *by the way* simultaneously informs its distinctive usage in a given context.

References

- By the way. 2018. In *Merriam-Webster Dictionary (online)*. Retrieved December 22, 2018, from <https://www.merriamwebster.com/dictionary/by%20the%20way>.
- By the way. 2018. In *Cambridge English Dictionary (online)*. Retrieved December 1, 2018, from <https://dictionary.cambridge.org/dictionary/english/by-the-way>.
- By the way. 2018. In *Longman Dictionary of Contemporary English (online)*. Retrieved November 31, 2018, from <https://www.ldoceonline.com/dictionary/by-the-way>.
- By the way. 2018. In *Collins English Dictionary (online)*. Retrieved December 1, 2018, from <https://www.collinsdictionary.com/dictionary/english/by-the-way>.
- By the way. 2018. In *Macmillan English Dictionary (online)*. Retrieved November 30, 2018, from https://www.macmillandictionary.com/dictionary/british/by-the-way_1.
- Bell, D. (1998). Cancellative discourse markers: A core/periphery approach. *Pragmatics*, 8, 515-541.
- Croucher, S.M. (2004). Like, you know, what I'm saying: A study of discourse marker frequency in extemporaneous and impromptu speaking. *National Forensic Journal*, 22(2), 38-47.
- Dailey, R.M. & Palomares, N. A. (2004). Strategic topic avoidance: an investigation of topic avoidance frequency, strategies used, and relational correlates. *Communication Monographs*, 71(4), 471-496.
- Fraser, B. (1988). Types of English discourse markers. *Acta Linguistica Hungarica*, 38(1-4), 19-33.
- Fraser, B. (2009). Topic Orientation Markers. *Journal of Pragmatics*, 41(5), 892-898.
- Halliday, M.A.K. & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Quirk, R. et al. (1985). *A comprehensive grammar of the English Language*. London: Longman.
- Schiffrin, D. (1987). *Discourse Markers*. Cambridge: Cambridge University Press.