



A Corpus-Based Pragmatic Analysis of Discourse Marker “*I Think*”

Lanping Li, Yang Pang

School of Foreign Languages, Donghua University, Shanghai, China

Email: 2053275693@qq.com

How to cite this paper: Li, L.P. and Pang, Y. (2022) A Corpus-Based Pragmatic Analysis of Discourse Marker “*I Think*”. *Open Access Library Journal*, 9: e9301. <https://doi.org/10.4236/oalib.1109301>

Received: September 11, 2022

Accepted: October 10, 2022

Published: October 13, 2022

Copyright © 2022 by author(s) and Open Access Library Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Prosperous development of corpus linguistics thrived the study on discourse markers, and among them, *I think* received a large share. Based on former researches, this paper adopted both quantitative and qualitative methods and made a contrastive analysis of the textual features of *I think*, choosing fiction and TV, these two various genres within COCA. The results demonstrate that there are sharp differences between these two genres in *I think* in terms of the frequency, position, collocation and function. There is still a certain correlation between where *I think* appears and the function it performs. In short, the new perspective study in this paper shed light on an investigation of other discourse markers.

Subject Areas

Cooperative Communications, Linguistics

Keywords

Discourse Marker, Corpus-Based, *I Think*, Pragmatic Function, Context

1. Introduction

Facing the advent of corpus linguistics in the 1960s, mass language information was available online, thus lots of research were conducted on figuring out the specific use of a certain word from a native corpus, especially discourse makers, aiming to shed light on the usage of non-native learners [1] [2] [3] [4] [5]. As a typical discourse marker, *I think* receives a great deal of discussion for decades and has been widely studied based on a corpus, involving kinds of domains, such as lexicology, second language acquisition even pragmatics [6] [7]. Among them, lots of studies were intended on the different functions of *I think* in native speakers [8] [9].

However, even though previous studies showed great importance to *I think* in various ways, fewer eyes were fixed on the comparison between genres. Moreover, it is known to all that most fiction can be adapted into a series of TV plays, which, to some extent, means that they are similar in styles, such as using similar plots or words. Based on this experience, the author will verify her observation in detail.

By far, the Corpus of Contemporary American English (COCA), as the most widely-used corpus, received worldwide acknowledgement for its large size and verified searching items. The corpus contains more than one billion words of data, including 20 million words each year from 1990-2019. In early 2020, it was dramatically expanded the scope and size and features of COCA to make it even more useful for researchers, teachers, and learners, which makes COCA the only corpus of English, that is, it has a wide range of genres. Thus, this paper will choose the two genres within COCA for further study.

Previous studies lacked attention to different genres, instead, researchers were still confined to comparative analysis of native and non-native speakers. Therefore, in this paper, it makes a contrastive analysis based on giants' shoulders, involving two various genres from COCA. More importantly, it explores the discourse marker *I think* by comparing the frequency, the position, the collocation and the function in two certain genres, fiction and TV, attempting to dig out their similarities and differences.

Generally speaking, the reflection upon *I think*, a typical discourse marker, lightens our way in second language acquisition, which can also mirror the pragmatic competence of English foreign language learners. Besides, this paper uses corpus-based analysis to investigate *I think*, covering two genres within COCA, which creates a new perspective of studying other discourse markers.

In order to make a pragmatic analysis of *I think* in two various genres, this paper covers five sections in total. The introduction tells the research background, the research purpose and significance, and also the structure of the holistic paper. The literature review delineates the study on *I think* at home and abroad, which indicates the research gap; then the taxonomy of function was also stated in this part, which paves the way for the latter part of this paper; then, detailed information about research tools, research methods and the principles used in this paper are described. Next section is the body part, in which it illustrates the differences and similarities of *I think* in fiction and TV in terms of the frequency, the position, the collocation and the function. At last, the conclusion shows the major findings, implications and limitations of this paper.

2. Literature Review

For decades, the research on *I think* can be roughly divided into two groups, according to their structural features. Just as Aijmer (1997) stated in a book named *I think-An English modal particle*, *I think* straddled grammar and discourse. Therefore, two groups mentioned above went to grammatical one and discursal one [10]. At the early stage, the researchers attached great significance to *I think*

in its grammatical use in reporting clauses, like Biber, Huang and O'keeffe. They were interested in the phenomena that there was always a *that* followed *I think* tightly, which was a symbol of reporting clause indeed. However, the flourishing development of Linguistics spurred further achievements in language, which argued that there was usually an omission of *that* next to *I think* in some sentences. This new finding challenged the research *I think* in its grammatical domain.

Meanwhile, numerous scholars also jumped into the studies of its discorsal use, holding that *I think*, as the main clause in a complement clause, functions a lot. Baumgarten and House (2010) supported that *I think* can express the speaker's points, belief and attitude in spoken discourse [11]. Moreover, Aijmer stressed its discorsal functions in his book, in which he pointed that *I think*, served as a DM, worked diversified as a link between contexts, allowing the discourse more coherent in between [12]. In addition, Qian Wang (2020) attempted to explore the different use of *I think* between ELF speakers and native speakers, in which he highlighted the discorsal functions of *I think*, such as topic changes of utterance, expressing contrasting views even online planning [13].

Admittedly, a great number of researches have been conducted on *I think* in different perspectives, however, relatively limited attention has been put on the analysis of discorsal use of various genres. Inspired by the recently published paper of Qian Wang, this paper will attempt to research further the function of *I think* in various genres.

2.1. Function Taxonomy of *I Think*

Holmes (1990) roughly distinguished the function of *I think* into *certainty* or *uncertainty* according to its contextual meaning between sentences. Later, Aijmer developed further and put forward: *tentative* and *deliberative* functions, in which tentative meant uncertainty while deliberative meant certainty. However, as mentioned above, this oversimplification of the function of *I think* was challenged later. Fortanet (2004) first distinguished the function taxonomy of *I think* based on its grammatical and discorsal use respectively, in which he named *epistemic* function and *interactional* function accordingly [14]. He proposed a taxonomy with six functions of *I think* in spoken academic English: 1) a basic epistemic function to express an opinion; 2) hedging functions to show vagueness, uncertainty or politeness; and 3) interaction functions as an approximator and a hesitation marker.

Qian Wang made a conclusion based on others' function taxonomy *I think*, and made a list below. 1) a basic lexical meaning, to express cognitive processes; 2) interaction functions to mark boundaries, do online planning, signal completion and seek a response; and 3) epistemic functions to hedge or boost the speaker's stance in both L1 and L2 English when they express their beliefs, opinions or subjective evaluations [13]. In addition, more and more scholars found that the *epistemic* function of *I think* weighed less because L1 and L2 seemed to use *I think* in a similar way, expressing similar attitude. Under this condition,

scholars jumped into researching its *discoursal* function as a DM.

Based on the taxonomy of *I think* mentioned above, the author makes a conclusion herself. No matter Holmes, Fortanet or Qian Wang, similarly, they mentioned several functions in **Table 1**.

In this paper, modelling on the taxonomy of others, this paper will cover its function as above. Until now, inspired by the Qian Wang's paper, this paper will attempt to figure out the following three research questions.

1), Are there any similarities or differences between *I think* in two various genres: fiction and TV?

2), What are the differences in terms of the frequency, the position, the collocation and the function?

3), Is there any relationship between the position of *I think* and the function it performs?

2.2. Data Collection and Research Methods

This paper chooses COCA as the source, in which it demonstrates the items of a certain word or phrase in various charts. In this finding, the author selects 50 texts in the fiction and TV column at a certain range of years randomly, thus the corpora are established. With the texts in hand, the current study can search *I think* in the corpus already made. First, once one clicks on the button of finding matching strings, all the concordance lines will occur, and the results can be saved. Secondly, it can use AntConc 3.5.7 (Windows), to draw upon techniques of corpus and text analysis in detail for its collocation and position. Lastly, for the function part, the author will make a judgement of its function according to the taxonomy function of *I think*.

AntConc 3.5.7 is a user-friendly software commonly used by linguists and capable of retrieving information that goes unnoticed, yielding illustrations of specific cases and identifying pragmatic markers that are difficult to discern. Overall, it has the capacity to produce sound empirical evidence for linguistic description and provide quantitative support for qualitative claims. Therefore, we choose Antcon as the research tool, in which it can search certain parameters correspondently in two texts already made, such as their frequency, position, collocate and functions. Based on the statistics, this paper will make a contrastive analysis, focusing on their differences.

Table 1. Function Taxonomy of *I think*.

-
1. Expressing certainty
 2. Expressing uncertainty
 3. Marking starting point/topic changes of utterance
 4. Expressing agreement
 5. Online planning
 6. Signaling completion and pursuing a response
 7. other functions unknown
-

In the following part, it will categorize functions of *I think* according to the taxonomy mentioned in the literature part. However, in this process, the author also needs to use major relevance theory as her theoretical foundation, which illuminates her way of distinguishing.

2.3. The Concept of Context

In *A New Concise Course in Linguistics for Students of English*, Dai Weidong mentions that context is considered as the knowledge shared by the speaker and the hearer. “Context determines the speaker’s use of language and also the hearer’s interpretation of what is said to him. Without such knowledge, linguistic communication would not be possible, and without considering such knowledge, linguistic communication cannot be satisfactorily accounted for in a pragmatic sense [15].” According to Sperber & Wilson (1995), relevance is defined in terms of processing effort and contextual effects. When new information is presented combined with existing assumption context, the contextual effect is aroused, which is also known as a cognitive effect. During the process of communication, the communicator combines the new information with the existing assumption with the aim to gain contextual implications. When the newly presented information achieves contextual effect, it is relevant in that context. At the same time, the higher the contextual effect is, the higher the relevance it is. That is to say, the relevance of utterances depends on their contextual effects.

2.4. The Concept of Communication

The process of communication involves both the speakers and the hearers, in which the speakers send their message to the hearers, and at the same time, the hearers should further infer the deep meaning within it, and then make a response. That is to say, “two information processing devices are involved. The first one is to modify the other’s physical environment, and the other is to encode. It means the speakers should say something already stored in the first device, making the informative language in tune with the context environment.” (Grice and David Lewis, 1992, 转自百度学术) For instance, in oral communication, when the speaker says something, he should observe the context environment built, as a result, the hearers can catch the correct meaning entertaining from the speakers. Moreover, in Aristotle’s viewpoint, the code model is the basis of all the communication theory, which involves encoding and decoding. Later, Grice and David Lewis propose the inferential model that “communication can be achieved by producing evidence and interpreting evidence.” (Grice and David Lewis, 1992, 转自百度学术) Then, Sperber and Wilson raise a different viewpoint that communication can be achieved through the ostensive-inferential communication, in which the communicator is associated with the ostension, and the audience is associated with inference.

2.5. The Concept of Inference

Human cognition tends to observe the CP and achieve the optimal relevance.

However, the phenomena of those who violate the communicative principle are not rare, therefore, making a correct inference is necessary. The inference is a process, in which the hearer should make a “local” response according to the “global” premise. For example, “What’s the matter with you?” is a global greeting to all, but it can be deduced according to the relationship and the situation locally, which will reduce our effort in making an inference. Inference can be reached from a syntactic or semantic point of view, which is the semantic relation of entailment and a syntactic relation of logical implication. From a semantic point of view, the meaning of sentences is of great significance. The speech from speakers should build a context environment, at the same time, it provides information. For the hearer, only by interpreting the meaning within the context environment from the speaker can he make an inference correctly. The more relevant the context environment is built, the less effort will be needed in making an inference. Secondly, from a syntactic point of view, the sentence structure will get more attention, especially the logical particles such as and, or and all. Because these small words can convey additional meaning except the conceptual one, which will disclose the relationship between sentences. In this way, the hearer can get the precise inference according to the logical partials.

3. A Corpus-Based Pragmatic Analysis of *I Think* in Fiction and TV

Based on the research methods and theoretical foundation in the last section, this part tries to explore the detailed information in response to its purpose. The compelling evidence covers overall frequency, the position, the collocation and the function of *I think* in fiction and TV in total.

3.1. The Overall Frequency of *I Think* in Fiction and TV

In response to the first question and finding the difference of *I think* in fiction and TV, the first item that it concerns is frequency. Due to the size inconsistency, the normed frequency was added here to standardize them, which makes it easy to compare. The specific statistics were showing in **Table 2**.

In **Table 2**, it is obvious that the normed frequency of TV is over 3 times that of fiction, which tells a sharp difference between the overall frequency of *I think* in TV and Fiction. In other words, *I think* occurs 3 times in TV texts than in fiction texts.

3.2. Position of *I Think* in Fiction and TV

In this part, the position of *I think* in Fiction and TV came next. In calculating this item, the author wants to test that whether there is any relation between the position of *I think* and the function it performs. With this aim, **Table 3** was shown below.

In **Table 3**, the difference also shocks the eyes because no *I think* occurs in the final clause in TV texts while there is almost 9 per cent in fiction texts. In fiction

texts, the frequency of *I think* appearing at the beginning of a sentence is almost 2 times more than appearing in the medial part of a sentence. In TV, the frequency of *I think* occurring at the beginning of a sentence or in the medial part counts almost 50% in total.

3.3. Collocation of *I Think* in Fiction and TV

Finishing the frequency and position, here comes the collocation of *I think* in fiction and TV in **Table 4**. In this part, it explores the difference in their collocation in L1 and R1, wishing to find more evidence.

Based on the statistics above, in R1, it is obvious that the pronoun *we* only accounts for 7.4% in fiction while 14.4% in TV. Besides, pronouns like *I* (37.2%) and *you* (14.4%) in TV almost occur 2 times more than in fiction. However, in L1, there is no sharp difference, but *yeah* here only occurs in TV, which well fits the feature of spoken communication.

In total, from the frequency of *it*, *you*, *she* and *we*, it can be roughly concluded that TV texts use a more informal voice while fiction texts, even adapted from TV, still stick to a relatively formal style.

Table 2. Overall frequency of *I think* in Fiction and TV.

	TV	Fiction
Raw frequency	118	108
Normed frequency	1.03	0.31

Table 3. Position of *I think* in Fiction and TV.

	Clause-Initial	Clause-Medial	Clause-Final
Fiction	65 (60%)	33 (31%)	10 (9%)
TV	60 (50.8%)	58 (49.2%)	0

Table 4. Collocation of *I think* in Fiction and TV.

Fiction	L1		Centre		R1	
but	15	14.3%	I think	it	20	18.6%
that	5	4.6%	I think	I	15	14.3%
and	3	2.8%	I think	she	9	8.3%
what	2	1.9%	I think	you	9	8.3%
actually	1	0.93%	I think	we	8	7.4%
TV	L1		Centre		R1	
can	8	6.8%	I think	I	44	37.2%
but	7	5.9%	I think	we	17	14.4%
yeah	5	4.2%	I think	you	17	14.4%
well	3	2.5%	I think	it	10	8.5%
and	2	1.7%	I think	the	7	5.9%

3.4. Functions of *I Think* in Fiction and TV

Based on all the information above, here moves to the most important part, to explore its function. In this part, the author should categorize *I think* with the principles aforementioned.

Based on the concept of context, communication and inference, the author categorizes *I think* in terms of their functions, which are showed in **Table 5**. In TV, most of *I think* is used to express the speaker's opinion, be certain or uncertain. Upon more pondering, it can find that there seems to be a relation between position and its function. In fiction, as shown in **Table 3**, there is 10% of *I think* occurs at the final clause, therefore it seemingly functions as signalling completion or pursuing a response. However, in TV, there is no *I think* occurs at the end of a clause, so there is no one to serve signalling completion or pursuing a response. Moreover, most of the *I think* occurred at the initial or medial part of a clause, thus the function of expressing certainty or uncertainty, online planning counts a bigger percentage. This correspondence between the position and its function implied the relation within them thoroughly.

3.4.1. Functions of *I Think* in Fiction

The analysis of the functions of *I think* suggests that there is a higher percentage of *I think* serving certainty or uncertainty even the online planning role. For example, *I think* in example 1 is preceded by an adverb, *Right*, which indicates the speaker's certainty. However, at the same time, the sentence ends with *I think*. In other words, *Right* creates a context for the author to make an inference of the speaker's attitude, which also indicates the completion of a sentence.

Example 1:

n't taken thirty minutes in ... like ... forever. # Right, I think. It's been ten years. Harris Alexander Pope. time traveler.

In addition, in example 2, *I think* is preceded by an adverb, *Personally*, which creates a certain context. Within such context, the speaker utters *I think* tightly, which can infer that *I think* demonstrates the uncertainty of the speaker's opinion here.

Table 5. Function of *I think* in fiction and TV.

Functions	Fiction (108)		TV (118)	
1. Expressing certainty	39	36%	59	50%
2. Expressing uncertainty	27	25%	23	19.5%
3. Marking starting point/topic changes of utterance	9	8.3%	11	9.3%
4. Expressing agreement	6	5.6%	12	10%
5. Online planning	13	12%	7	6%
6. Signaling completion and pursuing a response	7	6.5%	0	0
7. other functions unknown	13	12%	6	5%

Example 2:

302016FICFantasySciFiABC something strange going on in the northwest corner of this continent? Personally, *I think* Portland might be the local focus. # We already know there's an east-to-west.

Moreover, *I think* in example 3 also shapes a new context, in which the ellipsis before and after it infers the hesitation or the online planning of the speaker.

Example 3:

On the line, and I felt ... I felt something ... *I think* ... It felt like my sternum pressed back and then passed through my spine.

3.4.2. Functions of *I Think* in TV

In this part, it will use some typical examples to illustrate the function of *I think* in TV in detail. Half the percentage in functions of *I think* in TV is dominated by certainty role. In example 4, *I think* is preceded by *Okay*, which shapes a context for the author to infer the role of *I think*. In fact, *Okay* indicates that the speaker is certain about his opinions or ideas in total.

Example 4:

Weight into it. - Bud, quit playing around. - *Okay, I think* I got it. I think I got it. - She hurt my nose.

In example 5, *I think* was preceded by *no*, which implies that the speaker disagrees with the idea mentioned above. On the other hand, *no* can be used to infer that the speaker is expressing his countering opinions to the hearers.

Example 5:

You sure it didn't just walk off on its own? No, *I think* the shoulders were a little broader. The jaw was a little more chiselled,

However, *I think* in example 6 was different from that in example 5. *I think* in example 6 is preceded by the ellipsis, which creates the context that *I think* used for online planning or indicates the hesitation of the speaker.

Example 6:

Oh, well, you see, Terr, I think that the ... *I think* the '70s symbolizes the very best in America. Take The Avengers, the

4. Discussion

All the examples above serve the telling evidence of the functions of *I think* in fiction and TV. More specifically, it can also draw the conclusion that the position and the function of *I think* has a close relation. Example 1 gives *I think* in the final clause, in which it functions a role in the completion of a sentence; Examples 2, 4 and 5 show *I think* in an initial clause, in which it serves the role of expressing the speaker's opinion, be it certainty or uncertainty or disagreement. Examples 3 and 6 describe *I think* in the medial clause, which it shows the hesitation or the online planning of the speaker's attitude. In order to see clearer, it can be seen in **Table 6**.

Table 6. Distribution of position and functions.

Initial-clause	Certainty/uncertainty/ disagreement
Medial-clause	the hesitation / the online planning
final-clause	completion

5. Conclusions

Firstly, this paper attempts to explore the similarities and differences between *I think* in fiction and TV, and the results tell a lot about itself. In TV, there is no *I think* which occurs at the final part of a clause, which may also be understandable because we also seldom use *I think* at the end of an utterance in communication. On the contrary, almost 10% of *I think* appears in the final part of a clause in fiction. This may be explainable for its feature: there are words in the paper, so the readers are easy to get the information even in the end part.

Secondly, this paper finds that there is a loose relationship between the position *I think* and the function it performed. Commonly, *I think* occurs at the end part of a clause usually indicating the completion of a sentence. And *I think* occurs after the ellipsis, especially in the medial part, serves the online planning role mostly. If *I think* occurs at the beginning part, it is almost used to express the certainty or uncertainty of the speakers, but usually this can be inferred from other discourse markers related in position, such as *but*, *and*, *okay*, etc.

Lastly, the sharp difference of *I think* in fiction and TV counters may experience, in which the author thinks that most fiction can be adapted into a series of TV plays, which, to some extent, means that they are similar in styles, such as using similar words or plots. The new perspective in this paper can illuminate the study of the other discourse markers.

Firstly, the corpora established here are not big enough. The author chooses only 50 texts each to create the corpus, so it may affect the accuracy of the statistics. Secondly, in the taxonomy of the function of *I think*, the author finishes through her own judgment based on concepts of context, communication and inference. Even though there is a theoretical framework, the whole mental process of the author herself cannot be overlooked.

Acknowledgements

This research is sponsored by the Shanghai Pujiang Program (21PJC004), the Fundamental Research Fund for Central Universities (LZB2021006), and the DHU Distinguished Young Professor Program.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] Huang, L.-F. (2020) A Corpus-Based Exploration of the Discourse Marker *Well* in Spoken Interlanguage. *Language and Speech*, **62**, 1-24.

- [2] Wu, Y., Wang, J. and Cai, Z. (2010) The Use of “I Think” by Chinese EFL Learners: A Study Revisited. *Chinese Journal of Applied Linguistics*, **33**, 3-23+127.
- [3] 董芳源. 《哈克贝利·费恩历险记》中语用标记的功能研究[D]: [硕士学位论文]. 上海: 上海外国语大学, 2019.
- [4] 冉永平. 话语标记语的语用学研究综述[J]. 外语研究, 2000(4): 8-14.
- [5] 卫乃兴. 中国学生英语口语的短语学特征研究——COLSEC 语料库的词块证据分析[J]. 现代外语(季刊), 2007, 30(3): 280-291.
- [6] Haselow, A. (2019) Discourse Marker Sequences: Insights into the Serial Order of Communicative Tasks in Real-Time Turn Production. *Journal of Pragmatics*, **146**, 1-18. <https://doi.org/10.1016/j.pragma.2019.04.003>
- [7] 唐小利. 话语标记语“I think”的韵律分析[J]. 齐齐哈尔师范高等专科学校学报, 2011(1): 67-68+96.
- [8] He, Y.D. (2021) The Pragmatic Functions of Shields “I Think, I Believe” in TV Chat Shows. *Open Access Library Journal*, **8**, Article No. e7523. <https://doi.org/10.4236/oalib.1107523>
- [9] 初爱霞. 基于语料库的话语标记语 I think 的语法化和元功能研究[J]. 语文学刊, 2016(2): 8-10+19.
- [10] Aijmer, K. (1997) I Think—An English Modal Particle. In: Swan, T. and Westvik, O.J., Eds., *Modality in Germanic Languages: Historical and Comparative Perspectives*, Mouton de Gruyter, Berlin, New York, 1-48.
- [11] Baumgarten, N. and House, J. (2010) “I Think” and “I Don’t Know” in English as Lingua Franca and Native English Discourse. *Journal of Pragmatics*, **45**, 1184-1200. <https://doi.org/10.1016/j.pragma.2009.09.018>
- [12] Aijmer, K. (2002) English Discourse Particles. John Benjamins, Amsterdam. <https://doi.org/10.1075/scl.10>
- [13] Wang, Q. (2020) A Corpus-Based Contrastive Analysis of *I Think* in Spoken Hong Kong English: Research from the International Corpus of English (ICE). *Australian Journal of Linguistics*, **40**, 319-345. <https://doi.org/10.1080/07268602.2020.1823817>
- [14] Fortanet, I. (2004) I Think: Opinion, Uncertainty or Politeness in Academic Spoken English? *RAEL: Revista Electronica de Linguica Aplicada*, **3**, 63-84.
- [15] 戴炜栋, 何兆熊. 简明语言学教程[M]. 上海: 上海外语教育出版社, 2018: 64.