

Article

A Corpus-based Analysis on Pragmatic Fossilization in English Learners' Acquisition of Discourse Markers

Yueqi Ma¹ and Yi Zhang^{1,*}
¹ Northwestern Polytechnical University, Xi'an, China

* Correspondence: Yi Zhang, Northwestern Polytechnical University, Xi'an, China

Abstract: The study aims to investigate pragmatic fossilization in English learners' acquisition of discourse markers in oral English. Pragmatic fossilization refers to the inability to completely master and apply the target language, so comparison between non-native speakers and native speakers on pragmatic competence is essential for assessing the level of fossilization. The comparative corpora is built on 2025 "FLTRP-ETIC Cup" English Public Speaking Contest of provincial level (NNS) and 2025 Extemporaneous Public Speaking Finals of Montana State (NS). Based on Fung and Carter's functional classification of discourse markers, the study makes use of Antconc4.3.1 and SPSS (V.27) for quantitative analysis. The results indicate that for 23 groups of valid discourse markers, NNS presents 639 times while NS presents 474 times. Through chi-square test, the result that displays $\chi^2 = 73.182$; $df = 22$; $p(2\text{-tailed}) < 0.001$, can manifest there is significant difference in the pragmatic use of DMs in two corpora, which indicates that pragmatic fossilization does exist. Seven typical DMs--and, I think, like, well, yeah, you know, and absolutely are used for qualitative analysis. Through detailed analysis with examples, the results indicate some performances of pragmatic fossilization in non-native speeches. Firstly, the learners neglect some functions of DMs; secondly, they can not achieve the level of natives on using specific functions; thirdly, they can not internalize some functions and not use them naturally, resulting in overproduction. Meanwhile, reasons for those performances are illustrated--lack of language environment and opportunities for practicing, domestic pedagogical orientation toward examinations, negative transfer from native language system. It is expected that the paper can provide examples and insights into understanding pragmatic fossilization of discourse markers in oral utterances.

Keywords: pragmatic fossilization; discourse markers; oral discourse

1. Introduction

In the field of Second Language Acquisition (SLA), interlanguage has long been a critical research focus since Selinker [1] introduced the term into the field. According to him, interlanguage referred to a dynamic and unique linguistic system developed by second language learners during the acquisition. Based on language input, it differed from both the native language and the target language, but gradually approximated the target language as the learner's proficiency improved. At the same time, Selinker [2] proposed that one of the most important features of interlanguage was fossilization, which posed a significant obstacle in language acquisition, severely restricting learners' proficiency and hindering their ability to further master and apply the language. For non-native language learners, they can not completely master second language, so there must be some linguistic forms or uses deviating from target language kept in their interlanguage [3]. In this sense, study on fossilization has attracted extensive attention in SLA in recent years. Scholars tended to explore phonetic representation [4], vocabulary [5], and syntactical fossilization [6], yet few focused on pragmatic aspect [7]. As pragmatic competence is closely related to the daily use of language and directly reflects the distance toward target language, comparison between non-native speakers and native speakers on pragmatic competence is essential for assessing the level of fossilization.

Received: 30 July 2026

Revised: 05 September 2026

Accepted: 20 September 2026

Published: 25 September 2026



Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

As a very common language phenomenon in daily conversation, discourse markers (DMs) have become the main focus for linguists from different fields. According to interlanguage pragmatics which applied research achievements of SLA to pragmatic level of interlanguage [8], pragmatic competence of second language learners could be understood as the competence to comprehend and conduct speech acts and regulate utterances. Discourse markers were agreed upon by a variety of approaches as syntactically independent segments that were used in discourse, which had little propositional meanings and exerted little influence on the semantic meaning of the utterance [9]. However, they could be used to effectively regulate cohesion and coherence of sentences and utterances, express propositional attitudes and help build positive interpersonal relationship [10]. In this sense, how to use DMs correctly can be regarded as a standard to assess the pragmatic competence of second language learners [11].

The paper aims to investigate pragmatic fossilization in English learners' acquisition of discourse markers in oral English by building comparative corpora, non-native speakers (NNS) and native speakers (NS). Through comparison, the article intends to explore the distributional features of DMs used in two corpus and to analyze the differences and reasons behind them, offering insights to understand fossilization phenomenon occurring in non-native English learning process.

In terms of theoretical value, this study contributes to the theoretical understanding of pragmatic fossilization by examining its manifestations in the acquisition of discourse markers among English learners. By employing a corpus-based approach, the research bridges gaps between second language acquisition theories and pragmatics, particularly in how internal and external factors interact to impede learners' pragmatic development. Besides, the findings may refine existing models of fossilization by incorporating empirical evidence on discourse markers.

In terms of practical value, by identifying specific discourse markers prone to fossilization, the study provides educators with targeted areas for intervention, enabling the development of focused instructional strategies to mitigate persistent pragmatic errors. The corpus-based methodology offers a model for diagnosing fossilization patterns, which can integrate the design of authentic teaching materials and assessment tools.

2. Literature Review

2.1. Pragmatic Fossilization

The American linguist Selinker was the first to propose the theory of fossilization within the interlanguage system of second language learners. Selinker offered a more recent definition of fossilization as "a cessation of learning before the interlanguage system reaches the target language norms, resulting in L2 learners acquiring only incomplete knowledge of the target language." This phenomenon referred to the stagnation of learners' interlanguage development before attaining native-like proficiency. Researches in second language acquisition primarily focused on fossilization within the linguistic structure of interlanguage, such as phonology, morphology, and syntax, where learners persistently acquired certain correct or erroneous forms, leading to deviations from target language norms. However, pragmatics dimension was unexplored. Based on previous understanding, interlanguage pragmatics advocated applying key SLA findings to the pragmatic dimension of interlanguage, particularly by extending the concept of transfer to the pragmatic aspects of interlanguage, focusing on the pragmatic negative transfer by native pragmatic rules and cultures on interlanguage pragmalinguistic and sociopragmatic knowledge and competence. In this sense, pragmatic transfer, along with other contributing factors, may lead to fossilization at the pragmatic level of interlanguage. Pragmatic fossilization manifested as a discrepancy between L2 learners' pragmatic knowledge and that of native speakers, preventing learners from achieving native-like competence [12]. When communicating in interlanguage, learners may lack certain necessary pragmatic abilities or exhibit incomplete development in pragmatic skills, failing to meet native-speaker standards, which could be observed when learners either

employed inappropriate expressions or rigidly adhered to certain correct or incorrect pragmatic functions of linguistic forms [13]. While pragmatic fossilization did not necessarily result in communicative failure, it may still impair communicative effectiveness to varying degrees, hindering the convergence of learners' interlanguage pragmatic systems with those of native speakers [14]. Researches have explored the systematic and persistent errors in learners' pragmatic competence, such as the overuse or underuse of pragmatic markers, fixed patterns of interaction [15], and the influence of language transfer and cultural differences [16]. Besides, some corpus-based studies have also played a crucial role in identifying and analyzing these fossilization patterns, providing empirical evidence of learners' pragmatic development [17, 18]. Despite these advances, the field remains limited. As for the objective, few focused on learners' competence to regulate utterances and correctly express propositional meanings. As for the method, while corpus-based methods offered rich data, they lacked diachronic or cross-cultural comparative studies to fully capture the complexity of pragmatic fossilization.

2.2. Pragmatic Functions of DM

Studies on functions of DM can be classified into different methods and schools. Quirk et al [19] classified four cohesion relations and eight forms and sort "you know" into the form of discourse deixis functioning as grammatical cohesion. Van Dijk [20] also claimed his classification of discourse markers and took "you know" as one example parenthesis serving as connection between meaning. Thirty years later, Fung and Carter [21] proposed their functional taxonomy (Table 1) on this issue, and there were four dimensions--interpersonal, referential, structural, and cognitive domains.

Table 1. Classification of DMs(Fung & Carter, 2007)

Domain	Type	DMs
Interpersonal Domain	Markers of Shared Knowledge	see, you see, you know, listen
	Markers of Attitude	well, really, I think, obviously, absolutely, basically, actually, exactly, sort of, kind of, like, to be frank, to be honest, just, oh
	Markers of Response	ok/okay, oh, right, alright/all right, yeah, yes, I see, great, oh great, sure
	Markers of Cause	because, cos
Referential Domain	Markers of Contrast	but, and, yet, however, nevertheless
	Markers of Conjunction	and
	Markers of Disjunction	or
	Markers of Result	so
	Markers of Digression	anyway
	Markers of Comparison	likewise, similarly
Structural Domain	Markers of Topic Initiation/Closing	now, ok/okay, alright/all right, well, let's start, let's discuss, let me conclude the discussion
	Markers of Sequencing	first/firstly, second/secondly, then, next, finally
	Markers of Topic Shift	so, now, well, and, what about, how about
	Markers of Inference/Summary	so
	Markers of Topic Continuation	yeah, and, cos, so
Cognitive Domain	Markers of Cognitive Processes	well, I think, I see, and

Markers of Reformulation	I mean, you mean, that is, in other words, what I mean is, to put it in another way
Markers of Elaboration	like, I mean, so
Markers of Hesitation	well, sort of
Markers of Monitoring	you know
Listener's Knowledge	

Based on previous classification of DMs, scholars intended to explore more and more functions of these linguistic tools, such as textual, expressive and interpersonal functions [22]. Subsequently, a variety of schools agreed that DMs were (meta) pragmatic. Discourse markers could serve different pragmatic functions when put in different situations. It could be connected with core concepts of conversation analysis. Yu and Wu claimed their understanding on meta-pragmatic functions of discourse markers like stressing the topic, expressing the attitude, confirming, opening a new topic [23]. Ran [24] classified pragmatic functions of "you know" generally, such as reminding of the meta-knowledge, that is, to confirm the cognitive mutuality and stressing the commonsense. Heng [25] provided a panoramic explanation on the pragmatic use of "and" in spoke English-interruption, inserting added message, and switching the turns. Ren [26] also categorized a new coding scheme with the comparison between Chinese L1 speakers and L2 learners' identification of functions of DM, and the sociocultural factors were unfolded. However, it is proved that few studies concentrate on the field of second language acquisition, especially interlanguage and fossilization, to explore the differences of the pragmatic use of DMs between language learners and natives in utterances. Comparison between two groups on this topic can be essential for the understanding of how learners use DMs differently from natives and what the reasons are.

3. Methodology

3.1. Research Questions

Based on the above literature review and research gap of pragmatic fossilization, the paper aims to investigate how Chinese English learners use DMs differently from native speakers in their impromptu speeches and what the external and internal reasons for pragmatic fossilization on DMs are. The paper intends to answer the following questions:

What are the features of discourse markers used in non-native speaker corpus and native speaker corpus?

What are the differences of pragmatic functions of discourse markers used by non-native speakers and native speakers?

What are the underlying reasons for this pragmatic fossilization phenomenon?

3.2. Research Instruments

3.2.1. AntConc 4.3.1

AntConc 4.3.1 is a widely used corpus analysis tool designed for processing and analyzing large text datasets. In this study, the original speech scripts from two corpus will be input into TXT version and then imported into Antconc to retrieve discourse markers according to operational definition. In addition, KWIC (Keyword in Context) of discourse markers used will be explicit in the process, which helps return back to original context to better understand the meaning and function of markers to make comparison.

3.2.2. SPSS (V.27)

IBM SPSS Statistics 27.0 is a widely used statistical analysis tool that enables to process and interpret quantitative data effectively. It provides a range of statistical techniques, including descriptive statistics, correlation analysis, regression models, and hypothesis testing, making it a powerful tool for social science research. In this study, comparison is elaborated through chi-square test to explore if there is any significant difference of the use of discourse makers between Chinese students' and native students' speeches (NNS & NS). Discourse markers that exist in both speeches will be tested by

cross-table(2*N). By conducting statistical tests, the study quantitatively validated the qualitative analysis, reinforcing the robustness of the research findings.

3.3. Research Data

The paper intends to build comparative corpora for detailed analysis--one is based on Chinese college students' impromptu speeches (NNS), and the other is from Montana and Kentucky college students' extemporaneous speeches (NS). To be more precise, NNS (54.24mins) is selected from the transcribed scripts of 2025 "FLTRP-ETIC Cup" English Public Speaking Contest of provincial level, which is regarded as one of the most convincing and highest-standard English contests in China; while NS (53.20mins) is chosen from 2025 Extemporaneous Public Speaking Finals of Montana State. In view of the fact that prepared speech is the oral form of written scripts, which is insufficient for the study of discourse markers because DMs are the main feature of oral utterances [27], the paper only extracts impromptu speech part. As for the level of contestants, for NNS, all the seventeen contestants are from first-class universities in China with certificate of CET-6 to show high-level of English proficiency; for NS, they are all highly-educated students and represent the best level to participate in finals. The detailed information of the comparative corpora is shown as following:

3.4. Research Procedures

Research procedures can be divided into four stages. The first step is to build the comparative corpora in TXT version and retrieve discourse markers based on Fung and Carter's classification (Table 1) and KWIC through AntConc4.3.1. The second step is to classify retrieved markers according to two corpora, making cross-tables (2*N), and to conduct chi-square test through SPSS (V.27) to confirm the significant difference of the frequency. Then, the third stage is to conduct quantitative analysis to explore the distributional features of used discourse markers in two corpus respectively, and to return back to original script to analyze different pragmatic functions of typical discourse markers for qualitative analysis. After that, the results obtained can be used to answer research questions and potential reasons for pragmatic fossilization of Chinese students will also be exploited and finally come to conclusion.

4. Results and Discussion

Chapter four presents results and discussion. Firstly (4.3), through Antcon4.3.1 for retrieving discourse markers based on Table 1 and SPSS (V.27) for chi-square test based on the built cross-table of frequency of DMs in two corpora, the distributional features of used DMs in NNS and NS will be manifested. Secondly (4.2), typical examples of different functions of DMs used in NNS and NS will be extracted to analyze how Chinese English learners use linguistic tools differently from native speakers to check level of pragmatic fossilization. Meanwhile, reasons to explain the phenomenon will be analyzed.

4.1. Distributional Features of DMs

The paper aims to investigate pragmatic fossilization of the use of DMs happening in non-native English learners in their oral utterances. According to Table 2, the general frequency of DMs in NNS is little bit more than native speakers, which can indicate that in this study, non-native speakers' general willingness and consciousness to use DMs in their impromptu speeches are very high. Although it is proven that this sub-system of interlanguage pragmatic competence is approximating native speakers, the possibility of overuse of DMs can not be neglected. In this sense, it is indicated that interlanguage pragmatic competence is closed related to the level of education, and the higher the level of learners is, the closer the competence is to the target language, but this trend may cause overuse. To be more precise, in this section, the comparative corpora will be processed through Antcon4.3.1 to retrieve discourse markers listing in Fung and Carter's Table (2.1), then the data will be used to build cross-table to conduct chi-square test to validate if there is any significant difference in the use of discourse markers in NNS and NS. The result will be a scientific proof for the pragmatic fossilization in Chinese learners' English

proficiency. Except for the markers that are not use in both corpora, the cross-table of the frequency of the used DMs in two corpora is shown as follows (Table 3).

Table 2. Information of Corpora

Table 3. Frequency of DMs in NNS and NS

Corpus	DMs	Length	Tokens	Frequency
NNS	639	54.24mins	7666	11.78
NS	474	53.20mins	6452	8.90
Frequency of DMs		NNS	NS	
and		260	204	
I think		57	16	
just		48	14	
or		43	15	
but		39	30	
so		38	45	
like		26	22	
actually		20	7	
first		17	12	
well		8	15	
that is		6	10	
second		6	1	
however		5	1	
yeah		5	0	
you know		4	6	
see		3	9	
to be honest		1	0	
firstly		1	0	
I see		1	0	
you see		0	1	
obviously		0	1	
absolutely		0	4	
exactly		0	3	
Totally		639	474	

After importing the table above into SPSS for chi-square test, the result that displays $\chi^2 = 73.182$; $df = 22$; p (2-tailed) < 0.001 , rejects null hypothesis and indicates that there is significant difference between non-native speakers and native speakers in using DMs in their speeches. It is proven that pragmatic fossilization does exist in learners' language learning. In Table 2, it is manifested that and, I think. just, or, but, like, actually, first, second, however, yeah, to be honest, firstly, I see, are used more than native speaker, while so, well, you know, see, you see, obviously, absolutely, exactly are less than native speakers. In view of the similar length of video, DMs in NNS is used more and much more frequently than NS, displaying the possibility of overproduction of DMs, which features one situation of fossilization, because the pragmatic competence of non-native speakers does not reach the standard of native speakers.

4.2. Comparison of Pragmatic Functions of DMs

This section presents qualitative analysis by selecting seven typical discourse markers with examples--and, I think, like, well, yeah, you know, and absolutely. The comparison of the seven ones displays discrepancy in NNS and NS, requiring specific examples to illustrate. The comparative analysis on these different usages in two corpora will help illustrate the specific performance of learners and reasons for pragmatic fossilization.

4.2.1. And

As the most used DM in both corpora, and can reflect learners' strong consciousness to use it in the speech. However, from Table 3, the frequency of and in NNS is much more than its frequency in NS, which leads to overproduction in learners' English. In the sense of Li' s standard to judge fossilization, there is discrepancy in the use of and. On the other hand, the pragmatic functions of and in NNS and NS are different. Based on Table 1, in NNS, this discourse marker is mainly exploited to function as marker of conjunction to connect several terms together or list terms at the same time. In Example (1), and was used twice to make a sentence, which contained too much information at one time, Chinese students tended to use and to connect different items together. However, in NS, and occurs 206 times and functions as marker of conjunction, topic changing, and continuation. Native speakers are able to use it more naturally than non-native speakers. In Example (2), and here was used to switch the topic of speaker' s friend, with preceding description of his poor life. The speaker wanted to show his surprise to his friend' s big change. In Example (3), when talking about agriculture programs, the speaker intended to call on audiences' unity, so she firstly described the serious situation, and then use and to continue her emotion and push her topic into the next stage. In this sense, native speaker have ability to use and more flexibly and naturally than non-native speakers. In NNS, non-native speakers simplify the function of and on the one hand, and lacks the ability to internalize the use of and when it functions as conjunction, which leads to overuse of it.

When taking the rules of English learning in non-native countries' schools into consideration, it can be explained that Chinese students fossilize the rule of conjunctions like and, but, so. In their internal system, and is for conjunction, but for contrast, and so for result, meanwhile, this rule is also applicable in Chinese, so the influence of native language can help fossilize the rule to solely regard and as a conjunction.

Example (1)

When changing the world, the opportunities and challenges spur up at an extraordinary speech surrounded by technologies, challenges and different ideas. (From NNS)

Example (2)

He's had multiple wrecks that have almost taken his life. And over the couple of years, I've seen a change in him. (From NS)

Example (3)

The politics behind it are unknown. Even the agriculture itself is unknown. And that's why we need these programs to help get us out of it. (From NS)

4.2.2. I Think

If the frequency of and is close in NNS and NS, the frequency of I think in nonnative speakers' English is unbalance comparing to native' speakers. In NNS, I think appears 57 times. Though it can be understood that it is a speech contest which requires contestants to express their opinions toward a topic, it is overused when comparing to native usage. After classifying, Chinese students tended to use I think as marker of attitude. Once they needs to express their perspectives, I think can be functional. In Example (4), then considering a topic about AI's influence on human life, the speaker internally judged it as an argumentative topic and naturally gave her stance. Other contestants almost did in the same way, especially in impromptu speech with social issues and it is found that there should be nearly twenty times of I think in one speaker's three minutes' speech. However, in NS, I think is used much more diversely though there is only sixteen times. Beside functioning as marker of attitude, I think can also function as marker of cognitive process according to Fung and Carter's table. In Example (5), when talking about the topic of agriculture program, the speaker used I think to display her mental process about it. Actually, it was strategic and vivid to use I think in this way, because even though it was a similar argumentative topic, the speaker did not directly show her perspective, but led the audiences step by step along with her cognitive process when facing this issue, that is, from speech contest to agriculture issue. The process was smooth, humorous, and natural,

which was not found in NNS. In this sense, non-native speakers master one classic usage of I think to show attitude, but they fossilize this function and use it deliberately.

The reason for this fossilization can be better explained by educational process. On the one hand, I think appears for the first time in junior high school's textbook, and at that time, teachers tell student when coming across topics requiring attitudes and opinions, they must use I think. Frankly, comparing with other phrases to express stance like from my perspective, as far as I am concerned, which requires recitation for examinations, I think is much easier and shorter when used in speeches. As time goes by, students gradually carve this function in their hearts. On the other hand, relevant practices are insufficient. Chinese student tend to answer questions starting with I think no matter what question it is, which becomes a flawless formula, but other situations to use I think are not considered.

Example (4)

That's really good question and I think it's kind of really a philosophical question and I believe it's not paradox. (From NNS)

Example (5)

I think about things like prepared public speaking, extemporaneous public speaking, as well as agri science fair projects and employment skills. (From NS)

4.2.3. Like

This marker is used similarly in frequency in NNS and NS, but there is also difference. According to calculation, non-native speakers use it more than native speakers, and in their scripts, then tended to use it as the marker as elaboration and exemplification. In Example (6), when speakers intended to explain one terminology, the best way is to exemplify what the term is by listing some examples. In NNS, except for for example as the most classic phrase to listing terms, like is regarded as the simplified and oral form of for example, so it is exhaustively used as marker of elaboration. However, in NS, like can be used in another way, that is, to express the speaker's attitude in the way of analogy. In Example (7), when talking about the issue of company competition, the speaker vividly used her feeling as a speech contestant to show her attitude that this experience can be beneficial. This function is not found in NNS. It is comforting that Chinese students notice the main function of like to replace for example, which can be regarded as the product of exam-oriented education, and are conscious about using like in their oral utterances, but they are stuck in one function and tend to use it solely in one occasion, which may lead to overproduction. It is found that several Chinese students tend to continuously use like in three minutes.

It can be explained that highly-educated non-native students tend to avoid using for example in their speech, because this phrase is not natural in oral form. They actively learn from the local oral expressions of natives and notice the use of like in standard English, so they tend to use it to replace the long phrase on purpose. However, lacking of enough practices to use like in other ways, they can not completely internalize the use of like, which results in unnatural use of it.

Example (6)

Manufacturing Hubs become a very serious and popular issues over the whole globe, especially with the emergence of countries like Vietnam and Bangladesh and Mexico. (From NNS)

Example (7)

Seven leadership development events and ID members competing in these events like myself have the opportunity to develop leadership and business skills such as public speaking. (From NS)

4.2.4. Well

As a typical discourse marker used in oral utterances, well can reflect the level of English in a way. In the study, from Table 3, natives tend to use well more than non-native speakers. In Fung's functional classification, well can serve five functions--marker of attitude, topic initiation, topic shift, cognitive process, and hesitation, which can explain

the popularity of well in natives' daily conversation. In both NNS and NS, the last function is not found. As for the non-native speakers, it is illustrated that they used well in their speeches to initiate the topic as in Example (8) and switch the topic as in Example (9). As for the native speakers, except using well from structural domain in Example (10), they intended to use it to display cognitive process about their layouts of the speech when considering what leadership is in Example (11), and exploited well to express attitude toward mental health issues in Example (12). It is comforting that non-natives in the study have accepted high education and have strong motivation to study English, because they make efforts to make their English as local as possible by learning structural functions of well as natives to replace some written forms in exams. However, due to the lack of direct language environment and opportunities to practice, they can not totally achieve the level of natives. In terms of frequency, well in NNS is less than in NS, which indicates that non-native speakers' consciousness to use it is not strong; in terms of pragmatic function, natives tend to use well more naturally and diversely reaching three domains in functional framework than non-natives who solely regard well as a tool to organize the structure of the utterances. In this sense, non-natives do not reach the standard of pragmatic competence, resulting in fossilization of the use of well.

Example (8)

Well, I would say I stand for the first two perspectives. (From NNS)

Example (9)

I think they will be the best organ qualified, qualified psychiatrists in the future, well, however many cases happened. (From NNS)

Example (10)

Well, today I'm going to talk to you about the farmer mentality, about the severity of this issue and about what policies are out there that can help with it. (From NS)

Example (11)

Being a state senator and bringing policy forward, absolutely. Well, this is an example of leadership right here. (From NS)

Example (12)

How do we avoid that? Well, the older generation is definitely a hard one to reach. (From NS)

4.2.5. Yeah

It is interesting that in the study only non-native speakers used yeah in their speeches, but in previous studies yeah is a powerful evidence for high proficiency. In view of the quality of speech into consideration, there is several minutes for question time in Chinese contest and students are required to answer them in one minute, while there is null in English contest. According to Fung's framework, there are two functions of yeah--marker of response and topic continuation. In NNS, it is found that except twice for response as in Example (13), the rest are used to build confidence for the speaker, which does not belong to the pragmatic function in the framework. In impromptu speech, the contents are not prepared in advance, almost eighty percent of the speech is extemporaneous, which presents great challenge for non-natives to express themselves smoothly. For speakers in the study, they are contestants who have accepted professional training of great state of mood, so some discourse markers can be used by them to build confidence for themselves and make them as calm as natives. Yeah becomes a good choice because it contains the semantic meaning of affirmation. Though this usage is not found in natives' speeches, which indicates that non-natives tend to fossilize this usage due to lack of confidence and native semantic meaning, these speakers' willingness to express themselves and motivation to make themselves as local as possible should be encouraged.

Example (13)

Okay, thank you for your question. Well, yeah it's true that a biography ticket couldn't make us speak well overnight. (From NNS)

Example (14)

Oh well ladies and gentlemen it's a very funny test, yeah, and nowadays master-taken has become a trend of one generation.

4.2.6. You Know

As one of the most used discourse markers in oral English, you know has been studied from different perspectives. In this study, native speakers are more familiar to it and more able to use it properly in their speeches than non-natives. According to Fung's framework, you know is powerful to regulate utterances and build positive interpersonal relationship between speakers as a marker of shared knowledge and monitoring listener's knowledge. In NNS, you know appears four times in the same way--confirming shared knowledge, because in Chinese students' subconsciousness, there are judges in the audiences who are senior to them, they have no ability to control the knowledge of judges but only exploit some information that is shared by both and can resonate the audiences as in Example (15) to confirm China's achievements in the world. However, in NS, speakers can use you know to control the listeners' knowledge and lead them to think about the topic as in Example (16), in which the speaker raise a question about health programs and she used you know to control audiences' minds to think about other issues happening in Washington, D.C. This function is not found in NNS because you know is an oral and local marker in daily conversation, which can only be mastered in language environment. Besides, though in Chinese there is a similar phrase *nizhidaoma* which has same semantic meaning with you know, Chinese people tend not to use it in formal situations. Nevertheless, strategically using this phrase can add good effects in speeches especially in English speeches. In this sense, non-natives solely notice one function of you know that can be used in speech for resonance, but they tend to fossilize this function and neglect other usages.

Example (15)

It's a problem of time because, you know, in Goubi Desert there are so many different solar plants and some wind turbines. (From NNS)

Example (16)

Is this becoming an issue for these programs, these health programs? Are they seeing, you know, are they able to advance with all of the issues that are occurring in D.C.

4.2.7. Absolutely

As same as exactly and obviously, absolutely only appears in NS, which means that non-natives have no consciousness to use it in their oral utterances. Absolutely can be a powerful tool to express attitudes. Chinese students have stored the knowledge in their minds, but only for written forms like compositions. In domestic English classes, absolutely is high-frequency word in examinations, so students tend to use it in their compositions to get higher grades, but they have no willingness to use it in oral expression out of lack of practices. In NS, absolutely is highly used to express speakers' attitude and feeling toward the topic as in Example (17). The speaker found that sports could provide unique leadership chances and he was determined to do so in early times. This usage is local and emotional, which is not found in NNS, resulting in non-natives can not achieve the standard of pragmatic competence in this sense.

Example (17)

Uniquely offer in these leadership skills that you think maybe some other organizations don't have to? Absolutely. So I've been involved in sports since I was very, very young. (From NS)

5. Conclusion

The study intends to investigate pragmatic fossilization of the use of discourse markers in oral speeches by selecting 2025 "FLTRP·ETIC Cup" English Public Speaking Contest of provincial level (NNS) and 2025 Extemporaneous Public Speaking Finals of Montana State (NS) as comparative corpora. Through the instruments of Antconc4.3.1 and SPSS (V.27), and comparative analyses with specific examples, the paper accomplished mixed research to answer research questions. As for the first research question, by

retrieval in Antconc, there are 23 groups of valid DMs, with 639 times in NNS and 474 in NS. Through chi-square test in SPSS, the result that presents $\chi^2 = 73.182$; $df = 22$; $p(2\text{-tailed}) < 0.001$, can manifest there is significant difference in the pragmatic use of DMs in two corpora, which indicates that pragmatic fossilization does exist. The result echoes previous study of comparative interviews. As for the second question, the paper selects seven typical DMs for specific analysis--and, I think, like, well, yeah, you know, and absolutely. Through detailed analysis with examples, there are some performances of pragmatic fossilization in non-native speeches. Firstly, the learners neglect some functions of DMs; secondly, they can not achieve the level of natives on using specific functions; thirdly, they can not internalize some functions and not use them naturally, resulting in overproduction. The low conscious perception resonates the research of Schmidt. Meanwhile, reasons for those performances are illustrated--lack of language environment and opportunities for practicing, domestic pedagogical orientation toward examinations, negative transfer from native language system. However, there are still some limitations of the study. Firstly, the selection of corpus is not so much scientific because the number of contestants in each video are not equal, which may influence the frequency of use of DMs. Secondly, the study solely exploits the functional framework by Fung and Carter, but more other pragmatic functions are also essential and can help investigate more situations and reasons. In a nutshell, in spite of some limitations, it is expected that the paper can set an example for fossilization study and provide insights into understanding pragmatic fossilization of discourse markers in oral utterances.

References

1. L. Selinker, "Interlanguage," *International Review of Applied Linguistics*, 1972.
2. L. Selinker, *Rediscovering Interlanguage*. London, U.K.: Longman, 1992.
3. L. R. Yang, L. Ji, and S. P. Li, "A Review of Research on Interlanguage Fossilization," *Journal of Tianjin Foreign Studies University*, 2015.
4. S. H. Zhu and G. F. Tao, "Cognitive Causes and Countermeasures of Fossilization in Second Language Acquisition," *Journal of Northeast Normal University (Philosophy and Social Sciences Edition)*, 2011.
5. L. J. Deng, "A Review of Research on the Development of Productive Vocabulary Ability in Second Language," *Foreign Languages and Their Teaching*, 2006.
6. H. Goad and L. White, "Ultimate Attainment in Interlanguage Grammars: A Prosodic Approach," *Second Language Research*, 2006.
7. Y. Tang, "A Review of Research on Interlanguage Fossilization in Second Language Acquisition in the Past Decade," *Comparative Study of Cultural Innovation*, 2024.
8. G. Kasper and S. Blum-Kulka, "Interlanguage Pragmatics: An Introduction," in *Interlanguage Pragmatics*, G. Kasper and S. Blum-Kulka, Eds. Oxford, U.K.: Oxford University Press, 1993.
9. D. Schiffrin, *Discourse Markers*. Cambridge, U.K.: Cambridge University Press, 1987.
10. Y. P. Ran, "A Review of Pragmatic Studies on Discourse Markers Abroad," *Foreign Language Teaching and Research*, no. 4, 2000.
11. Q. L. Li, "A Preliminary Study on Pragmatic Fossilization of Discourse Markers in English Learners: Based on the Investigation and Analysis of Real Spoken Language Corpus," *Journal of PLA University of Foreign Languages*, 2004.
12. Z. H. Han, *Fossilization in Adult Second Language Acquisition*. Clevedon, U.K.: Multilingual Matters, 2004.
13. R. Ellis, *Understanding Second Language Acquisition*. Oxford, U.K.: Oxford University Press, 1985.
14. W. C. Chen, "A Review of Research on Fossilization of Foreign Language and Second Language Learners," *Foreign Language Teaching Theory and Practice*, 2010.
15. J. Jenkins, *The Phonology of English as an International Language*. Oxford, U.K.: Oxford University Press, 2000.
16. G. Kasper and R. Schmidt, "Developmental Issues in Interlanguage Pragmatics," *Studies in Second Language Acquisition*, 1996.
17. M. Cook and A. J. Liddicoat, "The Development of Comprehension in Interlanguage Pragmatics: The Case of Request Strategies in English," *Australian Review of Applied Linguistics*, 2002.
18. D. Schiffrin, *Discourse Marker Research and Theory: Revisiting And Approaches to Discourse Particles*. Leiden, The Netherlands: Brill, 2006.
19. R. Quirk, *A Comprehensive Grammar of the English Language*. London, U.K.: Longman, 1985.
20. T. A. Van Dijk, *Text and Context*. London, U.K.: Longman, 1977.
21. L. Fung and R. Carter, "Discourse Markers and Spoken English: Native and Nonnative Use in Pedagogic Settings," *Applied Linguistics*, 2007.

22. K. Aijmer, *English Discourse Particles: Evidence from a Corpus*. Amsterdam, The Netherlands: John Benjamins Publishing Company, 2002.
23. Y. X. Wu and G. D. Yu, "A Meta-pragmatic Analysis of Discourse Markers," *Foreign Language Education*, 2003.
24. Y. P. Ran, "A Pragmatic Increment Analysis of the Discourse Marker 'You Know'," *Journal of PLA University of Foreign Languages*, 2002.
25. R. Q. Heng, "A Pragmatic Account of the Discourse Marker 'And' in Conversational Interaction," *Shandong Foreign Language Teaching Journal*, no. 4, pp. 23–25, 2005.
26. X. Chen and W. Ren, "Functions, Sociocultural Explanations and Conversational Influence of Discourse Markers: Focus on Zenme Shuo Ne in L2 Chinese," *IRAL: International Review of Applied Linguistics in Language Teaching*, vol. 62, no. 4, pp. 1649–1677, 2024.
27. A. H. Jucker and S. W. Smith, "And People Just You Know Like 'Wow': Discourse Markers as Negotiating Strategies," in *Discourse Markers: Descriptions and Theory*, A. H. Jucker and Y. Ziv, Eds. Amsterdam, The Netherlands: John Benjamins, 1998.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of Publisher and/or the editor(s). Publisher and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.