# A Study of Self-repair Markers in Conversation by Chinese English Learners<sup>\*</sup>

Lihong Quan

School of English and Education, Guangdong University of Foreign Studies, Guangzhou, China Email: sallyquanli@126.com

Yanmei Zheng

Guangdong University of Foreign Studies, Guangzhou, China

Abstract—The present study examines the differences in the use of three self-repair markers well, I mean and maybe between native speakers and Chinese English learners. The corpora of BNC (British National Corpus) and SECCL (Spoken English Corpus of Chinese Learners) are employed in the study. The research results indicate, that the Chinese English learners underuse well and I mean, but overuse maybe, compared with native speakers. The detailed analysis, supported by statistical tests, proves the role of "Pragmatic Fossilization" as an element to consider in second language learning and teaching.

*Index Terms*—self-repair marker, pragmatic fossilization, corpora

### I. INTRODUCTION

One of the central findings of earlier work on repair concerned the preference for self-correction. Schegloff, Jefferson and Sacks (1977, p.361-382) noted a strong "empirical skewing" which resulted in the vastly more common occurrence of self over other-repair. Of the self-repair category, same-turn self-repair (STSR) is most common and has gained a considerable amount of attention in recent years (Fox et al., 1995; Wouk, 2005; Laakso and Sorjonen, 2010). However, research on the means for initiating this type of repair is still scarce. Often the initiators have merely been mentioned in the study and most of the mentions deal with prosodic cut-offs, whereas other means have remained unexplored (Laakso and Sorjonen, 2010). Other non-lexical initiators mentioned in prior studies are pauses, sound stretches and vocalizations such as *um\uh\er*, etc. (Schegloff et al., 1977, p. 367; Levelt, 1983). A few English lexical initiators have also been mentioned in prior studies (Schegloff et al., 1977; Levelt, 1983; Laakso and Sorjonen, 2010). These include words such as *or*, *well* and *sorry* or discourse markers such as *you know* and *I mean*. In some prior studies they are commonly referred to as self-repair markers (Chen, 2005; Wang, 2007). Self-repair markers (hereafter SRMs) belong to the general category of discourse markers. The speaker utilizes them to signal to the addressee that there is trouble and he/she is going to repair it. SRMs are short in form and may convey little semantic meanings. But they may serve as discourse lubricants and ensure interactions to go smoothly. They are important interactional devices (Biber et al., 1999).

The purpose of this study is two fold. First, through a close investigation of SRMs, it aims to find how differently SRMs are used by Chinese English learners (hereafter L2 learners), compared with NSs. Second, based on the research results, it aims to examine whether or not pragmatic transfer from Chinese to English is evident in the use of SRMs by L2 learners and whether pragmatic fossilization is the fundamental cause of the differences between L2 learners and NSs in the use of SRMs. This will also foster a better understanding of appropriate ways for L2 learners to use SRMs and also help teachers develop their learners' pragmatic awareness and sensitivity of employing discourse markers, including SRMs.

## II. RESEARCH BACKGROUND

Among previous research on SRMs, Bois' (cited in Svartvik, 1980) discussion is insightful in that it classified SRMs into four types and used the term "editing" to mean editing terms or discourse markers: reference editing such as *that is*, nuance editing such as *rather*, mistake editing such as *I mean*, and claiming editing such as *well*. This classification describes the form-function relationship of SRMs, but there is no doubt that it may not cover the overall forms of SRMs.

According to Schiffrin (1987), virtually every discourse marker (hereafter DM) has a basic function, but in interaction it may serve a wide range of functions. A typical example is that of *well*, which has received considerable attention in recent years specifically in the area of pragmatic functions. Previous research (Svartvik, 1980; Schiffrin, 1987; Ran, 2003) substantiated that *well* was endowed with a function of self-repairing and it is used when the speaker

<sup>\*</sup> This study was funded by a grant (No. 09YJA740030) from the Ministry of Education, China.

feels hesitant in conversation or tries to self-repair while making statements. The previous research on *I mean* found that it had a basic function of serving as SRMs (Schiffrin, 1987; Brinton, 1996; Tree & Schrock, 2002). *I mean* can contribute to repair in at least three ways. Firstly, it explicitly forewarns upcoming adjustments to what has just been said including speaker's modification, expansion or clarification of the prior utterance. Secondly, it can also be used to substitute a pause or to avoid a break in fluency caused by the pause. Thirdly, it serves to acquire more time as the speaker completes the conversation, including planning what to say, selecting words or restarting a false-started utterance (Tree & Schrock, 2002). According to Schiffrin (1987), *I mean* encourages listeners to focus more on speakers' thoughts, and this view is in accordance with the proposal that *I mean* is more speaker-oriented. The previous research also found that Chinese English learners tended to overuse *maybe* as SRMs (Chen, et al., 2005; Wang, 2007). Validity testing should be administered to provide empirical evidence to support this claim.

Previous researchers used CIA (Comparative Interlanguage Analysis) to explore the different utilization patterns of DMs between native and non-native speakers (Weinert, 1998; Trillo, 2002; Fuller, 2003a; Li, 2004; Hellermann & Vergun, 2007; Li& Chen, 2007; Wang & Wang, 2007; Cuenca, 2008; Liao, 2009). The research showed, in general, non-native speakers (hereafter L2 learners) tended to underuse *well* and *I mean* as SRMs. The previous studies did find some general patterns of SRMs, yet there is still a lack of detailed investigation on the patterns of SRMs. And comparative studies of the patterns of SRMs between NSs and L2 learners are even fewer. So it is difficult to generalize whether judgments such as "underuse" or "overuse" are ubiquitous and universal among L2 learners. And if the findings are evidently proved, what then would be the underlying causes of them?

Based on the previous research, the current study will compare the different use of three SRMs—well, I mean and maybe in conversation between NSs and L2 learners. The following research questions will be addressed:

(1) In general, what is the difference in patterns of SRM use when comparing L2 learners and NSs?

(2) What is the difference in patterns of the sub-categories of SRM use when comparing L2 learners and NSs?

## III. RESEARCH METHODOLOGY

#### A. Data Collection

The L2 learner data is obtained from the Spoken English Corpus of Chinese Learners (SECCL) (Wen, etc., 2005). SECCL, a learner spoken English corpus in China, has been developed by a team headed by Professor Qiufang Wen from The Foreign Language Education Research Center in China. The corpus sample size is designed to encompass more than 1 million tokens and expected to provide researchers with valuable and authentic data on Chinese College learners' spoken English. The samples are randomly selected from the TEM (Test for English Majors, which is a nationwide oral test for English majors) and the period of time for the selected tests ranges from 1996 to 2006. By doing so, the corpus developers wish to ensure the representativeness of the samples in the corpus. In SECCL, there are three kinds of oral activities: story retelling, monologue and role-play. In the present study, we only utilized the transcribed data of role-play since our intention is to observe the utilization patterns of SRMs in conversation. We select the data composed of about 93,000 tokens from SECCL. The data of NSs are from BNC (British National Corpus), and the texts from the TV interviews are randomly chosen with the total token of 93, 131. In this case, the observed corpus and the reference corpus have almost the same number of tokens. The fact that the sizes of two corpora are the same enhances the validity of measurement in the study. As for the topics, the two corpora share some similar topics, such as family life, sports and mass media.

## B. Data Coding

THE CATEGORIZATION AND EXEMPLIFICATION OF SRMs							
Categories		Examples					
Repetition marker	Syllabic R	well th well there is two of them involved at the time, but they					
-		(BNC-fl6-Eating disorders.txt)					
	One-word R you know temptation you just <i>I well I</i> I've found that						
		faith comes in (BNC-hmp-Jonathan Cowap Morning Show.txt)					
	Two-word R	There's a, well there's a thin line between it, being a very posit					
		experience, and(BNC-fl6-Eating disorders.txt)					
	Multi-word R	They cannot pay all the tuitions for their children, maybe for their					
		children, they, they will still, still (SECCL-04-128-13A.txt)					
Information RM	Information	Social experience is a good thing. But you know, I mean the study is, is,					
	supply/replacement	er, I think is more important for us now. (SECCL-04-199-08A.txt)					
Error RM	Lexical R	I can't well I don't know I I don't think you did. (BNC-d8y-Museum					
		society meeting. Txt)					
	Morphological R	He didn't er he, I mean, he wasn't responsible for their children.					
		(SECCL-04-128-16B.txt)					
	Syntactic R	What did you I, well I just didn't realize it was anything like that.					
		(BNC-flf-Mental health.txt)					
Appropriate RM	replacement	You can't er, concentrate, I mean you can't focus on the black sid					
	-	(SECCL-05-005-30B.txt)					
	insertion	But if you find a proper job, I mean, a proper part- time job,					
		(SECCL-05-005-30B.txt)					

TABLE 1

Considering the categories of self-repair, we follow both Levelt's model (1983) and van Hest's model (1996a, 1996b), of which the former was developed for NSs and the latter for L2 learners. We modify them for the fulfillment of the purpose of the present study (see table 1). We divided self-repair into four general types: self-repetition, error repair, information repair and appropriate repair. Self-repetition is sub-divided into 4 sub-categories: syllabic repetition, one-word repetition, two-word repetition and multi-word repetition. Under this category, a wrong syllable, an erroneous word or an inappropriate syntactic structure is selected. Information repair is related with the speaker's problem in conveying information (Under Levelt's system of classification, it was called different repair since the current message was replaced by a different one). We sub-divided it into information supply and information replacement. Error repair concerns repair on lexical, morphological and syntactic levels. The last category is appropriate repair, which concerns the manner of repair. This type of repair aims to make utterances more appropriate under the condition that the previous information is not changed. It is further divided into insertion and replacement in this study. Following the above categorization, We separate SRMs into four categories accordingly: self-repetition marker (It was sub-divided into repair markers for syllabic repetition, one-word repetition, two-word repetition and multi-word repetition), error repair markers (repair markers which concerns with error repair on lexical, morphological and syntactic levels), information repair markers (repair markers which signal information supply or replacement) and appropriate repair markers (repair markers which signal insertion or replacement).

The analytical procedure was as follows: first, we used WordsSmith Tools 5.0 to calculate the frequencies of three SRMs used by both NSs and L2 learners in two corpora. Second, we examined the concordance lines closely to exclude the other uses of the three DMs, such as the use of *well* in "very *well*", and *maybe* used as an adverbial, and the use of *I mean* in "Do you know what *I mean*"and "You know what *I mean*", etc.. After doing this, we employed a foreign teacher (He holds a doctorate in English literature from a certain Canadian university, and has taught English in China for two years.) to reexamine the materials to confirm whether a discourse marker was used as an SRM or not. By doing so, we hoped to position ourselves as to most effectively understand and interpret our findings. As previous research on the use of maybe as SRMs was minimal, the current study was tentative in its course. By observing the concordance lines from SECCL, We found that L2 learners used *maybe* as SRMs frequently. But this was a rare case in BNC.

#### IV. RESULTS

	•	O TERMEL DID II	incrition of i	51(1)15			
	Well	Well			Maybe	Maybe	
	BNC	SECCL	BNC	SECCL	BNC	SECCL	
Total frequency	443	273	184	47	66	388	
Log likelihood	-40.51**	-40.51**			+253.37**	+253.37**	
Frequency	41	6	177	34	6	39	
Percentage (%)	9.3	2.2	96.2	72.3	9.1	10.1	
Log likelihood	-29.21**	-29.21**		-99.40**		+27.09*	
Result	underuse	underuse		underuse		overuse	

TABLE 2 OVERALL DISTRIBUTION OF SRMS

Table 2 showed the overall distribution of three SRMs. The results suggested that the percentages of *well* for NSs and L2 learners be 9.3% and 2.2% respectively. Log likelihood equaled -32.09(p<0.01), which indicated there was a significant difference between NSs and L2 learners in the use of *well* as SRMs. The L2 learners tended to underuse *well* as SRMs, compared with NSs. As for *I mean*, the percentages of NSs and L2 learner were 92.3% and 72.3% respectively, with the log likelihood of -99.40 (p<0.01). This also indicated that there was a significant difference between NSs and L2 learners in the use of *I mean* as SRMs. Compared with NSs, L2 learners tend to underuse *I mean* as SRMs. The result also showed that the percentages of *maybe* by L2 learners and NSs are 10.1% and 9.1% respectively. Finally according to log likelihood (+ 27.09 (p<0.05)), there was a significant difference between NSs and L2 learners in their use of *maybe* as SRMs. L2 learners tended to overuse *maybe* as SRMs.

SRM		Well		I mean		Maybe	
		BNC	SECCL	BNC	SECCL	BNC	SECCL
Repetition	Syllabic RM	3	0	0	0	0	0
RM(RRM)							
	One-word RM	3	1	0	0	0	10
	Two-word RM	2	0	1	2	0	5
	Multi-word RM	1	0	2	1	0	2
Total		10	1	3	3	0	17
Information RM	Information supply	10	2	160	12	1	2
(IRM)	RM/replacement RM						
Total		10	3	162	16	2	3
Error RM(ERM)	Lexical RM	3	0	4	3	0	6
	Morphological RM	6	0	2	1	0	1
	Syntactic RM	2	1	2	3	0	4
Total		11	1	8	7	1	11
Appropriate RM	Replacement RM	6	1	2	7	1	4
(ARM)	Insertion RM	4	0	2	1	1	4
Total		10	1	4	8	2	8
Z score		z=-2.8, p=0.005		z=-1.44, p=0.886		z=-2.52, p=0.012	

TABLE 3 DISTRIBUTION OF SRMS BY SUB-CATEGORY

Table 3 showed the distributions of SRMs under the sub-categories. The statistic results indicated a significant difference in terms of *well*, with a z-score of -2.81(p=0.005). But the result showd no significant difference in the use of *I mean* by NSs and L2 learners under the sub-categories (z = -1.44., p=0.886). Since *maybe* tends to be overused by L2 learners, the result was diversified. *Maybe* was mainly used in repetition as SRMs with a total number of 17, in error repair with a total number of 11. The total number of *maybe* used by NSs as SRMs was only 6. Hence, the statistic results indicated a remarkable difference between NSs and L2 learners in the use of *maybe* as SRMs (z=-2.52, p=0.012).

To sum up, the present research had the following major findings while referring to the research questions: (1) In general, L2 learners and NSs used the three SRMs in a significantly different way. L2 learners tended to underuse *well* and *I mean* while overuse *maybe*; (2) Under the sub-categories, NSs and L2 learners had remarkable differences in the use of *well* and *maybe*, but no significant difference was found in the use of *I mean*. However, an interesting pattern showed that NSs tended to employ *I mean* mainly for information supplement or replacement while L2 learners used it to signal different kinds of self-repair, either for signaling self-correction or for achieving appropriateness of the utterances. (3) One more striking finding was that the NSs tended to use a "discourse marker+ *maybe*" pattern, whereas L2 learners tend to employ *maybe* directly as SRMs; (4) Further observations on the sub-categories indicate that, in BNC, *well* was almost equally deployed as RRMs (10), IRMs (10), ARMs (10) and ERMs (11), whereas, *I mean* was mostly deployed as IRMs (162). As for *maybe* in SECCL, it was deployed mainly as RRMs (especially one-word RRMs, with a total number of 17), less as ERMs (11), ARMs (8) and least as IRMs (3). Hence, the aforementioned findings may help specify the character of repair operations, for example, the ways in which an element of the prior talk ccould be self-repaired.

#### V. DISCUSSION

Next, we began with a closer look at the examples from two corpora which may demonstrate some striking patterns of three SRMs, which might have been exhibited in the aforementioned results. Example (1) and (2) below illustrated the practice of *well* as SRMs.

(1) It changed so much that you co committed an offence that ended up you in prison, in prison for, *well* you got a three year sentence. (BNC: hv1-Central Weekend Live.txt)

(2) After all I will try to manage it, because I think in the US, *well*, look at the things in US, when the kids gets 18 years old..... (SECCL: 04-199-31A.txt)

In example (1), the speaker replaced the previously uttered information by using *well* as an information replacement marker. In example (2), the speaker supplied more information by using *well*. The data from Table 3 further showed a strong tendency that NSs use *well* more often as DMs or SRMs than L2 learners. We would further explain this result in later sections. Example (3) and (4) illustrated the practice of *I mean* as SRMs.

(3) Right, now, what has this got to do with eating disorders? *I mean*, where do eating disorders come from? (BNC: fl6-Eating disorders.txt)

(4) Yes, I, I, read the report from the newspaper this morning. And I think, er, it's very wonderful. *I mean*, it's very competitive for them to compete for the position. (SECCL: 05-005-30B.txt)

In example (3), the speaker used *I mean* to make adjustments to what had just been said. In example (4), the speaker used *I mean* to signal the instantly upcoming self-repair for what had just been said. We found an interesting pattern that, in BNC, *I mean* was mainly used as information repair markers while in SECCL *I, mean* served, not only as information repair markers, but also as error or appropriate repair markers. In both cases, *I mean* was used to explicitly indicate what kind of repair is to follow. We would further interpret this kind of idiosyncratic usage in later sections. Example (5), (6)

and (7) illustrated the practice of maybe as SRMs.

(5) You can communicate with the *maybe* the guests of the hotel and just improve your abilities in some aspects. (SECCL: 05-005-03B.txt)

- (6) A: Are you wealthy?
  - B: Wealthy? I wish I was. Er I mean one maybe one day I will if if it makes a hit eh.....
- (BNC: hmd-BBC Radio Nottingham.txt)
- (7) A: If they're stupid enough to be homeless!
  - B: Well *maybe* sometimes it's not their fault. Most of the time people don't want to ...
  - (BNC: fld-Families.txt)

In example (5), *maybe* served as SRMs to initiate the upcoming self-repair. But in example (6), speaker B first used *I* mean as an initiator of self-repair, then he/she uses maybe to finish the self-repair implicitly. In example (7), speaker B first used well as an initiator of self-repair and uses maybe to implicitly expressed his/her idea. In both cases, maybe served not as explicit SRMs, rather it was used as a supplementary part of SRMs. As we further observed the concordance lines from BNC, we found that NSs tended to use a "discourse marker+ maybe" pattern. But we did not discover this striking pattern in SECCL. We found L2 learners tended to use maybe directly as SRMs.

In interpreting the aforementioned findings, we assumed that pragmatic transfer and overgeneralization were the major causes of learners' overuse or underuse of SRMs. In recent years, there had been a convergence of SLA and pragmatics in what had been termed interlanguage pragmatics. Kasper (1992, p. 203) defined interlanguage pragmatics as "the branch of second language research which studies how non-native speakers... understand and carry out linguistic action in a target language, and how they acquire L2 pragmatic knowledge." Kasper (1992) expanded the scope of the Interlanguage Pragmatics and put forward the notion of pragmatic transfer, which refers to the influence exerted by learners' pragmatic information. Some other notions, such as pragmatic fossilization in SLA have also been applied to the pragmatic aspect of language development. Trillo (2002) addressed fossilization from a pragmatic perspective and put forward the notion of pragmatic fossilization from a pragmatic perspective and put forward the notion of pragmatic development. Trillo (2002) addressed fossilization from a pragmatic perspective and put forward the notion of pragmatic fossilization in SLA have also been applied to the pragmatic aspect of language development. Trillo (2002) addressed fossilization from a pragmatic perspective and put forward the notion of pragmatic fossilization, which was defined as the phenomenon by which a non-native speaker systematically used certain forms inappropriately at the pragmatic level of communication. We would further explored how pragmatic transfer and overgeneralization exerted influence over L2 learners' use of SRMs in the following part.

As for *well* and *I mean*, some previous studies supported our findings (Wong, 2000; Tree & Schrock, 2002; Trillo, 2002; Liao, 2009). Liao (2009) found that Chinese English learners had an overall infrequent use of *well* and *I mean* in conversation. One of the possible explanations was that there were no countparts of *well* and *I mean* in Mandarin. In Mandarin, the equivalent marker to *well* as a delay device would be "um." As for *I mean*, the countpart in Mandarin was "Wo de yisi shi" (my meaning is) which contained five Chinese characters with five syllables, and thus was never regarded as DMs in Mandarin. Based on the above analysis, Liao came to a conclusion that the influence of L1 transfer might explain why these DMs were employed at a lower rate by Chinese English learners. Liao's research had another striking finding in which there was a high rate of *yeah* as SRMs. This supported Wong (2000)'s findings in which she observed the function of same-turn repair by her non-native participants whose native language was Mandarin. This usage of *yeah* had been found to be extremely rare in native discourse, yet non-native speakers used this *yeah* to resolve what was problematic or troublesome about the utterance. The studies mentioned above gave solid evidence to the present findings. The discrepancy between NSs and L2 learners in the use of *I mean* under the sub-categories demonstrated that L2 learners had not yet acquired the meaning of this SRM. They might overgeneralize its usage.

As for *maybe*, the previous studies also supported our findings (Wu, 2009; Hu, 2010). Wu's corpus-driven study, which compared the usage of two adverbs –*probably* and *maybe* by NSs and L2 learners, showed that NSs tended to use a "discourse marker+*maybe*" pattern to initiate self-repair, whereas Chinese L2 learners tended to use *maybe* directly as SRMs. Similarly, Hu's study found that Chinese English learners tended to overuse *maybe* when they tried to express epistemic possibility in English and he further suggested that *maybe* is perhaps entrenched in learners' cognitive system. Our possible explanation of this idiosyncratic usage was that the counterpart of *maybe* in Mandarin was "ye xu" or "da gai", which were frequently employed in oral Mandarin. As the present research results showed, L2 learners had not yet acquired the use of *I mean* and *well* as SRMs. Instead they tended to use *maybe* as SRMs in resolving what was problematic or troublesome about the utterance. Finally, we postulated that L2 learners in the study were reserved in making use of lexical items. When they were hesitant in choosing lexical items, they preferred to use some semantically familiar items like *maybe* as vocalized fillers to stall for time in speech organization processes.

As for the last aforementioned finding, we attempted to have the following interpretations. Previous studies indicated that sprinkling *I mean* into speech had been thought to provide three types of interpresonal information: information about the speaker, information about the situation, such as its formality or intimacy and information about the level of politeness (Tree & Schrock, 2002). Hence, *I mean* can be used as information management displays. This was in accordance with the present finding that *I mean* was mostly deployed as IRMs (In this study the total number is 162) by NSs. As for *may* (*maybe* is utilized mainly as RRMs (17) and ERMs (11), less as ARMs (8) and IRMs (3) by L2 learners in the present study), the finding may partly be a sign of L2 learners' language proficiency. L2 learners tended to focus on the accuracy of lower-level linguistic forms, such as lexical,

morphological and syntactical items rather than on the appropriateness of language expressions. There could be another interpretation for this. Repair is a speech activity during which speakers locate and replace a prior information unit. Because they focus on prior information, repairs achieve information transitions—forcing speakers to adjust their orientation to what has just been said before they respond to it in upcoming talk (Scheffrin, 1987). In this kind of repairing process, SRMs tended to be utilized to fulfill the information management tasks. Hence, the aforementioned findings may also show L2 learners' lack of the awareness of SRMs as information management displays. Further validity testing is needed to provide empirical evidence to support this claim.

Next, we attempted to further explore the reasons behind these remarkable differences between NSs and L2 learners in the use of SRMs. According to Painter (1999), NSs followed what was called a "function-to-form" developmental process, where the needed to communicate precedes the use of form; whereas non-native speakers follow a "form-to-function process", based on the learning of certain items which are usually contextualized at different subsequent stages. Based on this statement, Trillo (2002) believed that "the form-to-function language development process that was characteristic of non-native speakers was the key factor in the appearance of certain aspects of language fossilization". Trillo (2002) further postulated that non-native speakers may find the learning process of form-to-function problematic because of the de-contextualized nature of L2 learning environments. In most cases L2 learners acquire some simplified and context-free register of the target language with no explicit relationship between form and function. Therefore, he suggested that it is essential to teach the pragmatic function of language items of a target language, ranging from lexicalized phenomena to the functions of DMs in conversation. Similarly, Weinert's (1998) findings of L2 learners' incompetence of using DMs led her to question whether L2 language learning materials may be biased toward written language, whether naturalistic language development and classroom teaching are separated and whether oral discourse coherence devices like DMs might be explicitly taught.

Based on the above arguments, we assumed that the fundamental cause of pragmatic fossilization was that L2 learners could not acquire the target language through natural language contact, but in most cases they followed a "form-to-function" learning process in which the learning materials are de-contextualized. As was stated by Trillo (2002, p. 770-783), "pragmatic fossilization appeared in L2 learners not because of a lack of competence in other linguistic areas such lexis, grammar, etc., but because there was delay in the presentation of the pragmatic variation that existed with respect to the way communication competence is acquired in the mother tongue". The lack of this communicative competence may lead to pragmatic fossilization, and in many cases, to pragmatic failure. Considering pragmatic failure, Blum-Kulka and Olshtain (1986) made the important point that pragmatic failure may also occur between NSs, but there was a good chance that the speaker would recognize the failure and the reason for it, and would therefore perform suitable repair. Non-native speakers, however, might be completely unaware of the reason for the failure and therefore unable to repair the interaction.

#### VI. CONCLUSION AND PEDAGOGICAL IMPLICATIONS

# A. Conclusion

The present study was a corpus-based study on the use of three SRMs: *well, I mean* and *maybe* by NSs and L2 learners. The availability of two representative corpora guaranteed the objectivity of the results and also might shed light on future research where variables of the corpora might be different. The findings indicated L2 learners overused *maybe* and underused *well* and *I mean* as SRMs, compared with NSs. The study then explored the fundamental reasons for the differences and postulated that these SRMs might be fossilized both in the quantity and the diversity of usage. The study also has its own contributions to the present literature of interlanguage pragmatics in that it not only examined pragmatic transfer from Chinese to English was evident in L2 learners' use of SRMs, but also verified that the cause of pragmatic fossilization in L2 was due to the "form-to-function" learning process, in which the learning materials are de-contextualized.

# B. Pedagogical Implications

First, regarding the phenomenon of pragmatic fossilization, we conclude if DMs, including SRMs, are retained and used as part of L2 learners' interlanguage system, then L2 learners must obtain repeated exposure to and rehearsal of those markers through conversational interaction either inside or outside of the classroom. Since L2 learners do not always have the opportunities to be immersed in an English-speaking environment outside of the foreign language classroom, pedagogical factors, including textbooks, teachers' roles and learner factors, exert much influence on this aspect. As pointed out in many studies, textbooks play a crucial role in syllabus design and lesson planning. Textbooks constitute the bread and butter of L2 learners' language learning experience (Lam, 2009). As is often the case, the descriptions of DMs in textbooks are far from satisfactory and many examples are invented and decontextualized (Wang, 2007; Lam, 2009; Quan, 2010). The learners may be denied access to these useful DMs in their learning processes. In order to present a more comprehensive picture to learners concerning how DMs are used in real-life contexts, textbook writers can introduce corpus evidence into textbooks using authentic examples. In this respect, a corpus-based and data-driven learning approach to learning DMs with the use of concordancers may be useful (Lam, 2009).

Next, direct teaching of DMs on the part of teachers may increase the usage by L2 learners. Teachers may not spend

much of their class time teaching DMs, but they can make learners aware of these markers and their pragmatic functions by giving language samples from everyday conversation between native speakers. Also, teachers can give learners structured time for pair work and small group interaction in class after giving learners' input of the use of DMs through video or movie watching. Such increased awareness and classroom practice may provide learners with gradual access to DMs in appropriate situations inside and outside of the classroom.

Finally, as for learner factors, we think L2 learners' acquisition of appropriate pragmatic forms depend much on learners' conscious learning. According to Gass and Houck (1999), one reason for the difficulty in obtaining appropriate knowledge of the pragmatics of L2 may be a human assumption about learning a L2, namely, what needs to be learned is the phonology, lexicon, and syntax of a L2. Pragmatics, speech acts or even more specific, DMs, are initially perceived to be universal. Therefore, L2 learners are not aware of those pragmatic aspects and they are unlikely to notice the subtle differences between target language and their native language. As such, noticing mismatches between L1 and L2 is a major driving force in L2 learning. How to raise L2 learners' awareness on pragmatic forms concerns the roles of attention and awareness in L2 learning. This needs further discussion in future research.

This study was not without its limitations. The major weakness of the study was that the sizes of two corpora were limited. Future research which employs large- scaled corpora is needed to give more supporting evidence for the present study.

#### NOTES

1. The following example is an instance of a cut-off as repair initiators in English:

[Schegloff, Jefferson and Sacks, 1977: 366]

Naomi: But c'd we- c'd I stay u: p?

Here, Naomi stops her utterance by cutting it off after the word we, marked with a dash at the end of the word. This cut-off functions as repair initiator. After that, she recycles the verb form c'd in her prior talk and then replaces we with I.

2. Although terms such as "editing terms or expressions (Levelt, 1983: 41-104; 1989), discourse markers or particles; (Schiffrin, 1987: 31-40; Aijmer, 2002), cue phrases (Heeman & Allen, 1999), and repair markers (Chen, 2005; Wang, 2007) are preferred by some researchers, we think the term SRMs can display the functions of those discourse markers more explicitly, and we will therefore use the term "self-repair markers" in this article.

3. In table 1, the italicized parts indicate the whole self-repair segments: trouble sources, SRMs and repair outcomes. By doing so, we intend to show the ongoing self-repairing processes clearly. For instance: *You can't... er, concentrate, I mean you can't focus on* the black side. In this sentence, "concentrate" is the trouble source, *I mean* serve as SRMs, and the repair outcome is "focus on." In this sentence, the speaker traced back to the trouble source by repeating "*You can't.*"

4. In the present study, if there is significant difference between NSs and L2 learners in their use of SRMs (p<0.05), then we use "+" to signal the state of "overuse" and "- "to signal the state of "underuse."

5. In BNC, *I mean* is mainly used as information repair markers, whereas in SECCL, *I mean* serves, not only as information repair markers, but also as error or appropriate repair markers.

## REFERENCES

- [1] Aijmer, K. (2002). English Discourse Particles. John Benjamins, Amsterdam/Philadelphia.
- [2] Biber, D., et al. (1999). Longman Grammar of Spoken and Written English. London: Pearson Education Limited.
- [3] Blum-Kulka, S. and Olshtain E. (1986). Too many words, length of utterance and pragmatic failure. SSLA, 8, 165-180.
- [4] Brinton, L. J. (1996). Pragmatic Markers in English: Grammaticalization and Discourse Functions. Berlin and New York: Mouton de Gruyter.
- [5] Chen, Liping, et al. (2005). Gender differences in self-repairs by Chinese non-English majors in their spoken English. Modern Foreign Languages, 3, 279-287.
- [6] Corder, S. P. (1967). The significance of learners' errors, International Review of Applied Linguistics 5, 161-169.
- [7] Corder, S. P. (1981). Error Analysis and Interlanguage. Oxford: Oxford University Press.
- [8] Cuenca, M. J. (2008). Pragmatic markers in contrast: The case of well. Journal of Pragmatics 40, 1373-1391.
- [9] Fox, B. & Jesperson R. (1995). 'A syntactic exploration of repair in English conversation'. In: Davis, P. W. (Ed.) Alternative Linguistics. Descriptive and Theoretical Modes. John Benjamins, 77-134.
- [10] Franch, P.B. (1998). On pragmatic transfer'. Studies in English Language and Linguistics. 0, 5-20.
- [11] Fuller, J. M. (2003a). Discourse marker use across speech contexts: a comparison of native and non-native speaker performance. *Multilingua*, 22, 185–208.
- [12] Gass, M. S. & Houck Noel. (1999). Interlanguage Refusals—a Cross-cultural Study of Japanese English. Berlin; New York; Mouton de Gruyter.
- [13] Gass, S. & Selinker L. (2008). Second Language Acquisition: An Introductory Course. (3<sup>rd</sup>. ed.). Laurence Erlbaum Associates, Inc..
- [14] Heeman, P and Allen J. F. (1999). Speech Repairs, Intonational Phrases and Discourse Markers: Modeling Speakers' Utterances in Spoken Dialog. *Computational Linguistics*, 25-4, 527-572.
- [15] Hellermann, J. & Vergun, J. (2007). Language which is not taught: The discourse marker use of beginning adult learners of English. *Journal of Pragmatics*. 39, 157-179.

- [16] Hu, Chunyu. (2010). 'Modal expressions of epistemic possibility in Chinese EFL learners' interlanguage'. Modern Foreign Languages, 4, 379-384.
- [17] Kasper, G. (1989). 'Variation in interlanguage speech act realization'. In S. Gass, Madden, D. Preston, & L.Selinker (eds.). Variation in Second Language Acquisition: Discourse and Pragmatics, Multilingual Matters. Clevedon and Philadelphia.
- [18] Kasper, G. (1992). Pragmatic Transfer'. Second language Research. 8(2), 203-231.
- [19] Laakso, M. & Sorjonen M. (2010). 'Cut-off or particle –devices for initiating self-repair in conversation'. Journal of Pragmatics. 42, 1151-1172.
- [20] Levelt, W. J. M. (1983). 'Monitoring and self-repair in speech'. Cognition, 14, 41-104.
- [21] Levelt, W. J. M. (1989). Speaking: From Intention to Articulation. Cambridge, MA: MIT Press.
- [22] Li Min & Chen, Xinren. (2007). 'Chinese English majors' acquisition of the discourse marker well: an empirical study'. *Foreign Language Teaching and Research*, 1, 21-25.
- [23] Li, Qiaolan. (2004). On pragmatic fossilization in English learners' acquisition of discourse markers—an investigation based on natural spoken English data. *Journal of PLA University of Foreign Languages*, 3, 53-57.
- [24] Liao, S. (2009). Variation in the use of discourse markers by Chinese teaching assistants in the U.S. *Journal of Pragmatics*, 41, 1313-1328.
- [25] Quan, Lihong. (2010). A study of anticipatory retracing in English conversation by Chinese university students'. *Foreign Language Learning Theory and Practice*, 2, 8-14.
- [26] Ran, Yongping. (1995). Pragmatic functions of WELL in discourse. Journal of Sichan International Studies, 3, 41-44.
- [27] Ran, Yongping. (2003). Pragmatic account of the discourse marker WELL'. Foreign Languages, 3, 58-64.
- [28] Schegloff, E.A. Jefferson G. & Sacks H. (1977). 'The preference for self-correction in the organization of repair in conversation'. *Language*, 53, 361-382.
- [29] Schiffrin, D. (1985). Conversational coherence: the role of well'. Language, 61, 640-67.
- [30] Schffrin, D. (1987). Discourse markers. Cambridge: Cambridge Press, 31-40/74.
- [31] Schmidt, R. (1993). 'Consciousness, Learning and Interlanguage Pragmatics'. In Kasper, G. &, S. Blum Kulka. (eds.) *Interlanguage Pragmatics*. Oxford University Press, 21 42.
- [32] Selinker, L. (1972). Interlanguage. International Review of Applied Linguistics, 10, 209-231.
- [33] Selinker, L. & Han Z.H. (2001). Fossilization: moving the concept into empirical longitudinal study'. In C. Elder, A. et al. (eds.) *Studies in Language Testing: Experimenting with Uncertainty*. Cambridge: Cambridge University Press, 276-291.
- [34] Tree, J.E. & Schrock J.C. (2002). 'Basic meanings of you know and I mean'. Journal of Pragmatics. 34, 727-747.
- [35] Trillo, R. J. (2002). The pragmatic fossilization of discourse markers in non-native speakers of English'. Journal of Pragmatics. 34, 770-783.
- [36] van Hest E. (1986a). Self-repair in L1 and L2 Production. Tilburg: University Press.
- [37] van Hest E. (1996b). Self-repair as a Measure of Language Proficiency'. Paper presented at the 18<sup>th</sup> Annual Language Testing Colloquium in Tampers. Finland.
- [38] Weinert, R. (1998). Discourse organization in the spoken language of L2 learners of German'. *Linguistic Berichte*, 176, 459-488.
- [39] Wang, Li & Wang Tongshun. (2008). A corpus-based analysis on the acquisition of pragmatic markers by Chinese learners of English'. *Modern Foreign Languages*, 3, 291-300.
- [40] Wang, Xiaoyan. (2007). A study of age differences in speech repair by English learners'. *Journal of PLA University of Foreign Languages*, 5, 65-71.
- [41] Wong, J. (2000). The token "yeah" in nonnative speaker English conversation'. *Research on Language and Social Interaction*, 33-1, 39–67.
- [42] Wouk, F. (2005). 'The syntax of repair in Indonesian'. Discourse Studies 7(2), 237-258.
- [43] Wu, Zhifang. (2009). A corpus-based comparative study of epistemic stance adverbials *probably* and *maybe*'. Journal of Jiannan University (Education Sciences), 29 (2),166-170.

**Lihong Quan** was born in Gansu Province, China in 1965. She obtained her Master's degree in Linguistics and Applied Linguistics in 2001 from Guangdong University of Foreign Studies. She was a visiting scholar at the Department of Linguistics and English Language in Lancaster University, the U. K. in 2010.

She is currently an associate professor in the School of English and Education, Guangdong University of Foreign Studies, Guangzhou, China. Her research interests include sociolinguistics, second language acquisition and conversation analysis. She is a member of Corpus Linguistics Society of China (CLSC).

**Yanmei Zheng**, born in Hainan Province, China in 1976, obtained her Master's degree in Linguistics and Applied Linguistics in 2001 from Guangdong University of Foreign Studies. She was an academic visitor at the School of Foreign Languages and International Studies in the University of Central Lancashire, U. K. in 2010.

She is currently a lecturer and a doctoral student in Guangdong University of Foreign Studies, Guangzhou, China. Her research interests include pragmatics, second language acquisition and translation. And now she is working on her P.H.D program in the field of written corrective feedback for L2 writing.