Studies in Second Language Learning and Teaching
Department of English Studies, Faculty of Pedagogy and Fine Arts, Adam Mickiewicz University, Kalisz
SSLLT 3 (1). 109-130
http://www.ssllt.amu.edu.pl

The development of cohesion in a learner corpus

Colleen A. Neary-Sundquist
Purdue University, USA
cnearysu@purdue.edu

Abstract

This study examines the use of cohesive devices (pragmatic markers and conjunctions) in a 24,000-word corpus of transcribed oral data from 47 learners and native speakers of English. Both of these cohesive devices increase with proficiency level, but not in the same way. Conjunction use seems to increase steadily, and only the differences between the highest and lowest proficiency levels were found to be statistically significant. Pragmatic marker use, however, remains fairly stable across the three lowest proficiency levels and rises drastically for the two highest proficiency levels, and the two higher proficiency levels are significantly different from the two lower levels in their use of pragmatic markers. The results are compared to native speaker rates of cohesive device use for the same tasks and under the same conditions.

Keywords: cohesive device, pragmatic marker, conjunction, cohesion, proficiency level

The production of coherent speech is an important part of the development of second language competence. It facilitates the act of communication, making it easier for the listener and speaker to understand and to be understood. One of the ways in which a coherent discourse is achieved is through the use of cohesive devices such as pragmatic markers (expressions such as *so,*

I think, kind of) and conjunctions.[1] These devices indicate the relationships (logical or interpersonal) between clauses or supra-clausal units of discourse.

These expressions are so important for communication that Crossley, Salsbury, and McNamara (2010) found that nonnative-like use or nonuse of lexical cohesive devices resulted in negotiation for meaning in conversations in the same way that other errors or misunderstandings would.

This study will examine the use of two types of cohesive devices, that is pragmatic markers and conjunctions, in oral data from second language learners of English. These data will also be compared with those of native speakers' of English performing the same task. Previous research that has looked at the use of cohesive devices has not divided them into the categories of pragmatic markers and conjunctions. But this division is supported by recent corpus-based grammars of English (Biber, Johansson, Leech, Conrad, & Finegan, 1999; and Carter & McCarthy, 2006). The fact that previous research has examined all cohesive devices together has possibly obscured the fact that they are acquired and used differently by language learners.

Background

Cohesive Devices

According to Halliday and Hasan (1976) and Halliday and Matthiessen (2004), expressions such as pragmatic markers and conjunctions facilitate the construction of cohesive discourse. Halliday and Hasan argue that the fundamental property of a text (which can be either written or spoken) is cohesion, a semantic property which "refers to relations of meaning that exist within the text, and that define it as a text" (p. 4). Cohesion is achieved through the use of cohesive relations. Halliday and Hasan outline four such relations: reference, substitution, ellipsis, and conjunction. Conjunction as a cohesive relation works through the use of conjunctions but also through various other "conjunctive items." The conjunctive items that Halliday and Hasan discuss include classic discourse markers such as *now, well*, and *I mean*. Halliday and Hasan therefore consider both pragmatic markers and conjunctions to be important vehicles for the construction of cohesion in discourse. Since Halliday and Hasan's (1976) study, a variety of further research has explored how pragmat-

[1] Terminology for these types of expressions varies greatly and will be discussed below. The term *cohesive devices* will be used to refer to pragmatic markers and conjunctions together. In this study, the term *pragmatic markers* will be used instead of the (also common) term *discourse markers*.

ic markers in particular establish connectivity between different units of discourse (Fraser, 1996; Knott and Dale, 1994; Redeker, 1991; Schiffrin, 1987).

Another view of the importance of cohesive devices such as pragmatic markers and conjunctions can be found from the perspective of second language learning, in the American Council on the Teaching of Foreign Languages (ACTFL) guidelines for conducting Oral Proficiency Interviews (OPI) (Swender 1999). These guidelines incorporate the creation of discourse into their rating criteria under the rubric of discourse competence. One of the aspects of the progression from the Novice through to Superior level is the ability to generate discourse at various levels: word-level discourse at the Novice level, sentence-level discourse at the Intermediate level, paragraph-level discourse at the Advanced level, and multi-paragraph discourse at the Superior level. The ACTFL thus considers the development of discourse to be intrinsically connected to the development of language proficiency. But the question of how learners accomplish this developmental progression remains. Certainly this issue is complex and involves the overall development of language proficiency. However, it is equally clear that in order to move from the sentence to the multi-paragraph level of discourse, learners must also be able to control a range of expressions and devices for linking their utterances into a coherent discourse. Such expressions can be broadly called cohesive devices, and they also form part of the ACTFL guidelines. The ACTFL OPI training manual (Swender 1999) defines cohesive devices in the following way:

> words and phrases that link ideas and move forward the action in some form of logical narrative order, whether the "narrative" is a story, a description, or a set of instructions. Adverbs and conjunctions serve most frequently as cohesive devices (words and phrases such as: *and, but, because, suddenly, in the first place, however*). They permit logical sequencing; they establish time-frames for actions and events; they create structures of meaning by establishing principal and supporting language units; they help create and sustain comparisons; they provide an oppositional linguistic framework for debate. (p. 100)

The ACTFL guidelines consider the correct usage of cohesive devices to be a hallmark of the Advanced and Superior levels. Therefore, an examination of the use of pragmatic markers and conjunctions, which are some of the vehicles which provide the means for linking ideas together, should show that more advanced learners use more cohesive devices.

Cohesion and Coherence

Halliday and Hasan (1976) and Halliday and Matthiessen (2004) discuss how cohesion is realized in texts (both written and spoken). It is important to note,

however, that cohesion is only one way in which texts are coherent, and that it alone does not guarantee coherence. As Tanskanen (2006) points out, it is possible to construct a cohesive text that is not coherent and a coherent text that does not display any overt cohesive elements. Widdowson (1979, p. 138) offers the following example of an exchange that is coherent without being cohesive:

> A: *That's the telephone.*
> B: *I'm in the bath.*
> A: *OK.*

This text can be understood as a coherent exchange, even though none of Halliday and Hasan's cohesive relations are specified. Similarly, a text can also be overtly cohesive without being coherent. Another example, taken from Enkvist (1978), illustrates this point:

> *The discussions ended last week. A week has seven days. Every day I feed my cat. Cats have four legs. The cat is on the mat. Mat has three letters.*

The sentences above show strong cohesion in that each sentence contains the same noun as the previous sentence. Despite these connections, however, this text is not coherent: The connection between the sentences is entirely on the surface, and they are not otherwise joined together in any meaningful way.

Although cohesion and coherence are not interchangeable, this does not mean that they are not both important for language use. That is, although a text can possibly be coherent without being cohesive for the purposes of definition, this does not mean that coherence is not generally realized through the use of cohesive ties in most texts. For this reason, exploring the various ways in which cohesion is manifested is still important for understanding how coherence operates.

There are two noticeable limitations in the research on cohesive devices to date that might lead to confusion over terminology and apparently contradictory results. First of all, there is a great deal of terminological overlap and varying definitions of the expressions in question. What one researcher might label a *discourse marker* might be listed by another researcher as a *linking adverbial* or vice versa. To take one example, *however* appears in Halliday and Hasan's (1976) list of adversative conjunctive items (p. 242), in Fung and Carter's (2007) study as a discourse marker, and in Murray (1997) as a connective.

To give some idea of the diversity of terms, one need only consider research on second language production of these expressions. They have been examined as *smallwords* (Hasselgreen, 2005), *discourse markers* (Fuller, 2003; Fung & Carter, 2007; Hellerman & Vergun, 2007; Redeker, 1990; Romero-Trillo, 2002), *cohesive devices* (Hinkel, 2001; Liu & Braine, 2005), *cohesive features*

(Zhang, 2000), *logical connectors* (Green, Christopher, & Mei, 2000), *lexical bundles* (Nesi & Basturkmen, 2006), *connectors* (Bolton, Nelson, & Hung, 2003; Granger & Tyson, 1996), and *connectives* (Ozono & Ito, 2003; Yeung, 2009).

It is worth noting that some of the differences in terminology in these investigations stem from whether they focus on cohesion in speech or writing. Studies that have investigated speech have more generally looked at *discourse markers*, while those that have looked at writing have more often focused on *connectives*. The writing-speaking distinction is not hard and fast; for example, Siepmann (2005) examines discourse markers in writing.

It is not clear that such a division between cohesive devices in speech and writing is warranted. To the extent that these studies relate in some way to Halliday and Hasan's (1976) work on cohesion, such a division between written and spoken modalities is also unfounded. On the contrary, Halliday and Hasan emphasize that their use of the word *texts* refers to both written and spoken language.

Moving beyond Halliday and Hasan, it is also difficult to justify any absolute distinction between speech and writing. This can be illustrated by considering the case of *kind of*, a common pragmatic marker that could be assumed to be infrequent in writing. A search of the Corpus of Contemporary American English shows that *kind of* is indeed much more frequent in the spoken register (844 occurrences per million words) than in the written registers (295 per million in fiction, 224 per million in magazines, 233 per million in newspapers, and 153 per million in academic writing), but it is still clearly present in the latter.

The second limitation apparent in previous research is that cohesive devices have not been separated into different categories according to their grammatical function. This is to some extent not surprising since Halliday and Hasan (1976) do not separate them (except of course, into additive, adversative, causal, and temporal functions). However, recent comprehensive corpus-based grammars of English do separate them (see below). The question of whether looking at all cohesive devices together might obscure important differences in their acquisition and use remains unanswered.

This study aims to address the limitations of previous research by looking at conjunctions and pragmatic markers separately, following the terminology and identification of expressions of Carter and McCarthy (2006).

Separation of Pragmatic Markers and Conjunctions

There are both theoretical and empirical grounds for separating these two types of expressions. First of all, recent corpus-based grammars of English

have separated these items into different categories (Biber, Johansson, Leech, Conrad, & Finegan, 1999; and Carter & McCarthy, 2006).

Carter and McCarthy (2006) distinguish between conjunctions and pragmatic markers. They define conjunctions as "items used to mark logical relationships between words, phrases, clauses and sentences" (p. 897). They subdivide the category of conjunctions into subordinating (such as *although, after, as, because, before, since,* and *when*) and coordinating conjunctions (*and, but,* and *or*).

Carter and McCarthy (2006) define pragmatic markers as "a class of items which operate outside the structural limits of the clause and encode speakers' intentions and interpersonal meanings" (p. 208). They subdivide pragmatic markers into discourse markers, stance markers, hedges, and interjections.

The fact that some influential theoretical sources separate pragmatic markers and conjunctions, while the research literature does not, indicates a disconnect between the theoretical and empirical approaches to discourse cohesion. It is of course possible that this disconnect is of little consequence, and that both types of expressions are acquired in much the same way. But it is also possible that the grouping of these two types has masked important differences in how pragmatic markers and conjunctions are acquired and used by second language learners.

In order to see whether this is the case, this study will subdivide cohesive devices into the categories of pragmatic markers and conjunctions. Carter and McCarthy's (2006) definition was chosen for use in this study based on several factors. The first, and most important of these, is that this definition was generated from a corpus-based investigation of actual language use. The corpus-based nature of the definition makes it particularly appropriate for this study since it likewise investigates a corpus of native and nonnative speech. Carter and McCarthy's definition is also a good choice for a working definition of pragmatic markers because the corpus it is based on encompasses both spoken and written data from different varieties of English. Since the definition was derived from actual uses and explained with a number of examples, it is also well suited for use in identifying expressions from a corpus of language in context.

The Study

Aim and Methodology

The aim of this study is to examine the use of cohesive devices in spoken English by learners at varying proficiency levels in comparison with native speakers. In order to examine the use of cohesive devices, this study utilizes a corpus of transcribed speech. A learner corpus such as this one is valid for the

investigation of phenomena such as the expression of cohesion in second language research because it captures continuous discourse in context (Cobb, 2003; Granger, 2009). The continuous stretches of learner speech allow us to examine the various devices that learners may be using (or not using) to create cohesive ties in their language. The corpus used in this study includes data from learners at multiple proficiency levels as well as native speakers. All of the subjects are performing the same tasks under the same conditions.

Previous research that has examined cohesion with multiple proficiency levels and native speaker comparison data is scant. Hasselgreen's (2005) work on Norwegian learners of English is one exception. Her focus, however, was not cohesion, but fluency, and she was examining something she termed *smallwords*, which are similar to pragmatic markers. Hasselgreen found statistically significant differences among native speakers and nonnative speakers at two different proficiency levels, with the native speakers using the most smallwords, and the lower proficiency group using the fewest smallwords.

The current study expands upon Hasselgreen's work by including subjects at a wider range of proficiency levels (four nonnative speaker groups), different language backgrounds (Korean and Chinese instead of Norwegian), and different types of tasks (monologic instead of dialogic tasks in Hasselgreen's study).

Research Question

The research question posed in this study, along with its corresponding hypotheses, is as follows:

> Research Question 1: How are cohesive devices related to proficiency level?
> Hypothesis 1: Learners at higher proficiency levels will use more conjunctions than learners at lower proficiency levels.
> Hypothesis 2: Learners at higher proficiency levels will use more pragmatic markers than learners at lower proficiency levels.

As discussed in the review of the literature, pragmatic markers and conjunctions can both be considered to belong to the category of cohesive devices, which function to create connections in oral discourse.

Since we can assume that the speech of learners at higher proficiency levels will be more complex and require more connecting devices, the hypotheses above argue that both pragmatic marker and conjunction use will be higher at higher proficiency levels. There are several additional reasons to argue that pragmatic markers in particular would be more frequent in the speech of learners at higher proficiency levels. Pragmatic markers are not usually taught in the classroom, and high proficiency level learners, who have

probably had some experience in the country where the language is spoken, will have already acquired them. Furthermore, lower-proficiency level learners presumably have to allocate much more attention to formulating their basic message and do not have any resources to spare to indicate additional interpersonal meaning or interpretive information.

Participants

Data from 47 subjects were analyzed. The nonnative speaker examinees were all graduate students and prospective teaching assistants at an American university. The nonnative speakers were grouped into four different proficiency levels, numbered 3-6. There were 10 participants at Level 3, 10 at Level 4, 10 at Level 5, 7 at Level 6, and 10 native speakers. All of the nonnative speaker examinees came from either a Chinese or Korean language background.[2] A table showing the language background and scores of each of the participants can be found in Appendix A.

The data for each level were evenly split between examinees with a Chinese and Korean L1 background; that is, there were 5 examinees with L1 Chinese and five examinees with L1 Korean in each group of 10 examinees. The group of 7 examinees at Level 6 was made up of 3 native Chinese speakers and 4 native Korean speakers. Level 6 had only 7 examinees because that number was the total available in the testing records.

The examinees all had some previous training in English and took the test to be certified to teach at the university level. No further information about the examinees' previous exposure to or study of English is available.

Design and Procedure

The data examined in this study come from a corpus collected from a semidirect test of oral proficiency (hereafter OPT) that is administered to prospective international teaching assistants. The OPT is composed of naturalistic tasks that the examinees are asked to perform which are computer-recorded. This data elicitation technique has several advantages for the present study. First of all, the fact that the conditions under which the data are gathered are

[2] These two language backgrounds were chosen based on several criteria. The first was that enough speakers of these languages take the exam in order to provide a range of subjects at every proficiency level. The second consideration was that these groups learned English as a foreign language, but their schooling did not take place in English, which might be the case for Indian learners of English. Two language groups, rather than one, were chosen in order to counterbalance the results against the possibility of transfer effects from the L1.

held constant enhances the comparability of the responses. The context is the same for all of the examinees, which allows the discourse they produce to be compared. Secondly, since the data are part of an oral proficiency exam, the samples have already been classified into different proficiency levels by trained raters. This allows for the comparison of learners at different proficiency levels, as well as comparison with native speakers who have also taken the exam for comparison purposes. Lastly, the naturalistic tasks allow for the examination of longer runs of speech in context, which is well-suited for examining cohesion.

There are ten tasks on the OPT. For this study, I selected four of the ten tasks for transcription and coding. The four tasks selected were "news," "personal," "passing information," and "telephone." These four tasks were chosen out of the ten available in order to provide a range of task types and levels of structure.

In the news task, the examinees are asked to give their opinion about a news item they have read. In the personal task, the examinees give a response to an open-ended audio question about their personal experience such as how they learned English, or who their favorite teacher is. In the passing-information task, the examinees relate some information that they have read to someone who has no knowledge of it, such as describing a job notice to someone they think might like to apply for the job. In the telephone task, the examinees leave a short message that they have heard for one of their office mates on that person's home answering machine.

The particular questions used in the different tasks from the OPT were not identical; rather, they came from several different forms of the test that are given regularly. Although one version of the test is no more difficult than another, this does not mean that they were identical for the purposes of this study. Since using different versions of the test allowed for access to a greater variety of data, the benefits outweighed the potential downside of this approach.

The exams are rated by two trained raters. When the two raters disagree, the exam is sent to a third rater to break the tie. The exams are given scores from 2 to 6. A set of descriptors for each level can be found in Appendix B. For the purposes of the program, a score of 5 or 6 is considered sufficient for the examinee to be certified to teach undergraduates as a teaching assistant. If an examinee receives a score of 3 or 4, they must enroll in a course in oral English for teaching assistants. At the end of this course they may either be certified to teach in the classroom or asked to repeat the course. A score of 2 indicates that the examinee's oral proficiency skills are not developed enough for them to benefit from the oral English course. Scores of 2 or 6 are rare. Scores of 2 are generally given only when the examinee is clearly overwhelmed by the demands of the task and gives little or no response. Scores of

6 are given primarily to examinees whose second language proficiency approaches native or near-native competence.

The oral exam responses were transcribed by the present author. The responses were then coded for discourse markers and conjunctions by the author and a colleague, who is also a native speaker of English and an experienced instructor of second languages. In order to identify discourse markers and conjunctions, the definition and lists of discourse markers from Carter and McCarthy (2006) were used as a guide (see an earlier section for a discussion of this definition). Interrater reliability was .94. In the case of a discrepancy in the identification of the cohesive devices, the two coders discussed the example in question and came to an agreement.

Results and Findings

Table 1 gives a summary of the data collected. The mean number of words, pragmatic markers, and conjunctions produced by the examinees at each level are presented, along with the corresponding standard deviations, minimums, and maximums. The mean number of total words, pragmatic markers, and conjunctions rose from Level 3 through to 6. For each category, Level 6 had the highest mean values. The native speaker (NS) group produced fewer mean total words, pragmatic markers, and conjunctions than the Level 6 group, but more than the Level 5 group.

Table 1 Descriptive statistics for pragmatic marker and conjunction use

Level	*N*	Mean no. of words (*SD*)	Min/Max	Mean no. of pragmatic markers (*SD*)	Min/Max	Mean no. of conjunctions (*SD*)	Min/Max
3	10	421 (79)	320/359	9 (5)	1 / 17	20 (7)	11/30
4	10	525 (144)	296/823	13 (10)	2 / 36	26 (12)	11/49
5	10	579 (122)	320/718	18 (10)	7 / 41	31 (12)	11/50
6	7	743 (116)	592/901	40 (21)	23 / 85	45 (12)	27/63
Native speakers	10	584 (215)	326/1008	31 (13)	8 / 48	42 (17)	21/75

Figure 1 shows the average percentages of pragmatic markers and conjunctions used by speakers at different proficiency levels. The rates of pragmatic marker and conjunction usage were calculated by dividing the total number of pragmatic markers or conjunctions used by a speaker by the total number of words. Figure 1 shows that the rate of conjunction use rose with proficiency level, with the native speaker group using conjunctions at the

highest rate. Pragmatic marker use also rose with proficiency level, with Level 6 and the native speaker group using pragmatic markers at the highest rate.

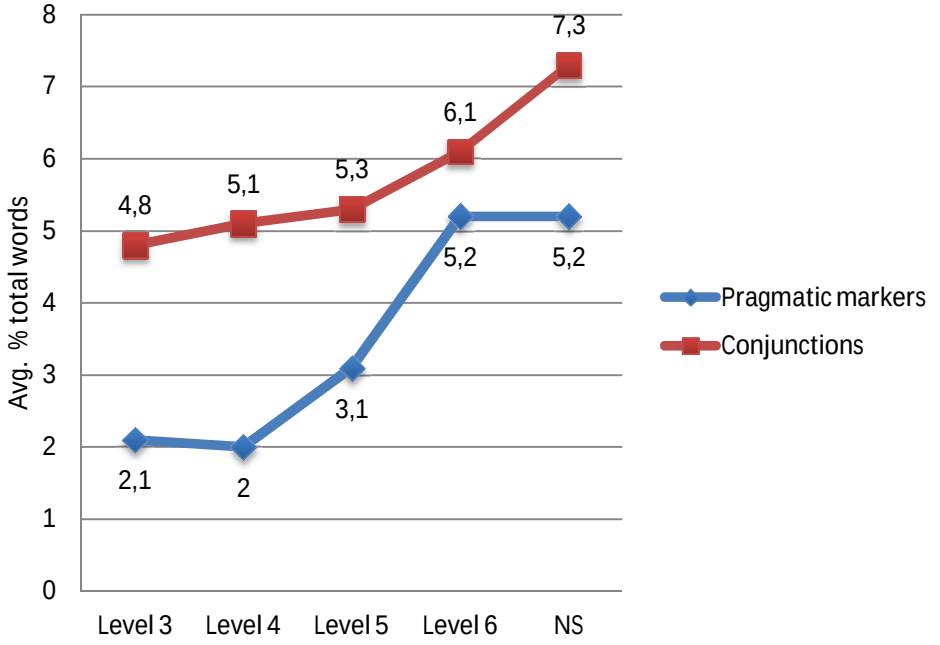

Figure 1 Pragmatic marker and conjunction use by proficiency level

The results above indicate that pragmatic marker and conjunction use rose with proficiency level. The question remains whether the differences in the use of pragmatic markers and conjunctions at different proficiency levels were statistically significant. In order to answer this question, two mixed-model ANOVAs were conducted, one on the pragmatic marker data and one on the conjunction data. The results show that proficiency level was a significant factor (p = .0001) in pragmatic marker use (F = 7.42). After proficiency level was found to be significant for the rate of pragmatic marker use, Tukey-Kramer pairwise comparisons were conducted to locate the source of the significance. As Tables 2 and 3 show, proficiency Levels 3 and 6, 3 and NS, 4 and 6, 4 and NS, and 5 and NS were significantly different from each other in their pragmatic marker use. The results of the post-hoc analysis show that the two lowest proficiency levels (3 and 4) were significantly different from the two highest proficiency levels (6 and the native speakers). In addition to this difference, the Level 5 group was also significantly different from the native speakers.

Table 2 Significant differences in pragmatic marker use

Proficiency level	Level 3	Level 4	Level 5	Level 6	Native speakers
Level 3				*	*
Level 4				*	*
Level 5					*
Level 6	*	*			
Native speakers	*	*	*		

* $p < .05$

Table 3 The post-hoc results for pragmatic marker use

Proficiency level	Proficiency level	Adjusted *p* value
3	6	.0023
3	NS	.0007
4	6	.0281
4	NS	.0119
5	NS	.0480

NS = native speakers

A separate ANOVA was conducted on the conjunction data to determine if the different proficiency levels differed significantly in their conjunction use. Table 4 below shows the results of the ANOVA, which indicate that proficiency level was a significant factor (p=0.0380) in conjunction use. After the ANOVA showed that proficiency level was a significant factor in the speakers' conjunction use, Tukey-Kramer pairwise comparisons were then conducted to locate the source of the significance. This post-hoc test showed that only two proficiency levels were significantly different from each other, namely, Level 3 and the group of native speakers.

Table 4 Significant differences in conjunction use

Proficiency level	Level 3	Level 4	Level 5	Level 6	Native speakers
Level 3					*
Level 4					
Level 5					
Level 6					
Native speakers	*				

* $p < .05$

Table 5 The post-hoc results for pragmatic marker use

Proficiency level	Proficiency level	Adjusted *p* value
3	NS	.0357

NS = native speakers

Discussion

The first point to note is that the examinees used more conjunctions than pragmatic markers at every proficiency level. For the lower levels, this gap is large, but it then narrows at Level 6, at which only a percentage point separates them. The gap then starts to widen again for the native speakers, as their use of conjunctions continues to rise in comparison with the nonnative speakers while the number of pragmatic markers remains flat. The trend for conjunction use rises as the proficiency level goes up, although the gain from Level 3, at 4.8%, to Level 5, at 5.3%, is small. A much more dramatic gain can be seen after Level 5; the Level 6 usage jumps almost a full percentage point from Level 5, and the native speaker group is more than a percentage point above the Level 6 group. Thus the overall pattern is that conjunction use rises with proficiency level, with the greatest gain occurring at the two highest proficiency levels. This confirms Hypothesis 1.

The pattern of pragmatic marker use also rises with proficiency level. Pragmatic marker use rises slightly from Level 3 to Level 4. The sharpest gain in pragmatic marker use occurs between Levels 5 and 6, with a jump of 2%. But pragmatic marker usage does not continue to rise for the native speakers; instead, their usage is identical to that of the Level 6 group. The pattern can be summarized as follows: the two highest and the two lowest proficiency groups are very alike in their pragmatic marker use, but that the middle level shows a dramatic rise in pragmatic marker use. This confirms Hypothesis 2.

The results presented above have showed that pragmatic marker and conjunction use rise with proficiency level. This result is not surprising, given that it is expected that pragmatic marker use would rise with proficiency level. The use of pragmatic markers shows that the speakers are able to provide additional information about their message, in addition to its basic meaning. This implies that they do not have to allocate as much attention to formulating their basic message and can instead give some attention to how they would like it to be interpreted. This is reflected in increased pragmatic marker use at higher proficiency levels.

What is interesting about these results is the fact that the different proficiency levels are more alike in their conjunction use than in their use of pragmatic markers. That is, there is less of a difference between the highest and lowest proficiency groups in the average rate of conjunction use than there is in pragmatic marker use. Furthermore, pragmatic marker use rises more suddenly as proficiency level rises, while conjunction use rises more gradually. This indicates that pragmatic marker use may be a more useful factor in discriminating between different proficiency levels.

The results have shown that the use of pragmatic markers distinguishes among learners at different proficiency levels to an extent that conjunction use does not. At this point, it is instructive to consider how this difference is manifested in two samples from the data, one from a nonnative speaker who was classified at Level 3, and one from a native speaker of English. In these samples, both subjects complete the same task, in which they were asked to pass on information about a flyer they saw advertising a French class to a friend of theirs who they knew was interested in learning French:

Example 1, Level 3 examinee:
Hello, my friend. I know you are looking for an elementary level French course. But, I found, the French course in X University, so, I want to let you know about that. Ah, X University will have the beginning French course…It is scheduled to start next week. The admission is on the first-come, first-served basis. No prior knowledge of French is required. And this class will be held every Thursday, from 1pm to 2pm, just one hour. But, this course will require regular attendance and participation, and the late registration fee is 25 dollars. You can pay by check or cash. If you wanna pay by check, you can mail your check to International Center, X Street, Town, State The zip code is XXXXX. And you have to include your name, and address, and contact information. If you wanta get further information, you can call XXX-XXX-XXXX. Good luck.

Example 2, Native Speaker examinee
Hey Jill I know that you were ah looking to take a French class because you're moving to Canada in a few months and looking for a job there. I just saw this bulletin ah I thought you might be interested in in ah this international center giving a French class. Ah, enrollment's beginning soon, and actually they're starting next week, so you might wanna come f--. It's on a first served, first come first served basis, but the neat thing about it is, you don't have any prior knowledge in French. And classes are held every Thursday from 1 to 2. And they do ask that you attend regularly, obviously if you don't you're not going to learn so much French. The fee is a little bit steep, 25 dollars, but that's not bad because it includes the materials that you'll need. So, ah if you're interested maybe, I don't know if you're interested in this particular class or not, but if you are, ah, you might want to check the international center on X Street, or write to them, on R—123 X Street, Town. And ah, you know, find some more information about it. If you want ah more information, you eve-ev-even call them. I think ah I have their number written down here. I think it's XXX-XXXX. So, it's it's an interesting thing. I think you might enjoy it and it, 'll give you at least a beginning level of French, and you'll find it's helpful when you get to Canada. Wish I could do it too.

The examples above show that while the two speakers are relaying the same information, the ways in which they indicate their knowledge of it and attitudes towards it varies. Table 6 compiles all of the information that both speakers presented and lists the wording that they used to introduce it or that they used while explaining it. Note that not all of these expressions were counted as pragmatic markers for the purposes of this study. If an expression was counted as a pragmatic marker, it is underlined in the transcripts above and in the table.

Table 6 Information presented and the wording used by a native speaker and a learner

No.	Information related	Wording used by the native speaker to introduce or nuance information	Wording used by the learner to introduce or nuance the information
1.	You've been looking for a French course	I know	I know
2.	This French course is scheduled to start next week.	Actually	
3.	First-come, first-served admissions		
4.	No prior knowledge of French is required.	But the neat thing about it is…	
5.	It meets every Thursday from 1-2.	And	And
6.	Regular attendance is required.	And…obviously	But
7.	The fee is $25.	A little bit	And
8.	Contact the international center for information.	Even…I think	

Although the comparison of only two samples of discourse from two speakers is limited in its generalizability, several interesting differences can be observed in the examples. The native speaker uses pragmatic markers or other expressions in communicating at several points where the learner uses nothing (items 2, 4, and 8 in Table 6). At two different points, both the learner and the native speaker use the same devices to introduce the information (items 1 and 5). What is particularly interesting is item 6, where the native speaker uses *and* to introduce the information that the course requires regular attendance and *obviously* to indicate that they think that this is a reasonable requirement: "*And they do ask that you attend regularly, obviously if you don't you're not going to learn so much French*." It may be that *obviously* is also inserted here to soften the existence of a requirement in a discourse that is generally meant to put the course in a positive light. The nonnative speaker chooses the conjunction *but* to introduce the same information: "*But, this course will require regular attendance and participation, and the late registration fee is 25 dollars*." In this sentence, the nonnative speaker subject uses *but* to introduce what could be considered the negative aspects of the course, or at least the points that do not in any way serve as inducements to the imagined interlocutor. The course costs

money and you cannnot just drop in whenever you want. As we have already seen, the native speaker attempted to soften this information with the use of pragmatic markers *obviously* and *a little bit*. The nonnative speaker may possibly also be attempting to use a softener in their use of *but* to introduce this information. This is not an incorrect use of *but*; however, its use here seems, at least to the present author, to lack some additional wording, such as *this course should be perfect for what you need, but you should know that attendance is required/ but it does have an attendance requirement*.

This task of passing on supposedly relevant information requires not only that the speaker relay the information, but also that the speaker explain why they think or know the information to be relevant to the person in question. Thus, there is a good deal of justification that usually goes into this response. Both speakers in the example above may have felt the same pragmatic need to relay information that might be perceived negatively by his interlocutor. But where the native speaker used a pragmatic marker, the nonnative speaker used a conjunction which is possibly the cohesive device that is closest in function to the meaning they would like to relay. This is an interesting possibility for further research into the relationship between these two cohesive devices.

Conclusions

Limitations

There are a number of important limitations to the current study. The first limitation is lack of agreement concerning the definition and identification of pragmatic markers. This study attempted to mitigate this difficulty by relying on a standardized list of pragmatic markers (Carter & McCarthy, 2006). However, it is important to note that other studies might find different results using a different definition and list of pragmatic markers.

A different issue comes up when we consider the identification of conjunctions. Although conjunctions do not present the same difficulties in identification as discourse markers do, neither are they a homogenous class. But distinctions within conjunctions, such as coordinating versus subordinating, could be important, especially when considering their role in creating cohesion between two utterances, and the extent to which particular tasks (narrating, describing) might favor the use of either type of conjunction.

A final limitation of this study is the testing context from which the data are drawn. Some previous research mentioned above was conducted using classroom data. The data used in this study comes from a semidirect oral proficiency test made up of discrete tasks with a few minutes of planning time.

The importance of the context of the data used in this study should not be underestimated. Shohamy (1994) found important differences in the language performance of learners on different types of oral proficiency interviews. Therefore, the testing context of the data in this study might have an effect on the language produced by the examinees. The effect of the context from which language learner data is gathered is often overlooked and offers a promising direction for future research.

Summary and Directions of Future Research

The results discussed above indicate that speakers at different proficiency levels are more easily and clearly distinguished by their pragmatic marker use than their conjunction use. That is, there are more significant differences between different proficiency levels involving pragmatic marker use than conjunction use. Even pragmatic marker use, however, does not appear to distinguish between speakers who are grouped into adjacent proficiency levels. In one respect, this is not surprising; we would expect that speakers who are close in proficiency level would also be close in other, more specific measures, such as the use of pragmatic markers. But this raises the question of what, if any, significant differences exist in their speech and, in turn, whether some of the finer distinctions that are made between different levels of speakers are in fact valid.

The results for conjunction use are more modest. Conjunction use rose with proficiency level, but there were fewer significant differences in conjunction use between proficiency levels. Only the difference between the highest and lowest proficiency groups was significant. This indicates that conjunction use is not as robust an indicator of differences in proficiency level as pragmatic marker use.

These results show that it is instructive to distinguish between pragmatic markers and conjunctions when investigating the use of cohesive devices by learners. When the two types of cohesive devices are examined together, this difference is obscured.

Why do pragmatic markers and conjunctions pattern so differently, if they are both types of cohesive devices? This could be due to a number of different factors. One possible contributing factor is instruction. Anecdotal reports indicate that pragmatic markers are not commonly taught or emphasized in the way that conjunctions are in ESL classrooms. This lack of attention to pragmatic markers leaves learners to notice and acquire them incidentally entirely from the input. More input or more contact with native speakers is then likely needed before these types of expressions are acquired. Another possible factor is the wider scope of pragmatic markers. Since pragmatic markers are understood to operate at a multiclause or sentence level and to encode interpersonal mean-

ings and the speakers' intentions, their use is both more complex and more optional than conjunctions, which mark logical relationships up to the sentence level. Encoding this additional pragmatic meaning is a luxury which lower-proficiency learners do not have, which would explain why there were more significant differences in discourse marker use among levels.

Moreover, the results reported here were made possible by the use of a corpus of learner speech, which highlights the efficacy of using computer learner corpora to generate and investigate research questions in the area of second language acquisition. Together, these factors offer several benefits in examining second language learner discourse. Especially in the area of discourse competence, more research into longer stretches of learner production is necessary to further refine the possible differences between the roles that pragmatic markers and conjunctions play in the development of learner competence.

References

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London: Longman.

Bolton, K., Nelson, G., & Hung, J. (2003). A corpus-based study of connectors in student writing: Research from the International Corpus of English in Hong Kong ICE-HK. *International Journal of Corpus Linguistics*, *7*(2), 165-182.

Carter, R., & McCarthy, M. (2006). *Cambridge grammar of English: A comprehensive guide. Spoken and written English. Grammar and usage*. Cambridge: Cambridge University Press.

Cobb, T. (2003). Analyzing late interlanguage with learner corpora: Quebec replications of three European studies. *Canadian Modern Language Review, 59*(3), 393-424.

Crossley, S., Salsbury, T. L., & McNamara, D. S. (2010). The role of lexical cohesive devices in triggering negotiations for meaning. *Issues in Applied Linguistics*, *18*(1), 55-80.

Enkvist, N. E. (1978). *Coherence, pseudo-coherence, and non-coherence*. Retrieved from ERIC database. (ED275188)

Davies, Mark. (2008-) *The corpus of contemporary American English: 425 million words, 1990-present*. Available online at http://corpus.byu.edu/coca/

Fraser, B. (1996). Pragmatic markers. *Pragmatics, 6*, 167-90.

Fuller, J. (2003). The influence of speaker roles on discourse marker use. *Journal of Pragmatics*, *35*(1), 23-45. doi:10.1016/S0378-2166(02)00065-6

Fung, L., & Carter, R. (2007). Discourse markers and spoken English: Native and learner use in pedagogic settings. *Applied Linguistics*, *28*(3), 410-439. doi:10.1093/applin/amm030

Granger, S. (2009). The contribution of learner corpora to second language acquisition and foreign language teaching. In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 13-32). Amsterdam: John Benjamins.

Granger, S., & Tyson, S. (1996). Connector usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes*, *15*(1), 17-27.

Green, C. F., Christopher, E. R., & Mei, J. L. K. (2000). The incidence and effects on coherence of marked themes in interlanguage texts: A corpus-based enquiry. *English for Specific Purposes*, *19*(2), 99-113.

Halliday, M. A. K., & Hasan, R. (1976). Cohesion in English. London: Longman.

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2004). *An introduction to functional grammar*. London: Arnold.

Hasselgreen, A. (2005). *Testing the spoken English of young Norwegians: A study of test validity and the role of "smallwords" in contributing to pupils' fluency*. Cambridge: Cambridge University Press.

Hellermann, J., & Vergun, A. (2007). Language which is not taught: The discourse marker use of beginning adult learners of English. *Journal of Pragmatics*, *39*(1), 157-179. doi:10.1016/j.pragma.2006.04.008

Hinkel, E. (2001). Matters of cohesion in L2 academic texts. *Applied Language Learning*, *12*(2), 111-132.

Knott, A. and Dale, R. (1994). Using linguistic phenomena to motivate a set of coherence relations. *Discourse Processes 18*, 35-62.

Liu, M., & Braine, G. (2005). Cohesive features in argumentative writing produced by Chinese undergraduates. *System*, *33*(4), 623-636.

Murray, J. D. (1997). Connectives and narrative text: The role of continuity. *Memory & Cognition, 25* (2), 227-236.

Müller, S. (2005). *Discourse markers in native and non-native English discourse*. Amsterdam: John Benjamins.

Nesi, H., & Basturkmen, H. (2006). Lexical bundles and discourse signalling in academic lectures. *International Journal of Corpus Linguistics*, *11*(3), 283-304.

Ozono, S., & Ito, H. (2003). Logical connectives as catalysts for interactive L2 reading. *System*, *31*(2), 283-297.

Redeker, G. (1990). Ideational and pragmatic markers of discourse structure. *Journal of Pragmatics*, *14*(3), 367-381.

Redeker, G. (1991). Linguistic markers of discourse structure. *Linguistics 29*, 1139-1172.

Schiffrin, D. (1987). *Discourse markers*. Cambridge: Cambridge University Press.

Siepmann, D. (2005). *Discourse markers across languages: A contrastive study of second-level discourse markers in native and non-native text with implications for general and pedagogic lexicography*. Abingdon: Routledge.

Shohamy, E. (1994). The validity of direct versus semi-direct oral tests. *Language Testing, 11*(2), 99-123.

Swender, E. (1999). *ACTFL oral proficiency interview tester training manual*. Yonkers, NY: American Council on the Teaching of Foreign Languages.

Tanskanen, S. K. (2006). *Collaborating towards coherence: Lexical cohesion in English discourse*. Amsterdam: John Benjamins.

Widdowson, H. G. (1979). *Explorations in applied linguistics*. Oxford: Oxford University Press.

Yeung, L. (2009). Use and misuse of "besides": A corpus study comparing native speakers' and learners' English. *System*, 37(2), 330-342. doi:10.1016/j.system.2008.11.007

Zhang, M. (2000). Cohesive features in the expository writing of undergraduates in two Chinese universities. *RELC Journal*, 31(1), 61-95.

APPENDIX A

List of participants

Native language	Native country	Overall score
Chinese	China	3
Chinese	China	3
Korean	Korea	3
Korean	Korea	3
Korean	Korea	3
Korean	Korea	3
Korean	Korea	3
Chinese	China	3
Chinese	China	3
Chinese	China	3
Chinese	China	4
Chinese	China	4
Chinese	China	4
Chinese	China	4
Chinese	China	4
Korean	Korea	4
Korean	Korea	4
Korean	Korea	4
Korean	Korea	4
Korean	Korea	4
Chinese	China	5
Chinese	China	5
Chinese	China	5
Chinese	China	5
Chinese	China	5
Korean	Korea	5
Korean	Korea	5
Korean	Korea	5
Korean	Korea	5
Korean	Korea	5
Chinese	China	6
Chinese	China	6
Chinese	China	6
Korean	Korea	6
Korean	Korea	6
Korean	Korea	6
Korean	Korea	6
	U.S.	
	U.S.	
	U.S.	
	U.S.	
	U.S.	
	U.S.	
	U.S.	
	U.S.	
	U.S.	
	U.S.	
	U.S.	

APPENDIX B

Descriptors for the Levels of the Oral Proficiency Test

Level 6	
Content	Delivery
Wide range of vocabulary	Smooth delivery
Complexity of sentence structure	Almost no pauses/ hesitations/ choppiness
Interpretative/summary statements	Thought expressed in one utterance
Some non-native usage	No problems with articulation
Meaning clearly expressed	Use of varied intonation and tone
Provision of a frame	
Economy of expression	

Level 5	
Content	Delivery
Somewhat unconventional words	Clearly non-native like delivery
Listener effort needed at times	Some pauses and choppiness, but comprehen-
Simple sentence construction	sion unobstructed
Well organized and coherent	Some sound substitutions
Meaning clear	Listener effort required at points

Level 4	
Content	Delivery
Dependence on the prompt	Ineffective repetition of words/phrases
Ineffective/abrupt transitions	Pauses/hesitations are more frequent
Omission of function words	Flat intonation
Systematic problems with bound morphology	Many identifiable articulation/pronunciation/
Topic shifts	stress problems
Lack of coherence	Pace interferes with comprehension
Weak organization	Close listener attention required
Repetition interferes with coherence	
Intended meaning unclear	
Lack of elaboration	

Level 3	
Content	Delivery
Misuse of particular words	Deliberate/ ineffective delivery
Problems with bound morphology	Frequent pauses/hesitations within phrasal
Frequent attempts to re-start/re-phrase without	boundaries
clarification	Ineffective attempts of interpretative state-
Unintended meaning	ments
Misunderstands prompt	Limitation of vocabulary