

The use of pragmatic markers across proficiency levels in second language speech

Colleen Neary-Sundquist

Purdue University, West Lafayette, IN, USA

cnearysu@purdue.edu

Abstract

This study investigates the use of pragmatic markers (PMs) by learners of English at varying proficiency levels. The study analyzes data from a university-level oral proficiency exam that categorized Chinese and Korean English-as-a-second-language (ESL) speakers into four proficiency levels and compares data with those of native speakers taking the same test. Findings indicate that PM use generally rises with proficiency level. The rates of PM use showed a dramatic increase between the highest and second-highest proficiency group. The highest proficiency ESL group used PMs at the same rate as native speakers. The study also found that the variety of different PMs used goes up steadily with proficiency level. These results are discussed in terms of their implications for understanding how second language learners' use of PMs develops.

Keywords: pragmatic marker, proficiency level, pragmatic competence, second language development, oral speech production

1. Introduction

Pragmatic markers (expressions such as *so*, *well*, and *you know*) have been shown to perform a number of important discourse and pragmatic functions in second language speech. In particular, as a subcategory of formulaic language in second language development, pragmatic markers facilitate the development of second

language fluency (Hasselgreen, 2004; Towell, Hawkins, & Bazergui 1996). They also allow for the communication of speakers' attitudes, how they intend their utterance to be interpreted, and the establishment of intersubjectivity with their interlocutor (Aijmer, 2013; Blakemore, 2008; Overstreet & Yule, 1997). As Halliday and Hasan (1976) and Schiffrin (1987) point out in their analyses of L1 pragmatics, pragmatic markers (PMs) are also one of the principal ways that coherent discourse is constructed (Halliday & Hasan, 1976; Schiffrin, 1987).¹

Despite the wide variety of functions and their importance in the development of second language fluency, they have received relatively little attention in second language acquisition research (Müller, 2005).² Previous research has examined the use of PMs in learner speech, but a number of important questions remain. In particular, little is known about the overall pattern of PM use across a wide spectrum of language learners. Although some L2 studies of PMs reveal important findings about individual markers (e.g., House, 2013; Müller, 2005; Romero-Trillo, 2002), few studies focus on the wider range of use of PMs, including those that may be used less frequently by some learners. Moreover, as Hasselgreen (2004) and Hellermann and Vergun (2007) note, the proficiency level of learners plays a role in the frequency and range of these types of expressions, although few studies break down proficiency levels into sub-levels while examining the larger set of PMs. Furthermore, few studies focus on the similarities and differences between highly advanced learners and native speakers with respect to the frequency and variety of PMs. Such information may tell us much about the interaction of proficiency level and pragmatic competence and whether learners progress more gradually or rapidly in the advanced stages of second language learning.

This study attempts to shed new light on these issues, examining PM use among English as a second language (ESL) learners at multiple proficiency levels compared to native speakers performing the same tasks. The rate of use and the range, or richness, of PM use will be explored from a quantitative perspective. The paper fills several gaps in the literature because it will examine the larger set of PMs rather than a small subset, and because it focuses on both the range and frequency of unique PMs across several sub-levels of proficiency. The results of statistical analysis reveal that the overall frequency of PMs increases at varying rates at each proficiency level and reaches a rate almost identical to

¹ Note that these authors use the term *discourse markers* rather than *pragmatic markers*, the term used throughout this paper. Further discussion of definitions used in this study is included in the background section on terminology.

² For a helpful overview of previous research on the development of L2 pragmatic competence in general, see Bardovi-Harlig (2013) or Ishida (2013), in which the study of pragmatic markers is mentioned in the larger context of L2 pragmatics research.

that of native speakers, although the range remains relatively limited at the most advanced stage of proficiency. This study will explore possible reasons for this pattern of use and point to future studies that may reveal more about the development of L2 pragmatic competence.

2. Review of the literature

2.1. Pragmatic markers: Terminology and definitions

Although PMs have received a great deal of attention in cross-linguistic studies over the last two decades, there is little consensus regarding the definition or appropriate terminology for this type of expression (Blakemore 2002, 2008; Romero-Trillo, 2012; Schourup 1999).³ In an early study investigating *discourse markers* in English, Schiffrin (1987) uses this term to refer to “sequentially dependent elements that bracket units of talk” (p. 31). Fraser (1999) also uses the term *discourse markers* but defines them differently as “linguistically encoded clues which signal the speaker’s potential communicative intention” (p. 168). Under the name *discourse particles*, they have been identified as particles that “are placed with great precision at different places in the discourse and give important clues to how discourse is segmented and processed” (Aijmer, 2002, p. 1). Alternatively, others such as Hasselgreen (2004) define these expressions as *small-words*, which she points out are phrases “occurring with high frequency in the spoken language, that help to keep our speech flowing, yet do not contribute essentially to the message itself” (p. 162). These various terms and definitions are offered here to give some idea of the variety of approaches to these expressions.

One of the most problematic aspects of classification deals with the difference between what is referred to as *pragmatic markers* and what others might label *discourse markers*. This terminological confusion arises, according to Romero-Trillo (2012), because some scholars make clear distinctions between the two terms while others use the term *pragmatic markers* to refer to a superordinate category under which discourse markers may be subsumed. In the present study, this more general term *pragmatic marker* will be used in order to avoid confusion with specific terms, including *discourse markers*, which refer to various subclasses of these expressions. In particular, the current study follows the definition and list used by Carter and McCarthy (2006), in which the

³ For further discussion of definitions and terminology of pragmatic markers, see Lenk (1997), Schourup (1999), Callies (2009), Romero-Trillo (2012) and Aimjer (2013). In-depth studies of particular pragmatic markers have been undertaken in languages as diverse as Spanish (Chodorowska-Pilch, 2008; Durán & Unamuno, 2001), Chinese (Feng 2008; Wang & Tsai, 2007), Japanese (Sasamoto, 2008), and English (Bell, 2010; Lenk, 1998; Schourup, 1999).

term *pragmatic marker* is used. This definition and the reasons for its use will be discussed in the methodology section of this paper.

Despite the lack of consensus concerning terminological conventions in previous research, according to Schourup (1999), there is general agreement on several fundamental characteristics of PMs. In his comprehensive overview, Schourup highlights three characteristics that most definitions of PMs share: connectivity, optionality, and non-truth-conditionality (pp. 1230-1232). Connectivity is the idea that PMs connect blocks of language to other blocks of language and that they signal the relationship between them. Optionality refers to the fact that these expressions are syntactically detachable from the utterance; they can be removed without disturbing the grammaticality of the utterance. Finally, PMs are said to not to affect the truth or falsity of an utterance. Therefore, although they may contribute somehow to the interpretation of an utterance, they are not an essential part of its meaning.

Carter and McCarthy (2006) provide a definition that takes these characteristics into consideration yet also places weight on the pragmatic contribution of these expressions. They discuss the connectivity and optionality of PMs, but they also highlight the importance of interpersonal meaning and the speaker's intentions in the definition. As part of their grammar of English based on a corpus of oral production data, Carter and McCarthy (2006, p. 208) define PMs as "a class of items which operate outside of the structural limits of the clause and which encode speakers' intentions and interpersonal meanings." The authors subdivide the class of PMs into discourse markers (including such expressions as *so*, *well*, and *anyway*), stance markers (*actually*, *I think*, *of course*), hedges (*kind of*, *maybe*), and interjections (*gosh*, *wow*). PMs are classified as grammatical options with social, contextual, and affective functions in spoken English.

From this definition, the importance of PMs for interpersonal communication is evident. These expressions allow speakers to communicate their intentions, indicate their attitudes towards information that is presented or received, and link their ideas for greater clarity. In other words, as Carter and McCarthy (2006) point out, PMs are a broad class of items that can provide structure and organization to utterances while indicating attitude, assertiveness, or reactions to discourse (p. 105).

In the next section, previous research relating to PMs in second language learning is discussed. The reader should keep in mind that the issues of the definition and identification make the comparison of different studies of PMs tentative. The literature review below outlines the findings of studies into the rate and range of use of PMs in second language research; however, the studies mentioned do not define or identify the markers in exactly the same way.

2.2. Pragmatic markers in second language studies

In most second language research on the rate of PM use, studies indicate that learners do not use PMs in their speech nearly as much as native speakers. Fung and Carter (2007) compared a corpus of elicited classroom data from intermediate-advanced learners of English in Hong Kong with the British English CANCODE corpus. They found that non-native speakers largely underuse PMs in comparison to native speakers. Over half of the PMs they examined were used less frequently by non-native speakers. However, they acknowledge that their results must be considered preliminary due to the fact that the conditions under which the data were obtained were significantly different for native and non-native speakers. Romero-Trillo (2002) found similar results in his study of the use of a set of PMs by native and non-native speakers of English. Analyzing corpora of spoken data from Spanish native and non-native learners of English, he determined that native speaker adults used markers such as *you know*, *I mean*, or *you see* at a significantly higher rate than non-native speaker adults.⁴ These findings are also supported by Müller's (2005) investigation of the use of *so*, *like*, *well*, and *you know* by native speakers and German learners of English. Analyzing data from a paired silent film retelling, she found overall significant underuse of these markers by learners (except for *well*).

Although it is clear from previous research that the total frequency of PMs by non-native speakers of English is lower than that of native speakers, much recent research focuses only on a small set of expressions. Fuller (2003), Hellerman and Vergun (2007), Müller (2005), and Romero-Trillo (2002) looked at three, four, five, and six individual PMs, respectively. Such detailed analyses provide insight into the use of specific PMs in particular contexts and allow us to explore how their use by ESL speakers differs from that of native speakers. However, in order to gain an overall picture of the frequency, distribution, and variety of PMs in learner speech, it is necessary to widen the search and investigate a greater number of expressions. Thus, this study attempts to fill a gap in previous research by investigating a larger variety of expressions from a comprehensive list of PMs.⁵

Most studies of PMs focused on a single proficiency level or fail to address the role of proficiency at all. Fung and Carter (2007) focus solely on intermediate-

⁴ In the case of *look* and *listen*, Romero-Trillo found no significant difference in the frequency of use by native versus non-native speakers.

⁵ Note that Fung and Carter (2003) and Hasselgreen (2004) looked at a larger number of PMs than most studies. Fung and Carter identified 23 pragmatic markers among the 100 most common English words and Hasselgreen looked at a total of 19 smallwords (her term for pragmatic markers). The current analysis expands on these larger-scale studies by focusing on even more expressions, including the 121 found in Carter and McCarthy's (2006) corpus-based grammar.

advanced learners and also provide no comparative data from native speakers and nonnative speakers performing the same task. Müller (2005) examined four specific markers in detail but did not consider overall rates of discourse marker use, nor did she examine different proficiency levels. Romero-Trillo (2002) analyzed data from both children (1st graders) and adult (university students) native and non-native speakers but did not split up the non-native speakers according to proficiency.

One of the few studies that addressed the issue of proficiency level in relation to a large set of PMs is Hasselgreen (2004). She focused on two groups of Norwegian learners of English (Norwegian A, described as "more fluent," and Norwegian B, or "less fluent") and determined that the frequency with which native speakers used smallwords was significantly higher than that of both non-native speaker groups. These groups were selected entirely on the basis of global grades on a speaking test (p. 160). Native speakers used 445 smallwords per 10,000 words while Norwegian A and B groups used them at a rate of 279 and 235 per 10,000 words, respectively. Hasselgreen noted that these differences were highly significant, with $p < .0001$ for native speakers versus Norwegian A and $p < .05$ for Norwegian A versus B (p. 171).

Hellermann and Vergun (2007) also examined the interaction of proficiency level with PM use, although they limited their analysis to a small set of expressions. They investigated video recordings of classroom interaction and interviews of 17 adult learners of English with no previous formal English language instruction. Focusing on the PMs *well*, *you know*, and *like*, they analyzed the interaction of PM use and proficiency level. Proficiency was classified according to the four levels of English language courses available to the learners: from absolute beginner (Level A) to advanced (Level D). Hellermann and Vergun's (2007) findings indicated that the use of PMs for *well*, *you know*, and *like* went up with each proficiency level, although they found very few examples, if any, at Level A and Level B (p. 167).

Data on the interaction of proficiency level with the rate of PM use are potentially revealing if we consider the possibility that there might be a developmental path in the acquisition of PMs. Both Scarcella (1983) and Hays (1992) speculated that learners acquire and use only certain PMs at lower levels of proficiency before acquiring others to use in more varied contexts at higher proficiencies. This suggestion is supported by Hasselgreen's (2004) study: She proposed three hypothetical stages of development in the use of PMs that are exemplified by each group of speakers in her study (lower proficiency learners, higher proficiency learners, and native speakers). Certain smallwords are underused by the lowest-level speakers while other expressions are only part of the repertoire of higher-proficiency speakers or native speakers (p. 222).

Related to the role of proficiency and the possibility of a developmental path of acquisition in the use of PMs is the issue of the extent to which learners use a variety of PMs. When we focus on just a small set, the diversity is not clear. Richness could be related to proficiency level and development. As discussed above, previous research has found that learners generally use PMs at a lower rate than native speakers. However, as we will see below, PMs are also often overused by non-native speakers.

Müller (2005) found that the markers *so*, *like*, and *you know* were underused by learners in comparison with native speakers. However, the marker *well* was not underused by non-native speakers. A full discussion of Müller's results is beyond the scope of this study since she also considered the effects of gender, age, relation between partners, role in the discourse, and time abroad on PM use. Relevant to this study is the fact that Müller's analysis found differences in the rates of use of particular functions of PMs. For example, after identifying 10 different functions for *you know* (five at the textual level, and five at the interactional level), Müller found that native speakers of American English used *you know* significantly more often than L1 German speakers on all of the textual functions and three out of the five of the interactional functions.

Hasselgreen (2004) also found a complex pattern of learners using different markers than native speakers to perform various communicative functions. In the category of hedges, for example, which "indicate the degree of vagueness or commitment" (p. 204), learners underused *just*, *sort of/ kind of*, *like*, and *a bit*. The "general extender" (Overstreet & Yule, 1997) type of hedge, exemplified by *and stuff/and everything*, was virtually absent from both Norwegian L1 groups. Within the same category of hedges, the learners used *I think* and *or something* at higher rates than native speakers. What is more, the higher-proficiency Norwegian group used more hedges than the lower-proficiency group, but not a greater variety. Romero-Trillo (1997) found that the markers *I mean*, *well*, *you know*, *you see* were underused but that *listen* was overused. He attributes this to transfer of Spanish *oye* 'listen,' which is used frequently in Spanish.

The current study aims to address several gaps in previous research by examining both the general rate of use as well as the range of use of PMs using data from learners at varying proficiency levels in comparison with native speakers. Previous studies have generally done one or more of these things, but not all at the same time. Hasselgreen (2004) is the only previous study that considered the proficiency level of the learners and had comparable native speaker data on a larger set of PMs. However, the current study differs from Hasselgreen's in three ways that will contribute to the understanding of the development of learner PM use. The current study has a higher number of finely-

grained proficiency levels (5 vs. Hasselgreen's 2), uses monologic data rather than dialogic, and investigates a much larger set of PMs (121 vs. 19).

It is important to consider both proficiency level and native speaker comparison data when trying to build an understanding of aspects of learner speech. Foster and Tavakoli (2009) called for the inclusion of native speaker comparison data in more research studies. They argued that the inclusion of data from native speakers performing the same tasks under the same conditions is important in order to distinguish more clearly task and processing effects from other effects. Likewise, data from learners at multiple proficiency levels allow for the investigation of how learner interlanguage develops.

3. Research Questions

This study was designed to investigate the acquisition of PMs by learners at various proficiency levels both in terms of the rate of their use as well as the type of expressions used. It therefore attempted to answer the following three research questions concerning the rate, frequency, and variety of PMs used by learners:

1. What is the rate of PM use by ESL speakers at different proficiency levels and how does this compare to that of native speakers performing the same speech tasks?
2. Do ESL speakers and native speakers use the same PMs most frequently, even if they are generally alike or different in their rate of PM use?
3. Do ESL speakers use a lesser variety of PMs than native speakers?

After the methodology section below, the results of the quantitative analysis will be presented in the order corresponding to the order of these three research questions.

4. Methodology

4.1. The oral proficiency test

The data for this study came from a computer-mediated, semi-direct test of oral proficiency (hereafter OPT) that is administered to prospective international teaching assistants.⁶ The exam consists of ten items that require examinees to record their responses by speaking into a computer microphone. After receiving

⁶ For more information on the methodology, participants, data coding, and testing procedure, see Neary-Sundquist (2013), where these issues are discussed in the context of a similar study.

a prompt for each item, examinees have 3 minutes to take notes and plan their response.⁷ Each response is limited to 2 minutes, after which time the recording is automatically stopped.⁸

For this study, four of the ten tasks were selected for transcription and coding, including the “news,” “personal,” “passing information,” and “telephone” tasks.⁹ In the news task, an opinion about a news item must be given. 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, examinees listen to a recorded message from one of their office mates and leave a short message on his/her home answering machine.

The particular questions used in the different tasks in the OPT were not identical; rather, they came from several different forms of the test that are given regularly. The internal consistency coefficients on the different versions of the test are very high, ranging from .96 to .98, and the single factor solutions were found in a study of the factor structure (Ginther, Dimova, & Yang 2010).

The exams are rated by two trained raters both of whom are also instructors in the university’s English proficiency program. When the two raters disagree, the exam is sent to a third rater to break the tie. The ten tasks in the exam are given a single holistic score that ranges from 2 to 6. A set of descriptors for each level can be found in Appendix A.

⁷ The exam item prompts include written, video, and audio input; the exact text of the prompts is secured and consequently not available to the public.

⁸ The monologic nature of the data used in this study could be considered a limitation in the analysis of PMs, due to their importance in interaction between the speaker and interlocutor. However, the testing circumstances under which the data were collected also offered a number of advantages. First of all, there was no variability introduced by the presence of an interlocutor. That is, the examinees were not more or less likely to employ a PM based on their judgment of the reactions of a conversation partner. Secondly, the testing environment allowed for the collection of larger amounts of data than have often been used in previous studies of PMs. Thirdly, since the data examined in this study came from a language test, they had already been subdivided into proficiency levels whose reliability had been independently established. This makes it possible to investigate the use of PMs by varying proficiency levels, an aspect of learner pragmatics that has not been previously examined in the literature.

⁹ These four tasks were chosen out of the ten available in order to provide tasks with a range of task types and levels of structure. The passing information and telephone tasks were considered more structured since they entail repeating specific information. The news and personal tasks were considered less structured since they are open-ended and the speaker has more leeway in structuring their response.

4.2. Participants

Data consisting of 184 speech samples were analyzed. The ESL speaker examinees were all graduate students and prospective teaching assistants at an American university. The ESL speakers were grouped into four different proficiency levels, numbered 3 to 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 ESL speaker examinees came from either a Chinese or Korean language background.¹⁰ The native speaker group consisted of speakers of American English who were also graduate university students. A table showing the language background and scores of the participants can be found in Appendix B.

Level 6 has only seven examinees because that number was the total available in the testing records. An additional limitation was that there were only six responses to the personal item for the native speaker group; the earlier version of the exam lacked this question.

The data for each level was evenly split between examinees with a Chinese and Korean L1 background; that is, there were 5 examinees with L1 Chinese and 5 examinees with L1 Korean in each group of 10. The group of 7 examinees at Level 6 was made up of 3 native Chinese speakers and 4 native Korean speakers.

4.3. Data preparation and coding

The oral exam responses were transcribed by the author using Carter and McCarthy's (2006) list of PMs. Their definition is appropriate for this study since it 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 non-native discourse in context. Furthermore, this definition of PMs is particularly useful for methodological reasons: Carter and McCarthy's classification is accompanied by a list of 121 expressions. Unlike other studies that more narrowly define PMs and focus exclusively on a small group of PMs, Carter and McCarthy's definition is more comprehensive and their list of appropriate expressions more extensive. Their list includes sub-categories of PMs, including "discourse markers" like *you know*, *well*, or *so*, "stance markers"

¹⁰ 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 participants at every proficiency level. The second consideration was that these groups learned English as a foreign language but their schooling does 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.

such as *actually*, *of course*, and *hopefully*, "hedges," for example, *I think*, *just*, and *kind of*, as well as "interjections" like *gosh*, *wow*, or *ouch*.¹¹

Two coders, the author and a colleague, analyzed the transcripts word-by-word for PMs. This methodology allowed for the identification of PM versus non-PM use of particular words. For example, *so* has multiple functions in English, some of which are not PMs (as an adverb or conjunction indicating purpose), and some which are PMs (with functions such as introducing a summary, shifting the discourse, or even holding the floor; see Buysse, 2009 and Buysse, 2012 for a thorough discussion of the various uses of *so*). Similarly, *also* can function as an adverb or as a PM that can signal a number of different discourse-related messages, including an additive function that is used when the speaker wants to legitimize the relevance of their utterance (see Waring, 2003 for further discussion on the functions of *also* in discourse). As an illustration of the multiplicity of uses, the passages below contain examples of *so* and *also* that were not counted as PMs, and further examples that were counted as PMs:

. . . she's very worried so she's wondering whether ah she could set up a special time ah with you to ah get some help on homework ah before the next exam. (Korean Level 6 speaker)

I have to take lessons, in English, and . . . many lessons in physics are also in English as well. (Chinese Level 3 speaker)

. . . So, I'd say that it's, it's a good idea to involve university students, and give them interaction with schoolchildren. (Chinese Level 6 speaker)

She asked to arrange a special time for her. If you can, she really appreciate you. Also, she worried about her homework because she did not do it well. (Korean Level 4 speaker)

The first examples of *so* and *also*, underlined above, were not counted as PMs. The second set of examples of *so* (*so I'd say*) and *also* were counted as PMs. This use of *so* occurred at the end of the speaking passage, after several reasons for the speaker's opinion were enumerated. It marked a summative statement. Prosody also sometimes gave clues to the identification of this type of *so* (see below).

A further advantage of the hand-coding methodology was that it was possible to allow for some variation in the list of expressions. For example, Carter and McCarthy (2006) list *to sum up* and *in the end* as PMs, but they do not include *in sum* or *all in all*. Since the coders were reading through all of the data rather than electronically searching for expressions, it was possible to identify

¹¹ For a comprehensive list of all PMs and typical examples, see Carter and McCarthy (2006).

such variations in wording. Pragmatic marker appears to be a highly idiosyncratic phenomenon in some cases, and this research methodology takes such individual variation into account.

Interrater reliability for the coding of PMs was .94. In the case of a discrepancy in the identification of the PMs, the two coders discussed the example in question and came to an agreement. Coding posed several challenges at times; as Aijmer (2013) pointed out, the meaning of PMs is inherently “flexible and fuzzy” (p. 15). However, identifying PMs was aided by the use of a well-defined list (Carter & McCarthy, 2006). In cases where the raters still had doubts about whether a particular expression was being used as a PM or not, listening to the actual recordings, rather than relying on the transcripts, was often helpful. Fung and Carter (2007, p. 413) noted several prosodic “clues” that distinguish PMs: “pauses, phonological reductions, and separate tone units.” Paying attention to such prosodic clues sometimes helped the coders to identify PM uses of some words and expressions. As Aijmer (2013) noted, however, the relationship between the formal properties of PMs, such as prosody, and their function and meaning is complex; prosodic features cannot be relied upon exclusively to identify PMs.

After the PMs were identified and counted, the total number of words used by each examinee was calculated. Non-lexical filled pauses such as *um*, *ah*, or *uh* were removed from the transcripts and not included in the total word count. The total number of words and the number of PMs used by each subject on each task were then entered into spreadsheets. The number of PMs was divided by the total number of words to calculate the percentage of PMs that was used by each subject. The resulting percentages were then averaged to give the average percentage of PM use for all of the responses at each level and for each task.

4.4. Statistical analysis

Descriptive statistics, including means, standard deviations, and 95% confidence intervals were calculated for the rates of PM use using Statistical Analysis Software, version 9.1.3. A Spearman rank order correlation was used to calculate the relationship between PM use and proficiency group since the OPT scale is ordinal. A Kruskal-Wallis test with Mann-Whitney post-hoc tests was performed to test the results for statistical significance.

5. Results

The quantitative results are presented in the following sub-sections according to the three main issues generated by the research questions. The data are followed

by a brief description of the results. The relationship between the results and the research questions will be addressed in the discussion section of the paper.

5.1. Rate of pragmatic marker use

The rate of PM usage was calculated by dividing the total number of words used by a speaker by the total number of PMs they used. Table 1 presents the descriptive statistics and 95% confidence intervals (CIs) for the rates of PM use for all five proficiency groups, along with the means and standard deviations of the number of words and PMs, and the mean rate of PMs for each level.

Table 1 Descriptive statistics for the rate of PM use

Level	<i>N</i>	Mean no. of words (SD)	Mean no. of PMs (SD)	Mean PM rate (SD)	Min	Max	Lower 95% CI	Upper 95% CI
3	10	421 (79)	9 (5)	0.021 (0.01)	0.003	0.035	0.01	0.03
4	10	525 (144)	13 (10)	0.025 (0.019)	0.006	0.066	0.01	0.04
5	10	579 (122)	18 (10)	0.031 (0.016)	0.012	0.064	0.02	0.04
6	7	743 (116)	40 (21)	0.052 (0.02)	0.037	0.094	0.04	0.07
NS	10	584 (215)	31 (13)	0.052 (0.014)	0.025	0.067	0.04	0.06

Note. NS = native speakers

The results show that PM use rose with proficiency level; however, the pattern was not entirely straightforward. Levels 3 and 4 were similar, at 2.1% and 2.5%. The sharpest gain in PM use occurred at levels 5 and 6, with a jump of 1% and then a further jump of over 2%. But PM usage did not continue to rise for the native speakers; instead, their usage was identical to that of the Level 6 group. The pattern showed that the two highest and the two lowest proficiency groups were very alike in their PM use but that the middle level showed a steeper rise in PM use.

Correlation tests and tests of statistical significance indicated that there was a positive correlation between PM use and proficiency level and that there were significant differences between levels. The Spearman correlation for the rate of PM use and proficiency level was .66. The results of a Kruskal-Wallis test were significant ($H = 22, 4 df, p = <.0001$). Mann-Whitney post-hoc pairwise comparisons showed that Level 3 was significantly different from Levels 5, 6, and native speakers, and that Levels 4 and 5 were each significantly different from Levels 6 and native speakers.

The results of the analysis of PM usage rate can be summarized as follows: The use of PMs correlated positively with proficiency level. Learners generally used PMs at a lower rate than native speakers, but advanced learners' rate of use did reach native-speaker levels. The use of PMs increased gradually from Levels 3 to 5 and rose rather abruptly between Levels 5 and 6.

5.2. Frequency of particular pragmatic markers

The second research question is: Which particular PMs were used most frequently both by learners and by native speakers? Table 2 shows the three most frequently used expressions for each level. Results indicate that the three most frequently used expressions were the same for the Level 3, 4, and 5 proficiency groups. *I think*, *so*, and *also* were the top three PMs in these three groups. The Level 6 group is noticeably different in this regard, with just one of these expressions in the top 3: *so*. The native speakers had two of the same expressions as the first three proficiency groups: *I think* and *so*. In terms of differences in the groups, Level 6 and the native speaker group stood out for their frequent use of *just*, while the Level 6 results were notable for the inclusion of *you know*.

Table 2 The three most frequently used PMs for each proficiency level

Level 3	Level 4	Level 5	Level 6	Native speakers
<i>I think</i>	<i>I think</i>	<i>I think</i>	<i>just</i>	<i>I think</i>
<i>so</i>	<i>also</i>	<i>so</i>	<i>so</i>	<i>so</i>
<i>also</i>	<i>so</i>	<i>also</i>	<i>you know</i>	<i>just</i>

Table 3 provides an overview of these five most common PMs in the corpus at each of the five proficiency levels. Both the raw figures for the number of PMs used and the percentage each PM represents are given. The percentages for each particular PM were obtained by dividing the number of occurrences of that marker by the total number of PMs. The data indicate that the higher proficiency levels, including Levels 5 and 6, along with the native speakers, had a more evenly distributed use of different PMs. Speakers at lower levels (Levels 3 and 4) used certain PMs, including *I think* or *so* or *also* for Level 4, at a much higher rate than other more common phrases like *just* or *you know*. Thus, the higher levels of proficiency showed a flatter distribution of the most frequently used PMs, while the two lower levels exhibited some spikes in the frequency of certain PMs.

Table 3 The frequency of particular PMs by proficiency level

PMs	Level 3	Level 4	Level 5	Level 6	NS
Total PMs (% of PMs)	92 (100)	134 (100)	181 (100)	281 (100)	310 (100)
<i>I think</i>	30 (32.6)	31 (23.1)	32 (17.7)	25 (8.9)	44 (14.2)
<i>so</i>	16 (17.4)	10 (7.5)	26 (14.4)	31 (11)	28 (9)
<i>also</i>	12 (13)	25 (18.7)	21 (11.6)	19 (6.8)	15 (4.8)
<i>just</i>	3 (3.3)	5 (3.7)	7 (3.9)	31 (11)	23 (7.4)
<i>you know</i>	1 (1.1)	5 (3.7)	14 (7.7)	29 (10.3)	10 (3.2)

Note. NS = native speakers

5.3. The variety of pragmatic marker use

In the previous section, results indicated that learners in the lower proficiency groups were overusing certain expressions such as *I think* in relation to the higher-proficiency speakers. If this is the case, then this lack of variety in their expressions should be apparent in more than just one expression. Figure 1 shows the number of different PMs used by learners in each proficiency group. The number of different expressions used at each level went up steadily from 25 PMs at Level 3 to 69 PMs among the native speakers. This shows that as learners acquire PMs, they also make use of a greater variety of them. However, it is possible that, although they were using a greater variety of expressions, these new expressions did not represent a large proportion of their overall PM use. In order to examine if this was the case, it is necessary to look at how much of the learners' PM use is accounted for by the most frequently used expressions at each level.

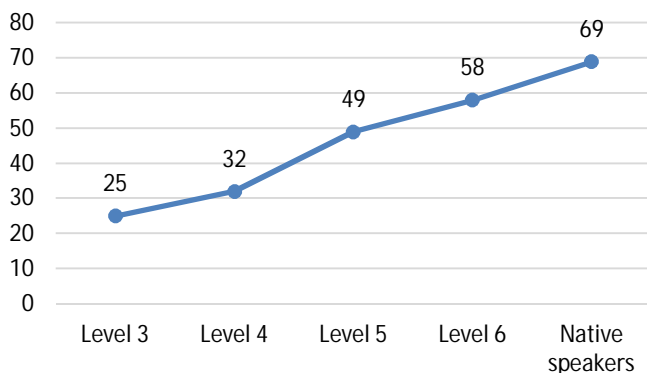


Figure 1 The number of unique PMs used by level

Figure 2 compares the use of the three most frequent PMs for each proficiency group with the remainder of the PM each group used. The data here show that the percentage of PMs accounted for by the most used expressions fell consistently with proficiency level. That is, PM use for the lower proficiency groups was composed of repeated use of a few expressions and a more limited use of a larger variety of expressions. The three most frequent expressions accounted for a majority of PM use for the Level 3 proficiency group and almost half for the Level 4 proficiency group.

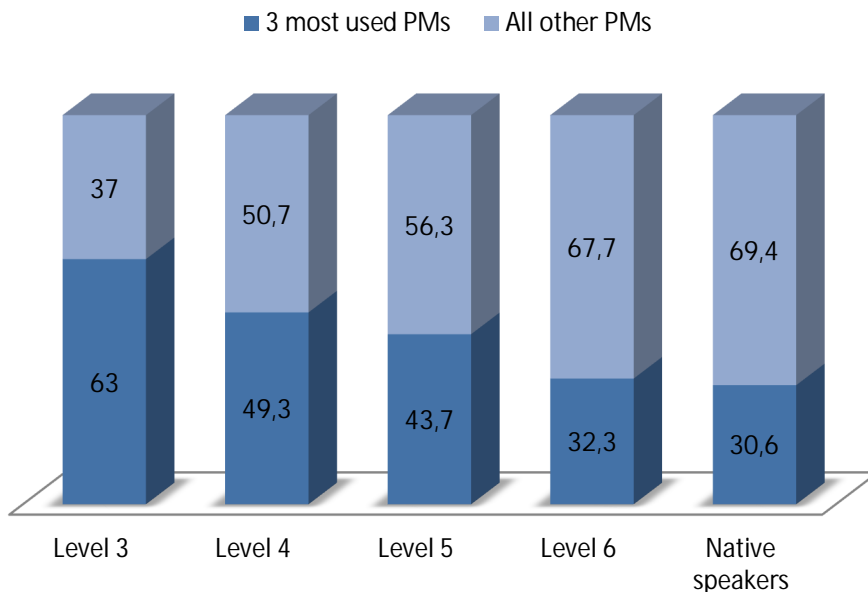


Figure 2 The percentage of PMs accounted for by the three most frequent PMs

6. Discussion

6.1. Frequency of pragmatic markers across proficiency levels

According to the results presented above, the data indicate that PM use rises with proficiency level and correlates positively with it. At first glance, the fact that this is the case is not surprising. It is expected that PM use would rise with proficiency level, based on previous research on PMs among second language learners (Hasselgreen 2004; Hellerman & Vergun, 2007). The use of PMs shows that the speakers are able to provide additional information about their message, in addition to its basic meaning. This may indicate 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 PM use at higher proficiency levels. Thus, in terms of the Research question 1 of this study, the data indicate a general rise in the frequency of PMs with increased proficiency.

What is surprising, however, is the abrupt rise in PM use between Levels 5 and 6. This seems to indicate that PM use makes sudden progress after a certain critical mass of general proficiency is reached. During the transition from Level 5 to Level 6, the learners' use of PMs jumps dramatically to native-like levels. This seems to indicate that this area of pragmatic development occurs fairly late for

learners. This could possibly be due to the fact, often mentioned in definitions of PMs, that these expressions are part of the message but not of the essential propositional meaning. Therefore, learners are only able to add these types of nuances to what they are saying when they do not have to devote as much attention to formulating the content of what they are saying. Looked at from the perspective of PMs and fluency, this could also indicate that the use of PMs is actually supporting or even enabling the speakers to become more fluent. However, this suggestion must be considered with caution in light of the fact that this study does not directly examine the relationship between PM use and fluency measured with temporal variables. The relationship between PMs and fluency is no doubt complex and provides rich opportunities for further study.

Although a sharp increase in the rate of PM use is seen between Levels 5 and 6, the use of PMs was still notable at Levels 3 and 4. Even at the fairly low proficiency level (Level 3), the examinees used 2.1% PMs. This is remarkable when we consider that PMs are not regularly taught in the classroom. It is likely that this pedagogical oversight is due to the fact that PMs are considered secondary, extraneous, and optional compared with the necessity of mastering categories such as verbs, nouns, and prepositions. It may also arise from the fact that PMs are considered too difficult to teach. With few formal instructional materials, teachers are left to rely on their own intuition, which they soon find inadequate to explain the intricacies of PM use.

6.2. The most frequent pragmatic markers at each proficiency level

In terms of Research question 2, the results indicate that the first three proficiency groups were remarkably similar with respect to the PMs they used most frequently. At least at this level of analysis, the results are somewhat surprising in how similar they are across groups; we might have expected at least some variation in the most frequent expressions. However, the pattern that emerges instead is one of close similarity among the first three levels and then a drastic difference at Level 6, with only one of the same expressions to be found among the three most frequent PMs. Interestingly, this pattern mirrors the pattern seen above for the frequency of PMs in general. In both cases, the first three proficiency groups cluster together, followed by a jump in use for Level 6 and the native speaker group. These results therefore also support the intriguing pattern that some significant change in the acquisition of PMs occurs between the Level 5 and Level 6 proficiency groups.

It is important, however, to note that the native speaker group has two expressions in common with the three lower proficiency groups; this argues against the significance of the fact that the Level 6 group differs markedly with

respect to the proficiency groups below it. Note that these expressions do not represent the most frequent expressions for all of the groups. Rather, they were included based on the fact that they appeared in the top three most frequent expressions for any one group.

As we can see in Table 3 above, the three highest proficiency groups, Level 5, Level 6, and native speakers, are quite similar in their rates of use of these five expressions. The Level 3 proficiency group seems to follow the same general trend as these three higher proficiency groups, although the rates of use are quite different for *I think*. The Level 4 proficiency group is the outlier here, with noticeably higher rates of use for *I think* and *also*.

What the two lower proficiency levels have in common is that they both use *I think* to a much greater extent than the three higher proficiency groups. There are several possibilities that would account for greater use of *I think*. It is possible that they are using this expression as a filler, much in the same way a speaker might repeat *umm*. Other possibilities are that they could be marking greater uncertainty about what they are saying, or they may be overusing this PM to compensate for the fact that they do not have a command of a greater variety of expressions.

Recall that Hasselgreen (2004, p. 222) also discussed the possibility of an order of acquisition for PMs based on which PMs the lower and higher proficiency groups in her study seemed to have mastered. Her results are not directly comparable to those in this study due to the differences in the types of tasks from which her data were drawn. However, she too observed two main stages of development that resemble those found in this study: Learners at lower levels of proficiency tend to overuse a small number of expressions, while those at higher levels expand to a greater variety that resembles the pattern of use among native speakers. The possibility of such changes in PM use would have important implications both for the study and teaching of PMs and requires further research. If it is true that there is an order of acquisition for PMs, this would not be the first pragmatic phenomenon that has shown evidence of such an order. Dittmar and Terborg (1991), for example, found an order of acquisition for expressions that encode modality. Similarly, Kasper and Rose (2002) suggested that the act of making requests proceeded through five stages for L2 learners.

As pointed out in the methodology section, this study relies on monologic data, which has some advantages in terms of research design, allows for the analysis of a large amount of data for cross-sectional comparison, and makes it possible to consider PM use by learners at independently established proficiency levels. However, another effect of the monologic data may be that certain markers, such as *you know*, are likely to be less frequent since there is no interlocutor present. However, it is interesting to note that many of the PMs

that signal the speaker's attitude or offer indications about how their utterances should be interpreted are still well represented in the results. For example, hedges such as *I think* were still quite frequent among learners at all levels. This suggests that even in a monologic environment, the speakers are still giving cues to an imagined interlocutor, at least in the circumstances under which the data were collected. Even learners at a low proficiency level feel the requirement to offer this type of pragmatic information, which may indicate some interesting avenues for future research into learner pragmatics.

6.3. Variety of pragmatic markers at each proficiency level

In terms of Research question 3 and the variety of PM use, ESL speakers use fewer unique expressions than native speakers. In contrast to the rate of use of PMs in general, which rose sharply between Levels 5 and 6, the number of unique expressions used rose more gradually from 25 expressions at Level 2 to 58 at Level 6. The largest gain in the number of expressions used occurred between Levels 4 and 5, where 17 new expressions were gained. Although the Level 6 group used PMs at the same rate as native speakers, they were still well behind them in the variety of expressions.

The data on the variety of PMs further suggests the interesting possibility that learners first attain command of a critical mass of expressions before they begin to use them at more target-like levels. This increase in the number of unique expressions used thus occurs earlier than the increase in the rate of PM use, which, we saw, occurred between Levels 5 and 6. The fact that an increase in the variety of expressions used precedes an increase in the rate of use is logical and might indicate that a critical mass of PMs must be mastered before ESL speakers might begin to use them at native-like levels.

One interesting possibility concerning the relationship between the rate and the range of PM use is suggested by research on second language vocabulary acquisition. Various studies of vocabulary acquisition have found a consistent relationship between general proficiency level and measures of lexical diversity (Daller, Van Hout, & Treffers-Daller, 2003; Malvern & Richards, 2002; O'Loughlin, 1995; Zareva, Schwanenflugl, & Nikolova, 2005). Lexical diversity has proven to have a more consistent positive correlation with proficiency level than other measures of vocabulary acquisition.

Lu (2012) found that a number of different lexical variation measures had significant correlations with the quality of oral English narratives produced by Chinese learners on proficiency test rankings. It was lexical variation, rather than lexical sophistication or density, that exhibited this correlation. Yu (2010)

also found a statistically significant positive correlation between lexical diversity and the ratings given to writing and speaking samples of learners of English.

The acquisition of PMs might proceed in a fashion that is similar to that of vocabulary, with learners showing an overreliance on a more limited number of expressions.

7. Conclusions, limitations, and directions for future research

This study provides new data on the general rate of PM use in the speech of ESL learners. The data indicate that the frequency of PMs increases gradually at first, increases more dramatically at the highest proficiency levels, and reaches a rate identical to that of native speakers. Learners at lower levels of proficiency tend to use only a small repertoire of PMs that includes *I think*, *so*, and *also*, while those at higher levels use more varied expressions at rates similar to those of native speakers. The richness of these expressions increases gradually across proficiency levels but never reaches native-like patterns of variation.

This study was limited by several important factors that offer indications of areas for future research. The data used in this study come from a semi-direct test of oral proficiency. The fact that the data come from a language test and are monologic (no interlocutor was present) may have a noticeable impact on the types of language produced. There are few studies that directly compare the same tasks being performed under both monologic and dialogic conditions, and those that have examined this have shown mixed results. Gan (2012) found that ESL learners produced more T-units and clauses and longer T-units and utterances on monologic rather than dialogic tasks. Michel, Kuiken, and Vedder (2007) found that the language produced by learners of Dutch on monologic tasks was more complex but less fluent and accurate than the language produced on dialogic tasks. Therefore, the monologic condition under which the data for this study were gathered may have a significant impact on the results. Further research should look at PM use in more naturalistic data and with a speaking partner. This is especially important given the interactive nature of PMs. As discussed above, the monologic data proved to be advantageous in many ways in the research design of this study; however, dialogic data would provide additional information concerning aspects of PM use that are not addressed in this study.

As discussed above, another limitation involved in the study of PMs is the lack of agreement about their function, status, and identification. The identification issue is particularly troublesome since it affects the comparability of different studies. This study has tried to minimize the problem by using a compre-

hensive, published list of PMs from a corpus-based grammar (Carter & McCarthy, 2006). However, the generalizability of the results here is still limited by the fact that most studies use different definitions, lists, and even terminology for PMs. Building an understanding of how learners use PMs will require not only more research but particularly research that makes it clear which types of expressions were counted as PMs.

The findings of this study that highly proficient non-native speakers use PMs at a rate, but not with a range, similar to native speakers also suggests areas for future research. There may be potentially instructive comparisons to be made between the use of PMs and the development of vocabulary in general. Research into second language vocabulary acquisition, as discussed above, has shown that lexical diversity correlates positively with proficiency level. It would be instructive to consider the development of PM use in comparison with the development of lexical diversity and to examine whether they proceed in a similar fashion. In fact, the development of PM use should be considered in light of the growth of other linguistic subsystems in learners' interlanguage.

In this vein, Bardovi-Harlig (1999) noted a lack of an acquisitional perspective in second language pragmatics and has called for a research agenda that re-focuses attention on the interlanguage in interlanguage pragmatics. One example she gives is the investigation of expressions that include *I think* and how the acquisition of these expressions might or might not be related to the acquisition of the grammatical knowledge necessary to use them. This is also a significant next step necessary for expanding our understanding of how PM use fits into the larger project of second language pragmatic development.

References

- Aijmer, K. (2002). *English discourse particles: Evidence from a corpus*. Amsterdam: John Benjamins.
- Aijmer, K. (2013). *Understanding pragmatic markers: a variational pragmatic approach*. Edinburgh: Edinburgh University Press.
- Bardovi-Harlig, K. (1999). Exploring the interlanguage of interlanguage pragmatics: A research agenda for acquisitional pragmatics. *Language Learning*, 49, 677-713.
- Bardovi-Harlig, K. (2013). Developing L2 pragmatics. *Language Learning*, 63(s1), 68-86.
- Bell, D. M. (2010). Nevertheless, still and yet: Concessive cancellative discourse markers. *Journal of Pragmatics*, 42, 1912-1927.
- Blakemore, D. (2002). *Relevance and linguistic meaning: The semantics and pragmatics of discourse markers*. Cambridge: Cambridge University Press.
- Blakemore, D. (2008). Discourse markers. In R. L. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 221-240). Malden, MA: Blackwell.
- Buyse, L. (2009). So as a marker of elaboration in native and non-native speech. In S. Slembrouck, M. Taverniers, & M. Van Herreweghe (Eds.), *From will to well: Studies in linguistics offered to Anne-Marie Simon-Vandenbergen* (pp. 79-91). Gent: Academia.
- Buyse, L. (2012). So as a multifunctional discourse marker in native and learner speech. *Journal of Pragmatics*, 44, 1764-1782.
- Callies, M. (2009). *Information highlighting in advanced learner English: The syntax-pragmatics interface in second language acquisition*. Amsterdam: John Benjamins.
- Carter, R., & McCarthy, M. (2006). *Cambridge grammar of English: A comprehensive guide*. Cambridge: Cambridge University Press.
- Chodorowska-Pilch, M. (2008). Verás in Peninsular Spanish as a grammaticalized discourse marker invoking positive and negative politeness. *Journal of Pragmatics*, 40, 1357-1372.
- Daller, H., Van Hout, R., & Treffers-Daller, J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied Linguistics*, 24, 197-222.
- Dittmar, N., & Terborg, H. (1991). Modality and second language learning. A challenge for linguistics. In T. Huebner & C. Ferguson (Eds), *Crosscurrents in second language acquisition and linguistic theories* (pp. 347-384). Amsterdam: John Benjamins.
- Durán, M. E., & Unamuno, V. (2001). The discourse marker a ver (Catalan, a veure) in teacher-student interaction. *Journal of Pragmatics*, 33, 193-208.
- Feng, G. (2008). Pragmatic markers in Chinese. *Journal of Pragmatics*, 40, 1687-1718.
- Fillmore, C. (2000). On fluency. In H. Rigggenbach (Ed.), *Perspectives on fluency* (pp. 43-61). Ann Arbor: University of Michigan Press.

- Foster, P., & P. Tavakoli (2009). Native speakers and task performance: Comparing effects on complexity, fluency, and lexical diversity. *Language Learning, 59*, 866-896.
- Fraser, B. (1999). What are discourse markers. *Journal of Pragmatics, 31*, 931-952.
- Fuller, J. (2003). The influence of speaker roles on discourse marker use. *Journal of Pragmatics, 35*, 23-45.
- Fung, L., & Carter, R. (2007). Discourse markers and spoken English: Native and learner use in pedagogic settings. *Applied Linguistics, 28*, 410-439.
- Gan, Z. (2012). Complexity measures, task type, and analytic evaluations of speaking proficiency in a school-based assessment context. *Language Assessment Quarterly, 9*(2), 133-151.
- Ginther, A., Dimova, S., & Yang, R. (2010). Conceptual and empirical relationships between temporal measures of fluency and oral English proficiency with implications for automated scoring. *Language Testing, 27*, 379-399.
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Hasselgreen, A. (2004). *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.
- Hays, P. R. (1992). Discourse markers and L2 acquisition. *Papers in Applied Linguistics-Michigan, 7*, 24-34.
- Hellermann, J., & Vergun, A. (2007). Language which is not taught: The discourse marker use of beginning adult learners of English. *Journal of Pragmatics, 39*, 157-179.
- House, J. (2013). Developing pragmatic competence in English as a lingua franca: Using discourse markers to express (inter) subjectivity and connectivity. *Journal of Pragmatics, 59*, 57-67.
- Ishida, M. (2013). Second Language Pragmatic Development. *The Encyclopedia of Applied Linguistics*.
- Kasper, G., & Rose, K. R. (2002). *Pragmatic development in a second language*. Oxford: Blackwell.
- Lenk, U. (1997). Coherence in spoken and written discourse: How to create it and how to describe it. In W. Bublitz, U. Lenk, & E. Ventola (Eds.), *Selected papers from the international workshop on coherence, Augsburg* (pp 1-17). Amsterdam: John Benjamins.
- Lenk, U. (1998). *Marking discourse coherence: Functions of discourse markers in spoken English*. Tübingen: Narr.
- Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Journal, 96*, 190-208.

- Malvern, D., & Richards, B. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing, 19*(1), 85-104.
- Michel, M. C., Kuiken, F., & Vedder, I. (2007). The influence of complexity in monologic versus dialogic tasks in Dutch L2. *IRAL: International Review of Applied Linguistics in Language Teaching, 45*(3), 241-259.
- Müller, S. (2005). *Discourse markers in native and non-native English discourse*. Amsterdam: John Benjamins.
- Neary-Sundquist, C. (2013). Task type effects on pragmatic marker use by learners at varying proficiency levels. *L2 Journal, 5*, 1-21.
- O'Loughlin, K. (1995). Lexical density in candidate output on direct and semi-direct versions of an oral proficiency test. *Language Testing, 12*(2), 217-237.
- Overstreet, M., & Yule, G. (1997). Fostering pragmatic awareness. *Applied Language Learning, 10*(1), 1-14.
- Romero-Trillo, J. (1997). Pragmatic mechanisms to obtain the addressee's attention in English and Spanish conversations. *Journal of Pragmatics, 28*, 205-221.
- Romero-Trillo, J. (2002). The pragmatic fossilization of discourse markers in non-native speakers of English. *Journal of Pragmatics, 34*, 769-784.
- Romero-Trillo, J. (2012). Pragmatic markers. *The encyclopedia of applied linguistics*.
- Sasamoto, R. (2008). Japanese discourse connectives dakara and sorede: A re-assessment of procedural meaning. *Journal of Pragmatics, 40*, 127-154.
- Scarcella, R. C. (1983). Developmental trends in the acquisition of conversational competence by adult second language learners. In N. Wolfson & E. Judd (Eds.), *Sociolinguistics and language acquisition* (pp. 175-183). Rowley, MA: Newbury House.
- Schiffrin, D. (1987). *Discourse markers*. Cambridge: Cambridge University Press.
- Schourup, L. (1999). Discourse markers. *Lingua, 107*(3-4), 227-265.
- Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics, 17*, 84-119.
- Wang, Y., & Tsai, P. (2007). Textual and contextual contrast connection: A study of Chinese contrastive markers across different text types. *Journal of Pragmatics, 39*, 1775-1815.
- Waring, H. Z. (2003). 'Also' as a discourse marker: Its use in disjunctive and dis-affiliative environments. *Discourse Studies, 5*, 415-436.
- Yu, G. (2010). Lexical diversity in writing and speaking task performances. *Applied linguistics, 31*, 236-259.
- Zareva, A., Schwanenflugel, P., & Nikolova, Y. (2005). Relationship between lexical competence and language proficiency: Variable sensitivity. *Studies in Second Language Acquisition, 27*, 567-595.

APPENDIX A

Descriptors for the levels of the oral proficiency test

Level	Content	Delivery
6	Wide range of vocabulary Complexity of sentence structure Interpretative/summary statements Some non-native usage Meaning clearly expressed Provision of a frame Economy of expression	Smooth delivery Almost no pauses/hesitations/ choppiness Thought expressed in one utterance No problems with articulation Use of varied intonation and tone
5	Somewhat unconventional words Listener effort needed at times Simple sentence construction Well organized and coherent Meaning clear	Clearly non-native like delivery Some pauses and choppiness, but comprehension unobstructed Some sound substitutions Listener effort required at points
4	Dependence on the prompt Ineffective/abrupt transitions Omission of function words Systematic problems with bound morphology Topic shifts Lack of coherence Weak organization Repetition interferes with coherence Intended meaning unclear Lack of elaboration	Ineffective repetition of words/ phrases Pauses/hesitations are more frequent Flat intonation Many identifiable articulation/ pronunciation/stress problems Pace interferes with comprehension Close listener attention required
3	Misuse of particular words Problems with bound morphology Frequent attempts to re-start/ re-phrase without clarification Unintended meaning Misunderstands prompt	Deliberate/ineffective delivery Frequent pauses/hesitations within phrasal boundaries Ineffective attempts of interpretative statements Limitation of vocabulary

APPENDIX B

Participant data

Number	Native language	Native country	Oral proficiency test score
5	Chinese	China	3
5	Korean	Korea	3
5	Chinese	China	4
5	Korean	Korea	4
5	Chinese	China	5
5	Korean	Korea	5
3	Chinese	China	6
4	Korean	Korea	6
10	English	USA	N/A

APPENDIX C

List of pragmatic markers from Carter and McCarthy (2006)

actually	I mean	not to say	that's to say
admittedly	I must admit	now	there again
also	I must say	obviously	third/ly
amazingly	I think	of course	to be frank
and	I'm afraid	oh	to be honest
and then	I'm sorry	okay	to put it another way
anyway	ideally	on top of that	to put it mildly/bluntly
apparently	if you ask me	ooo	to sum up
arguably	if you like	oops	to tell you the truth
as I say	in a manner of speaking	or rather	ugh
as I was saying	in fact	ouch	understandably
as it were	in general	perhaps	undoubtedly
basically	in other words	predictably	unfortunately
by any chance	in the end	presumably	well
certainly	in the X place	probably	what's more
clearly	indeed	putting it mildly/bluntly	where was i
confidentially	just (about)	really	wow
'cos	just think	remember	you know
damn	kind of	right	you see
doubtless	last of all	rightly	
essentially	lastly	roughly	
finally	like	sadly	
fine	like	second/ly	
first(of all/ly)	listen	seriously	
for a start	literally	so	
fortunately	look	so to speak	
frankly	maybe	sort of	
going back to	mind you	strictly speaking	
good	naturally	surely	
gosh	next	surprisingly	
great	no doubt	thankfully	
honestly			
hopefully			