# Japanese EFL Learners' Acquisition of Discourse Markers: A Comparative Analysis of Spoken Corpora and English Textbooks

**A Dissertation**

**Submitted to the University of Tsukuba**

**In Partial Fulfilment of the Requirements for**

**the Degree of**

**Doctor of Philosophy in Linguistics**

**Kazunari SHIMADA**

**2014**

## Abstract

Discourse markers are words and phrases for speakers or writers to make their message more accessible to others and manage the discourse. Particularly in speech, non-native speakers can strategically use discourse markers to make up for their limited English language proficiency and facilitate communication (e.g., Flowerdew & Tauroza, 1995; House, 2013; Williams, 1992). In other words, discourse markers are important lexical items for improving learners' communicative competence, and hence their acquisition is an essential part of English language learning.

However, despite the importance of discourse markers in spoken English, there have been only a limited number of studies investigating how non-native speakers acquire and use them, or exploring the relationship between the input and output of discourse markers in second and foreign language learning contexts. To rectify the inadequacies, this research focuses on Japanese EFL learners' speech and examines the features and process of acquiring discourse markers.

The work consists of five related studies using the framework of corpus-based multiple comparisons. The first two studies aim to identify features of Japanese EFL learners' discourse marker use. Study 1 compared the use of discourse markers by Japanese EFL learners and native English speakers. To verify the results of Study 1, Study 2 first conducted a replication analysis using different spoken corpora, and then compared Japanese learners' discourse marker use with that of other English language learners of different first language backgrounds.

The following two studies concern factors affecting the acquisition of discourse markers in the Japanese EFL context. Study 3 applied a contrastive analysis to English-Japanese parallel data and included a small-scale experiment in which a picture

description task was used to investigate first language transfer in the acquisition and use of discourse markers by Japanese learners. Study 4 examined the presentation of discourse markers in Japanese EFL textbooks for junior high and high school students and explored how input from textbooks can affect Japanese learners' acquisition of discourse markers.

Additionally, Study 5 investigated the presentation of discourse markers in EFL textbooks for Japanese college students and adult learners as a means of examining how textbook designers should take into consideration the features of Japanese English learners' interlanguage, especially in regard to discourse marker acquisition.

The findings of Studies 1 and 2 indicate that Japanese learners may overuse or underuse certain discourse markers compared to native speakers and other non-native speakers of English. Additionally, the findings of Studies 3 and 4 suggest that Japanese learners may acquire discourse markers under the influence of first language transfer and input data from textbooks, and these factors may be the reasons behind certain features of their discourse marker use. Moreover, the findings of Study 5 suggest that learners' proficiency level may be a key factor in determining the frequency and kind of discourse markers used in textbooks. Based on the findings of the research, some suggestions for language teaching are also provided along with further implications for designing textbooks as primary input resources.

**Acknowledgements**

I wish to express my gratitude to all people who helped me in the completion of this research. First, I would like to specially thank Professor Akira Kubota, my supervisor and dissertation committee chair, for his constant guidance and support in the writing of this dissertation. His comments enlightened me about the perspectives of materials development based on second language acquisition research. He also gave me many opportunities to learn the theory and practice of language teaching.

I am also extremely grateful to the other dissertation committee members: Professor Hirosada Iwasaki, for his helpful advice on corpus analysis and academic writing, Professor Akiyo Hirai, who taught me how to analyse and interpret statistical data, and Professor Yukio Tono, for his invaluable comments from the perspective of learner corpus research as an external committee member.

I also wish to thank Professor Yuji Ushiro. He gave me constructive comments on my research design and the overall discussion.

I am indebted to my colleagues at Tokyo Keizai University, especially to Aika Miura, who helped me collect learners' speech data and gave insightful comments in various stages of the dissertation writing.

Special thanks go to my fellow PhD students, in particular Yasuhiro Okuyama, and graduates, Katsuyuki Konno, Mayumi Kobayakawa, Hideki Iimura, Sachiyo Takanami, and Yuki Shibahara, for fruitful discussions and friendship over the years.

I gratefully acknowledge all the university students who participated in the study and provided their speech data. Without their participation, this study would not have been possible.

I am grateful as well to the National Institute of Information and Communications

Technology (NICT), which has given permission to use the NICT-JLE Corpus data for my research.

Finally, I would like to thank my friends and family for their constant support. Above all, I wish to express my endless gratitude to my wife Hanae for her encouragement and understanding. Without her support, the dissertation project could never have been completed.

## Table of Contents

## List of Tables

Page

**Chapter 2**



**Chapter 3**



**Chapter 4**



**Chapter 5**

**Chapter 6**

**Chapter 7**

## List of Figures

Page

**Chapter 2**



**Chapter 3**



**Chapter 4**



**Chapter 5**

**Chapter 6**



**Chapter 7**



**Chapter 8**

## List of Appendices

## List of Abbreviations

| Abbreviation | Meaning |
|---|---|
| ANOVA | analysis of variance |
| BNC | British National Corpus |
| CA | contrastive analysis |
| CANCODE | Cambridge and Nottingham Corpus of Discourse in English |
| CEFR | Common European Framework |
| CHILDES | Child Language Data Exchange System |
| CIA | Contrastive Interlanguage Analysis |
| DM | discourse marker |
| EFL | English as a foreign language |
| FED | frequency of English DMs |
| ICLE | International Corpus of Learner English |
| IL | interlanguage |
| LINDSEI | Louvain International Database of Spoken English Interlanguage |
| LLR | log-likelihood ratio |
| LOB | Lancaster-Oslo-Bergen |
| LOCNEC | Louvain Corpus of Native English Conversation |
| LOCNESS | Louvain Corpus of Native Speaker Essays |
| L1 | first language |
| L2 | second language |
| NICT | National Institute of Information and Communications Technology |
| NL | native language |
| NNS | non-native speaker |
| NP | noun phrase |
| NS | native speaker |
| OL | original language |
| RQ | research question |
| SL | source language |
| SST | Standard Speaking Test |
| TL | translated language (Figure 2.3), target language (Figure 2.5) |
| TTR | type-token ratio |

# Chapter 1

# Introduction

## 1.1 Background of the Research

Discourse markers (DMs) are lexical items whose pragmatic and syntactic functions play a crucial role in speech communication: Speakers use them to create textual coherence in interaction, as well as to express their own feelings or stances (Carter & McCarthy, 2006). For example, *OK/okay*, *really*, and *right* are used to respond to a speaker's utterance and to suggest agreement, alignment, or active listening. *But*, *first*, and *then* serve to organise discourse structure. Words like these are tools that enable speakers to convey their meanings to their listeners.

Additionally, even if spoken sentences or phrases are grammatically correct, the lack of DMs may make it difficult to attract listeners' attention in a polite way (Romero-Trillo, 2002) and may create a negative impression of being uncollaborative or awkward in conversation (Svartvik, 1980). In other words, the lack of DMs may result in communication failure in the interaction between speaker and listener.

Therefore, DMs are of special importance to non-native speakers (NNSs), who can use them to compensate for limited English language proficiency and to improve their production and comprehension of messages. With regard to production, for example, House (2013) pointed out the importance of DMs for NNSs. Her qualitative study revealed that NNSs strategically used DMs such as *yes*, *yeah*, *so*, and *okay* to express their own views. Concerning listener's comprehension, Flowerdew and Tauroza (1995) found that second language (L2) learners comprehended a lecture with the use of DMs better than that without their use. However, while Watts (1989) suggests that native speakers (NSs) can unconsciously make use of DMs in their speech, NNSs have

difficulty using DMs. In pedagogical settings, authentic DM inputs may cause cognitive overload of learners. That is to say, too much emphasis should not be placed on native-like use of DMs.

Along with recognising the importance of DMs, considerable interest has emerged in the roles and functions of individual DMs such as *because*, *oh*, and *well* (e.g., Blakemore, 2002; Fraser, 1999, 2009; Schiffrin, 1987). Additionally, the development of corpus linguistics has enabled data-driven quantitative and qualitative analyses of the use of DMs by NSs of English (e.g., Lenk, 1998; McCarthy & Handford, 2004). However, a relatively limited amount of research has been conducted concerning DM use in terms of L2 acquisition, especially in the Japanese English as a foreign language (EFL) context. Positioned against this contextual background, this research focuses on DM use in the speech data of Japanese English learners.

### 1.2 Organisation of the Dissertation

This dissertation consists of eight chapters. The first chapter gives the background and significance of the work, and outlines the organisation of the dissertation. Chapter 2 reviews studies on (a) definition and taxonomy of DMs in spoken English, (b) spoken DMs and L2 acquisition, (c) DM use in the first language (L1) and L2, (d) DM use in EFL textbooks as input resources, and (e) frameworks for corpus-based interlanguage analysis. Based on this review, some limitations and problems of the previous literature are identified, and the purpose of the current research is defined in the last part of the chapter.

Chapter 3 describes Study 1. This study examines the use of DMs by Japanese learners and NSs of English. In order to identify which markers should be provided to learners, the analysis is based on the comparison of three different corpora: a spoken

corpus of Japanese EFL learners, and the adult and child speech data of native English speakers. Moreover, some features of the development of DMs are discussed through the results of the analysis.

Chapter 4 describes Study 2, which was conducted to explore the features of the use of DMs in the speech of Japanese learners of English. The chapter is divided into two major parts. The first part is devoted to the replication of the results of Study 1. The replication analysis is carried out using the frequency of DMs in two datasets of 15-minute interviews with Japanese EFL learners and NSs of English. The second part identifies the similarities and differences in DM use between Japanese learners and other English language learners with different L1 backgrounds. In order to distinguish the features of DM use by Japanese learners from those of non-native English learners, a comparative analysis is conducted using an international database that includes interviews produced by English learners with six different L1 backgrounds.

Chapter 5 presents two cross-linguistic studies to investigate the influence of L1 transfer on DM use in L2. Study 3A applies contrastive analysis (CA) to English-Japanese parallel data: English speech data extracted from the Japanese EFL learner corpus used in Study 1 and their Japanese translations. Through the comparison of English and Japanese DMs, the analysis examines correspondences between the two languages. In Study 3B, a small-scale experiment using a picture description task is performed to explore L1 effects on the use of English DMs by Japanese EFL learners. Based on the findings of Study 3A, Study 3B compares the learners' L1 and L2 utterances under the same condition and investigates how the use of Japanese DMs influences the use of English DMs.

Chapter 6 presents Study 4, which focuses on the presentation of DMs in Japanese EFL textbooks for junior high and high school students and explores how

textbooks, as primary input data, play a role in Japanese EFL learners' speech. The chapter is also divided into two major parts. The first part investigates the presentation of DMs in textbooks according to their target levels. The second part compares the results of the first part with spoken DM use by Japanese EFL learners in order to explore how the input of DMs can be transferred into learners' output.

Chapter 7 describes Study 5, which was conducted to investigate the presentation of DMs in EFL textbooks for Japanese college students and adult learners and explore how the textbooks take into account Japanese English learners' interlanguage. In other words, this study examines the language of textbooks, which accounts for a large portion of the input data for language learners, and considers what kind of and how many DMs should be incorporated in textbooks.

The final chapter summarises the findings of these corpus-based studies with concluding remarks and pedagogical implications for language teaching and materials development. Some limitations of the work and suggestions for further research are also discussed at the end of the chapter.

# Chapter 2

# Literature Review[1]

## 2.1 Discourse Markers (DMs)

### 2.1.1 Definition of DMs

DMs have been defined in various ways by researchers and given a variety of labels: cue phrases, discourse connectives, discourse operators, discourse particles, pragmatic markers, pragmatic operators, and semantic conjuncts among others (Fraser, 2009). The use of such a wide range of labels implies, as Fung and Carter (2007) note, that researchers have diverse interests in DMs, and from various viewpoints. However, there is generally a consensus that they mainly serve pragmatic and syntactic functions in discourse (e.g., Hellermann & Vergun, 2007).

Schiffrin (1987), whose study has been regarded as one of the most extensive on pragmatic functions of DMs, defines DMs as "sequentially dependent elements which bracket units of talk" (p. 31). In other words, she considers DMs to be contextual coordinators for integrating certain components of talk, and establishing or maintaining a relationship between the speaker and hearer.

(1) Zelda: Are you from Philadelphia?

Sally: *Well* I grew up uh out in the suburbs. And then I lived for about seven years up in upstate New York. And then I came back here t'go to college.

(Schiffrin, 1987, p. 106)

In example (1), Sally uses *well* as a signal that she cannot give a clear answer to Zelda's

yes-no question—in other words, that her pragmatic contribution is at odds with her interlocutor's expectations. Thus, as Schiffrin points out, the marker *well* plays the role of contextual coordinator, marking a juncture between a speaker's intention and a hearer's interpretation.

Additionally, Schiffrin uses the term DMs to cover expressions such as *I mean*, *you know*, *oh*, and *like*, and nonverbal signals including a gaze. These items do not serve essential syntactic functions; rather, they are optional devices through which speakers can shape their utterances to affect hearers' knowledge.

(2) a. *I mean* I may be wrong, but I'm—*I mean* that's what I'm—that's my opinion.

b. We have some *y'know*.

(Schiffrin, 1987, pp. 34–35)

Their predominant function is a pragmatic one. The markers in examples (2a) and (2b) play a role in indicating the speakers' intention to keep conversation going and help the hearers focus on the upcoming words. Markers such as those in (2a) and (2b) are ubiquitous in everyday spoken English, despite their syntactic irrelevance. In regard to this point, Brinton (1996) states that DMs often occur as "a feature of oral rather than of written discourse" (p. 33), and they are optional items with little or no proposition.

On the other hand, Fraser (1999) defines DMs as "a class of lexical expression drawn primarily from the syntactic classes of conjunctions, adverbs, and prepositional phrases" (p. 931). He addresses their syntactic functions and considers DMs to be linguistic items signalling a relationship between two segments of discourse. Fraser argues that a DM must be included as an integral syntactic part of its next discourse

segment. The items are italicised in (3):

(3) a. Jones died last night. *But* he had been very ill for a long time.

b. I went to Boston first *and* later on, went to Cape Cod.

c. The water wouldn't boil, *so* we couldn't make any tea.

(Fraser, 2009, p. 294)

In other words, the purpose of each marker in examples (3a), (3b), and (3c) is to make coherent links between one discourse segment and another.

Like Fraser, Blakemore (2002) also considered DMs to be devices to signal the relationship between two discourse segments and to constrain a selection of interpretation or assumption of the hearer or receiver.

(4) a. Tom can open Ben's safe. So he knows the combination.

b. Tom can open Ben's safe. After all, he knows the combination.

(Blakemore, 2002, p. 79)

In sentence (4a), the proposition in the first segment (i.e., *Tom can open Ben's safe*) leads to the conclusion in the second segment (i.e., *he knows the combination*), while, in sentence (4b), Tom's ability expressed in the first segment is caused by applying his knowledge in the second segment. In short, DM use can change the interpretation of a sentence.

Therefore, Schiffrin's (1987) definition of DMs is a broader one than Fraser's (1999, 2009) or Blakemore's (2002) definition, and her model illustrates features of the spoken mode in more detail.

### 2.1.2 Taxonomy of DMs in Spoken and Written English

DMs are used as linking devices in spoken English, and some also help to organise written texts (Carter & McCarthy, 2006). In other words, DM use may vary according to the mode of discourse: spoken or written.

Biber (1988) examined NS spoken and written texts of approximately 960,000 words across 23 genres such as press reportage, editorials, face-to-face conversation, and telephone conversation.[2] His corpus-based analysis revealed that the speech data contained five markers, *well*, *now*, *anyway*, *anyhow*, and *anyways*, more frequently than the written data, while the speech data had the marker *because* less frequently than the written data. In short, the frequency of DMs may be influenced by the difference of mode: spoken vs. written language.

Berman, Ragnarsdóttir, and Strömqvist (2002) qualitatively compared texts produced by English NSs with written expository texts and found that there were significant differences between the spoken and written mode in DM use by English NSs.

(5) And uh *anyway* we *kind of* fell *you know* lost touch after high school. But that I don't know why that came to mind.

(6) *I think* people should take a moment and think about why the problem is occurring before taking on an automatic adversary role.

(Berman, Ragnarsdóttir, & Strömqvist, 2002, p. 270)

In example (5), an utterance by an English-speaking graduate student, markers such as *anyway*, *kind of*, and *you know* serve interactive functions with the hearer. Contrastively, example (6), the same student's writing, contains no such interactive types of markers,

while it begins a new discourse segment with the marker *I think*. Berman, Ragnarsdóttir, and Strömqvist (2002) thus point out that these examples may reflect the features of spoken and written discourse.

In view of these differences, some researchers build taxonomies of DMs in spoken and written mode. In regard to the spoken mode, Fung and Carter (2007) proposed a functional paradigm of DMs based on their analysis on both NS and NNS spoken data. In the framework, 57 items were categorised into four discourse functions: interpersonal, referential, structural, and cognitive (see Table 2.1).

Table 2.1
*A Functional Paradigm of DMs in Speech*

| Category | Discourse Functions and Markers |
|---|---|
| Interpersonal | Marking shared knowledge, indicating attitudes, or showing responses:<br>*absolutely, actually, basically, exactly, great, I see, I think, just, kind of, like, listen, obviously, oh, oh great, OK/okay, really, right/alright, see, sort of, sure, to be frank, to be honest, well, yeah, yes, you know, you see* |
| Referential | Indicating relationship between utterances:<br>*and, anyway, because/'cause, but, cos, however, likewise, nevertheless, or, similarly, so, yet* |
| Structural | Organising or managing the direction of conversations:<br>*and, finally, first, firstly, how about, let me conclude the discussion, let's discuss, let's start, next, now, OK/okay, right/alright, second, secondly, so, then, well, what about, yeah* |
| Cognitive | Denoting thinking process, or reformulating utterance:<br>*and, I mean, I see, I think, in other words, like, sort of, that is, to put it in another way, well, what I mean is, you know* |

*Note.* Adapted from "Discourse markers and spoken English: Native and learner use in pedagogic settings," by L. Fung and R. Carter, 2007, *Applied Linguistics, 28,* p. 418. Some DMs such as *and, I think*, and *well* have multiple functions in discourse.

To summarise their definitions, on interpersonal use, DMs such as *I see* and *OK/okay* are used for showing responses to the listener, and items such as *you know* and *you see* enable the speakers to share their knowledge. Several DMs with the referential function are conjunctions that join utterances, while DMs with the structural function are used to organise or manage the direction of conversations. In the cognitive category[3], speakers can pause to consider their words by using DMs such as *well* and *I think*, and can also reformulate their utterance with some items such as *I mean*.

In transactional spoken discourse, DMs can be also assigned to one of two organisers: micro- and macro-markers, according to the definitions of Chaudron and Richards (1986) and Nattinger and DeCarrico (1992) (see Table 2.2). Micro-markers are often used when speakers process lower-level information in discourse, while macro-markers are signals of higher-level information. Micro-markers mainly facilitate the process of understanding small segments in discourse, while macro-markers allow listeners to predict the likely development of discourse by using their prior knowledge. According to these definitions, examples of micro-markers are *and*, *just*, and *sort of*, while macro-markers are *first*, *how about*, and *then*.

Table 2.2
*Micro- and Macro-markers*

| Organisers | DMs |
|---|---|
| Micro-markers | *absolutely, actually, and, anyway, basically, because/'cause, but, cos, exactly, great, I mean, I see, I think, just, kind of, like, likewise, listen, nevertheless, now, obviously, oh, oh great, OK/okay, or, really, right/alright, see, similarly, so, sort of, sure, to be frank, to be honest, to put it in another way, well, yeah, yes, yet, you know, you see* |
| Macro-markers | *finally, first, firstly, how about, however, in other words, let me conclude the discussion, let's discuss, let's start, next, second, secondly, that is, then, what about, what I mean is* |

Concerning the written mode, Hyland and Tse (2004) propose a framework of metadiscourse markers based on their analysis of an L2 postgraduate corpus (see Table 2.3). Metadiscourse markers are divided into two categories: interactive and interactional resources. The former involves signals that organise discourse through the text and make the reader understand the writer's intention. The interactive category consists of five subcategories: transitions, frame markers, endophoric markers,

Table 2.3
*A Framework of Metadiscourse Markers in Academic Texts*

| Category | Function | Examples |
|---|---|---|
| Interactive resources | | |
| Transitions | Express semantic relation between main clauses | *in addition*, *but*, *thus*, *and* |
| Frame markers | Refer to discourse acts, sequences, or text stages | *finally*, *to conclude*, *my purpose here is to* |
| Endophoric markers | Refer to information in other parts of the text | *noted above*, *see Fig*, *in section 2* |
| Evidentials | Refer to source of information from other texts | *according to X/(Y, 1990)*, *Z states* |
| Code glosses | Help readers grasp functions of ideational material | *namely*, *e.g.*, *such as*, *in other words* |
| Interactional resources | | |
| Hedges | Withhold writer's full commitment to proposition | *might*, *perhaps*, *possible*, *about* |
| Boosters | Emphasize force or writer's certainty in proposition | *in fact*, *definitely*, *it is clear that* |
| Attitude markers | Express writer's attitude to proposition | *unfortunately*, *I agree*, *surprisingly* |
| Engagement markers | Explicitly refer to or build relationship with reader | *consider*, *note that*, *you can see that* |
| Self-mentions | Explicit reference to author(s) | *I*, *we*, *my*, *our* |

*Note.* Adapted from "Metadiscourse in academic writing: A reappraisal," by K. Hyland and P. Tse, 2004, *Applied Linguistics*, *25*, p. 169.

evidentials, and code glosses. The latter represents engaging and involving the reader in the writer's argument. Like the interactive category, the interactional category also consists of five subcategories: hedges, boosters, attitude markers, engagement markers, and self-mentions.[4]

The markers are defined as "devices writers use to explicitly organise their texts, engage readers, and signal their attitudes to both their material and their audience" (Hyland & Tse, 2004, p. 156). Thus, metadiscourse markers can be regarded as DMs in written texts.

## 2.2 Spoken DMs and L2 Acquisition

### 2.2.1 DMs Used by Native and Non-native English Speakers

With the widely recognised importance of DMs in spoken discourse, there have been several studies examining the use of DMs by language learners. In the earlier studies, the comparative approach was often used to identify differences in DM use between native and non-native English speakers.

Romero-Trillo (2002) investigated the development of DMs in Spanish learners and NSs of English. His analysis, using spoken data of children and adults, revealed that there was little difference between native and non-native children in their use of certain markers, including *I mean*, *well*, and *you know*, but non-native adults used these features less frequently than native adults did. He also found that Spanish children overused *listen* due to the influence of their mother tongue: The frequency of the equivalent word was high in Spanish conversations.

Likewise, Müller (2004) compared the use of DMs by German EFL speakers with their use by American NSs. Her research revealed similar findings to those of Romero-Trillo (2002). She pointed out that the frequency of DMs used by German

learners was influenced by the frequency of the translation equivalents in their L1. In particular, she noted that German speakers used *well* much more frequently than American speakers did, but they used *so* much less frequently. As part of the discussion, Müller pointed out that both DMs were translated as the German adverb *also*, and that German speakers might have a preference for *well* in order to avoid confusing English *so* and German *so*.

Aijmer (2004) conducted corpus-based analyses revealing significant differences in the distributions of certain DMs in NS and NNS speech. Aijmer (2004) found that Swedish learners of English overused *I don't know* in order to signal uncertainty or hesitation. In addition, Aijmer (2011) revealed that, although Swedish learners used *well* more frequently than NSs of English did, they underused some of the marker's functions. She found that Swedish learners frequently used *well* for buying more time to plan what to say, while they underused it for certain functions such as self-repair and clarification. In this respect, Aijmer (2011) suggested that Swedish learners' use of *well* may be influenced by the absence of equivalent markers in their L1.

Similarly, Fung and Carter (2007) showed that learners in Hong Kong underused many markers, such as *right*, *yeah*, *well*, and *you know*, compared to the frequencies found in British NS data. Their study also found that NSs used a wider variety of DMs while NNSs overused some markers. However, Fung and Carter did not classify the roles of some items into either discourse functions or other grammatical ones due to the limitation of their computer software. For instance, some items such as *well* and *so* are used as not only DMs, but also adverbs without the role of discourse functions. Thus, there seem to be statistical problems if a computer program merely counts all instances of the words.

The comparative studies mentioned above shed light on differences in the use of

DMs by NSs and NNSs, but there is room for further research to explore why learners under- or overuse certain DMs. Previous studies such as Romero-Trillo (2002), Müller (2004), and Aijmer (2011) suggested that learners may under- or overuse some items under the influence of their L1, but the L2 corpora were not cross-linguistically compared with L1 corpora. Additionally, some learner corpora may not be of sufficient size to analyse the linguistic features.[5] Although, as Leech (1991) argues, data validity or reliability should not be judged only by the size of a corpus, the question remains as to whether the learner corpora in their research are an appropriate size.

In research on DM use by Japanese learners of English, some researchers have had more interest in DM use in written texts than that in spoken ones. Someya (2001) calculated the frequency of DMs in business letters written by Japanese learners and NSs according to Hyland's list (Hyland & Tse, 2004; see Table 2.3), a categorical list of metadiscourse markers in academic writing. His analysis revealed that Japanese learners used some markers in certain functional categories more frequently than NSs did.

Similarly, Kobayashi and Yamada (2008) based their comparative analysis on Hyland's framework to categorise metadiscourse markers. They utilised data from the JEFLL Corpus (Tono, 2007), a collection of more than 10,000 Japanese English learner free compositions, and the NICT JLE Corpus (Izumi, Uchimoto, & Isahara, 2004), a spoken corpus of more than 1,200 Japanese EFL learners. The results of the analysis indicated there were significant differences between written and spoken modes in the use of metadiscourse markers. Moreover, Kobayashi (2009) compared the use of metadiscourse markers in the JEFLL corpus with that in a corpus of native English essays, and his quantitative and qualitative analyses confirmed significant differences in the use of metadiscourse markers between Japanese learners and NSs.

To my knowledge, only one comparative study investigated differences in the use

of spoken DMs between Japanese learners and NSs of English. Miura (2011) used Japanese learners and NS spoken data from the NICT JLE Corpus to compare the frequency of DMs used by Japanese learners of English to those of NSs. The results revealed that certain markers such as *well*, *I mean*, *kind of*, and *like* were underused by novice and lower-level learners.

As described above, there has been a growing interest in the distribution of DMs found in Japanese learner corpora; however, few studies have investigated DM use in Japanese learners' speech by comparing data of the former with NS spoken data.

### 2.2.2 Acquisition of DMs by Non-native English Speakers

Despite a growing interest in the use of DMs by non-native English speakers, there have been only a limited number of studies examining the L2 acquisition process of DMs. Hellermann and Vergun (2007) gathered speech data of adult learners of English for five years and investigated their use of three DMs, *you know*, *like*, and *well*. The 17 learners who had no previous formal education in English took community college English classes for non-native English speakers living in the United States but were not provided with explicit instruction on DM use. The results of data analysis revealed that, although the learners did not often use the three DMs, some of them tended to use the marker *well* which occurred in their instructors' talk. Additionally, the results suggested that students who stayed longer in the United States used DMs more frequently.

Polat (2011) also conducted a longitudinal research on the development of DM use by an immigrant learner in the United States. The 25-year-old learner was a NS of Turkish and was not taught DM use explicitly. Polat investigated the learner's DM use and development over one year and found that the development pattern differed

depending on the item. At the beginning of the year, while the learner used *you know* frequently, he did not use *like*. However, as the frequency of *you know* decreased, the frequency of *like* increased. Six months into the data collection, the learner used *you know* with approximately the same frequency as *like*. Polat indicated that the "acquisition of discourse markers may depend on individual features of each marker, such as interactional salience, as well as each learner's particular pragmatic needs" (p. 3755). In short, as in Hellermann and Vergun (2007), Polat suggested that interaction with NSs may affect DM use by NNSs.

Only a few studies have empirically investigated acquisition of DMs in the speech of Japanese English learners. Hays (1992) described the acquisition of DMs used by Japanese college students of various English proficiency levels. His analysis of the spoken data revealed that while the markers *and*, *but*, and *so* were frequently used, *you know* and *well* were rarely uttered. In other words, his results indicate that the Japanese learners had greater difficulties acquiring pragmatic markers such as *you know* and *well*.

Yamamoto (2010) investigated the effect of the study abroad experience on the acquisition of *so* used by 27 Japanese college students with intermediate and upper-intermediate proficiency levels. The results showed that the learners used the interactional and textual functions[6] of *so* in their speech more frequently than before the six-week study abroad program in Australia or New Zealand. Additionally, Yamamoto (2012) conducted similar research on Japanese English learners' usage of *because*. The results suggested that the six-week study abroad program may boost the frequency of *because* in the learners' utterances.

Therefore, Yamamoto (2010, 2012) indicated that Japanese learners of English may also develop their DM use in English-speaking countries. However, although Hays (1992) showed some features of the acquisition of DMs by Japanese learners of English,

few empirical studies have revealed how they can acquire DMs in the Japanese EFL context.

### 2.3 Comparative Studies of L1 and L2 DMs

As mentioned in section 2.2.1, some researchers suggested that learners may use English DMs under the influence of their L1. To investigate the effect of L1 on L2 use, L1 DM use needs to be cross-linguistically compared with L2 DM use. However, there have been a few studies addressing this issue.

Sankoff, Thibault, Nagy, Blondeau, Fonollosa, and Gagnon (1997) analysed interview data from 17 adult Anglophone speakers of Montreal French and investigated the effect of their L1 on French DM use. They interviewed the participants in French, with follow-up interviews in English conducted several weeks later. The instances of DMs in a 15-minute speech from both interviews were counted, and the frequency of DMs was compared between the two languages. The results revealed that the participants often used French DMs such as *tu sais* 'you know,' *alors* 'so,' *comme* 'like,' and *bien* 'well,' as well as the equivalent English markers such as *you know*, *so*, *like*, and *well*. Additionally, the L1 English speakers preferred to use the French DM *comme* not only as an exemplification marker, but also as an adverb to fill a pause.

(7) Comment est-ce que je peux *comme* prendre un petit promenade après?
[How can I *like* take a little walk afterwards?]

(Sankoff et al., 1997, p. 205)

However, French NSs do not make use of *comme* in the way exemplified in (7). Sankoff et al.'s study, therefore, suggests that Anglophone speakers of Montreal French may use

French DMs under the influence of their L1.

Based on Sankoff et al.'s (1977) research design, Liu (2013) investigated the effect of Mandarin Chinese on English DM use by Chinese speakers of English. In Liu's study, the participants were graduate students at the University of Florida who had an advanced level of English proficiency, which is equivalent to a paper-based TOEFL® score of 607 to 657. A 15-minute English interview was conducted to elicit their English DM use, followed by a 15-minute Chinese interview two or three weeks later. Additionally, for comparison of English DM use, five English NSs were interviewed in English for the same amount of time. The frequency analysis of English and Chinese DMs revealed that, while 18 English markers were used in the English interviews, 12 Chinese translation equivalents of them were used in the Chinese interviews. The cross-linguistic study also found that the usage of three Chinese DMs was transferred to that of the corresponding English DMs. For example, the participants often used *I think* in medial or final position when they deliberatively stated their opinion.

(8) yeah I like "Forest Gump" movie. It's really kind of life style different. But it really encouraging people to achieve their own goal *I think*.

(Liu, 2013, p. 159)

The equivalent Chinese marker *wo juede* was also used in a position similar to the one occupied by *I think* in example (8). However, English NSs used *I think* as a marker of tentative function at the end of a sentence, while they used *I think* with the deliberative function at the beginning of a sentence. Therefore, in terms of L1 transfer of DM use to L2, Liu's (2013) study supports the findings of Sankoff et al. (1977).

Additionally, He (2001) investigated the effect of L1 Cantonese on use of the

English DM *so* by Chinese English learners. Her study compared *so* in spoken and written English by Chinese English learners with the Cantonese equivalent *gum* in spoken conversations. The results indicated that its frequency, sentence position, and associated patterns may be influenced by the usage of *gum*. For example, the learners often used the marker *so* followed by *therefore*, *because*, *if*, *now*, and *after* in their writing.

(9) … I was very lazy in my study. *So after* the test, they could smiled aloud.

(He, 2001, p. 46)

In the spoken Cantonese data, *gum* was often associated with the equivalents of *therefore*, *because*, *if*, *now*, and *after*. Therefore, He suggested that the patterns in spoken Cantonese may be transferred to the learners' L2 writing. However, although He mentioned that the English DM *so* can be translated as *suoyi*, *yushi*, *name*, and *gum* in spoken Cantonese, she did not analyse the other three equivalents, *suoyi*, *yushi*, and *name*.

On the other hand, there has been few studies concerning the effect of Japanese on English DM use by Japanese learners of English. One exception is Mizutani (2001), who compared connective expressions in spoken Japanese and English. She collected English-Japanese and Japanese-English translations of approximately 20,000 words from various speech data such as dialogues in textbooks and interviews, and analysed connective expressions, including DMs, in the two languages. For example, Mizutani illustrated a correspondence between Japanese connectives for introducing reasons and their equivalent translated expressions in English (see Table 2.4).

Table 2.4

*Relationship Between Japanese Connectives for Introducing Reasons and the English-Translated Expressions*

| | *(da)kara* | *dadesu-mono* | *ndenode* | *mon-dakara* | *karada* | *te* | Total |
|---|---|---|---|---|---|---|---|
| *because* | 9 | 1 | 1 | | | | 11 |
| *it's because* | 4 | | | | | | 4 |
| *because* | 1 | | 2 | | | | 3 |
| *so* | 14 | 1 | 3 | | | 1 | 19 |
| *so that* | | | | | | 1 | 1 |
| *since* | 8 | | 1 | | | | 9 |
| *as* | 1 | | | | | | 1 |
| *that's why* | | | | | 1 | | 1 |
| *and* | 2 | | 4 | 1 | | | 7 |
| none | 55 | 10 | 4 | 1 | | | 70 |
| other | 8 | 3 | 1 | | | | 12 |
| Total | 102 | 15 | 16 | 2 | 1 | 2 | 138 |

*Note*. Adapted from *Zoku niche-ei hikaku hanashi kotoba no bunpo* [Grammar of spoken language: Comparison between Japanese and English] by N. Mizutani, 2001, p. 100. Tokyo: Kurosio. If it is impossible to translate the Japanese connectives into English ones, they are labelled *none*.

Table 2.4 shows the raw frequency of Japanese connectives in the speech data, but some items such as *and, because,* and *so* are translated into multiple equivalent expressions in English.

Likewise, Onodera (2004) counted the frequency of conjunctions in a transcribed corpus of Japanese speech (see Table 2.5). According to the survey in 1982, four Japanese words equivalent to *and* accounted for 26 per cent of the total occurrence of conjunctions in the Japanese speech data. Additionally, Onodera regarded three Japanese conjunctions as items which can be translated into the English marker *but.*

Table 2.5

*Frequency of Conjunctions in Spoken Japanese (1982)*

| | | | | | | |
|---|---|---|---|---|---|---|
| 1. | *sorede* | 219 | (*and*) | 4. | *dakara* | 299 (16.4%) (*so*) |
| | *soide* | 28 | 539 (29.6%) | | | |
| | *de* | 290 | | 5. | *datte* | 82 (4.5%) (*'cause*) |
| | *nde* | 2 | | | | |
| 2. | *ja* | 212 | (*well then*) | 6. | *sorekara* | 76 (4.1%) (*then*) |
| | *jaa* | 99 | 311 (17.1%) | | | |
| 3. | *demo* | 262 | (*but*) | | | |
| | *dakedo* | 41 | 304 (16.7%) | | | |
| | *dakedomo* | 2 | | | | |

*Note*. Adapted from *Japanese discourse markers* by N. O. Onodera, 2004, p. 6. Amsterdam: John Benjamins. Total number of frequency of 160 conjunctions = 1819 (100%).

Thus, the previous studies imply that there may be a great deal of complexity in the relationship between Japanese and English DMs, although Liu (2013) and He (2001) adopted a one-to-one correspondence between Chinese and English DMs to investigate L1 transfer. As far as I know, however, there has been no detailed study comparing Japanese DMs with English DMs in spoken data.

To study learners' L1 transfer, Tono (2002, 2004) proposes the framework of a multiple comparison approach using a learners' L1 corpus (see section 2.5.3 in this chapter). That is to say, his framework adopts comparisons of learners' L1, their L2 interlanguage, and the target language corpora in order to identify learners' overuse, underuse, or misuse of a certain linguistic item. Tono investigates Japanese learners' acquisition of verb subcategorisation frame patterns in English writing and indicates that the multiple comparisons are conducive to the exploration of learners' language development. Equally, a contrastive linguistic approach might help investigate the development of Japanese learners' DM use.

## 2.4 DM Use in EFL Textbooks as Input Resources

### 2.4.1 Influence of Textbook Materials on Learners' Output

While most previous studies have stressed the importance of DMs in spoken discourse, limited research efforts have been put towards investigating how the input of DMs can be transferred into learners' output, especially with regard to how the presentation of DMs in textbooks reflects learners' speech data. Noteworthily, Müller (2004) examined four German EFL textbooks and German learners' speech data and pointed out that German learners' overuse of the marker *well* could be induced by high frequencies of its DM use in textbooks. She also suggests that textbook contents play a central role in the EFL teaching environment if most English teachers are NNSs of English and rely on materials in textbooks.

Similarly, Lam (2010) also used corpus data to examine how DMs are presented in teaching materials. The study compared the use of DMs in the Hong Kong Corpus of Spoken English with the presentation in a database collected from 15 textbooks for upper-secondary students in Hong Kong. Lam found that there were wide discrepancies in the use of *well* between the textbook database and the speech data uttered by successful English users, including advanced L2 English speakers with L1 Cantonese in Hong Kong. In other words, Lam focused on the mismatch between the textbooks and the speech data but did not discuss the influence of the textbooks in the speakers' DM use.

Although research examining the relationship between input and output of DMs is limited, Tono (2002) investigated the influence of textbook materials on learners' output in terms of verbs. He compared the use of verbs in Japanese EFL learners' essays with the use of these verbs in Japanese junior high and high school textbooks. The results revealed that Japanese learners preferred verbs which were more frequently used in

textbooks. Additionally, Ota, Kanatani, Kosuge, and Hidai (2003) found that words in textbooks covered approximately 70 per cent of words which Japanese junior high school students uttered in a speech contest. With regard to the coverage of vocabulary in textbooks, Tono (2011) revealed that 77.98 per cent of the most frequent 1,000 words in the British National Corpus (BNC) were found in Japanese high school textbooks. As Müller (2004) notes, therefore, these findings suggest that language data which EFL learners encounter in textbooks can be a vital portion of their speech and writing.

### 2.4.2 Materials Design and Learners' Acquisition of DMs

To analyse linguistic features such as DM use in textbooks, most previous studies conducted data-driven quantitative and qualitative analyses of the differences between the language of textbooks and native English speech data.

Scotton and Bernsten (1988) compared natural conversations with dialogues found in ESL textbooks and highlighted the disparity in the sequence structure. Their conversation analysis showed that naturally occurring direction-giving dialogues were more complex than textbook dialogues in the structure of the exchange. They pointed out that the sequence of exchanges in textbooks was broken down into only three parts: asking for directions, giving directions, and expressing thanks to the direction giver. In contrast, naturally occurring interactions contained more steps such as repeated questions, formulaic expressions, and confirmations. Additionally, spoken DMs such as *okay* and *well*, fillers, pauses, and incomplete sentences were found to be common features of natural conversations. Therefore, Scotton and Bernsten suggest that more attention in L2 instruction should be given to natural conversation data and discourse patterns.

Carter (1998) made use of the Cambridge and Nottingham Corpus of Discourse in

English (CANCODE), a collection of spoken British English, to evaluate dialogues in an EFL textbook published in the United Kingdom. In his analysis, spoken data from a hairdressing salon were compared with similar textbook dialogues, and the results revealed the following linguistic features in the CANCODE data: frequent use of DMs, ellipsis, vague language, and ungrammatical expressions. By contrast, the textbook dialogues consisted almost entirely of complete sentences without contentless utterances such as *er* and *mm*.

Additionally, Cullen and Kuo (2007) investigated the extent to which the features of spoken grammar in natural English discourse were reflected in EFL textbooks. The results of the analysis revealed that the British ELT materials developers attempted to include various features of spoken grammar such as ellipsis and the use of past progressive reporting verbs, DMs, and *less* instead of *fewer* with countable nouns. However, Cullen and Kuo maintain that "this is inadequate for many learners, particularly those for whom the development of oral fluency in informal interactions with native speakers is an important goal" (p. 361).

These studies based on corpus data also indicate that there is a large gap between the language of textbooks and naturally occurring speech data. Moreover, most of the previous studies argue that textbooks should reflect natural usage and discourse patterns in order to provide learners with more opportunities to be exposed to naturally occurring language data.

The problem here is that the textbook analyses based only on the language data of NSs lack the perspective of learners' interlanguage. NS corpora can be a helpful reference for materials writers with regard to various linguistic features such as lexis, grammar, and discourse, but the authentic texts[7] "cannot ensure fully effective EFL learning and teaching, mainly because they contain no indication of the degree of

difficulty of words and structures for learners" (Granger, 1998, p. 7). Despite the importance of learners' language acquisition, relatively little attention has been given to designing textbooks that take into consideration learners' interlanguage data.

In one of the very few studies on the subject, however, Miura (2009) focused on some DMs, such as *I mean*, *I guess*, *just*, and *like*, included in NS corpus-based English textbook series[8] and investigated how the markers were used in the NICT JLE Corpus, which contains Japanese EFL learners' speech data. From the results of frequency analysis and qualitative observation of DMs, she found a gap between the textbooks and Japanese learner speech in the use of some DMs and pointed out that the textbooks may not necessarily be appropriate for learners' target levels. For example, the results revealed that Japanese learners with novice or lower intermediate proficiency levels used certain markers such as *I mean* and *I guess* less frequently than those with upper intermediate or advanced proficiency levels, although the two DMs are regarded as important items in the beginner and high beginner courses in the textbook series.

Accordingly, the previous studies described in this section suggest that the presentation of DMs in textbooks should be analysed and considered from the perspective of learners' DM acquisition.

### 2.4.3 Models of Input Data for Learners

In considering the type and frequency of DMs that should be incorporated into EFL textbooks, the main issue concerns the norm of the language. Some researchers have questioned the textbook analyses based on the language data of NSs on the ground that English is used as a lingua franca in international contexts and that authentic texts may not always provide non-native learners with input data that are important for their language acquisition. In other words, some critics raise issues regarding considering the

language data of NSs to be an ideal model that EFL learners should aim for, partly because approximately one in four of the world's population, including NNSs, use English as a global lingua franca (Crystal, 2003), and partly because NS data do not cover non-native varieties of English and aspects of interlanguage that cause difficulty or errors in learning.

Feng and Byram (2002) claim that teaching materials should include texts produced by speakers who use English as a lingua franca as well as those by NSs in order to develop learners' intercultural perspectives. In short, Feng and Byram call into question native-speaker language norms. Additionally, McDonough and Shaw (2003) argue that some content of teaching materials needs to be localised according to learning contexts.

From the viewpoint of second language acquisition, Kubota (1997) suggests that the language of textbooks should be input data appropriate for EFL learners and contribute to their output in the learning process. He indicates that, if learners are provided with authentic texts, they may not always include essential linguistic items needed for production. Kubota argues, therefore, that modified input data should be used in teaching materials in consideration of learners' proficiency and requirements. Richards (2006) also argues that the use of authentic materials is not always needed in materials design for second language learning given the level of learners' understanding. He points out that authentic texts are generally too difficult for beginner or low proficiency learners because they contain many low frequency words and phrases, and complex sentence structures. Richards stresses, therefore, that it is important for materials writers to provide "learners with a repertoire of well selected vocabulary, sentence patterns and grammar, as well as a stock of communication strategies" (p. 22).

Moreover, many researchers point out that non-standard and grammatically

unacceptable usages are often included in the English NS conversations (e.g., Carter & McCarthy, 2006; Rühlemann, 2009). In designing textbooks, these features of NS English should be also taken into account. However, little pedagogical rationale has been given for making textbook language appropriate for EFL learners.

Accordingly, although it seems difficult to resolve the issue of what the model of input data for EFL learners should be, Granger (1998) suggests that a comparison of NS and NNS data could help in considering input data for learners and designing teaching materials. As reviewed in section 2.2.1, the corpus-based comparison has not provided a conclusive approach to the input issue, but some implications have been discussed in the studies of DMs.

## 2.5 Frameworks for Corpus-based Interlanguage Studies

### 2.5.1 Contrastive Interlanguage Analysis

When language instructors or materials writers consider input data for learners, they can refer to various linguistic features such as lexis, grammar, and discourse through corpora. However, can NS corpora suit their purpose of exploring input data appropriate for non-native learners? The answer to this question may depend on language instructors' or materials writers' views on the language of teaching materials, but NS corpora do not have information enough to know the linguistic features of non-native learners' interlanguage.

In light of the features of non-nativeness, Granger (1996, 1998, 2002) proposed Contrastive Interlanguage Analysis (CIA), which makes two types of comparisons:

1. Comparison of NS and NNS data,
2. Comparison of different NNS data.

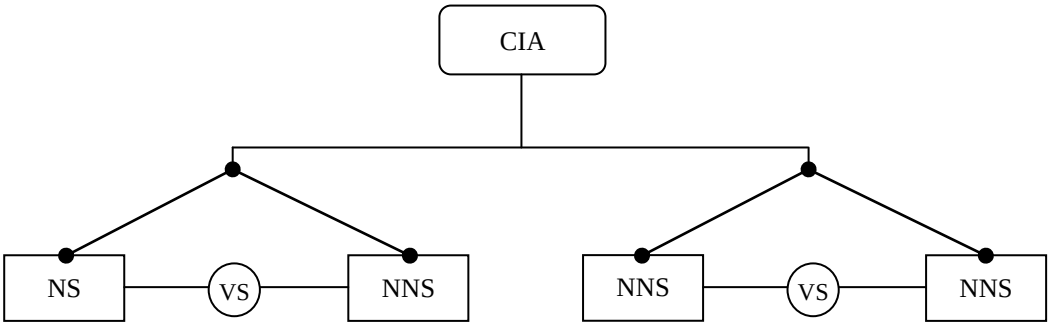



*Figure 2.1*. Contrastive Interlanguage Analysis (Granger, 2002, p. 12).

Figure 2.1 illustrates the framework of CIA. NS/NNS comparisons reveal learners' overuse, underuse, and misuse of features in their speech or writing. NNS/NNS comparisons are made in one of two ways: (a) comparison of interlanguages by speakers of the same L1 in terms of age or proficiency level and (b) comparison of interlanguages by speakers of different mother tongues (e.g., L1 Spanish, L1 French, L1 Swedish). This approach enables us to have a better understanding of interlanguage development as well as to identify some of the difficulties that learners face in the process of language acquisition. Thus, CIA has been used in a great deal of corpus research.

Granger and Tyson (1996) applied the CIA approach to their study on learner connector usage in EFL essay writing. They compared French EFL learners' writing data from the International Corpus of Learner English (ICLE)[9] with those of NSs from the control corpus, the Louvain Corpus of Native Speaker Essays (LOCNESS). The results revealed that while the French learners overused some connectors such as *indeed*, *of course*, and *moreover*, they underused other connectors such as *however*, *though*, and *therefore*. However, contrary to their initial hypothesis, there was a small difference between the French learners and native English speakers in the overall frequencies of

connectors. In addition, Granger and Tyson conducted a further study into the sentence-initial use of *however*, *indeed*, and *so* by English language learners with different L1 backgrounds (i.e., French, Dutch, and Chinese) and NSs of English, using the three subcorpora of ICLE. They found that there were differences in the sentence-initial use not only between the French learners and NSs, but also among the NNS subcorpora. Therefore, their studies based on CIA contribute to the identification of features in French EFL learners' connector usage.

Similarly, Petch-Tyson (1998) used four subcorpora of ICLE and NS corpora to explore the use of first- and second-person pronouns and explicit referential expressions in argumentative essays. The results revealed that L1 Swedish, L1 Finnish, L1 Dutch, and L1 French learners used these linguistic items to make their essays visible to readers more frequently than NSs of English did.

Osborne (2008) also mainly used ICLE as well as a corpus of essays written by L1 French students at the University of Savoy to examine learners' grammatical errors such as omission of third person *–s*, inappropriate adverb placement, and plural use of mass nouns. The results showed that non-native English learners were often confused as to the grammar in the following:

(10) Drugs are an issue which arouse strong feelings.

(Osborne, 2008, p. 81)

In example (10), the learner seems to use *arouse* instead of *arouses* in the relative pronoun clause because the head word is plural. Osborne identified the blending effect by using the international learner corpora.

On the other hand, learners' spoken data have also been analysed in the

framework of CIA. A number of findings have been reported which suggest that the Louvain International Database of Spoken English Interlanguage (LINDSEI; Gilquin, De Cock, & Granger, 2010)[10] can help researchers examine learner language. Kaneko (2003) used three subcorpora of LINDSEI (i.e., L1 Japanese, L1 Chinese, and L1 French subcorpora) to investigate NNSs' use of emotional expressions such as *angry*, *surprised*, *nervous*, and *sad*. The results revealed that the Japanese learners preferred *surprised*, while they used *anger* and *grief* expressions less frequently than the other non-native English learners. Additionally, De Cock (2011) compared LINDSEI Chinese, French, and German subcorpora with the control NS corpus, the Louvain Corpus of Native English Conversation (LOCNEC) in terms of the use of positive and negative evaluative adjectives such as *beautiful*, *pretty*, *angry*, and *ugly*. She found that NNSs and NSs differed in their use of phrase patterns with these adjectives, although the patterns used by German speaking learners were closer to those used by NSs.

The previous studies using learner corpora such as ICLE and LINDSEI indicate that CIA can provide researchers with valuable information on the linguistic features of learner speech. By adding learners' interlanguage data to the language data of NSs, it is possible to consider whether or not input data for learners are appropriate. In other words, CIA could help in improving materials for EFL learners if materials designers not only rely on their intuition, but also place emphasis on linguistic information retrieved from NS and learner corpora (Granger, 1998).

However, when a learner corpus is compared with a NS corpus within the framework of CIA, terms such as overuse, underuse, and misuse should be interpreted in a cautious manner. Leech (1998) insists that they should be interpreted as not normative judgments but just relative differences between two databases and that a finding in CIA does not always contribute to the development of teaching materials. In

short, he warns against over-interpretation of the results of comparative analysis.

In response to Leech's warning, Granger (2009) stresses that features of learner language found by the CIA approach should be sifted through learner needs, teaching objectives and teachability. Figure 2.2 shows the path towards pedagogical implications from learner corpus research.

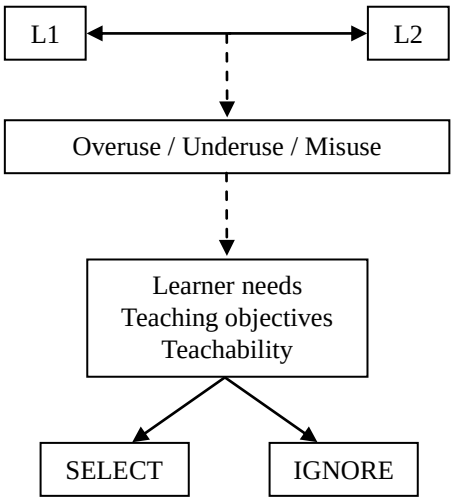



*Figure 2.2.* From learner corpus analysis to language teaching (Granger, 2009, p. 23).

Accordingly, CIA can find quantitative differences between NS and NNS data (i.e., L1 and L2), but, as Granger (2009) points out, it is important that they be interpreted in view of variables in the teaching and learning context.

### 2.5.2 Integrated Contrastive Model

As mentioned above, CIA helps researchers to know the linguistic features of learners' interlanguage. However, Granger (1996, 1998) indicates that the results obtained from CIA do not explain why learners overuse or underuse a linguistic item, although they can lead to a hypothesis. For example, CIA cannot confirm a hypothesis

that a certain item is underused by learners under the influence of their L1. Thus, Granger stresses that contrastive analysis (CA) between L1 and L2 should be performed as well.

In order to identify evidence supporting hypotheses in CIA, therefore, Granger (1996) proposed the Integrated Contrastive Model. The model provides a framework for integrating CA data into CIA data (see Figure 2.3).

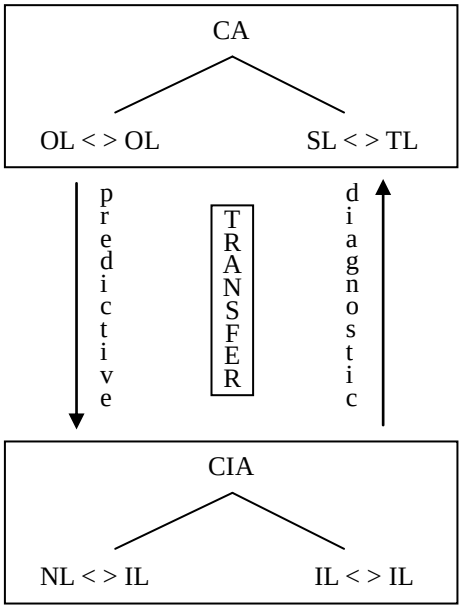



*Figure 2.3.* Integrated Contrastive Model (Granger, 1996, p. 47).

In this model, four types of comparisons are combined as follows:

1. Comparison of different original language data (i.e, original language A vs. original language B),
2. Comparison of source language and translated language data,
3. Comparison of NL and IL data (i.e., NS vs. NNS data),

4. Comparison of different IL data (i.e., NNS vs. NNS data).

While CA consists of comparisons 1 and 2, CIA consists of comparisons 3 and 4. In the framework, findings in CIA data can be diagnosed by CA data. Additionally, "CA data helps analysts to formulate predictions about interlanguage which can be checked against CIA data" (Granger, 1996, p. 46).

Using the Integrated Contrastive Model, Gilquin (2000/2001) investigated verbal and adjectival causative constructions in English and French. She mainly focused on French EFL learners' use of causative constructions with English *make* in their writing and explored how their interlanguage was influenced by French equivalents of the causative constructions. The results of comparing L1 French subcorpora of ICLE with NS essays from LOCNESS revealed that although French learners did not overuse or underuse verbal causative constructions such as *make me laugh*, they significantly underused adjectival causative constructions such as *make life easier* (i.e., *make* + NP + comparative adjective). In order to diagnose the findings from the CIA data, then, Gilquin performed CA using English-French parallel corpora. The results indicated that French learners may underuse the adjectival causative constructions because the frequency of the French equivalent constructions is low in French translations. In short, her study illustrates the viability and importance of the Integrated Contrastive Model to identify L1 transfer.

However, Gilquin (2000/2001) also points out some limitations of Granger's (1996) Integrated Contrastive Model. Gilquin argues that both predictions and diagnosis using CA and CIA data are only hypothesis, partly because mismatches between L1 and L2 do not always affect learners' L2 performance, and partly because learners' errors may result from various sources such as teaching effects and learners' L2 learning

strategy. Thus, the model was fine-tuned by Gilquin to remedy these limitations (see Figure 2.4).

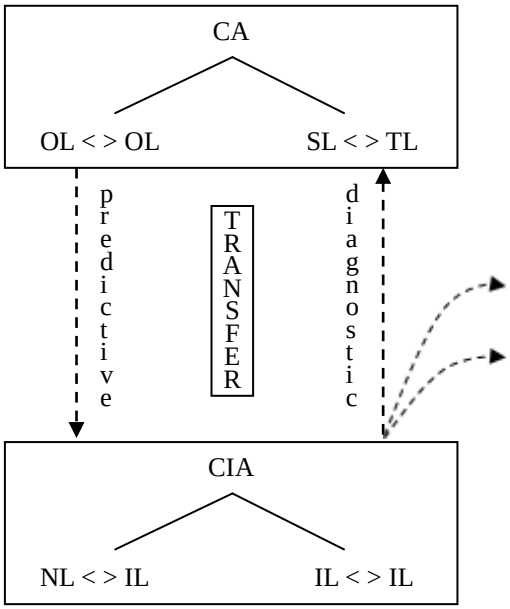



*Figure 2.4.* Integrated Contrastive Model (Gilquin, 2000/2001, p. 100; based on Granger, 1996, p. 47).

In the revised model, the two carved arrows from CIA were added to indicate other sources than L1 transfer. Additionally, all the arrows were changed to broken lined ones because the prediction and diagnosis in the model would be hypotheses.

Some researchers also demonstrated the viability of the Integrated Contrastive Model to identify L2 learners' linguistic features. Paquot (2008) used the framework to investigate L1 influence on the use of exemplifying lexical items in L2 academic writing. She compared the frequencies of each item in five subcorpora of ICLE with those in LOCNESS and found that NNSs with L1 French, Spanish, Dutch, German, and Polish used *for example* and *for instance* more frequently American NSs did. She explained that the NNSs were likely to overuse *for example* partly because the

exemplifying expression can be translated into its direct equivalent in each L1 and partly because it is "typically emphasised in instruction and teaching materials" (p. 110). She also pointed out that the overuse of *for instance* could be attributed to learning effects in the classroom because *for instance* can be seen as a synonym of *for example* in teaching materials. Additionally, Paqout compared the L1 French subcorpus of ICLE with French NS essays and suggested that French EFL learners may overuse *Let's take the example of* because they prefer the equivalent expression *Prenons l'exemple de* in French writing.[11] Thus, Paquot's study using the Integrated Contrastive Model empirically identifies an L1 influence on L2 writing performance. However, her explanation about the influence of teaching materials needs to be verified by examining language data in teaching materials.

Likewise, Hasselgård and Johansson (2011) also applied the Integrated Contrastive Model to their investigation of Norwegian EFL learners' interlanguage. They used the L1 Norwegian subcorpus of ICLE, LOCNESS, and Norwegian NS data to examine the use of *seem* and found that Norwegian EFL learners may overuse expressions with *seem* such as *it seems* under the influence of their L1. However, Hasselgård and Johansson mentioned that there was "a mismatch of genres and/or writer proficiency" (p. 55) between the learner corpora and the Norwegian NS data available in their study. In other words, Hasselgård and Johansson point out the difficulty of controlling homogeneity of text or task types and learner variables in corpora under the framework of the Integrated Contrastive Model.

### 2.5.3 Tono's Multiple Comparison Approach

The Integrated Contrastive Model offers a framework to investigate features of learners' L2 acquisition, including L1 transfer. However, as was shown in the previous

section, the model does not deal with the influence of teaching materials as input resources. Therefore, Gilquin's (2000/2001) revised model includes two carved arrows which indicate factors affecting L2 acquisition other than L1 transfer (see Figure 2.4). In this respect, Tono (2002, 2004) proposed multiple comparisons of L2 learner corpora, L1 corpora, and the corpus of language used in EFL textbooks as target language (TL). In other words, Tono's model was designed to investigate learners' L2 acquisition in terms of factors such as learner variables (e.g., learners' age or proficiency level), L1 transfer, and L2 input (i.e., the effects of English textbooks).

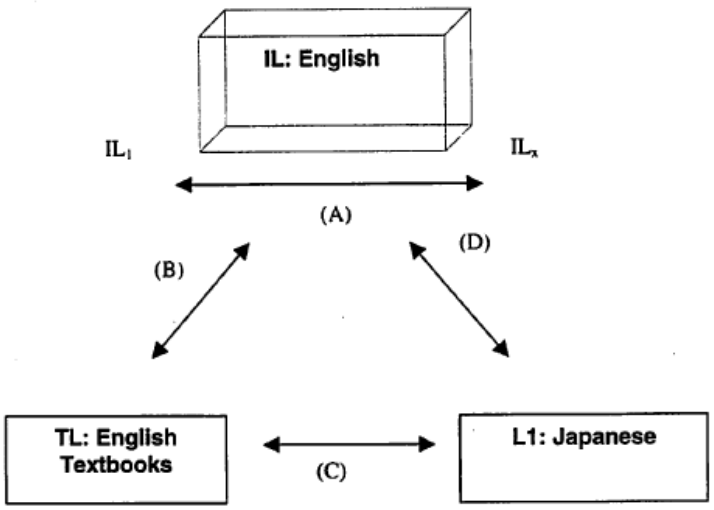



*Figure 2.5.* Multiple comparison of L1, TL, and IL corpora (Tono, 2004, p. 53).

Figure 2.5 illustrates the framework of the multiple comparisons. Comparison (A) is intended to investigate different interlanguage data, which is also included in CIA and the Integrated Contrastive Model. A noteworthy difference between Tono's multiple comparison approach and the Integrated Contrastive Model is that language data from textbooks, instead of NS corpora, are included in the corpus-based comparison model to identify features of learners' interlanguage. In other words, comparison (B) can play a

role in investigating the influence of teaching materials on learners' L2 acquisition, and comparison (C) can "provide significant information on the influence of the source language on the acquisition of the target language" (Tono, 2004, p. 53). Moreover, as in the Integrated Contrastive Model, comparison (D) focuses on the investigation of L1 transfer. Based on the multiple comparison model, Tono designed and developed JEFLL Corpus (Tono, 2007), a collection of Japanese English learner compositions.

In sum, with the development of corpus linguistics, some frameworks have helped researchers design their studies and find features of learners' interlanguage in terms of various linguistic items.

### 2.6 Limitations and Problems in Previous Studies

As was reviewed above, the previous studies have raised some important questions on the following five points: (a) definition and taxonomy of DMs in spoken English, (b) spoken DMs and L2 acquisition, (c) DM use in L1 and L2, (d) DM use in EFL textbooks as input resources, and (e) frameworks for corpus-based interlanguage analysis

First, DMs have been variously defined and conceptualised by researchers. Although, as is summarised in Hellermann and Vergun (2007), DMs mainly serve pragmatic and syntactic functions in discourse, it is difficult to clarify the scope of the definition of DMs in spoken English. Additionally, there have been only a limited number of taxonomies of spoken DMs.

Second, the study of DM use has mostly focused on the difference between native and non-native English speakers in terms of frequency and functions. The findings in the comparative studies have indicated there are significant differences in DM use by NSs and NNSs. However, there is a need for further research to explore why NNSs

under- or overuse certain DMs in their speech and how they acquire DMs in classroom settings, which is different from a native-speaking context. In this respect, some studies examining the L2 acquisition process of DMs have been carried out, but few empirical studies have revealed how Japanese learners can acquire spoken DMs in the Japanese EFL context.

Third, a few studies have addressed the issue of L1 transfer of DM use in L2 acquisition, but there has little study concerning the effect of Japanese on English DM use by Japanese EFL learners. In addition, there has been no detailed investigation of the correspondence relationship between Japanese DMs and English DMs in speech, although Mizutani (2001) and Onodera (2004) compared some Japanese DMs with their English-translated equivalents.

Fourth, although some researchers have pointed out the influence of teaching materials, especially textbooks, on learners' L2 acquisition, only limited research efforts have been made to understand how the presentation of DMs in textbooks reflects learners' speech data. In other words, it is necessary for further research to explore the relationship between the input and output of DMs. Additionally, there is no question that textbooks are primary input resources for language learners in classroom settings, but few studies have considered the type and frequency of DMs that should be incorporated into textbooks for Japanese EFL learners. Moreover, there is little pedagogical rationale for making textbook language as input resources appropriate for EFL learners. Thus, designing the presentation of DMs in textbooks should be considered from the perspective of Japanese learners' DM acquisition and the norm of the language.

With regard to identifying features of L2 acquisition and considering input data for learners, some researchers have proposed and improved corpus-based analysis models. These models have been used to identify features of learners' interlanguage in

linguistic items other than DMs. However, as pointed out in Hasselgård and Johansson (2011), it may be difficult for researchers to select corpus data for the models. This is because researchers need to use multiple corpora with homogeneity of text or task types and learner variables when they apply the corpus-based analysis models to their interlanguage studies including L2 acquisition research of spoken DMs. This is the fifth point to be noted in reviewing the previous literature.

## 2.7 Current Research

### 2.7.1 Purposes of the Current Research

In the light of the limitations and problems of the previous literature, the current research focuses on Japanese EFL learners' DM use and investigates the features of their acquisition of DMs in speech (Purpose 1). Additionally, the current research also examines the language of textbooks as primary input resources and explores the relationship between input and output of DMs in the Japanese EFL context (Purpose 2). With regard to DMs involved in analysis, it mainly draws on Schiffrin's (1987) definition and Fung and Carter's (2007) functional paradigm of DMs because they portray the scope of the definition and taxonomy of spoken DMs.

### 2.7.2 Framework of the Current Research

To attain the two purposes above, the current research adopts the framework based on the Integrated Contrastive Model (Gilquin, 2000/2001; Granger, 1996) and Tono's (2002, 2004) multiple comparison approach. As was reviewed in section 2.5.2, the Integrated Contrastive Model does not deal with the effect of textbooks as input resources. Thus, it should be complemented by the textbook corpus-based comparison in Tono's approach. Figure 2.6 illustrates the framework of the current research.

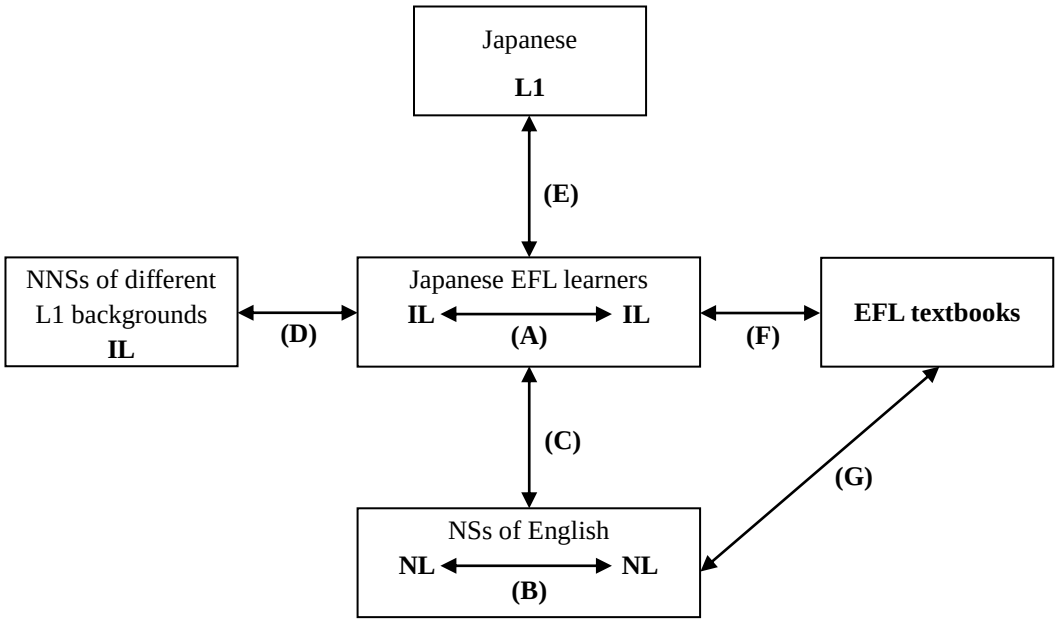



*Figure 2.6.* Framework of the current research.

The framework consists of seven comparisons to attain the two purposes of the current research. In other words, comparisons (A), (B), (C), (D), (E), and (F) are designed to identify features of Japanese EFL learners' acquisition of DMs. Additionally, comparisons (F) and (G) are added to explore the relationship between input and output of DMs, and to consider what type of and how many DMs should be incorporated in textbooks. Comparison (F) is involved with both of the purposes.

Comparison (A) is intended to investigate the features of the development of DMs in Japanese EFL learners. This comparison is based on a part of CIA: a comparison of interlanguage of the same L1 in terms of the speaker's age or proficiency level. Additionally, comparison (B) plays the role of investigating the features of the development of DMs in NSs of English. The current research, therefore, can identify the difference between Japanese learners and NSs of English in the development of DMs. The difference can be explained by comparison (C). Moreover, comparison (D) is also a

part of CIA: a comparison of interlanguages of different mother tongues. In other words, the four comparisons are performed to overcome the second limitation to the previous studies, which was described in section 2.6.

Comparison (E) is designed to overcome the third limitation. That is to say, the issue of L1 transfer in DM use is addressed here. The comparison is based on a CA approach in the Integrated Contrastive Model.

Comparison (F), based on Tono's (2002, 2004) model, is conducted to overcome the fourth limitation of the previous studies. The results of the comparison can help reveal how the language of textbooks can affect Japanese EFL learners' acquisition of DMs. Additionally, the findings can help the researcher consider what DM input is appropriate for Japanese EFL learners.

Comparison (G) also addresses the limitation of the previous studies that relates to textbook language. In other words, the norm of the language should be considered in designing the presentation of DMs in textbooks. Thus, comparison (G) is performed to investigate how different the presentation of DMs in EFL textbooks is from their distribution in NS speech and to examine what norm is adopted in designing EFL textbooks.

Taken together, the current research addresses five studies:

1. DM use by Japanese EFL learners and native English speakers (i.e., comparisons (A), (B), and (C)),
2. DM use by non-native and native English speakers (i.e., comparisons (C) and (D)),[12]
3. L1 transfer in DM use (i.e., comparison (E)),
4. DM use in Japanese EFL textbooks for junior high and high school students (i.e., comparison (F)),

5. DM use in EFL textbooks and materials design (i.e., comparisons (F) and (G)).

Through the results of the five studies, I will illustrate the features of Japanese EFL learners' acquisition of spoken DMs and provide some suggestions for language teaching and materials design. Additionally, I will discuss whether the framework using learner corpora and the language data of teaching materials is useful for identifying features of learners' L2 acquisition.

Finally, it should be mentioned that research questions will be presented at the beginning of each study. The research questions will be set based on the seven comparisons in Figure 2.6 and will be answered to attain the purposes of the current research.

**Endnotes**

1. Some portions of this chapter have been published in Shimada (2011, 2012, 2013, 2014).
2. Biber (1988) mainly used the Lancaster-Oslo-Bergen Corpus of British English (LOB Corpus), a collection of printed texts such as press reportage, editorials, and fictional works, and the London-Lund Corpus of Spoken English, a collection of British English spoken texts including face-to-face conversations, telephone conversations, and spontaneous speeches.
3. Although *sort of* is regarded as a DM in the cognitive category, *kind of* is not included in the category.
4. In Fung and Carter's (2007) interpersonal category, *I think*, *just*, *kind of*, *like*, and *sort of* can be regarded as hedges. Additionally, *absolutely*, *actually*, *exactly*, and *obviously* can be categorised into boosters.

5. Most of the learner corpora used in the previous studies contained fewer than 50,000 words. Due to the small size, some kinds of markers may be infrequent in the corpora. Although corpus size is a controversial issue in corpus linguistics, Pearson (1998) argues that a one million-word corpus is sufficient for special purpose investigations.
6. In Yamamoto (2010), the interactional function marker *so* is used to hold or give the floor in Japanese English learners' speech, and the textual function *so* is used to organise their message in discourse. The two functions found by Yamamoto are included in the structural function defined by Fung and Carter (2007) (see Table 2.1).
7. Authenticity is still an ambiguous and complicated concept in language learning, but, as Lee (1995) observed, it can be divided broadly into two types: text authenticity and learner authenticity. Text authenticity refers to the quality of language used in the text. That is to say, authentic texts are defined as "real language not intended for the non-native learners" (Porter & Roberts, 1981, p. 37). On the other hand, learner authenticity is dependent mainly on learners' interpretation of the texts, so it seems difficult to measure the degree of authenticity.
8. Miura (2009) analysed the *Touchstone* series (e.g., McCarthy, McCarten, & Sandiford, 2006a). This series is a corpus-based, four-level (i.e., beginning, high beginning, low intermediate, and intermediate levels) EFL course.
9. The corpus is composed of argumentative essays written by university students with 16 different L1 backgrounds (i.e., Bulgarian, Chinese, Czech, Dutch, Finnish, French, German, Italian, Japanese, Norwegian, Polish, Russian, Spanish, Swedish, Turkish, and Tswana). As of 2014, ICLE version 2 (Granger, Dagneaux, Meunier & Paquot, 2009), which is available on CD-ROM, contains 3.7 million words.

10. See section 4.2.1 in Chapter 4.
11. Paquot (2008) also indicated that French EFL learners may overuse *Let's take the example of* because they prefer the first person plural imperative in their L1 writing for academic purposes. Thus, she concludes that the learners' preference "does not conform to English academic writing conventions but rather to French academic style" (p. 115).
12. As mentioned in Chapter 1, in Study 2, comparison (C) is also performed to re-examine the findings obtained in Study 1.

# Chapter 3

## Study 1: Discourse Marker Use in Japanese EFL Learners and Native English Speakers[1]

### 3.1 Purpose of Study 1 and Research Questions

As reviewed in the last chapter, there have been several studies examining the difference in DM use between native and non-native English speakers. The findings in the previous studies have indicated that significant differences exist between NSs and NNSs in the use of spoken DMs. However, few studies have been conducted to investigate how Japanese learners acquire and use spoken DMs in the Japanese EFL context, which is different from native-English speaking community.

In the first study, therefore, I will investigate the use of DMs in Japanese EFL learners and NSs of English. In order to detect quantitative differences in the distribution of the pragmatic devices between NNSs and NSs, the present study is based on multiple comparisons of three corpora: a spoken corpus of Japanese EFL learners, and the adult and child speech data of native English speakers.

Based on the CIA approach, the present study addresses the following research questions (RQs). RQ1-1 corresponds to a comparison of interlanguages of the same L1, and RQ1-2 does to a comparison of NS and NNS data.

RQ1-1: To what degree does Japanese learners' proficiency in English have an effect on their use of DMs in conversations?

RQ1-2: How different is the development of DMs in Japanese learners from its development in NSs of English?

Therefore, the purpose of Study 1 is to explore the features of the development of DMs both in Japanese learners and in native English speakers through the answers to the two questions.

## 3.2 Method

### 3.2.1 Databases

In order to make multiple comparisons in the present study, three different databases were used as follows:

1. Japanese EFL learners: the NICT JLE Corpus (Izumi, Uchimoto, & Isahara, 2004),
2. NS adults: the BNC XML Edition,
3. NS children: Child Language Data Exchange System[2] (CHILDES) (MacWhinney, 2000).

The NICT JLE Corpus is a Japanese EFL learners' speech database, which consists of more than 1,200 interviews from an English proficiency test called the Standard Speaking Test.[3] The transcribed data of 1,536,000 words were originally divided into nine proficiency levels, but, in the present study, they were analysed at three proficiency levels: lower (levels 1–3), middle (levels 4–6) and higher (levels 7–9) proficiency groups. Unlike most of the learner corpora used in previous studies, the corpus seems to be of sufficient size for investigating the features of learners' interlanguage.

With regard to NS data, corpora of two different age groups were selected to examine the development of DMs in L1. Adult speech data of 1,051,215 words were taken from the spoken part of the BNC XML Edition. The data, which were extracted

from 166 subcorpora of the part, were restricted to dialogues among British and American speakers aged 25–59. In subcorpus searches, BNCweb (CQP-Edition)[4] hosted by Lancaster University was also used as a supplementary tool. Child speech data were collected from five subcorpora of CHILDES: Belfast, Fletcher, Carterette, HSLLD, and Warren. The 168,135 words in them, uttered by British and American children aged 4–8, were turned into a database for comparative analysis in the study.

### 3.2.2 Procedure

#### 3.2.2.1 Extraction of DMs from Corpora

The present study adopts 57 markers, based on the functional paradigm of DMs that Fung and Carter (2007) proposed (see Table 2.1 in Chapter 2), partly because the items were selected by extracting from both NS and NNS spoken data, and partly because the framework draws on Schiffrin's (1987) definition of DMs. In other words, while Fraser (1999) and Blakemore (2002) limit the scope of the definition of DMs to syntactic elements, Schiffrin's notion allows expressions independence of syntax to be DMs. Examples of them include *I mean*, *you know*, *oh*, and *like*; they are regarded as the most common items in everyday spoken language (Carter & McCarthy, 2006).

In the previous studies (e.g., Kobayashi & Yamada, 2008), Hyland and Tse's (2004) framework of metadiscourse markers was used to investigate Japanese EFL learners' speech data from the NICT JLE Corpus, but the present study did not use the framework because Hyland and Tse designed it based on their analysis of academic writing texts.

In the first procedure, the total word count of each set of data was established and the frequency analysis of the 57 DMs was carried out using WordSmith Tools 5.0, the NICT JLE Corpus Analysis Tool 1.0[5], and CLAN 25. In calculating the frequency,

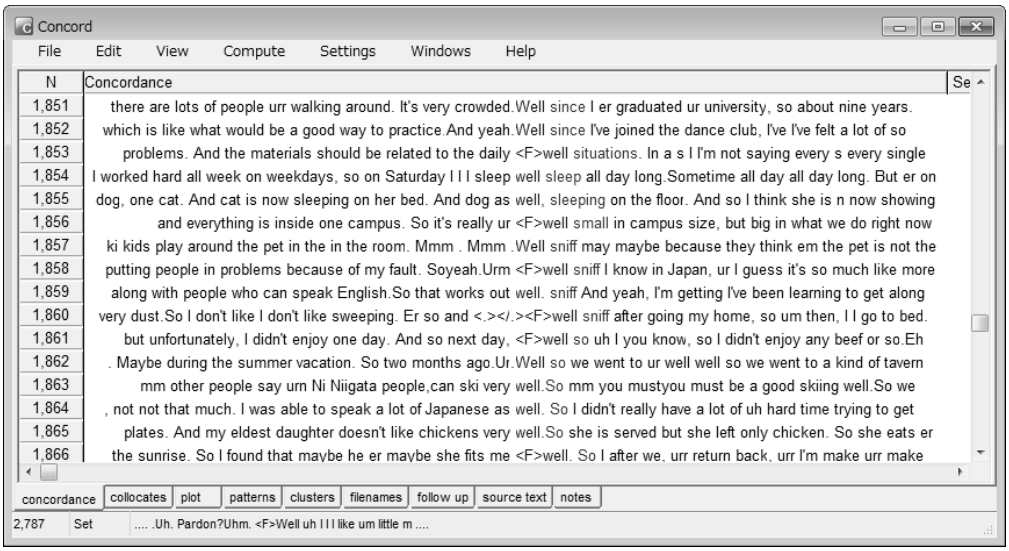



*Figure 3.1.* Concordance results on WordSmith Tools display.

concordance lines were also viewed to confirm whether features were in fact DMs (see Figure 3.1). Some examples are listed below:

The use of DMs:

Erm *well* first of all er I'll I'll come back to that.

(The BNC XML Edition, JA4 91)

I have a little bit cold, *so* I have a stomachache.

(The NICT JLE Corpus, 00374.stt)

The use of non-DMs:

Because I couldn't sleep *well* yesterday.

(The NICT JLE Corpus, 00048.stt)

I watch it at five o'clock every time and it's *so* good.

(CHILDES, Carterette, third.cha)

The categorization was carried out by the author. In order to test the reliability of the coding, a post-hoc intra-coder reliability check was conducted based on Müller (2004) at an interval of about three years. Despite the long interval, the simple agreement rate of the coding of *like*, *so*, and *well* was 96%, 91%, and 96%, respectively.[6] Thus, the reliability of the coding process is considered high.

Next, DMs occurring in each corpus were sorted into the four discourse functions that Fung and Carter (2007) proposed. Additionally, they were assigned one of two organisers: micro- and macro-markers according to the definitions of Chaudron and Richards (1986) and Nattinger and DeCarrico (1992).

#### 3.2.2.2 Statistical Data Analysis

Lastly, the present study performed several statistical data analyses[7] using the frequency of DMs. In order to provide an answer to RQ1-1, the frequency of each marker in the NICT JLE Corpus was calculated at each proficiency level, and was then standardized as percentage points and frequency per 10,000 words. The chi-square test was also conducted for the purpose of finding significant differences in the distribution of DMs among the levels of learner proficiency. The chi-square value was calculated using raw frequencies. Furthermore, the Kruskal-Wallis test and Steel-Dwass test were performed to compare the frequency of DMs by category among the three proficiency groups. However, it was quite difficult to classify some DMs, such as *and*, *so*, and *well*, with multiple discourse functions by category. Therefore, although there was a methodological problem, the statistical tests were conducted using all tokens of each DM.[8]

As for RQ1-2, the first procedure was identical to that performed for RQ1-1. The standardized frequency of DMs in the BNC and CHILDES was computed in order to

compare the results of RQ1-1. Following the calculation, the study made a multi-comparison of the frequency of DMs among three corpora: the NICT JLE Corpus, the BNC, and CHILDES. Additionally, the log-likelihood ratio[9] and chi-square value were used to detect significant differences in the use of DMs among the three corpora.

In addition, two-way between-subjects ANOVAs, with mother tongues (i.e., Japanese and English) and levels of English language development (i.e., higher and lower) as independent variables, were conducted to test for differences in the use of the markers in the categories of discourse functions and micro-/macro-markers. Figure 3.2 graphically presents the design of the comparative analysis conducted in the present study. In this design, a lower proficiency group of Japanese learners is considered on the same level as NS children. Likewise, a higher proficiency group of Japanese learners is paralleled with NS adults.

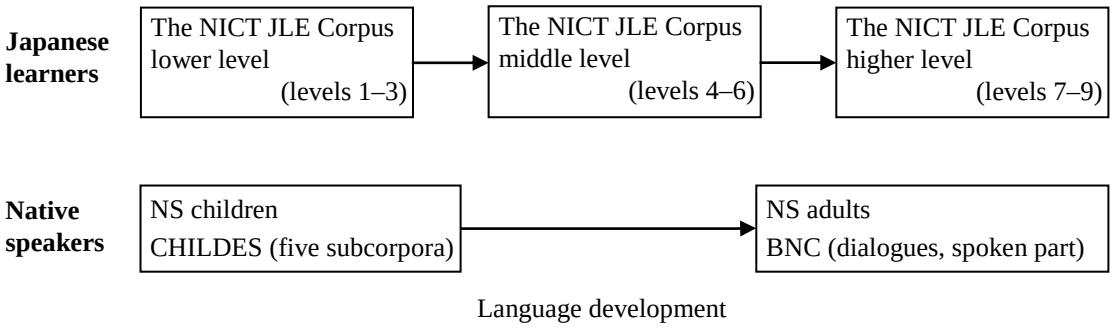



*Figure 3.2.* Design of the comparative analysis in terms of language development.

Although the NICT JLE Corpus has information on each learner's age, the present study did not apply the variable to the comparative analysis. That was partly because it was difficult to classify the speech data into two or three groups according to the learner's age, and partly because many adult learners were not at a higher proficiency

level while some younger learners were categorised into the higher level groups (i.e., levels 7–9).

### 3.3 Results and Discussion

#### 3.3.1 RQ1-1: Effect of Japanese Learners' Proficiency in English on Their Use of DMs

Prior to data analysis for RQ1-1, lexical statistical features of Japanese EFL learners' speech data were described. Table 3.1 lists the number of words (i.e., tokens), different words (i.e., types), and the standardized type-token ratio (TTR) in the NICT JLE Corpus. The speech data were classified into three proficiency levels. The standardized TTR increased, though only gradually, as Japanese learners' proficiency in English improved. Thus, the rise in the ratio indicates the increase in lexical richness.

Table 3.1
*Lexical Statistical Features of Japanese EFL Learners' Speech Data*

| | Total | Lower | Middle | Higher |
|---|---|---|---|---|
| Subjects | 1,281 | 260 | 848 | 173 |
| Tokens | 1,536,000 | 208,245 | 1,045,395 | 282,360 |
| Types | 19,099 | 6,613 | 14,835 | 7,531 |
| Standardized TTR | 28.84 | 27.21 | 28.50 | 30.80 |

*Note.* The total number of words was counted including fillers such as *ah*, *er*, *oh*, and *uh-huh*. Standardized TTR was computed for every 1,000 words.

Next, Table 3.2 shows the frequency of each marker in the NICT JLE Corpus at each proficiency level. Based on a frequency analysis of DMs in the overall data, chi-square tests were performed on 25 items with a more than 0.01 per cent distribution rate. With exceptions such as *OK/okay*, *I see*, and *first*, there were significant differences in the frequency of DMs among the levels of learner proficiency.

Table 3.2

*Frequency of DMs in the NICT JLE Corpus at Each Level of Learner Proficiency*

| | | Frequency per 10,000 words | | | | |
|---|---|---|---|---|---|---|
| DMs | Categories | Lower | Middle | Higher | Total (%) | Chi-square |
| *and* | Ref/Str/Cog/Micro | 203.65 | 248.56 | 293.95 | 2.508 | 406.459*** |
| *so* | Ref/Str/Micro | 56.66 | 133.34 | 127.99 | 1.220 | 857.839*** |
| *yes* | IP/Micro | 125.91 | 93.75 | 50.93 | 0.902 | 798.679*** |
| *yeah* | IP/Str/Micro | 65.93 | 77.21 | 80.04 | 0.762 | 35.938*** |
| *but* | Ref/Micro | 35.39 | 68.98 | 103.45 | 0.708 | 804.935*** |
| *OK/okay* | IP/Str/Micro | 44.90 | 47.52 | 48.20 | 0.473 | 3.147 |
| *I think* | IP/Cog/Micro | 15.56 | 43.51 | 67.29 | 0.441 | 733.252*** |
| *or* | Ref/Micro | 14.12 | 34.41 | 41.29 | 0.329 | 291.748*** |
| *oh* | IP/Micro | 37.50 | 30.23 | 28.62 | 0.309 | 35.775*** |
| *because/'cause* | Ref/Micro | 11.96 | 31.25 | 42.14 | 0.306 | 361.615*** |
| *you know* | IP/Cog/Micro | 1.54 | 14.20 | 55.11 | 0.200 | 2275.644** |
| *well* | IP/Str/Cog/Micro | 2.98 | 8.33 | 51.88 | 0.156 | 2951.979** |
| *just* | IP/Micro | 3.41 | 11.97 | 37.93 | 0.156 | 1192.806** |
| *then* | Str/Macro | 5.76 | 13.46 | 29.75 | 0.154 | 529.161*** |
| *really* | IP/Micro | 0.96 | 6.66 | 42.46 | 0.125 | 2545.122** |
| *actually* | IP/Micro | 0.43 | 5.90 | 16.79 | 0.072 | 521.336*** |
| *I mean* | Cog/Micro | 1.34 | 4.71 | 19.48 | 0.070 | 806.269*** |
| *how about* | Str/Macro | 4.61 | 7.21 | 1.63 | 0.058 | 125.065*** |
| *I see* | IP/Cog/Micro | 3.17 | 4.46 | 2.27 | 0.039 | 3.933 |
| *right/alright* | IP/Str/Micro | 1.49 | 2.61 | 8.32 | 0.035 | 234.756*** |
| *sure* | IP/Micro | 0.67 | 1.61 | 4.29 | 0.020 | 101.509*** |
| *first* | Str/Macro | 2.11 | 1.55 | 1.66 | 0.016 | 3.352 |
| *kind of* | IP/Micro | 0.10 | 1.68 | 10.13 | 0.015 | 593.295*** |
| *finally* | Str/Macro | 0.58 | 1.58 | 1.88 | 0.015 | 14.978** |
| *like* | IP/Cog/Micro | 0.10 | 0.85 | 4.07 | 0.013 | 199.901*** |

*Note.* IP = interpersonal; Ref = referential; Str = structural; Cog = cognitive. **$p < .01$. ***$p < .001$.

Additionally, the frequency of several items including *but*, *you know*, *well*, and *I mean* increased as learners' language proficiency improved, while the frequency of using *yes* and *oh* decreased.

In addition to the above analysis, the Kruskal-Wallis test was performed to compare the frequency of 25 items by category among the three proficiency groups. Since some markers have multiple functions, they were included in each category. With regard to the four discourse functions, the test indicated a significant difference ($H$ (2) = .028, $p < .05$) in the use of DMs in the interpersonal category among the levels of proficiency. There was also a significant difference ($H$ (2) = .026, $p < .05$) in the micro-marker category. Non-parametric post-hoc comparisons by the Steel-Dwass test revealed that the higher proficiency group used DMs in the interpersonal and micro-marker categories more frequently than the lower group ($p < .05$). However, there was no significant difference in the distribution of DMs in other categories.

To sum up, these results suggest that the diversity and quantity of DMs in speech grow as Japanese EFL learners' proficiency improves and that the frequency in use increases with a few exceptions. In particular, learner proficiency has an effect on the frequency of DMs in the interpersonal category. Thus, the results imply that many lexical devices for interaction are used for higher level learners to facilitate effective communication.

### 3.3.2 RQ1-2: Differences in the Development of DMs Between Japanese EFL Learners and NSs of English

In order to answer RQ1-2, multiple comparisons based on the results of RQ1-1 were made among three databases: the NICT JLE Corpus, CHILDES, and the BNC. If the occurrence rate of DMs was 0.01 per cent or below in each database, the items were not included in the analysis.

#### 3.3.2.1 Comparison Between NS Children and NS Adults

In section 3.3.1, the use of DMs in speech data of Japanese EFL learners was examined at three proficiency levels, and some features of the development of DMs were revealed. This section, in turn, investigates the development of DMs in L1 by comparison of two different age group corpora: CHILDES and the BNC.

Table 3.3 provides the statistics of DMs used in CHILDES and the BNC. In this comparative analysis, the log-likelihood ratio as well as the chi-square value was applied to the two databases to clarify the differences between them. If the log-likelihood ratio is represented as positive numbers, native children used the marker more frequently than native adults. On the contrary, if the ratio is negative, native adults used the item more frequently.

The results obtained from chi-square tests revealed that there were significant differences in the frequency of most markers between the two corpora. Additionally, the log-likelihood ratio indicated that native adults often used interpersonal or cognitive function markers such as *right/alright, I mean*, and *sort of*, while native children preferred relatively simple types of DMs such as *and*, *yeah*, *then*, and *oh*. Thus, with increasing age, NSs are considered to develop communication strategies, including showing responses, modifying remarks, and planning what to say by using DMs.

#### 3.3.2.2 Comparison Between Japanese EFL Learners and NS Children

In the previous sections of this chapter, the use of DMs both in L1 and in L2 was individually investigated in terms of language development. The remainder of the chapter is devoted to the results and discussion of comparative analysis between NNS and NS spoken data. As in Romero-Trillo's (2002) comparative analysis, NS spoken data of children as well as adults are used in the present study.

Table 3.3

*Comparisons of DMs in CHILDES and the BNC*

| DMs | Frequency per 10,000 words | | LLR | Chi-square value |
|---|---|---|---|---|
| | CHILDES | BNC | | |
| *and* | 374.04 | 236.65 | 957.761 | 1098.634*** |
| *yeah* | 101.64 | 48.83 | 603.587 | 724.653*** |
| *then* | 63.88 | 29.29 | 421.838 | 510.862*** |
| *because/'cause* | 56.32 | 25.72 | 375.259 | 454.691*** |
| *oh* | 39.73 | 19.60 | 221.360 | 263.221*** |
| *see* | 2.44 | 0.05 | 132.323 | 219.660*** |
| *first* | 1.49 | 0.04 | 76.983 | 127.941*** |
| *like* | 19.86 | 15.36 | 17.266 | 18.408*** |
| *finally* | 1.01 | 0.49 | 5.747 | 6.832* |
| *kind of* | 2.74 | 1.90 | 4.568 | 4.990* |
| *what about* | 3.45 | 2.74 | 2.434 | 2.575 |
| *yes* | 42.82 | 40.73 | 1.523 | 1.548 |
| *well* | 30.04 | 34.37 | −8.328 | 8.091** |
| *how about* | 0.71 | 1.56 | −8.601 | 7.195** |
| *just* | 34.50 | 40.48 | −13.587 | 13.128*** |
| *now* | 10.41 | 14.17 | −16.199 | 15.074*** |
| *really* | 8.68 | 13.18 | −25.794 | 23.405*** |
| *anyway* | 1.37 | 4.06 | −36.253 | 28.517*** |
| *absolutely* | 0.48 | 2.52 | −38.094 | 27.084*** |
| *but* | 47.16 | 59.46 | −40.063 | 38.128*** |
| *you see* | 3.45 | 7.80 | −45.884 | 47.733*** |
| *exactly* | 0.36 | 2.66 | −48.620 | 32.891*** |
| *so* | 15.17 | 25.62 | −74.031 | 65.651*** |
| *I see* | 0.12 | 3.84 | −102.566 | 62.499*** |
| *I think* | 34.73 | 53.88 | −114.992 | 104.236*** |
| *you know* | 24.74 | 48.31 | −208.467 | 179.440*** |
| *OK/okay* | 9.75 | 27.48 | −229.053 | 182.411*** |
| *or* | 16.77 | 39.41 | −248.468 | 205.467*** |
| *actually* | 0.95 | 14.95 | −350.918 | 234.797*** |
| *sort of* | 1.07 | 16.13 | −374.768 | 234.305*** |
| *I mean* | 9.99 | 37.67 | −436.019 | 329.136*** |
| *right/alright* | 7.61 | 42.47 | −662.984 | 469.296*** |

*Note.* If the occurrence rate of DMs was zero per cent in either corpus, they were excluded from this analysis due to the impossibility of computing the log-likelihood ratio (LLR). *$p < .05$. **$p < .01$. ***$p < .001$.

Table 3.4

*Comparisons of DMs in the NICT JLE Corpus and CHILDES*

| | Frequency per 10,000 words | | | |
|---|---|---|---|---|
| DMs | NICT-JLE | CHILDES | LLR | Chi-square value |
| *so* | 121.96 | 15.17 | 2368.190 | 1568.642*** |
| *OK/okay* | 47.29 | 9.75 | 693.710 | 492.048*** |
| *yes* | 90.24 | 42.82 | 479.068 | 401.683*** |
| *or* | 32.92 | 16.77 | 149.356 | 126.575*** |
| *but* | 70.76 | 47.16 | 410.589 | 124.160*** |
| *actually* | 7.15 | 0.95 | 134.855 | 104.701*** |
| *how about* | 5.83 | 0.71 | 114.058 | 74.588*** |
| *I see* | 3.88 | 0.12 | 106.289 | 61.118*** |
| *sure* | 1.97 | 0.30 | 34.984 | 23.537*** |
| *I think* | 44.09 | 34.73 | 32.704 | 30.855*** |
| *really* | 12.47 | 8.68 | 19.703 | 17.963*** |
| *finally* | 1.50 | 1.01 | 2.736 | 2.473 |
| *kind of* | 3.02 | 2.74 | 0.422 | 0.411 |
| *first* | 1.65 | 1.49 | 0.245 | 0.239 |
| *anyway* | 0.80 | 1.37 | −4.909 | 5.691* |
| *sort of* | 0.57 | 1.07 | −5.075 | 6.034* |
| *you know* | 20.00 | 24.74 | −15.715 | 16.684*** |
| *I mean* | 6.97 | 9.99 | −17.303 | 19.113*** |
| *oh* | 30.92 | 39.73 | −34.577 | 37.138*** |
| *right/alright* | 3.51 | 7.61 | −52.597 | 65.118*** |
| *yeah* | 76.20 | 101.64 | −115.251 | 125.665*** |
| *what about* | 0.38 | 3.45 | −119.901 | 210.119*** |
| *see* | 0.10 | 2.44 | −125.568 | 246.900*** |
| *well* | 15.61 | 30.04 | −154.403 | 185.387*** |
| *you see* | 0.21 | 3.45 | −158.162 | 301.482*** |
| *just* | 15.58 | 34.50 | −249.302 | 311.388*** |
| *because/'cause* | 30.64 | 56.32 | −254.698 | 302.412*** |
| *now* | 0.67 | 10.41 | −465.486 | 881.030*** |
| *and* | 250.81 | 374.04 | −784.241 | 898.769*** |
| *like* | 1.34 | 19.86 | −871.948 | 1641.555*** |
| *then* | 15.41 | 63.88 | −1194.371 | 1766.592*** |

*Note.* *$p < .05$. ***$p < .001$.

Table 3.4 shows comparisons of DMs used in the NICT JLE Corpus and in the subcorpora of CHILDES. The analysis was carried out using the same method as that used in Table 3.3.

The comparisons revealed that Japanese EFL learners overused some DMs such as *but*, *OK/okay*, *or*, *so*, and *yes*. Among the items, it was noticeable that the log-likelihood ratio of *so* was the most positive. As some researchers (e.g., Müller, 2004; Romoro-Trillo, 2002) pointed out, the frequency of the item may be influenced by learners' L1. That is to say, it seems to be easy for many Japanese learners to use the marker because there are some similarities between English *so* and Japanese *so*[10] in usage and pronunciation. For example, some learner data such as (1a) and (2b) indicate that English *so* and Japanese *so* are mixed up in order to continue an utterance.

(1) a. Uum er my room is ee tatami room, *so* de rokujou. Mm. *So*. Mm maa uhm…

(The NICT JLE Corpus, 00446.stt)

b. It's a tough question. Er ah *so* ne ah *so* uu there is a spaghetti.

(The NICT JLE Corpus, 00061.stt)

As with English *so*, Japanese *so* is also used to signal the sequence of discourse segments in the thinking process. Sadanobu (2002) indicates that Japanese *so* has the discourse function of a filler although it also plays a major role in giving back channel feedback in casual conversations.[11] Some examples in italics follow:

(2) a. Raishu no, *so* (nee), kayou ni kite moraimashouka?

[Next week . . . , *so* could you come here next Tuesday?]

b. Sankasha wa kaijyou ni bisshiri de, *so* (nee), hyakunin gurai ita to omoimasu.

[The hall was filled with participants. *So*, I think there were about one hundred.]

(Sadanobu, 2002, pp. 92–93)

Japanese *so* in (2a) is used to buy some time for considering when the speaker is available. Likewise, *so* in (2b) also serves to pause for a moment to think what to say. These examples show that Japanese *so* has something in common with English *so* in discourse management as fillers. Therefore, Japanese English learners may prefer to use *so*, as a referential or structural marker, instead of other DMs, while native children often use *and* and *then* as well as *so*.

#### 3.3.2.3 Comparison Between Japanese EFL Learners and NS Adults

This section examines differences in the use of DMs between Japanese English learners and NS adults. Many previous studies have compared some linguistic features in learners' speech with those in NS spoken data. However, as described in Chapter 2, few studies have attempted to investigate the distribution of DMs in Japanese learners' speech by comparing spoken data of NS adults.

Table 3.5 shows comparisons of DMs used in the NICT JLE Corpus and in the spoken data of the BNC. The analysis was carried out using the same method as that used in Tables 3.3 and 3.4. In the BNC data, the frequency of each of the 34 markers was greater than 0.01 per cent, but 12 of the items were used with a less than 0.01 per cent distribution rate in the NICT JLE Corpus. In short, the results indicate that native adults use a wide variety of DMs in speech.

Additionally, the frequency analysis revealed that native adults often used interpersonal or cognitive function markers such as *you know*, *right/alright*, *I mean*, *well*, and *sort of*. These markers, as Aijmer (2004) points out, have the pragmatic function of confirming whether the hearer is following the speaker's utterances, or of denoting planning, preparation, or revision in the thinking process. Thus, native adults are considered to use interpersonal or cognitive markers in order to manage conversational interaction. The examples are as follows:

(3) a. The er, things in erm … *you know*, *I mean*, that was horrendous!

(The BNC XML Edition, F7E 429)

b. Er, we're looking at, to whether … the branch and the manager, *sort of* be … take a more active role including the … the branch sector.

(The BNC XML Edition, FLS 520)

On the other hand, markers such as *OK/okay*, *so*, *yeah*, and *yes* were used more frequently by Japanese EFL learners than by NS adults. The comparisons are similar to those between Japanese learners and NS children.

Table 3.5

*Comparisons of DMs in the NICT JLE Corpus and the BNC*

| DMs | Frequency per 10,000 words | | LLR | Chi-square value |
|---|---|---|---|---|
| | NICT-JLE | BNC | | |
| *so* | 121.96 | 25.62 | 8182.942 | 7053.124*** |
| *yes* | 90.24 | 40.73 | 2339.126 | 2196.667*** |
| *yeah* | 76.20 | 48.83 | 743.722 | 723.040*** |
| *OK/okay* | 47.29 | 27.48 | 652.110 | 626.512*** |
| *how about* | 5.83 | 1.56 | 316.452 | 278.271*** |
| *oh* | 30.92 | 19.60 | 315.021 | 304.761*** |
| *first* | 1.65 | 0.04 | 229.794 | 162.686*** |
| *but* | 70.76 | 59.46 | 122.036 | 121.384*** |
| *sure* | 1.97 | 0.52 | 107.906 | 94.769*** |
| *finally* | 1.50 | 0.49 | 63.920 | 57.577*** |
| *because/'cause* | 30.64 | 25.72 | 53.287 | 52.799*** |
| *and* | 250.81 | 236.65 | 51.283 | 52.374*** |
| *kind of* | 3.02 | 1.90 | 31.535 | 30.417*** |
| *I see* | 3.88 | 3.84 | 0.022 | 0.022 |
| *really* | 12.47 | 13.18 | −2.508 | 2.520 |
| *however* | 0.46 | 1.31 | −55.867 | 57.024*** |
| *basically* | 0.61 | 1.73 | −72.910 | 74.433*** |
| *or* | 32.92 | 39.41 | −73.154 | 74.150*** |
| *I think* | 44.09 | 53.88 | −123.137 | 125.093*** |
| *exactly* | 0.49 | 2.66 | −216.468 | 215.255*** |
| *what about* | 0.38 | 2.74 | −266.389 | 260.408*** |
| *anyway* | 0.80 | 4.06 | −312.788 | 312.294*** |
| *in other words* | 0.06 | 2.14 | −338.376 | 299.074*** |
| *actually* | 7.15 | 14.95 | −359.049 | 368.091*** |
| *absolutely* | 0.05 | 2.52 | −413.435 | 360.530*** |
| *then* | 15.41 | 29.29 | −558.015 | 572.380*** |
| *obviously* | 0.09 | 4.26 | −696.995 | 608.013*** |
| *well* | 15.61 | 34.37 | −922.899 | 947.346*** |
| *you see* | 0.21 | 7.80 | −1237.590 | 1092.741*** |
| *just* | 15.58 | 40.48 | −1472.303 | 1509.462*** |
| *you know* | 20.00 | 48.31 | −1550.895 | 1592.402*** |
| *cos* | 0.31 | 10.37 | −1615.574 | 1435.560*** |
| *like* | 1.34 | 15.36 | −1838.248 | 1744.618*** |
| *now* | 0.67 | 14.17 | −2027.949 | 1849.385*** |

(continued)

Table 3.5 (continued)

| DMs | Frequency per 10,000 words | | LLR | Chi-square value |
|---|---|---|---|---|
| | NICT-JLE | BNC | | |
| *sort of* | 0.57 | 16.13 | −2445.503 | 7053.124*** |
| *I mean* | 6.97 | 37.67 | −3042.346 | 3032.253*** |
| *right/alright* | 3.51 | 42.47 | −5183.350 | 4907.297*** |

*Note.* ***$p < .001$.

### 3.3.2.4 Comparative Analysis in Terms of Language Development

In this analysis, two-way between-subjects ANOVAs were performed to test for differences in the use of DMs in each category. To minimize the effect of a certain item on the overall results, *and* and *well*, which have three discourse functions, were not included in the analysis. As for subjects in the ANOVAs, a lower proficiency group of Japanese learners is considered on the same level as NS children, while a higher proficiency group of Japanese learners is treated in the same way as NS adults.

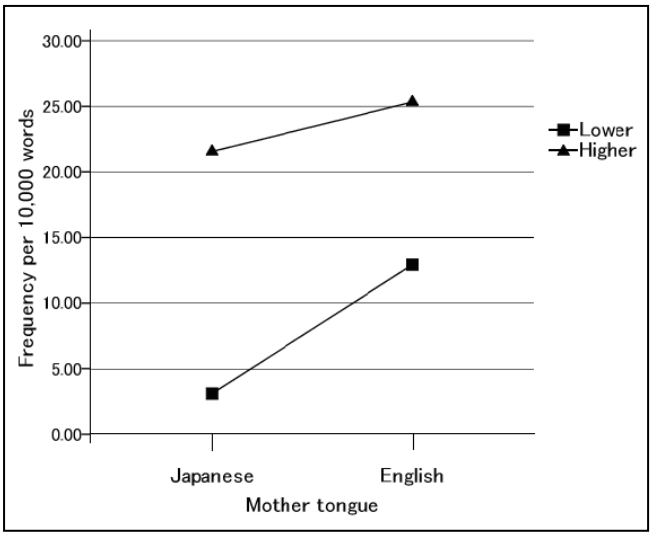



*Figure 3.3* Frequency of DMs in the cognitive category.

Results of the ANOVAs revealed that there was no significant interaction between mother tongue and levels of English language development in any category of discourse functions or micro-/macro-markers. However, concerning the cognitive function category, the ANOVA showed significant main effects for language development as presented in Table 3.6. Additionally, Figure 3.3 shows the

Table 3.6

*Results of a Two-way ANOVA for Mother Tongue and Language Development (in the Cognitive Category)*

| Source | *SS* | *df* | *MS* | *F* | *p* | $\eta^2$ |
|---|---|---|---|---|---|---|
| Mother Tongue | 324.072 | 1 | 324.072 | .891 | .355 | .04 |
| Language Development | 1666.249 | 1 | 1666.249 | 4.580 | .043* | .19 |
| MT × LD | 64.031 | 1 | 64.031 | .176 | .679 | .01 |
| Error | 8731.654 | 24 | 363.819 | | | |

*Note.* $*p < .05$. MT = mother tongue. LD = levels of English language development.

frequency of DMs in the cognitive category. It is inferred from the result that lower level speakers have difficulty in acquiring cognitive markers such as *I mean*, *sort of*, and *you know*, regardless of their mother tongue.

### 3.4 Summary of Study 1

The present study quantitatively investigated the use of DMs in Japanese EFL learners and NSs of English. The results revealed that there were differences and similarities in the distribution between the two language groups. Additionally, some qualitative observations on their speech data illustrated some of the features of DM used both by Japanese learners and by NSs.

In regard to the answer to RQ1-1, the frequency analysis of DMs in the NICT JLE Corpus revealed that the levels of Japanese EFL learners' proficiency had a significant effect on their DM use in conversations. That is to say, as learner's proficiency improved, a greater variety of DMs were used, and the frequency increased, with a few exceptions such as *I see*, *first*, *oh*, and *yes*. In addition to the analysis, statistical tests revealed a significant difference in the distribution of interpersonal function markers among the levels of Japanese English learners' proficiency.

Therefore, the results indicate that although lower level learners make a limited use of DMs in conversation, higher level learners use a larger of discourse devices as part of a repertoire of communication strategies. In particular, interpersonal markers such as *really*, *right/alright*, and *you know* are often used for higher level learners to keep the conversational ball rolling.

Regarding the answer to RQ1-2, multiple comparisons of the NICT JLE Corpus, CHILDES, and the BNC revealed that the distribution and development of DMs used by Japanese learners differed substantially from those used by NSs of English. Altogether, learners used DMs less frequently than native children and adults, but they overused some markers such as *OK/okay*, *so*, and *yes*. However, the factor of mother tongue seemed to have an insignificant effect on acquiring cognitive markers.

Additionally, a qualitative analysis on the marker *so* suggested that the frequency of the item may be influenced by the similarities between English *so* and Japanese *so* in usage and pronunciation. It may be easy for Japanese EFL learners to use *so* rather than other markers such as *then* because of the accessibility.

In summary, results of this study suggest that the frequency of DMs is influenced by the levels of language development: The more proficient speakers become in their linguistic performance, the more frequently they can use a variety of DMs, regardless of their mother tongue. However, there are significant differences in the distribution of DMs among Japanese learners and native children and adults.

Thus, these differences need to be taken into consideration when instructors or materials developers provide learners with DMs as language input. In particular, difficult or infrequent items for learners should be carefully selected to match the learners' interlanguage level.

This study has four limitations. First, there are differences in tasks and situation

of data collection between corpora. For example, some macro-markers such as *let's start* and *let's discuss* were infrequently used in the NICT JLE Corpus. Some tasks in the oral tests are likely to have an effect on the frequency of some markers. However, as pointed out in Hasselgård and Johansson (2011), it seems to be quite difficult to find spoken data of learners collected under the same conditions or tasks as native corpora in the design of the present study.

Second, the present study employed the categorical analysis of micro- and macro-markers, but only four macro-markers were included in the frequency analysis of the NICT JLE Corpus (see Table 3.2). In speech, unlike in writing, learners may not choose to use macro-markers.

Third, although learners' misuse of some items may boost the frequency, the errors were not focused on in the analyses. Further qualitative research, therefore, is necessary to explore the acquisition of DMs by Japanese English learners.

Finally, the present study did not analyse a particular role of multi-function markers such as *so* and *right/alright*. In other words, while learners may prefer a function of each DM, they may not make use of another function in their speech. Therefore, the learners' preference may influence the frequency of multi-function markers. In further chapters, the details of functions of these DMs will be investigated quantitatively and qualitatively.

**Endnotes**

1. An earlier version of this chapter has been published in Shimada (2011).
2. Programs and database of CHILDES are downloadable at http://childes.psy.cmu.edu/.
3. The Standard Speaking Test (SST), developed by American Council on the Teaching of Foreign Languages and ALC Press, is an interview test to measure the oral

proficiency of Japanese EFL learners. SST consists of five stages: (a) warm-up questions, (b) single picture description, (c) role-play with the interviewer, (d) story narration on four or six pictures, and (e) wind-down questions (http://www.alc.co.jp/edusys/sst/e/index.html; Izumi, Uchimoto, & Isahara, 2004).

4. http://corpora.lancs.ac.uk/BNCweb/home.html
5. The software is included in Izumi, Uchimoto, and Isahara (2004).
6. The post-hoc intra-coder reliability check was conducted using the speech data in the NICT JLE Corpus. The coding in the BNC and CHILDES was carried out in the same way as that in the NICT JLE Corpus was.
7. In this dissertation, most of the statistical data analyses were done using IBM SPSS Statistics 19.0 for Windows. With regard to the Steel-Dwass test, I used the web-based statistical analysis program developed by MEPHAS, a research team at Osaka University. The program is available at http://www.gen-info.osaka-u.ac.jp/testdocs/tomocom/s-d.html. Additionally, as for calculating effect sizes, I used Mizumoto's (2009) calculation sheet.
8. In the subsequent chapters, tokens of some multi-function markers, such as *right* and *so*, were counted in randomly sampled speech data and were classified by functional category.
9. The log-likelihood ratio (LLR) has been recently used an alternative to the chi-square test in order to make word-frequency comparisons between two different sized corpora. LLR means the relative differences between the two items: The larger the ratios are regardless of positive or negative number, the greater the differences are. If the log-likelihood ratio for comparing two datasets on 1 degree of freedom is ±3.84 or more, the difference between them is statistically significant at a 5 per cent significance level (Rayson, Berridge, & Francis, 2004). Additionally, it is noted that

LLR can be also calculated using raw frequencies, not standardized ones.

10. Japanese *so* can be spelled as *sou* in some transliteration schemes. However, in the present study, I use *so* instead of *sou* because *so* is used for representing Japanese *sou* as well as English *so* in the NICT JLE Corpus.
11. Sadanobu (2002) defines Japanese *so* in several functions or parts of speech as follows: (a) anaphora, (b) affirmative responses, (c) expressions showing doubt, (d) expressions showing approval, and (e) fillers. Among these usages, (b) and (d) are used to give back channel feedback in casual conversations. According to Sadanobu's definition, (b), (c), (d), and (e) are also categorized as interjections.

# Chapter 4

## Study 2: Discourse Marker Use in Non-native and Native English Speakers[1]

### 4.1 Purpose of Study 2 and Research Questions

Study 1 conducted a corpus-based analysis of English DM use by Japanese learners and NS children and adults. The results revealed that as speakers' proficiency improved, they used many items more frequently, regardless of their L1. However, the quantitative analysis confirmed significant differences in the distributions of DMs between Japanese learners and NSs. One of the notable findings was that Japanese learners overused relatively simple types of DMs such as *OK/okay*, *so*, and *yes*[2].

Most studies on learners' use of spoken DMs have revealed that learners use certain items much more or less frequently than NSs do. However, the difference in frequency between NS and NNS speech data cannot fully explain the features of DM use in interlanguage—that is, researchers have not yet determined whether the differences are due to the specific influences of individual L1 backgrounds, or whether they are common to language learners in general. In order to address the issue, as Granger (2002) argued, it is necessary to construct a comparison of learner languages that incorporate speakers of different L1 backgrounds.

In addition, many comparative studies are based on disparate databases. For example, Study 1 compared three spoken corpora, but there were considerable differences in the ways the data were collected. In that study, the Japanese learner corpus comprised a collection of interviews from a speaking test, while the speech data of NS children and adults were extracted from naturally occurring conversations in daily situations. These different situations may affect how speakers use DMs to facilitate communication, and different types of data collection may generate different results.

The present study explores features of DM use in the speech data of Japanese learners of English. Using the methods of CIA, the study addresses the following research questions (RQs):

RQ2-1: How do usage levels of spoken English DMs by Japanese learners differ from those of NSs of English?

RQ2-2: How do usage levels of spoken English DMs by Japanese learners differ from those of other English language learners with different L1 backgrounds?

RQ2-1 is intended to replicate Study 1 using homogeneous databases. RQ2-2, on the other hand, is designed to investigate the similarities and differences in DM use between Japanese learners and other English language learners of various L1 backgrounds (see Figure 4.1).

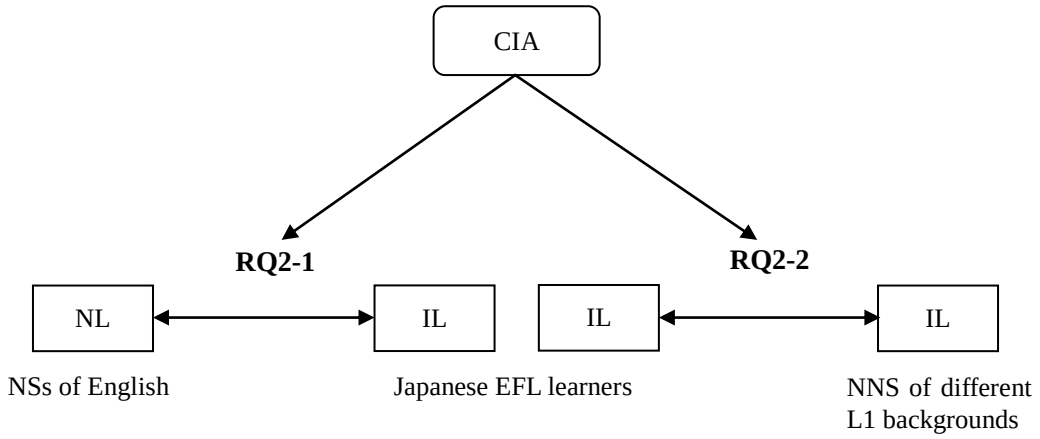



*Figure 4.1.* Design of the present study (based on Granger, 1996, p. 44; 2002, p. 12).

## 4.2 Method

### 4.2.1 Databases

In order to make comparisons based on the CIA approach, the present study used two corpus databases. Data from EFL learners were from the Louvain International Database of Spoken English Interlanguage (LINDSEI; Gilquin, De Cock, & Granger, 2010), and data from native English speakers were from the NICT JLE Corpus (Izumi, Uchimoto, & Isahara, 2004). The former database, LINDSEI, is a spoken corpus consisting of interviews produced by university undergraduates with different L1 backgrounds. All are higher intermediate and advanced learners of English. The spoken corpus consists of 11 subcorpora, classified according to learners' L1, and the data collection was performed using the same procedure for all subcorpora. Each interview lasts about 15 minutes and contains three tasks: (a) warm-up questions on a set topic (e.g., the most impressive country they have visited, their favourite film/play), (b) free/informal discussion with the interviewer, and (c) a picture description. The present study drew on six of the subcorpora, which are characterised in Table 4.1 below.

Table 4.1
*Number of Interviews and Words per Subcorpus*

| L1 subcorpus | Language family | *n* of interviews | *n* of words |
|---|---|---|---|
| Japanese (JP) | Asian | 51 | 37,126 |
| Chinese (CH) | Asian | 53 | 63,542 |
| Dutch (DU) | Germanic | 50 | 79,652 |
| German (GE) | Germanic | 50 | 85,950 |
| French (FR) | Romance | 50 | 91,402 |
| Spanish (SP) | Romance | 50 | 64,804 |
| Total | | 304 | 422,476 |

*Note*. Adapted from *LINDSEI: Louvain international database of spoken English interlanguage* by G. Gilquin, S. De Cock, and S. Granger (Eds.), 2010, p. 25. Louvain-la-Neuve, Belgium: Presses universitaires de Louvain.

Each subcorpus is made up of about 50 interviews, but the number of words in Japanese subcorpus is much lower than that in the other subcorpora.[3]

NS data from the NICT JLE Corpus (i.e., NICT-NS) consist of 20 interviews (94,845 words) produced by American speakers aged 20–24. Each interview lasts about 15 minutes. The interview tasks are also similar to those of LINDSEI, including warm-up questions, a picture description task, and a role-play with the interviewer. Table 4.2 summarises design structure of each database. The present study therefore aims to address gaps in earlier work, ensuring the homogeneity of databases in order to permit an effective comparison of NS and NNS speech.

Table 4.2
*Comparison of Design Structure Between LINDSEI-JP and NICT-NS*

| | Databases | |
|---|---|---|
| Variables | LINDSEI-JP | NICT-NS |
| Data genre | Interview<br>(about 15 minutes) | Interview<br>(about 15 minutes) |
| Participants | University undergraduates<br>(average age: 19.55) | American speakers aged 20–24 |
| Mother tongue | Japanese | English |
| Tasks | (a) warm-up questions on a set topic<br>(b) free/informal discussion with the interviewer<br>(c) description of four pictures | (a) warm-up questions<br>(b) single picture description<br>(c) role-play with the interviewer<br>(d) story narration on four or six pictures<br>(e) wind-down questions |

*Note.* Based on Gilquin, De Cock, and Granger (2010) and Izumi, Uchimoto, and Isahara (2004).

In the previous studies (e.g., De Cock, 2011), LOCNEC was used as the control NS corpus to examine linguistic features of NNSs' spoken discourse in LINDSEI. However,

the present study does not use LOCNEC because the NS corpus has not been released to the public.[4]

### 4.2.2 Procedure

The present study focuses on the 57 DMs listed in Fung and Carter's (2007) functional paradigm, which embraces the features of DMs in spoken English. In the first procedure, the corpus analysis software WordSmith Tools 5.0 was used to obtain frequencies for each of the 57 items. As in Study 1, concordance lines were also viewed to differentiate words used as DMs from those playing other grammatical roles. Some examples are as follows:

The use of DMs:

They are advertising by the week, *so* I found it.

(The NICT JLE Corpus, N_file00006.stt)

… *well* first of all it's her expression she's got this really sour expression.

(LINDSEI-GE050)

The use of non-DMs:

… I … wouldn't be able to come back *so* early.

(LINDSEI-FR006)

… but now I cannot speak English very *well*.

(LINDSEI-JP051)

The categorization was carried out by the author. In order to test the reliability of the coding, as was conducted in Study 1, a post-hoc intra-coder reliability check was conducted at an interval of about two years. Despite the long interval, the reliability of

the coding of *like*, *so*, and *well* was 94%, 99%, and 98%, respectively. Thus, the reliability of the coding process is considered high.

The present study did not classify DMs into the categories of micro- and macro-markers. As mentioned in the last chapter, there seems to be little comparison between micro- and macro-markers because learners do not often use macro-markers in their speech.

Statistical analyses of the frequencies of DMs were conducted to answer RQ2-1 and RQ2-2. The raw frequency of each item was standardized as a frequency per 10,000 words, and then used to calculate the log-likelihood ratio and chi-square value for comparison between corpora of different sizes. In corpus studies, while chi-square tests have been often performed to compare word frequencies across corpora, log-likelihood tests are considered to have higher reliability than other statistical methods when comparing different-sized datasets (Rayson & Garside, 2000). As mentioned in the last chapter, when comparing two datasets with a single degree of freedom, significance is statistically tested by the log-likelihood ratios.[5] If the log-likelihood ratio is ±3.84 or more, a significant difference exists between the two datasets at a 5% significance level (Rayson, Berridge, & Francis, 2004).

In addition to these quantitative analyses, the study included qualitative observations about the context, situation, and discourse function of spoken DMs. These observations serve to complement the quantitative analyses, providing vital details on the functions of DM use in actual learner speech.

## 4.3 Results and Discussion

### 4.3.1 RQ2-1: Comparisons of DM Use Between Japanese EFL Learners and NSs of English

In order to answer RQ2-1, a comparative analysis was conducted using the frequency of DMs in two subsets of speech data: the Japanese subcorpus of LINDSEI (i.e., LINDSEI-JP) and the NS subcorpus of the NICT JLE Corpus (i.e., NICT-NS). Table 4.3 below provides the standardized frequency of each marker, the log-likelihood ratios, and chi-squared values. If the occurrence rate of DMs was 0.01% or below in each database, the items were not included in the analysis.

Chi-square tests revealed that significant differences existed between the two databases in the frequencies of 21 out of 27 DMs with an occurrence rate of more than 0.01%. Additionally, log-likelihood ratios were added to the results obtained with the chi-square tests. If the ratio applied to the two databases was +3.84 or more, the item was considered to be used more frequently in LINDSEI-JP than in NICT-NS. On the other hand, when the ratio was −3.84 or less, the item was considered to be used less frequently in the Japanese learner data. The tests revealed that Japanese learners more frequently used relatively simple markers such as *yes*, *so*, and *I think*, while they used some interpersonal or cognitive markers such as *like*, *really*, *you know*, *kind of*, and *I mean* less frequently than NSs of English. Therefore, the results support those of Study 1 and previous studies (e.g., Hays, 1992; Miura, 2011), in finding that there was a significant discrepancy between Japanese learners and NSs of English in the frequency of DMs.

Table 4.3

*Comparisons of DM Use Between Japanese EFL Learners (LINDSEI-JP) and NSs of English (NICT-NS)*

| DMs | Categories | Frequency per 10,000 words | | LLR | Chi-square value |
|---|---|---|---|---|---|
| | | LINDSEI-JP | NICT-NS | | |
| *yes* | IP | 71.92 | 14.55 | 248.791 | 287.012*** |
| *so* | Ref/Str | 206.86 | 133.38 | 88.213 | 95.000*** |
| *I think* | IP/Cog | 88.35 | 51.66 | 54.020 | 58.292*** |
| *but* | Ref | 145.72 | 101.22 | 44.215 | 46.994*** |
| *now* | Str | 13.47 | 3.58 | 35.907 | 40.969*** |
| *first* | Str | 2.96 | 0.11 | 21.678 | 23.961*** |
| *finally* | Str | 2.96 | 0.74 | 8.470 | 9.684** |
| *yeah* | IP/Str | 86.46 | 72.54 | 6.599 | 6.817** |
| *and* | Ref/Str/Cog | 420.46 | 398.02 | 3.297 | 3.464 |
| *because/'cause* | Ref | 47.68 | 46.29 | 0.109 | 0.111 |
| *I see* | IP/Cog | 1.08 | 1.48 | −0.326 | 0.311 |
| *or* | Ref | 50.10 | 54.09 | −0.811 | 0.806 |
| *exactly* | IP | 2.15 | 3.48 | −1.622 | 1.507 |
| *anyway* | Ref | 1.08 | 2.32 | −2.356 | 2.090 |
| *basically* | IP | 0.27 | 4.32 | −20.173 | 13.780*** |
| *oh* | IP | 7.54 | 21.30 | −34.107 | 29.021*** |
| *then* | Str | 15.35 | 38.91 | −53.065 | 46.000*** |
| *right/alright* | IP/Str | 0.27 | 11.07 | −60.590 | 38.787*** |
| *OK/okay* | IP/Str | 22.90 | 59.25 | −83.548 | 72.304*** |
| *actually* | IP | 4.85 | 27.94 | −86.724 | 66.491*** |
| *I mean* | Cog | 2.15 | 25.73 | −110.554 | 77.784*** |
| *well* | IP/Str/Cog | 5.39 | 37.32 | −128.558 | 96.303*** |
| *kind of* | IP | 5.39 | 41.12 | −148.569 | 110.000*** |
| *just* | IP | 10.77 | 77.39 | −271.486 | 203.074*** |
| *you know* | IP/Cog | 4.31 | 64.32 | −294.673 | 203.503*** |
| *really* | IP | 8.62 | 78.13 | −304.263 | 221.379*** |
| *like* | IP/Cog | 28.82 | 140.65 | −390.444 | 308.967*** |

*Note*. If the occurrence rate of DMs was zero per cent in either corpus, they were excluded from this analysis due to the impossibility of computing the log-likelihood ratio (LLR). IP = interpersonal; Ref = referential; Str = structural; Cog = cognitive. ** $p < .01$. *** $p < .001$.

### 4.3.2 RQ2-2 Comparisons of DM Use Between Japanese EFL Learners and Other English Learners

This section addresses RQ2-2, comparing DM frequencies in NNS speech from the Japanese subcorpus with the five other subcorpora of LINDSEI (i.e., LINDSEI-OTHERS). Table 4.4 provides comparisons of the frequency of DMs. As in the analysis of the previous section, if the occurrence rate of a given DM was 0.01 per cent or below in each database, the item was not included in the analysis.

The results of chi-square tests revealed that while Japanese learners often used some items such as *so* and *but*, they also used 14 out of 27 DMs less frequently than other non-native English learners did. These findings were supported by tests of log-likelihood ratios. Interpersonal or cognitive function markers such as *well*, *really*, *you know*, *I mean*, and *just* were used less frequently by Japanese learners than by other English learners. Thus, the significant differences in the frequencies of DMs may represent the features of Japanese learners' DM use.

On the other hand, the results given in Table 4.4 reveal no significant differences between the two databases in the frequency of seven items: *exactly*, *kind of*, *or*, *OK/okay*, *anyway*, *cos*, and *basically*. There were only small differences between learners' respective frequencies of three markers—*and*, *yes*, and *right/alright*—although the differences were significant at a 5% significance level. In short, it was notable that Japanese learners used some items just as frequently as other non-native English learners. Among these items, the use of *kind of*, *OK/okay*, *basically*, *yes*, and *right/alright* may be regarded as the features of DM use in NNSs' interlanguage because the frequency of the five items differed significantly between Japanese learners and NSs of English (see Table 4.3).

Table 4.4

*Comparisons of DM Use Between Japanese EFL Learners (LINDSEI-JP) and Other Non-native English Learners (LINDSEI-OTHERS)*

| DMs | Categories | Frequency per 10,000 words | | LLR | Chi-square value |
|---|---|---|---|---|---|
| | | LINDSEI-JP | LINDSEI-OTHERS | | |
| *so* | Ref/Str | 206.86 | 96.04 | 315.280 | 397.358*** |
| *but* | Ref | 145.72 | 119.45 | 18.157 | 19.430*** |
| *now* | Str | 13.47 | 8.15 | 9.638 | 11.130** |
| *finally* | Str | 2.96 | 1.09 | 7.093 | 9.470** |
| *first* | Str | 2.96 | 1.17 | 6.292 | 8.234** |
| *and* | Ref/Str/Cog | 420.46 | 394.14 | 5.815 | 6.164* |
| *OK/okay* | IP/Str | 22.90 | 19.05 | 2.456 | 2.591 |
| *kind of* | IP | 5.39 | 4.88 | 0.173 | 0.178 |
| *exactly* | IP | 2.15 | 2.13 | 0.001 | 0.001 |
| *or* | Ref | 50.10 | 55.35 | −1.749 | 1.711 |
| *anyway* | Ref | 1.08 | 2.36 | −3.025 | 2.484 |
| *cos* | Ref | 4.31 | 6.90 | −3.859 | 3.414 |
| *basically* | IP | 0.27 | 1.32 | −4.363 | 3.057 |
| *yes* | IP | 71.92 | 84.57 | −6.790 | 6.553* |
| *right/alright* | IP/Str | 0.27 | 2.15 | −9.283 | 6.050* |
| *I think* | IP/Cog | 88.35 | 109.15 | −14.451 | 13.799*** |
| *yeah* | IP/Str | 86.46 | 111.48 | −20.767 | 19.613*** |
| *like* | IP/Cog | 28.82 | 44.56 | −21.778 | 19.507*** |
| *actually* | IP | 4.85 | 14.07 | −28.085 | 21.731*** |
| *oh* | IP | 7.54 | 18.42 | −28.653 | 23.000*** |
| *because/'cause* | Ref | 47.68 | 73.26 | −34.943 | 31.434*** |
| *then* | Str | 15.35 | 33.61 | −42.937 | 35.367*** |
| *just* | IP | 10.77 | 47.72 | −145.738 | 104.410*** |
| *I mean* | Cog | 2.15 | 31.30 | −164.463 | 100.366*** |
| *you know* | IP/Cog | 4.31 | 39.91 | −182.483 | 117.121*** |
| *really* | IP | 8.62 | 57.53 | −227.775 | 153.006*** |
| *well* | IP/Str/Cog | 5.39 | 70.01 | −357.270 | 221.268*** |

*Note*. If the occurrence rate of DMs was zero per cent in either corpus. They were excluded from this analysis due to the impossibility of computing the log-likelihood ratio (LLR). IP = interpersonal; Ref = referential; Str = structural; Cog = cognitive. *$p < .05$. **$p < .01$. ***$p < .001$.

However, these data do not address differences in DM use within the category. LINDSEI-OTHERS, and distributions within individual subcorpora could boost or lower the overall frequency (see Appendix 4-A). To provide a clear picture, the frequencies of 12 DMs mentioned in this section were also compared across the six subcorpora of NNS speech. The further comparison was made to confirm whether the use of *so*, *but*, *well*, *really*, *you know*, *I mean*, and *just* exhibited the features of Japanese learners' speech, and whether the use of *yes*, *kind of*, *right/alright*, *basically*, and *OK/okay* reflected the features of DM use in NNSs' interlanguage.

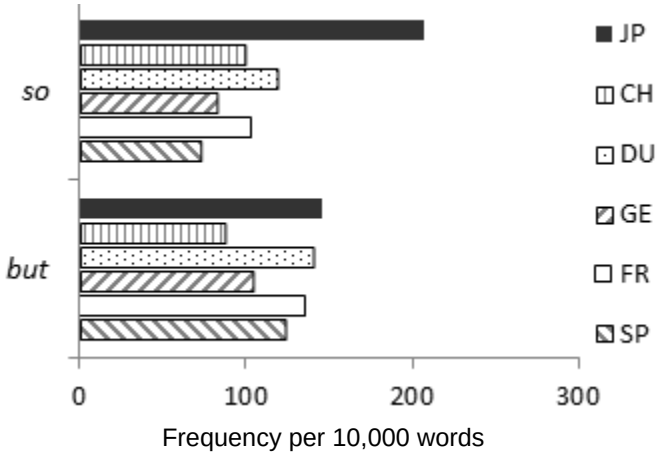



*Figure 4.2*. Frequency of *so* and *but* in each subcorpus of LINDSEI.

Figure 4.2 shows the frequency of *so* and *but* in each subcorpus. While *so* was used in the Japanese subcorpus substantially more frequently than in any other non-native subcorpus, only small differences existed among subcorpora in the frequency of *but*. Thus, the results confirm that the marker *so* is overused by Japanese learners, and that the lower usage levels of *but* in the Chinese and German subcorpora lower the overall frequency of

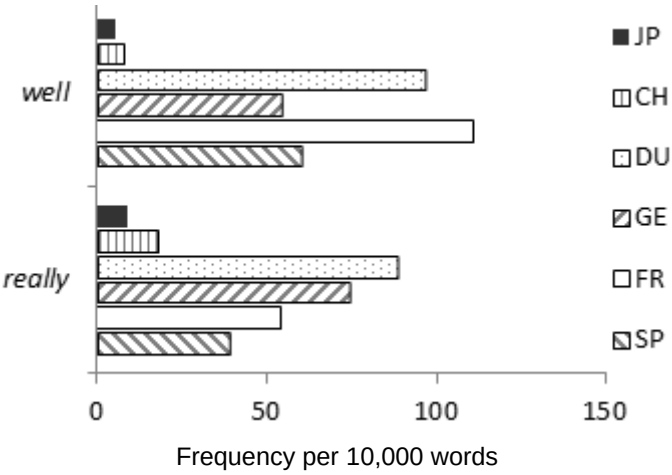



*Figure 4.3*. Frequency of *well* and *really* in each subcorpus of LINDSEI.

LINDSEI-OTHERS.

Figure 4.3 compares the frequency of *well* and *really* in each subcorpus. The analysis revealed that both Japanese and Chinese learners of English used the two markers notably less frequently than other non-native English learners. In other words, the results suggest that English learners whose L1 belongs to an East Asian language family may be more likely to underuse the markers *well* and *really*.

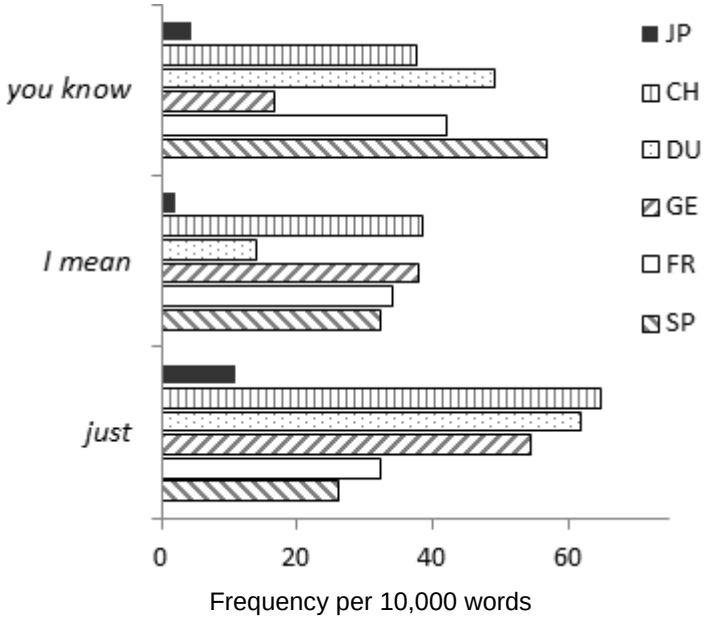



*Figure 4.4*. Frequency of *you know*, *I mean*, and *just* in each subcorpus of LINDSEI.

Figure 4.4 shows the frequency of *you know*, *I mean*, and *just* in each subcorpus. The analysis revealed that Japanese learners used the three markers less frequently than other non-native English learners. In other words, the results display a marked tendency for Japanese learners to underuse the interpersonal or cognitive function markers. These distinguishing features may be found only among Japanese learners of English; that is, they

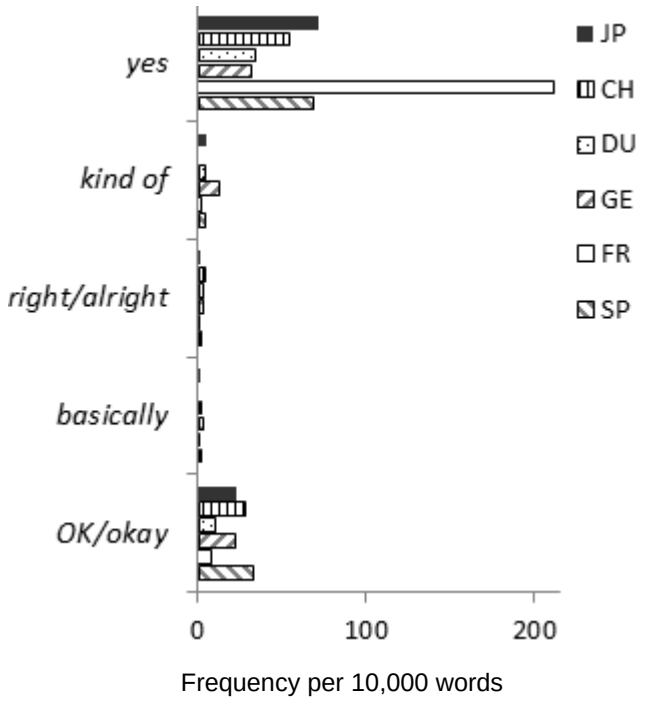



*Figure 4.5*. Frequency of *yes*, *kind of*, *right/alright*, *basically*, and *OK/okay* in each subcorpus of LINDSEI.

may not be shared by non-native English learners with different L1 backgrounds.

Figure 4.5 shows the frequencies of *yes*, *kind of*, *right/alright*, *basically*, and *OK/okay* in each subcorpus. The marker *yes* generally displays small differences among the subcorpora except for in the French subcorpus, where it was quite frequent indeed. On the other hand, the three markers *kind of*, *right/alright*, and *basically* were infrequently used in all six subcorpora. The general overuse of *yes* and the low frequencies of *kind of*, *right/alright*, and *basically* may be common to learners of English. With regard to the frequencies of *OK/okay*, the figure shows that there is a considerable variability among the subcorpora.

In short, while simple items such as *yes* may be preferred by NNSs, items such as *kind of*, *right/alright*, and *basically* may be more difficult for them to acquire.

### 4.3.3 Analysis of Discourse Functions of the Marker *so*

Study 1 and previous studies such as Hays (1992) and Miura (2011) suggest that Japanese learners may underuse certain pragmatic markers such as *well*, *I mean*, and *you know*, but they may frequently use simple types of markers such as *so* and *yes*. The present study yielded similar findings and distinguished features particular to Japanese learners from those seen in the speech of other NNSs. To investigate the acquisition of DMs in Japanese learners' speech, however, it is important to explore why they under- or overuse some items. To that end, this section focuses on the marker *so*, which is overused by Japanese learners.

According to Fung and Carter's (2007) framework, the marker *so* has two discourse functions, referential and structural. While the referential marker *so* serves a syntactic function to signal a relationship between one discourse segment and another, the structural marker *so* has some pragmatic functions such as a signal of summarizing

opinions and topic shifts. In the present study, tokens of *so* were classified by functional category: referential, structural, or the others. The following are illustrative examples of *so* extracted from the speech data of LINDSEI:

(1) Referential: I don't think I pronounce it very well, *so* I am a bit embarrassed …

(LINDSEI-SP015)

(2) Structural: … I think that's Julia Roberts. *So* that's all.

(LINDSEI-CH019)

(3) Structural: *So* what do you think of the city Guangzhou?

(LINDSEI-CH045)

(4) Other: … I always use bus *so* untto[6] … my nearest station is Ujiie Station.

(LINDSEI-JP005)

In example (1), the speaker uses the referential marker *so* in order to establish a cause-and-effect link between the first clause and the second one. In example (2), the speaker tries to mark the conclusion of the topic by using the structural marker *so*. The speaker in example (3) changes the topic to the listener's impression of the city Guangzhou by using the structural marker *so*. In example (4), however, the marker *so* is neither referential nor functional; instead, it seems to be used as a filler, which can provide time for the speaker to think about what to say next. The usage of *so* denotes the similar discourse management as found in the qualitative observation in Study 1 (see section 3.3.2.2 in Chapter 3).

Figure 4.6 shows the percentages for the three types of *so* (referential, structural, other) in the randomly sampled speech data, which comprise 10 interviews from each subcorpus. The coding of the functional categories was carried out by the author. As in

the categorization of DMs described above, a post-hoc intra-coded check was conducted for the three subcorpora, LINDSEI-JP, -CH, and -DU (i.e., 30 interviews) at an interval of about two years. The overall agreement rate was 93%. Thus, the reliability of this analysis is considered high.

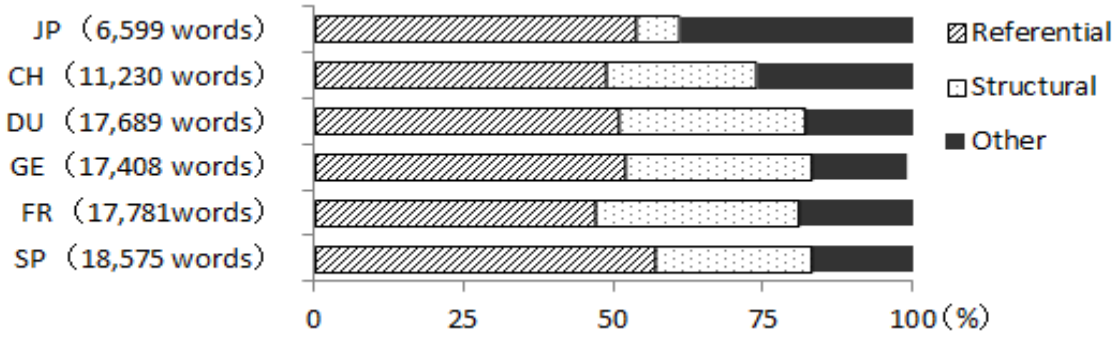



*Figure 4.6*. Percentages for the three types of *so*.

The results given in Figure 4.6 reveal that the proportion of the structural marker *so* was very low in the Japanese subcorpus. The third class of *so*, which is neither referential nor structural in function (i.e., other) was used more frequently by Japanese English speakers than by any other subcorpus group. The use of *so* as a filler may boost the frequency of the marker in Japanese English learners' speech.[7]

### 4.4 Summary of Study 2

This study employed CIA to investigate the use of DMs in the speech data of Japanese learners of English. The results illuminate some features of these speakers' DM use.

This study's first research question (RQ2-1) asked how frequencies of DMs in the speech of Japanese learners compare to those of NSs of English. Frequency analysis revealed significant differences between Japanese learners and NSs of English in the

frequency of many DMs. Japanese learners overused some simple markers such as *yes*, *so*, and *I think*, yet they underused certain interpersonal or cognitive function markers such as *like*, *really*, *you know*, *kind of*, and *I mean*. These findings corroborate those of previous studies, and they indicate that Japanese learners may have more difficulty acquiring particular pragmatic markers. These findings have important implications for language instructors, who may improve their students' interactional L2 skills as well as their linguistic ones through instructional focus on DMs.

The second research question (RQ2-2) asked how usage levels of English DMs by Japanese learners compare to those of English learners with different L1 backgrounds. Frequency analyses revealed both similarities and differences between Japanese learners and other non-native English learners in their use of DMs. While Japanese learners used *so* much more frequently than other non-native learners, they also used certain interpersonal or cognitive function markers such as *you know*, *I mean*, and *just* much less frequently. In other words, certain features of their DM use are distinguishable from those of non-native English learners generally. This suggests the need for language instructors and materials writers to carefully provide Japanese learners with language input according to the characteristics of their interlanguage. For example, language instructors and materials writers should provide infrequent and difficult items, such as interpersonal or cognitive markers, at an intermediate or advanced proficiency level. Additionally, they should furnish Japanese learners with opportunities to use as many kinds of easy-to-use items as possible at a lower level.

This study has two basic limitations. First, qualitative observations indicated that Japanese learners might use *so* as a filler, but this analysis has been far from exhaustive; more work on qualitative patterning is thus needed. As Romero-Trillo (2002), Müller (2004), and Aijmer (2011) have suggested, Japanese learners' under- or overuse of DMs

may be a result of the influence of their L1. Second, some tasks to elicit speech may have an effect on learners' DM use. For example, a picture description task may not lend itself to the use of interpersonal markers such as *really* and *just*. Further chapters will analyse learners' speech data from a qualitative perspective and investigate why Japanese learners may display different tendencies in English DM use from other non-native English learners.

**Endnotes**

1. A different version of this chapter first appeared in *JALT Journal*, *36*, 2014 (Shimada, 2014), published by the Japan Association for Language Teaching.
2. According to the online English Vocabulary Profile (http://www.englishprofile.org/), the markers *OK/okay*, *so*, and *yes* are classified into the Common European Framework (CEFR) level A1 or A2. Therefore, these markers can be regarded as easy items for English learners.
3. As Pritchard (1995) points out, Japanese learners of English may prefer slow, careful speech and take a long pause before answering a question. If so, the interaction style may have a negative effect on fluency in speech production. However, LINSDEI does not contain audio data and does not provide the information necessary to find out why the Japanese students produced a much smaller number of words than any of the other non-native English learners.
4. As of February 2014, LOCNEC has not been released yet.
5. The tests of the log-likelihood ratios are also called G-tests. The author combined the five subcorpora into one group and ran log-likelihood tests to compare the frequency of DMs between LINDSEI-JP and LINDSEI-OTHERS.
6. The Japanese word *untto* is approximately equivalent to the English marker *well*.

7. In the Japanese subcorpus, *so* as a filler was ubiquitous, although the frequency was not fully examined. Study 1 also points out that the filler usage may contribute to Japanese learners' overuse of the marker. The present study confirms those earlier findings.

# Chapter 5

## Study 3: L1 Transfer on Discourse Marker Use

### 5.1 Purpose of Study 3 and Research Questions

Study 2 made a comparative analysis of English DM use in speech data of non-native and native English speakers and explored some features of Japanese EFL learners. The results revealed that Japanese learners used some simple markers such as *yes*, *so*, and *I think* more frequently than NSs of English did, while they used certain interpersonal or cognitive function markers such as *like*, *really*, *you know*, *kind of*, and *I mean* less frequently. Additionally, the comparative study found that Japanese learners used *so* much more frequently than other English language learners with different L1 backgrounds, while they also used *you know*, *I mean*, and *just* less frequently. However, the study did not identify why Japanese EFL learners overused or underused certain DMs.

As mentioned in Chapter 2, some researchers suggest that NNSs' DM use may be influenced by the use of the translation equivalents in their L1. Sankoff et al. (1997) compared French DMs with English DMs in spoken data and found that L1 English speakers used a certain French marker influenced by their L1. Likewise, Liu (2013) also found that the usage of three Chinese DMs was transferred to that of the equivalent English DMs in Chinese English learners' speech. However, there has been no detailed study investigating the effect of Japanese on English DM use by Japanese learners of English.

The present study explores the influence of L1 transfer on English DM use by Japanese learners of English. The study addresses the following research questions (RQs):

RQ3-1: What Japanese DMs correspond to spoken English DMs?

RQ3-2: Is Japanese DM use transferred to English DM use in Japanese EFL learners' speech?

Using contrastive analysis (CA), RQ3-1 compares DMs in spoken English with those in spoken Japanese and investigates the relationship between English and Japanese DMs. RQ3-2, on the other hand, is designed to examine whether Japanese DM use has an influence on Japanese EFL learners' acquisition of DMs.

In order to answer the two research questions, this chapter consists of two crosslinguistic studies. The first study (i.e., Study 3A) addresses RQ3-1 using English-Japanese parallel data. Based on the findings of the parallel corpus-based study, in the second study (i.e., Study 3B), a small-scale experiment is carried out to answer RQ3-2.

## 5.2 Study 3A (RQ3-1): Contrastive Analysis of English and Japanese DMs

### 5.2.1 Method

#### 5.2.1.1 Databases

In order to make comparisons using CA, the present study used English-Japanese parallel corpus data. Spoken English data from Japanese EFL learners were from the NICT JLE Corpus (Izumi, Uchimoto, & Isahara, 2004) used in Study 1, and the parallel data were from the Japanese translation. As mentioned in Chapter 3, the NICT JLE Corpus consists of more than 1,200 interviews in English, and the transcribed data were divided into nine subcorpora according to learners' English proficiency levels. Additionally, 127 of the interviews were translated into Japanese for the purpose of investigating L1 transfer in L2 acquisition. The translation was made by a research

assistant who was experienced in corpus markup and translation work (K. Uchimoto, personal communication, March 22, 2014).[1] In the present study, five examples of interview data were randomly sampled from each of the subcorpora except for level 1, and the 40 English-Japanese parallel data were used to answer RQ5-1.

Before conducting data analysis, lexical statistical features of the Japanese EFL learners' speech data and the Japanese translations were computed using WordSmith Tools 5.0 and KH Coder 2.0, a quantitative analysis tool of Japanese language. KH Coder 2.0 calculates tokens and types on the basis of the results of analysing lexical morphemes. Tables 5.1 and 5.2 list tokens, types, type-token ratio (TTR)[2], and words per sentence in the parallel data.

Table 5.1
*Lexical Statistical Features of 40 Japanese EFL Learners' Speech Data*

| | Total | L2 | L3 | L4 | L5 | L6 | L7 | L8 | L9 |
|---|---|---|---|---|---|---|---|---|---|
| Subjects | 40 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| Tokens | 40,363 | 1,430 | 2,665 | 4,585 | 5,141 | 5,206 | 7,080 | 6,697 | 7,559 |
| Types | 3,021 | 413 | 511 | 763 | 938 | 941 | 1,159 | 1,057 | 1,164 |
| TTR | 7.48 | 28.88 | 20.68 | 16.64 | 18.25 | 18.08 | 16.37 | 15.78 | 15.40 |

*Note.* Speech data at level 1 were excluded from this analysis because they were not translated into Japanese. L = level.

Table 5.2
*Lexical Statistical Features of the Translation of Japanese EFL Learners' Speech Data into Japanese*

| | Total | L2 | L3 | L4 | L5 | L6 | L7 | L8 | L9 |
|---|---|---|---|---|---|---|---|---|---|
| Subjects | 40 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| Tokens | 51,248 | 1,915 | 3,304 | 4,880 | 6,401 | 6,544 | 9,172 | 8,563 | 10,469 |
| Types | 3,281 | 397 | 545 | 775 | 967 | 1,011 | 1,268 | 1,137 | 1,269 |
| TTR | 6.40 | 20.73 | 16.50 | 15.88 | 15.11 | 15.45 | 13.82 | 13.28 | 12.12 |

*Note.* Speech data at level 1 were excluded from this analysis because they were not translated into Japanese. L = level.

The tokens in both English and Japanese data increased as Japanese learners' English proficiency improved. The total number of tokens in English was fewer than that in the translation of English into Japanese.

#### 5.2.1.2 Procedure

Based on the results in Study 1, the present study focuses on the 25 DMs with a more than 0.01 per cent distribution rate in the NICT JLE Corpus (see Table 3.2 in Study 1). First, WordSmith Tools 5.0 was used to count raw frequencies for each of the 25 items. As for the items playing other grammatical roles besides those of DMs, the concordance lines were also examined to obtain frequencies for only DMs.[3] For example, the following italicized words were excluded from this analysis.

> In free time, urrm I *like* to read books. (The NICT JLE Corpus, 00286.stt)
>
> And you should know the person. Who do *you know* call?
>
> (The NICT JLE Corpus, 00286.stt)

Next, ParaConc 1.0 (Build 269), a bilingual or multilingual concordancer, was used to confirm the translation equivalents (see Figure 5.1). In the parallel concordance program, an original text (e.g., English text) is aligned with the counterpart text (e.g., Japanese text). For example, if the marker *but* is searched in the program, the following parallel texts are displayed on the computer screen.

> Well before I got married. I used to go urr skiing. *But*, nowadays, I don't.
>
> [Kekkon mae wa yoku sukii e itta mono deshita. *Demo* ima wa iku koto wa arimasen.] (The NICT JLE Corpus, 00287.stt)

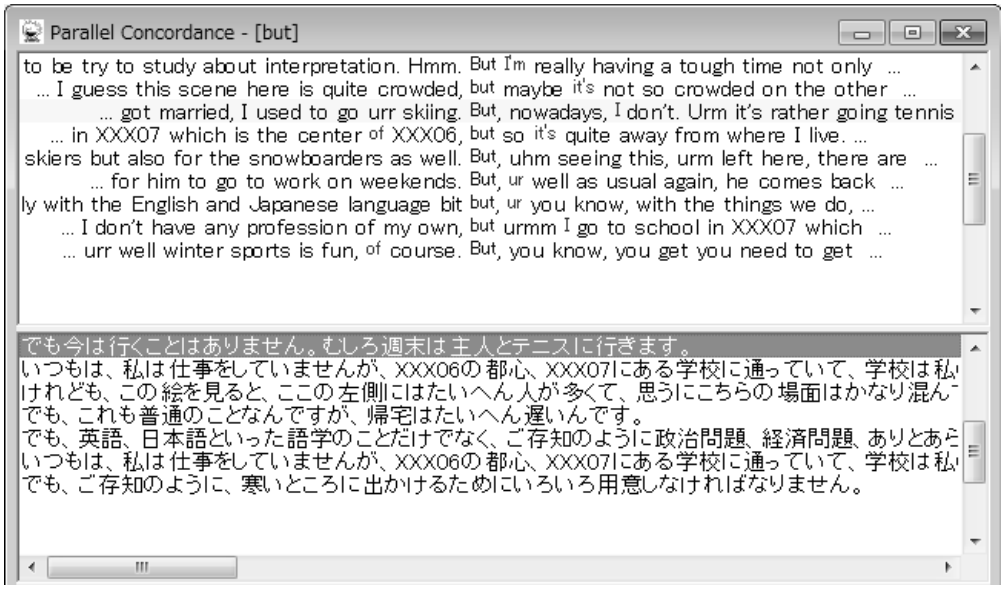



*Figure 5.1*. Concordance results on the ParaConc display.

In the parallel texts, the marker *but* is translated into the equivalent Japanese marker *demo*. As illustrated in the above example, parallel concordance lines were viewed to identify translation equivalents.

### 5.2.2 Results and Discussion

Table 5.3 shows the raw frequencies of each English DM and the translation equivalents. The results reveal three major findings. First, all English DMs except for *I see* were translated into two or more Japanese equivalents. Among the most frequent English DMs, *and* and *so* were translated into many Japanese expressions. The results support the findings of Mizutani (2001) and Onodera (2004).

Second, some Japanese DMs such as *sorede*, *ga*, and *dewa* correspond to different English DMs. For example, the Japanese marker *sorede* is equivalent to *and*, *so*, and *then*. These three English DMs serve structural functions such as continuing the topic and signalling sequential relationships in discourse (Fung & Carter, 2007). The marker *sorede* also plays a similar role in Japanese conversation.

Table 5.3

*Raw Frequency of English DMs and the Japanese Translation Equivalents*

| English DMs | Categories | FED | Frequency of the Japanese translation equivalents | | |
|---|---|---|---|---|---|
| *and* | Ref/Str/Cog | 1179 | *soshite* 57 | *sorede* 28 | *soreto* 19 |
| | | | *de* 12 | *soreni* 11 | *ga* 4 |
| | | | *sorekara* 2 | *ato* 1 | *demo* 1 |
| | | | *dewa* 1 | *node* 1 | *to* 1 |
| | | | none 1041 | | |
| *so* | Ref/Str | 596 | *desukara* 87 | *sorede* 54 | *node* 15 |
| | | | *soredewa* 9 | *nanode* 7 | *ittemireba* 3 |
| | | | *soredeshitara* 3 | *sou yatte* 3 | *dewa* 3 |
| | | | *dakara* 1 | *demo* 1 | *sou desune* 1 |
| | | | none 409 | | |
| *but* | Ref | 417 | *demo* 215 | *ga* 150 | *keredomo* 11 |
| | | | *shikashi* 6 | *keredo* 1 | none 34 |
| *yes* | IP | 379 | *hai* 338 | *sou desu* 17 | *sou desune* 5 |
| | | | *ee* 1 | *sou nandesuka* 1 | none 17 |
| *yeah* | IP/Str | 240 | *hai* 182 | *sou desu* 17 | *sou desune* 12 |
| | | | *ee* 7 | *aa* 3 | *haa* 2 |
| | | | *sou desuka* 1 | *sou nandesu* 1 | none 15 |
| *OK/okay* | IP/Str | 203 | *wakarimashita* 25 | | *hai* 12 |
| | | | *ii desuka* 4 | | *jyunbi dekimashita* 4 |
| | | | *ii desuyo* 3 | | *kekko desu* 3 |
| | | | *kamaimasen* 2 | | *yoroshii desuka* 2 |
| | | | *daijyoubu* 1 | | *doumo* 1 |
| | | | *ii desune* 1 | | none 145 |
| *or* | Ref | 166 | *toka* 37 | *ka* 20 | *aruiwa* 12 |
| | | | *mata* 7 | *soretomo* 2 | none 88 |
| *because/'cause* | Ref | 161 | *node* 47 | *kara* 25 | *to iimasunomo* 19 |
| | | | *desukara* 8 | *nazenara* 4 | *tame desu* 1 |
| | | | *to iu riyuu de* 1 | none 56 | |
| *oh* | IP | 121 | *maa* 14 | *oo* 4 | *ara maa* 1 |
| | | | none 102 | | |
| *I think* | IP | 100 | *to omoimasu* 56 | | *sou omoimasu* 6 |
| | | | *omoimasu ga* 3 | | *to omouno desuga* 3 |
| | | | *omoimasu kedo* 1 | | *to omouno desu* 1 |
| | | | *watashi ga omouni* 1 | | *watashi ga omoimasuni* 1 |
| | | | *wo kangaemasu* 1 | | *wo kangaetemimasu* 1 |
| | | | none 26 | | |

(continued)

Table 5.3 (continued)

| English DMs | Categories | FED | Frequency of the Japanese translation equivalents | | |
|---|---|---|---|---|---|
| *then* | Str | 87 | *sorede* 33 | *sorekara* 9 | *soredewa* 5 |
| | | | *soshite* 3 | *sonogo* 1 | none 36 |
| *well* | IP/Str/Cog | 65 | *hai* 2 | *maa* 1 | *sou desunee* 1 |
| | | | none 61 | | |
| *just* | IP | 45 | *chotto* 15 | *choodo* 5 | none 25 |
| *actually* | IP | 42 | *jitsuwa* 14 | *jissai* 8 | *jitsu no tokoro* 4 |
| | | | *jissai no tokoro* 4 | *jissainiwa* 2 | *jissaiwa* 2 |
| | | | *jitsu wo iuto* 1 | none 7 | |
| *you know* | IP/Cog | 40 | *anou* 12 | *gozonji no youni* 10 | |
| | | | *gozonji desyouga* 8 | *owakari desyouga* 8 | |
| | | | *gozonji no toori* 1 | *nanto iuka* 1 | |
| *like* | IP/Cog | 35 | *no youna* 13 | *no youni* 11 | *mitaina* 3 |
| | | | *nado* 2 | *toka* 1 | none 5 |
| *sure* | IP | 23 | *ii desuyo* 7 | *mochiron desu* 5 | *hai* 4 |
| | | | *sou desu* 2 | none 5 | |
| *I mean* | Cog | 20 | *tsumari* 8 | *sono* 3 | *ie* 1 |
| | | | none 8 | | |
| *finally* | Str | 18 | *kekkyoku* 7 | *tsuini* 4 | *saisyuutekini* 2 |
| | | | *saigoni* 1 | none 4 | |
| *kind of* | IP | 18 | *chotto* 7 | *nanka* 1 | *maa* 1 |
| | | | none 9 | | |
| *really* | IP | 14 | *hontouni* 7 | *sounan desuka* 1 | *hontou desuka* 1 |
| | | | *jitsu wa* 1 | *soukai* 1 | none 3 |
| *first* | Str | 9 | *saisyowa* 3 | *mazu* 3 | *saisyo* 1 |
| | | | *mazu saisyoni* 1 | *mazu daiichini* 1 | |
| *how about* | Str | 6 | *(wa) dou desuka* 1 | *(wa) ikaga desuka* 1 | |
| | | | (wa) dounan desuka 1 | none 3 | |
| *right/alright* | IP/Str | 5 | *ii desuyo* 1 | *sou desuka* 1 | *sounan desu* 1 |
| | | | *chigaimasuka* | none 1 | |
| *I see* | IP/Cog | 4 | *wakarimashita* 4 | | |

*Note.* FED = frequency of English DMs; IP = interpersonal; Ref = referential; Str = structural; Cog = cognitive. If it is impossible to find Japanese translation equivalents, they are labelled *none*.

(1) One day a last week, I went to the station to take a train. *And* the … on the platform, the man who were standing in front of me hit me.

[Senshuu no aru hi, watashi wa densha ni noru tame ni eki ni ikimashita. *Sorede*, puratto houmu de watashi no mae ni ita otoko no hito ga watashi ni butsukatte kimashita.] (The NICT JLE Corpus, 00008.stt)

(2) … this year, she entered university. *So* she talked about her university life.

[… imouto wa kotoshi daigaku ni shingaku shimashita. *Sorede* kanojo wa jibun no gakusei seikatsu ni tsuite hanashimasu.]

(The NICT JLE Corpus, 00121.stt)

In examples (1) and (2), both the English DMs and the Japanese marker *sorede* are used to demonstrate a continuation of the topic. Likewise, the English marker *then* in Example (3) also corresponds to the Japanese structural function marker *sorede*.

(3) My husband mother … to … will come to visit my house suddenly. *Then*, I have to go out.

[Shujin no haha ga kyuuni tazunetekuru koto ni narimashita. *Sorede*, gaishutsu shinakereba narimasen node.] (The NICT JLE Corpus, 00006.stt)

Additionally, like the referential function marker *but*, the marker *and* is also used to introduce a discourse segment that contrasts with the previous segment. In Japanese, the marker *ga* serves this function. Some examples are shown below.

(4) … we er put this urm the brightening star on his hair, and it's the light urm urr goes on and off on and off. So he's right in the jogging, *and* urm I I feel a

little bit ashamed.

[… watashi tachi wa chichi no kami no ke ni kirakira hikaru hoshi wo tsuketa no desu ga, sono hikari wa tenmetsu surundesu. Sorede chichi wa jyogingu chu demo daijyoubu nano desu *ga*, watashi wa chotto hazukashii desu.]

(The NICT JLE Corpus, 00286.stt)

(5) I was supposed to go to the dinner party, *but* I couldn't …

[Dinaa paatii ni oukagai suru koto ni natte ita no desu *ga*, ikenaku narimashite …] (The NICT JLE Corpus, 00318.stt)

These examples indicate that a Japanese DM can correspond to multiple English DMs if the English DMs have a similar function in discourse.

Third, in the parallel corpora, most occurrences of *and*, *so*, and *well* were labelled *none* in Table 5.3. In other words, it was often the case that it was impossible to find out which Japanese DMs were equivalent to these English DMs. In regard to *so*, as found in the previous chapters, the use of *so* as a filler[4] was frequently identified in the speech data.

(6) … the boy erm heard something um somethings. *So* … he he founds the sound is er from the strange box.

[… sono otoko no ko wa nanika monooto wo kikimashita. Kare wa sono oto ga minarenai hako kara hasserareteiru noni kizukimashita]

(The NICT JLE Corpus, 00226.stt)

In example (6), *so* is not translated into Japanese, because it serves as neither a referential function marker nor a structural one in discourse. Similarly, the usage as a filler was also found in *and* and *well* as follows:

(7) One day last week. Ah *and* we ah I explain?

[Senshu no aru hi. Watashi ga setsumei suruno desuka?]

(The NICT JLE Corpus, 00121.stt)

(8) … he is a *well* like called ordinary salary-paid man.

[… kare wa iwayuru hutsuu no sararii man desu.]

(The NICT JLE Corpus, 00287.stt)

In (7) and (8), these English fillers are also not translated into Japanese.

Moreover, it is partly because Japanese DMs may not even be used in the discourse context where these English DMs are used.

(9) … they make bookings to urm shipping careers like us. *And* ur *so* we take bookings, and ur we export cargoes.

[… karera wa watashi tachi noyouna kaiun gyousha ni yoyaku wo iremasu. Watashi tachi wa yoyaku wo ukete yushutsu shimasu.]

(The NICT JLE Corpus, 00397.stt)

The speaker in (9) uses the referential marker *and*, and then changes it to the referential marker *so* in order to connect the first discourse segment with the second more appropriately. However, the translated sentences do not include any equivalent Japanese DMs. In short, some English DMs may often not translate into Japanese expressions.

### 5.2.3 Overview of the Findings in Study 3A

Regarding the answer to RQ3-1, the quantitative and qualitative analyses revealed that there was not a one-to-one correspondence between English DMs and Japanese

DMs. Therefore, although Liu (2013) and He (2001) adopted a one-to-one correspondence in exploring the effect of Chinese DM use on the use of English DMs by Chinese learners, their method cannot apply to the investigation of L1 transfer on Japanese EFL learners' DM use.

Figure 5.2 illustrates the relationship between English DMs and Japanese DMs as found in the present study. Each English DM corresponds to multiple Japanese DMs, while some Japanese DMs (i.e., E and F) correspond to different English DMs (i.e., A and B). However, it is often the case that English DMs are not translated into Japanese expressions, partly because they are used as fillers (i.e., grey-coloured part of each English DM).

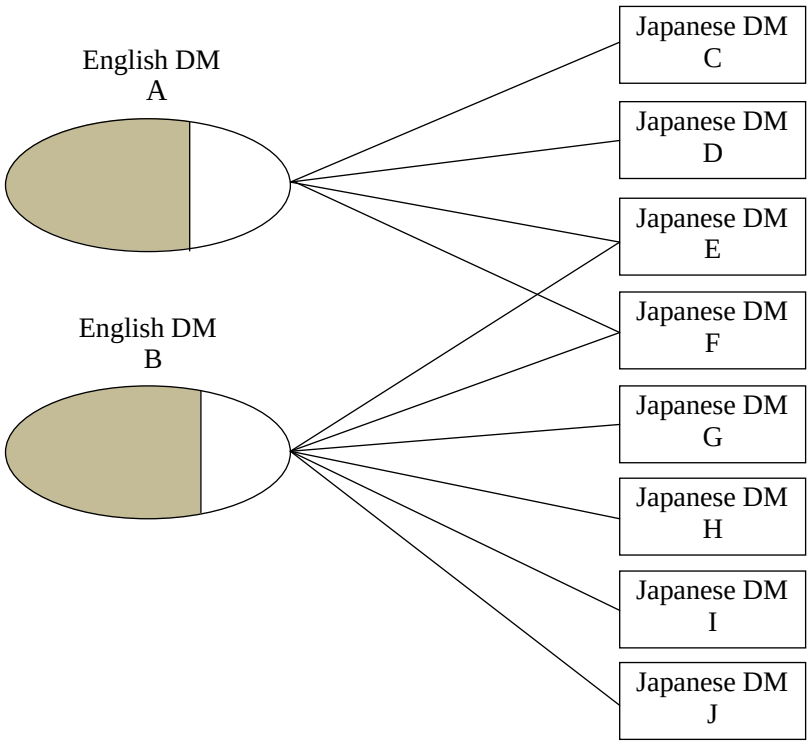



*Figure 5.2*. Relationship between English DMs and Japanese DMs.

## 5.3 Study 3B (RQ3-2): L1 Effects on the Use of English DMs by Japanese EFL Learners

### 5.3.1 Method

#### 5.3.1.1 Participants

Participants were 30 undergraduates from a private university in Tokyo who took an elective English class in the spring and fall semesters (from April 2012 to January 2013). They were assigned into two classes (i.e., Classes 1 and 2) and were taught by two different instructors per class. All the students in both classes were selected based on their application and initial interview test. The screening process, therefore, ensured that they were at a lower-intermediate or intermediate level of English proficiency, which is equivalent to a TOEIC® score of 365 to 600. They majored in economics, business administration, or law.

A small-scale experiment was performed on the 30 participants. Three participants were eliminated from the data analysis: One participant was an international student, and the others' speech data were not recorded clearly due to technical problems. Thus, 27 participants' data were analysed in the present study.

#### 5.3.1.2 Materials

In the present experiment, a picture description task was given to the participants using two kinds of four-frame picture strips (see Appendices 5-A and 5-B). The comic strips were taken from the pre-first grade level of the Eiken Test in Practical English Proficiency. In the task, the participants were required to narrate a story based on each comic strip. Some spoken learner corpora such as the NICT JLE Corpus used oral narrative tasks, and the present study adopted similar task material in order to elicit the participants' DM use in their speech.

### 5.3.1.3 Procedure

First, a comic strip was presented on the computer screen, and the participants were given one minute to plan the story of the comic strip. They were required to describe the pictures orally in Japanese for two minutes, and narrate the story in English for two minutes. Then, the participants performed the same tasks with the other comic strip, but in English to describe the pictures and in Japanese to narrate the story. Finally, they narrated the second comic strip in English for two minutes. The third narration task was designed to improve the students' speaking performance in the class, but the speech data were not used for the analysis in the present study.

The presentation orders of the two comic strips (i.e., Comic Strips A and B) were counterbalanced between the participants to eliminate any ordering effect. Therefore, Class 1 ($n$ = 15) was given Comic Strips A and B in that order, while Class 2 ($n$ = 12) was given the two comic strips in reverse order. The procedure is illustrated in Figure 5.3.

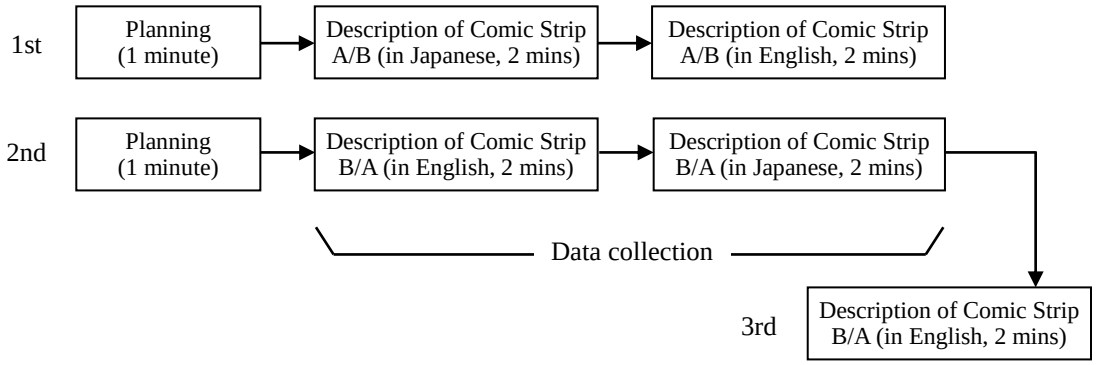



*Figure 5.3*. Procedure of the picture description task.

The one-shot experiment was conducted as part of the regular classroom activities. With careful consideration of the course schedule in each class, the experiment in Class 1 was carried out in July 2012, and that in Class 2 was in October 2012.

In the data collection, the participants' speech was recorded with a headset microphone, and the audio recordings were transcribed by two professional transcribers. Before conducting data analysis of DM use, as in Study 3A, lexical statistical features of the learners' speech data both in English and in Japanese were computed using WordSmith 5.0 and KH Coder 2.0.

Table 5.4
*Lexical Statistical Features of Japanese College Students' Speech Data in English*

| | | Comic strip A | | Comic strip B | |
|---|---|---|---|---|---|
| | Total | Class 1 | Class 2 | Class 1 | Class 2 |
| Subjects | 54 | 15 | 12 | 15 | 12 |
| Tokens | 4,843 | 1,243 | 1,158 | 1,257 | 1,185 |
| Types | 472 | 203 | 192 | 220 | 216 |
| Tokens/Subject | 89.69 | 82.87 | 96.50 | 83.80 | 98.75 |
| Type-token ratio | 9.75 | 16.33 | 16.58 | 17.50 | 18.23 |

Table 5.5
*Lexical Statistical Features of Japanese College Students' Speech Data in Japanese*

| | | Comic strip A | | Comic strip B | |
|---|---|---|---|---|---|
| | Total | Class 1 | Class 2 | Class 1 | Class 2 |
| Subjects | 54 | 15 | 12 | 15 | 12 |
| Tokens | 9,093 | 2,412 | 2,146 | 2,655 | 1,880 |
| Types | 620 | 305 | 238 | 295 | 256 |
| Tokens/Subject | 168.39 | 160.8 | 178.83 | 177.00 | 156.67 |
| Type-token ratio | 6.82 | 12.65 | 11.09 | 11.11 | 13.62 |

*Note*. Token and types were calculated on the basis of the results of analysing lexical morphemes.

Tokens per subject in Table 5.4 show that the students in Class 2 narrated both stories more fluently than those in Class 1. While the average TOEIC® score in Class 1 was 445.67, that in Class 2 was 506.67 ($t$ (25) = −2.949, $p$ = .007, $d$ = 1.14). The difference

of their English proficiency levels may affect their speaking performance. However, the table also shows that there were small differences between the two comic strips in tokens per subject, but this difference was not significant, $t$ (26) = −.43, $p$ > .05, $r$ = .09. Therefore, the difference of the comic strips had little effect on the number of English words uttered by the participants. Additionally, the token per subject in Table 5.5 shows that it was easier for the participants to narrate a story in Japanese than in English.

As for the analysis of DM use, the method was the same as that used in section 5.2.1.2 in Study 3A. The present study also focuses on the 25 English DMs with a more than 0.01 per cent distribution rate in the NICT JLE Corpus. The equivalent Japanese DMs are analysed based on the results of Study 3A.

### 5.3.2 Results and Discussion

#### 5.3.2.1 Frequency of English DMs

Table 5.6 displays the results of the frequency analysis in the students' speech data in English. The frequency of each marker was standardized as the frequency per 1,000 words. As mentioned in the previous section, there were significant differences between the two classes in the tokens to describe both comic strips, but, in this section, the participants are not divided into two groups because this analysis does not focus on DM use at each English proficiency level (i.e., lower or higher level).

The students used 11 items of English DMs for describing Comic Strip A and 10 items for Comic Strip B. They preferred the referential and structural function markers such as *and, so, but, then,* and *because*. Some examples are shown below.

(10) … their their mother wish they had more space *because* the difficult the cleaning is difficult *and* they decided to big house more than now has.

Table 5.6

*Frequency of English DMs in Japanese College Students' Speech in the Picture Description Task*

| DMs | Categories | Frequency per 1,000 words | | |
|---|---|---|---|---|
| | | Comic Strip A (2,401 words) | Comic Strip B (2,442 words) | Total (4,843 words) |
| *and* | Ref/Str/Cog | 22.07 | 23.34 | 22.71 |
| *so* | Ref/Str | 16.66 | 10.65 | 13.63 |
| *but* | Ref | 7.08 | 10.24 | 8.67 |
| *then* | Str | 3.75 | 5.73 | 4.75 |
| *because* | Ref | 2.92 | 1.64 | 2.27 |
| *oh* | IP | 0.83 | 2.87 | 1.86 |
| *next* | Str | 1.25 | 1.23 | 1.24 |
| *I think* | IP/Cog | 1.67 | 0.00 | 0.83 |
| *finally* | Str | 0.83 | 0.41 | 0.62 |
| *yes* | IP | 0.42 | 0.82 | 0.62 |
| *yeah* | IP/Str | 0.83 | 0.00 | 0.41 |
| *or* | Ref | 0.00 | 0.41 | 0.21 |

*Note*. IP = interpersonal; Ref = referential; Str = structural; Cog = cognitive.

(11) A few days later finally the couple umm get a dog. *Then* shopper recommend them to to enter the pet insurance …

In example (10), the speaker uses two referential markers, *because* and *and*, to connect a discourse segment with the preceding one. In other words, these referential markers serve as a connecting device to control the flow from one picture to another, and help to make a plot of the story of Comic Strip A. Similarly, the structural marker *then* in example (11) is used to control the sequence of two behaviour units in the second picture of Comic Strip B: the couple's purchase of a dog and the salesclerk's recommendation to buy pet insurance.

Additionally, the referential marker *but* plays an important role in leading to the ending of the story as follows:

(12) Six months later the family … lived can live a big house … *but* uh children uhh never play together.

(13) … dog dog get wells *but* they get uh bill bill sheet gohyaku $500. They feel uh ... They feel pretty bad.

In examples (12) and (13), the use of the marker *but* emphasises the contrast between the preceding comic scenes and the unhappy endings. Thus, the narration task using comic strips may elicit the use of referential and structural function markers to organise the plot of the story.

On the other hand, the students used interpersonal function markers such as *I think*, *yes*, and *yeah* less frequently. The use of these DMs may be unsuited for monologic tasks such as picture descriptions because interpersonal markers often occur between speakers.

#### 5.3.2.2 Frequency of Japanese DMs

Section 5.3.2.1 revealed that the students used 12 English DMs in the picture description task. This section, in turn, focuses on the translation equivalents (i.e., Japanese DMs) in the students' speech data in Japanese. Based on the results in Table 5.3, the frequency of each Japanese DM[5] was counted to explore the influence of L1 transfer on the English DMs (see Table 5.7). The frequency of each marker was a standardized frequency per 1,000 words.

The results revealed that three or more kinds of Japanese DMs were used as the counterpart of *and*, *so*, *but*, *then*, *because*, or *yeah* each. For example, the eight items, *soshite*, *ga*, *de*, *node*, *to*, *sorede*, *soreni*, and *demo*, were used to perform the function of *and*, which was the most frequent English DM in the students' speech (see Table 5.6).

Table 5.7

*Frequency of Japanese DMs in Japanese College Students' Speech in the Picture Description Task*

| English DMs | Japanese DMs | Frequency per 1,000 words | | |
|---|---|---|---|---|
| | | Comic Strip A (4,557 words) | Comic Strip B (4,534 words) | Total (9,091 words) |
| *and* | *soshite* | 5.92 | 3.31 | 4.62 |
| | *ga* | 1.10 | 6.40 | 3.74 |
| | *de* | 1.76 | 0.66 | 1.21 |
| | *node* | 1.32 | 0.88 | 1.10 |
| | *to* | 0.44 | 0.44 | 0.44 |
| | *sorede* | 0.00 | 0.44 | 0.22 |
| | *soreni* | 0.22 | 0.00 | 0.11 |
| | *demo* | 0.00 | 0.22 | 0.11 |
| *so* | *node* | 1.32 | 0.88 | 1.10 |
| | *sorede* | 0.00 | 0.44 | 0.22 |
| | *nanode* | 0.22 | 0.22 | 0.22 |
| | *demo* | 0.00 | 0.22 | 0.11 |
| *but* | *ga* | 1.10 | 6.40 | 3.74 |
| | *shikashi* | 1.97 | 1.99 | 1.98 |
| | *keredo* | 0.00 | 0.22 | 0.11 |
| | *demo* | 0.00 | 0.22 | 0.11 |
| *then* | *soshite* | 5.92 | 3.31 | 4.62 |
| | *sonogo* | 0.00 | 2.21 | 1.10 |
| | *sorede* | 0.00 | 0.44 | 0.22 |
| *because* | *node* | 1.32 | 0.88 | 1.10 |
| | *kara* | 0.66 | 0.44 | 0.55 |
| | *nazenara* | 0.22 | 0.22 | 0.22 |
| *finally* | *kekkyoku* | 0.00 | 0.22 | 0.11 |
| | *tsuini* | 0.22 | 0.22 | 0.22 |
| *yes* | *ee* | 0.22 | 0.66 | 0.44 |
| | *hai* | 0.22 | 0.44 | 0.33 |
| *yeah* | *aa* | 0.00 | 0.88 | 0.44 |
| | *ee* | 0.22 | 0.66 | 0.44 |
| | *hai* | 0.22 | 0.44 | 0.33 |
| *or* | *ka* | 0.22 | 0.00 | 0.11 |

*Note*. Japanese DMs which were equivalent to *oh* and *I think* were not used in the speech data.

In other words, the various uses of Japanese DMs may influence English DM use by the Japanese students.

Additionally, the results also revealed that *sosite* and *ga* were frequently used in the speech data. The high frequency of items such as *soshite* and *ga* may influence the frequency of the equivalents in English. An example is shown below.

(14) Sono yoru sono kazoku wa motto ookii motto hiroi ie ni kaeyou to keikaku wo shita. *Sosite* hantoshi go ee … karera wa ookii ie wo tate …
[That night the family planned to live in a larger, more spacious house. *And* half a year later … they built the large house …]

(15) That night the wife talk to his his husband and why don't you enlarge your house. *And* 6 months later umm they decide to move.

While (14) is a description of Comic Strip A in Japanese, (15) is the same picture description in English. Both were uttered by one student.[6] Although a few differences existed between his two narrations in the word choice, he used *soshite* and *and*, respectively, to connect the first discourse segment with the second one.

Moreover, as mentioned in Study 3A, *soshite* and *ga* corresponded to multiple English DMs. The marker *soshite* was used to perform the function of *and* and *then*. Likewise, the marker *ga* served as the counterpart of *and* and *but*. Therefore, the use of these Japanese DMs may have an influence on the use of multiple English DMs.

On the other hand, the results in Table 5.7 did not confirm the effect of Japanese DM use on the use of the English DM *so*. In other words, the students used the Japanese equivalents except for *node* less frequently, while they used the marker *so* frequently in their English speech. However, the findings of Study 2 indicated that Japanese EFL

learners may overuse *so* as a filler. To that end, therefore, a further qualitative analysis was conducted to explore why the students preferred *so* in their narration work.

As in Study 2, tokens of the English DM *so* were classified by functional category. In the present study, the proportion of *so* used as fillers, or their errors or mistakes, was also calculated (see Table 5.8).

Table 5.8
*Raw Frequency of* so *in Japanese College Students' Speech in the Picture Description Task*

| | Raw frequency of *so* | | | | |
|---|---|---|---|---|---|
| Speech data | Ref | Str | Fillers | E or M | Total |
| Comic Strip A | 18 (45.0) | 7 (17.5) | 14 (35.0) | 1 (2.5) | 40 (100.0) |
| Comic Strip B | 12 (46.2) | 4 (15.4) | 8 (30.8) | 2 (7.7) | 26 (100.0) |
| Total | 30 (45.5) | 11 (16.7) | 22 (33.3) | 3 (4.5) | 66 (100.0) |

*Note.* In calculating the frequency, concordance lines were viewed to confirm the discourse functions. Ref = referential; Str = structural; E or M = students' errors or mistakes. Percentages appear in parentheses.

The results given in Table 5.8 revealed that, although the proportion of the structural marker *so* was very low, the use of *so* as a filler boosted the frequency of the marker in the students' speech. Thus, they support the findings of Study 2.

Next, Table 5.9 shows which Japanese words correspond to these filler usages. The results suggest that Japanese filler expressions such as *aa, ee*, and *uun* may be transferred into English *so* as a filler. Some examples are shown below.

(16) *Ee* rokkagetsu go ni buji … uun … kansei shite …

[*Er* 6 months later the house was successfully completed …]

(17) *So* 6 months later the the house is the house is … perfect …

Table 5.9

*Raw Frequency of the Japanese Equivalents of the Filler* so

| | Raw frequency of the Japanese equivalents | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Speech data | *aa* | *ee* | *etto* | *sou* | *soushite* | *uun* | none | Total |
| Comic Strip A | 1 | 5 | 0 | 1 | 2 | 1 | 4 | 14 |
| Comic Strip B | 0 | 3 | 1 | 0 | 0 | 0 | 4 | 8 |
| Total | 1 | 8 | 1 | 1 | 2 | 1 | 8 | 22 |

*Note.* If it is impossible to find Japanese equivalents, they are labelled *none*.

(18) Hitokumi no kappuru ga niwa ni imashita. *Ee* kare wa inu wo kawanai to … tei … kanojyo ni teian shite …

[There was a couple in a yard. *Er* he suggested to her that they get a dog ...]

(19) There one couple in a park, *so* he pro propose that why don't get a dog …

As in examples (14) and (15), examples (16) and (17) show the comparison of a student's narration in Japanese and in English.[7] In (16), the student used the Japanese filler *ee*[8] in order to think about what to say next, while she used *so* for the same purpose in (17). Likewise, as shown in (18) and (19), another student also used *ee* and *so* as fillers in her narration work.

In short, Japanese EFL students may tend to use the marker *so* in the same way as they use fillers such as *aa*, *ee*, and *uun* in Japanese speech, although native English speakers do not use *so* as a filler.

### 5.3.3 Overview of the Findings in Study 3B

The answer to RQ3-2 is yes; the contrastive analyses in the present study indicate that Japanese DM use may be transferred to English DM use in Japanese EFL learners' speech. From the quantitative results, the use of various Japanese DMs may be

transferred into the use of some English DMs such as *and*, *so*, *but*, *then*, *because*, and *yeah*. Additionally, the use of Japanese DMs *soshite* and *ga* may influence the use of the same function English DMs such as *and*, *but*, and *then*. Moreover, the qualitative analysis revealed that the English DM *so* was used to serve the same function as Japanese fillers such as *aa*, *ee*, and *uun*. Although the present study conducted a limited observation of certain DMs, these findings may be part of the evidence of L1 effects on Japanese EFL learners' DM use.

### 5.4 Summary of Study 3

The present study attempted to answer two research questions to explore influence of L1 transfer on English DM use by Japanese learners of English.

Regarding the answer to RQ3-1, the analyses using English-Japanese parallel corpus data revealed that there was not a one-to-one correspondence between English DMs and Japanese DMs. In other words, most English DMs were translated into a wide variety of Japanese DMs, while some Japanese DMs corresponded to different English DMs. The results support the findings of Mizutani (2001) and Onodera (2004), but they do not indicate that Japanese EFL learners may acquire English DMs as a result of their L1 influence.

Based on the results of the first research question, therefore, the small-scale experiment was conducted to answer RQ3-2 on the influence of L1 transfer. As mentioned in the previous section, the results suggest that the frequency of some referential and structural function markers such as *and*, *so*, and *but* may be influenced by the use of Japanese DMs or fillers. In short, Japanese EFL learners may overuse these English DMs, especially *so*, because of the influence of their L1.

Thus, the findings contribute to answering part of the questions raised in Studies 1

and 2, but the present study leaves some of them unanswered. One of the questions was why Japanese learners underuse certain DMs such as *you know*, *I mean*, and *just*, but the answer was unclear because the experiment in Study 3B did not elicit the participants' use of these DMs. In addition, another question concerns why Japanese learners prefer *so* rather than other markers when they try to think what to say next. For further research, it would be necessary to address these questions.

This study has two major limitations. First, Japanese learners' English speech in the NICT JLE Corpus is translated into accurate Japanese expressions. The translation[9] does not reflect the learners' errors and mistakes because they are corrected to expressions suitable in the context. Therefore, the L1 transfer of the misuse of DMs is out of the present study's scope. Additionally, there was a little difference between the formal-style translation and the college students' speech (i.e., casual-style speech) in Japanese expressions. Due to the difference, some casual-style Japanese DMs such as *nde* and *dakedomo* (see Table 2.5 in Chapter 2) were not included in the analyses of the present study.

The second limitation involves the speech data analysed in Study 3B. The speech data may not be sufficient to investigate the features of Japanese students' DM use both in English and in Japanese. As mentioned in Chapter 2, the data size should be larger in order to produce more reliable and valid findings. In addition, the speech data should be collected with dialogue tasks as well as monologic narrative ones because picture description may be unfit for eliciting the use of interpersonal function markers.

Despite these limitations, the findings in the present study indicate that the contrastive analyses may be complementary to those based on the CIA method used in Studies 1 and 2. In other words, the integrated approach (i.e., Integrated Contrastive Model; see section 2.5.2 in Chapter 2) may be useful to investigate L2 learners' DM

acquisition under their L1 influence.

**Endnotes**

1. There may be some potential for bias in the choice of translated words due to a translator's work.
2. In the present study, TTR is not standardized because KH Coder 2.0 does not have a function to calculate the standardized ratio. TTR is considered to be sensitive to text length (Wolfe-Quintero, Inagaki, & Kim, 1998). However, the measure was used because there were not big differences between each data set in the number of words.
3. The coding of DMs was carried out in the same way as that in Studies 1 and 2. Therefore, I did not conduct the reliability check. The same applies hereinafter.
4. Researchers are divided over whether or not lexical fillers are regarded as part of DM functions. For example, Schiffrin (1987) accepts filler expressions as DMs, but Fraser (1999, 2009) disagrees with Schiffrin's definition. In regard to *so*, Redeker (2006) and House (2010) focus on the former and confirm the use of the marker *so* to fill a pause in spontaneous talk. The present study also regards the filler *so* as one of the DM functions.
5. The Japanese expressions equivalent to English DM *next* were excluded from the analysis of Study 3A. Therefore, *next* was not analysed in Study 3B, although the DM was uttered by the students in the picture description task.
6. The student was in Class 2. He described Comic Strip A in English, followed by a narration in Japanese. In the qualitative observation, therefore, the sequence does not contribute to the influence of Japanese on his English DM use. On the other hand, the students in Class 1 explained the comic strip in reverse order. Their

performance in English may have been affected slightly by the order.

7. While the speaker in examples (16) and (17) was in Class 2, the speaker in (18) and (19) was in Class 1. In these examples, both speakers narrated in English first.
8. The Japanese filler *ee* is also equivalent to various English fillers such as *urr*, *um*, and *ah*, or the cognitive function marker *well*.
9. The translation is called back-translation in the learner corpus. However, back-translation originally means putting a translated text into the original language again (Izumi, Uchimoto, & Isahara, 2004).

# Chapter 6

# Study 4: Discourse Marker Use in Japanese EFL Textbooks for Junior High and High School Students[1]

## 6.1 Purpose of Study 4 and Research Questions

As reviewed in Chapter 2, some researchers have pointed out that language textbooks play an important role in L2 acquisition. With regard to learners' DM acquisition, Müller (2004) examined four German EFL textbooks and German learners' speech data and pointed out that the learners' DM use may be influenced by the presentation of DMs in the textbooks. However, there have been a limited number of empirical studies that have focused on the influence of textbooks on NNSs' acquisition of linguistic items including DMs.[2]

Positioned against this contextual background, the present study examines the presentation of DMs in Japanese EFL textbooks for junior high and high school students, and it explores the influence of textbooks on Japanese learners' DM use in their speech. The following research questions (RQs) are addressed:

RQ4-1: How does the presentation of DMs in textbooks differ according to grade levels?

RQ4-2: In speech, to what degree do Japanese learners of English use DMs which appear in textbooks?

RQ4-1 is intended to investigate what kind of and how many DMs are incorporated into textbooks according to the target grades, and RQ4-2 is intended to compare the results of RQ4-1 with spoken DM use at each proficiency level of Japanese learners.

## 6.2 Method

### 6.2.1 Databases

In Japan, EFL textbooks for junior high and high school students are authorized by the Ministry of Education, Culture, Sports, Science and Technology (MEXT). They are published and distributed by Japanese textbook companies. Tables 6.1, 6.2, and 6.3 show the market share.

Table 6.1
*Market Share of Japanese Junior High School English Textbooks*

| Rank | Textbook title | Publisher | Sales numbers | Market share |
|---|---|---|---|---|
| 1 | *New horizon English course* | Tokyo Shoseki | 1,635,094 | 45.1 |
| 2 | *New crown English series* | Sanseido | 730,839 | 20.2 |
| 3 | *Sunshine English course* | Kairyudo | 612,182 | 16.9 |
| 4 | *Total English* | Gakko Tosho | 359,212 | 9.9 |
| 5 | *One world English course* | Kyoiku Shuppan | 234,413 | 6.5 |
| 6 | *Columbus 21 English course* | Mitsumura Tosho | 52,184 | 1.4 |
| Total | | | 3,623,924 | 100.0 (%) |

*Note.* Based on Watanabe (2009). The sales numbers represent the number of copies sold in the school year 2010–2011.

Table 6.2
*Market Share of English I Textbooks for Japanese High School Students*

| Rank | Textbook title | Publisher | Sales numbers | Market share |
|---|---|---|---|---|
| 1 | *Crown English series I* (New ed.) | Sanseido | 129,392 | 10.1 |
| 2 | *All aboard! English I* | Tokyo Shoseki | 104,825 | 8.1 |
| 3 | *Power on English I* | Tokyo Shoseki | 97,810 | 7.6 |
| 4 | *Pro-vision English course I* (New ed.) | Pearson Kirihara | 81,438 | 6.3 |
| 5 | *Big dipper English course I* | Suken Shuppan | 74,819 | 5.8 |
| 6 | *Vista English series* (New ed.) | Sanseido | 67,873 | 5.3 |
| | The other 30 titles | | 730,407 | 56.8 |
| Total | | | 1,286,564 | 100.0 (%) |

*Note.* Based on Jiji Press (2011). English I is a compulsory elective subject which is generally taken by the first year students in high school. The sales numbers represent the number of copies sold in the school year 2011–2012.

Table 6.3

*Market Share of English II Textbooks for Japanese High School Students*

| Rank | Textbook title | Publisher | Sales numbers | Market share |
|---|---|---|---|---|
| 1 | *Crown English series II* (New ed.) | Sanseido | 119,909 | 10.6 |
| 2 | *Power on English II* | Tokyo Shoseki | 86,173 | 7.6 |
| 3 | *Pro-vision English course II* (New ed.) | Pearson Kirihara | 84,631 | 7.5 |
| 4 | *All aboard! English II* | Tokyo Shoseki | 72,212 | 6.4 |
| 5 | *Big dipper English course II* | Suken Shuppan | 66,257 | 5.9 |
| 6 | *Unicorn English course II* (New ed.) | Buneido | 55,501 | 4.9 |
| | The other 30 titles | | 685,634 | 57.1 |
| Total | | | 1,170,317 | 100.0 (%) |

*Note.* Based on Jiji Press (2011). English II is an elective subject which should be taken by the second or third year students in high school. The sales numbers represent the number of copies sold in the school year 2011–2012. *Vista English series II*, which ranks fourth in the original list, was included in the other 30 titles because it consists of two books.

To conduct an analysis of DMs in textbooks, a database was compiled from 25 EFL textbooks: 15 textbooks for junior high school students and 10 textbooks for high school students (see Table 6.4). Based on these market share data, the five top-ranked textbook series for each grade level were selected for the textbook database. The five textbook series for junior high school students accounted for 98.6 per cent of the total share. On the other hand, the five series for English I and II accounted for 37.9 and 38.0 per cent, respectively, because 36 textbook series published by 16 companies competed in the market for high school English textbooks.

The database comprised dialogues, reading passages, and spoken monologues extracted from the 25 textbooks and teachers' manuals. Word lists and exercises such as incomplete sentences were not included in the compiled data. All the texts for the database were scanned by the use of optical character recognition and converted to electronic text data.

Table 6.4

*Japanese EFL Textbook Database*

| Textbook title | Acronym | Publication year and publisher | |
|---|---|---|---|
| *New horizon English course 1* | NH1 | 2011 (2005) | Tokyo Shoseki |
| *New crown English series 1* | NC1 | 2011 (2005) | Sanseido |
| *Sunshine English course 1* | SU1 | 2011 (2005) | Kairyudo |
| *Total English 1* | TE1 | 2011 (2005) | Gakko Tosho |
| *One world English course 1* | OW1 | 2011 (2005) | Kyoiku Shuppan |
| *New horizon English course 2* | NH2 | 2011 (2005) | Tokyo Shoseki |
| *New crown English series 2* | NC2 | 2011 (2005) | Sanseido |
| *Sunshine English course 2* | SU2 | 2011 (2005) | Kairyudo |
| *Total English 2* | TE2 | 2011 (2005) | Gakko Tosho |
| *One world English course 2* | OW2 | 2011 (2005) | Kyoiku Shuppan |
| *New horizon English course 3* | NH3 | 2011 (2005) | Tokyo Shoseki |
| *New crown English series 3* | NC3 | 2011 (2005) | Sanseido |
| *Sunshine English course 3* | SU3 | 2011 (2005) | Kairyudo |
| *Total English 3* | TE3 | 2011 (2005) | Gakko Tosho |
| *One world English course 3* | OW3 | 2011 (2005) | Kyoiku Shuppan |
| *Crown English series I* (New ed.) | CR1 | 2011 (2006) | Sanseido |
| *Power on English I* | PO1 | 2011 (2006) | Tokyo Shoseki |
| *Pro-vision English course I* (New ed.) | PV1 | 2011 (2006) | Pearson Kirihara |
| *All aboard! English I* | AA1 | 2011 (2006) | Tokyo Shoseki |
| *Big dipper English course I* | BD1 | 2011 (2006) | Suken Shuppan |
| *Crown English series II* (New ed.) | CR2 | 2011 (2007) | Sanseido |
| *Power on English II* | PO2 | 2011 (2007) | Tokyo Shoseki |
| *Pro-vision English course II* (New ed.) | PV2 | 2011 (2007) | Pearson Kirihara |
| *All aboard! English II* | AA2 | 2011 (2007) | Tokyo Shoseki |
| *Big dipper English course II* | BD2 | 2011 (2007) | Suken Shuppan |

*Note.* CR1, PO1, PV1, AA1, and BD1 are textbooks for English I. Also, CR2, PO2, PV2, AA2, and BD2 are textbooks for English II. Years when the textbooks were authorized appear in parentheses.

Before conducting data analysis for the two research questions, lexical statistical features of the textbook database were computed using a concordancer, WordSmith Tools 5.0. Tables 6.5 and 6.6 list tokens, types, standardized type-token ratio (TTR), and words per sentence in each textbook for each grade level.

Table 6.5

*Lexical Statistical Features of the Textbooks for Japanese Junior High School Students*

| | Total | NH1 | NC1 | SU1 | TE1 | OW1 | NH2 | NC2 |
|---|---|---|---|---|---|---|---|---|
| Tokens | 81,963 | 3,324 | 3,799 | 5,517 | 2,277 | 4,785 | 6,313 | 5,540 |
| Types | 4,531 | 685 | 701 | 852 | 491 | 779 | 1,246 | 1,018 |
| Standardized TTR | 35.47 | 32.25 | 32.80 | 29.60 | 27.00 | 29.13 | 35.95 | 36.00 |
| Words per sentence | 5.77 | 4.94 | 4.37 | 4.64 | 4.04 | 4.25 | 6.18 | 5.48 |

| SU2 | TE2 | OW2 | NH3 | NC3 | SU3 | TE3 | OW3 |
|---|---|---|---|---|---|---|---|
| 7,853 | 3,908 | 6,146 | 5,626 | 6,294 | 8,577 | 4,736 | 7,268 |
| 1,211 | 819 | 1,160 | 1,214 | 1,155 | 1,297 | 977 | 1,301 |
| 37.70 | 34.40 | 37.24 | 36.22 | 36.77 | 34.52 | 38.67 | 37.75 |
| 6.20 | 5.46 | 5.84 | 7.47 | 6.77 | 6.76 | 6.00 | 7.51 |

*Note.* Standardized TTR was computed for every 1,000 words.

Table 6.6

*Lexical Statistical Features of the Textbooks for Japanese High School Students*

| | Total | CR1 | PO1 | PV1 | AA1 | BD1 | CR2 |
|---|---|---|---|---|---|---|---|
| Tokens | 90,000 | 12,405 | 6,000 | 9,486 | 3,694 | 6,111 | 13,045 |
| Types | 6,958 | 2,058 | 1,346 | 1,726 | 962 | 1,512 | 2,274 |
| Standardized TTR | 40.51 | 39.50 | 41.65 | 39.25 | 42.05 | 40.50 | 40.68 |
| Words per sentence | 10.06 | 9.03 | 8.75 | 11.57 | 6.55 | 8.50 | 10.61 |

| PO2 | PV2 | AA2 | BD2 |
|---|---|---|---|
| 8,960 | 15,856 | 5,262 | 9,181 |
| 1,849 | 2,753 | 1,203 | 2,039 |
| 40.59 | 40.60 | 40.33 | 41.96 |
| 11.36 | 14.17 | 7.59 | 9.63 |

*Note.* Standardized TTR was computed for every 1,000 words.

In the tables, similar features are seen in standardized TTR and words per sentence, although the figures vary slightly from title to title.

Besides the textbooks, the NICT JLE Corpus (Izumi, Uchimoto, & Isahara, 2004) was also used to answer RQ4-2. The learner corpus comprises more than 1,200

interviews from the Standard Speaking Test (SST), an interview test to measure the oral proficiency of Japanese English learners. The corpus data were originally divided into nine subcorpora according to learners' proficiency levels: novice low (level 1) to advanced (level 9). In the present study, as in Study 1, the learners' speech data were analysed at three proficiency levels (see Table 3.1 in Chapter 3).

### 6.2.2 Procedure

As in Studies 1 and 2, the present study also focuses on 57 DMs in Fung and Carter's (2007) functional paradigm, because it embraces not only syntactic properties of DMs but also those pragmatic properties common in spoken English. In the paradigm, 57 DMs were selected and classified into four functional categories: interpersonal, referential, structural, and cognitive (see Table 2.1 in Chapter 2).

In the first procedure, using WordSmith Tools 5.0, the number of occurrences of each marker was counted in the textbook database and the Japanese learner speech data. As for the items playing other grammatical roles besides those of DMs, the concordance lines were also examined to obtain frequencies for only DMs. The coding process was the same as that used in Studies 1, 2, and 3. For example, the following italicized words were manually filtered out:

> But my brother Ryo plays very *well*. (*New Horizon English Course 3*, p. 13)
>
> … if my suitcase or bag is *so* heavy, I prefer ah putting my bag down on the platform. (The NICT JLE Corpus, 00199.stt)

Next, statistical data analyses were conducted using the frequency of DMs to address the two research questions. The raw frequency of each item was standardized as

a frequency per 10,000 words and was used to calculate chi-square values for identifying the significant differences in the frequency between data sets. Frequency analyses can help find information on items more frequently encountered in input data and learners' acquisition process of linguistic forms (e.g., Ellis & Barkhuizen, 2005; Leech, 2011). Therefore, as in the previous studies, the analyses mainly drew on the frequency information.

For RQ4-1, a correspondence analysis was performed to determine the kinds of items used in textbooks at each grade level and to explore similarities and differences among 25 textbooks in the presentation of DMs. For RQ4-2, the log-likelihood ratios, as well as chi-square values, were calculated for comparison between different sized databases. Moreover, from the perspective of language acquisition, the frequency of DMs at each proficiency level of learners was compared with that in the textbooks.

In addition to these quantitative analyses, some qualitative aspects of Japanese learners' use of DMs in their speech were observed. Particular attention was paid to the degree to which Japanese learners accurately use DMs they encounter in textbooks.

## 6.3 Results and Discussion

### 6.3.1 RQ4-1: Presentation of DMs in Japanese EFL Textbooks

This section investigates the presentation of DMs in EFL textbooks for Japanese junior high and high school students. The results of the frequency analysis are displayed in Table 6.7 (see Appendices 6-A, 6-B, and 6-C). If the occurrence rate of DMs was 0.01 per cent or below in the textbook database, the items were not included in the analysis. In the overall database, the frequency of each of the 26 markers was higher than 0.01 per cent, but four of the items were not used in the textbooks for junior high school grade 1. The results show that some interpersonal or cognitive function markers

Table 6.7

*Frequency of DMs in EFL Textbooks for Japanese Junior High and High School Students*

| | | Frequency per 10,000 words | | | | | | |
|---|---|---|---|---|---|---|---|---|
| DMs | Categories | JHS1 | JHS2 | JHS3 | HS1 | HS2 | Total | Chi-square |
| *and* | Ref/Str/Cog | 67.00 | 79.30 | 112.92 | 149.88 | 142.05 | 118.80 | 140.662*** |
| *but* | Ref | 34.01 | 56.45 | 53.54 | 54.91 | 55.06 | 52.57 | 14.916** |
| *yes* | IP | 121.82 | 63.84 | 50.77 | 13.79 | 12.24 | 41.35 | 530.563*** |
| *or* | Ref | 15.23 | 17.81 | 12.00 | 24.14 | 30.21 | 21.57 | 38.751*** |
| *oh* | IP | 49.23 | 29.91 | 17.85 | 7.69 | 9.75 | 18.84 | 164.213*** |
| *I think* | IP/Cog | 2.03 | 45.03 | 27.69 | 9.02 | 11.09 | 18.61 | 189.997*** |
| *just* | IP | 6.09 | 14.78 | 21.54 | 20.69 | 14.53 | 16.28 | 24.017*** |
| *how about* | Str | 21.32 | 29.57 | 20.31 | 10.61 | 5.35 | 15.35 | 88.667*** |
| *so* | Ref/Str | 9.64 | 22.18 | 19.38 | 11.94 | 13.38 | 15.29 | 20.945*** |
| *then* | Str | 10.66 | 12.43 | 14.15 | 16.45 | 17.02 | 14.83 | 5.924 |
| *because/'cause* | Ref | 1.02 | 12.77 | 14.15 | 20.69 | 13.38 | 13.61 | 37.157*** |
| *OK/okay* | IP/Str | 22.33 | 20.83 | 14.46 | 7.69 | 3.25 | 11.57 | 80.411*** |
| *well* | IP/Str/Cog | 6.09 | 16.47 | 13.23 | 11.14 | 9.75 | 11.46 | 13.738** |
| *I see* | IP/Cog | 11.17 | 16.13 | 17.23 | 3.71 | 3.44 | 9.19 | 45.477*** |
| *really* | IP | 13.20 | 10.75 | 8.00 | 8.22 | 7.27 | 8.90 | 7.303 |
| *like* | IP/Cog | 0.00 | 5.71 | 5.85 | 10.88 | 11.85 | 8.08 | 32.873*** |
| *now* | Str | 4.57 | 8.74 | 7.69 | 5.31 | 8.41 | 7.21 | 5.922 |
| *however* | Ref | 0.00 | 1.68 | 1.85 | 10.61 | 10.13 | 6.05 | 58.239*** |
| *sure* | IP | 8.63 | 9.74 | 9.23 | 3.45 | 1.15 | 5.52 | 42.215*** |
| *right/alright* | IP/Str | 16.75 | 6.38 | 5.23 | 1.86 | 2.87 | 5.29 | 63.796*** |
| *yeah* | IP/Str | 7.61 | 1.34 | 2.15 | 1.59 | 1.53 | 2.33 | 27.261*** |
| *finally* | Str | 0.00 | 1.68 | 2.15 | 2.92 | 2.49 | 2.09 | 5.983 |
| *first* | Str | 2.03 | 2.69 | 3.08 | 1.86 | 0.76 | 1.92 | 6.841 |
| *you know* | IP/Cog | 2.03 | 2.02 | 1.85 | 1.06 | 1.53 | 1.63 | 1.341 |
| *actually* | IP | 0.00 | 1.34 | 0.31 | 0.80 | 3.63 | 1.57 | 22.103*** |
| *great* | IP | 5.08 | 1.34 | 1.85 | 1.33 | 0.38 | 1.57 | 19.389** |

*Note*. IP = interpersonal; Ref = referential; Str = structural; Cog = cognitive. JHS 1, 2, and 3 = English textbooks for junior high school grade 1, 2, and 3; HS 1 = English I textbooks for high school; HS 2 = English II textbooks for high school. ** $p < .01$. *** $p < .001$.

such as *you know* and *I mean*[3] are rarely used in the textbooks, although the frequency of *just* is relatively high. In Chapter 2, Study 2 revealed that these markers were underused by Japanese EFL learners. Therefore, the lack of input from textbooks may have an influence on Japanese learners' use of these DMs.

Table 6.7 also shows that a wider variety of DMs appear in textbooks for the second grade of junior high school students and above. The results of the chi-square tests revealed that significant differences existed among the grade levels in the frequency of 20 DMs. Additionally, while interpersonal function markers such as *yes, I think, oh, OK/okay,* and *I see* were used more frequently in junior high school textbooks than in high school textbooks, referential function markers such as *and, or, because,* and *however* were used more frequently in high school textbooks than in junior high school textbooks.

As Fung and Carter (2007) explain, DMs in the interpersonal function category are used to facilitate interaction between speakers, mainly in dialogue. For example, the marker *oh* in dialogue (1) serves a function of showing a response to speaker A's utterance.

(1) A: I play the piano.
B: *Oh,* you play the piano. (*One world English course 1*, p. 85)

In short, interpersonal markers often appear in spoken language rather than in written language. However, as shown in examples (2) and (3), referential markers often appear both in spoken and in written language.

(2) A: Why do you like it?
B: *Because* the songs are good. (*All aboard! English I*, p. 17)

(3) We had to think about the local situation, *because* looking at the situation through Western or Japanese eyes could lead us to make wrong decisions.
(*Crown English series II*, p. 37)

Therefore, to investigate in detail the features of the presentation of DMs in the textbooks, frequency analysis of these interpersonal and referential markers was also conducted in two modes: dialogue and the others (e.g., reading passages). As shown in Table 6.8, the results revealed that while the interpersonal markers were more frequently used in dialogue, the referential markers, with the exception of *however*, were often used in both modes. In other words, junior high school students can often encounter easy-to-use interpersonal markers such as *yes*, *I think*, *oh*, *OK/okay*, and *I see*, through dialogue in textbooks. Additionally, at the high school level, students can receive much input regarding the referential markers such as *and*, *or*, and *because* from both spoken and written language, especially by reading passages from textbooks.

Table 6.8
*Frequency of Some Interpersonal and Referential Markers in the Textbook Database*

| DMs | Categories | Frequency per 10,000 words | | | | Total |
|---|---|---|---|---|---|---|
| | | Junior High School | | High School | | |
| | | Dialogue (38,451 words) | the Others (43,512 words) | Dialogue (16,409 words) | the Others (73,591 words) | |
| *and* | Ref/Str/Cog | 49.15 | 125.48 | 99.95 | 155.45 | 118.80 |
| *yes* | IP | 148.50 | 5.52 | 62.77 | 1.77 | 41.35 |
| *or* | Ref | 18.20 | 11.95 | 18.28 | 29.76 | 21.57 |
| *oh* | IP | 51.49 | 10.57 | 33.52 | 3.40 | 18.84 |
| *I think* | IP/Cog | 40.05 | 17.01 | 36.57 | 4.35 | 18.61 |
| *because/'cause* | Ref | 11.44 | 9.65 | 14.63 | 16.85 | 13.61 |
| *OK/okay* | IP/Str | 35.63 | 3.68 | 24.38 | 0.82 | 11.57 |
| *I see* | IP/Cog | 32.77 | 0.00 | 19.50 | 0.00 | 9.19 |
| *however* | Ref | 0.26 | 2.30 | 0.61 | 12.50 | 6.05 |

*Note.* IP = interpersonal; Ref = referential; Str = structural; Cog = cognitive.

Finally, to explore similarities and differences among the 25 textbooks, a correspondence analysis was performed based on the frequency of DMs in the textbooks. The purpose of the analysis was to investigate the extent to which the

presentation of DMs varied from textbook to textbook. The analysis was performed on 26 markers with more than 0.01 per cent occurrence rate in the textbook database. In the analysis, the two categorical variables, the 25 textbooks and 26 DMs, were represented as points on two scatter plots: the row and column point plots (see Figures 6.1 and 6.2; Appendices 6-D and 6-E).

To output the scatter plots, the first two dimensions out of 24 dimensions were selected based on the percentage of inertia, or the proportions of variance explained (see Appendix 6-F). The two dimensions explained 50.5% and 11.1% of the total inertia. The singular values[4] were .433 and .203 respectively, and the chi-square value between the two dimensions was significant ($\chi^2$ (600, $N$ = 650) = 2609.073, $p$ < .001). These statistical values indicate that an association exists between the row and column point plots, and the two plots can display the relationship among categories. In the two-dimensional maps, if categories are closely positioned, similarities may exist between them. On the other hand, if the distance between categories is far, differences may exist in their relationships (Clausen, 1998).

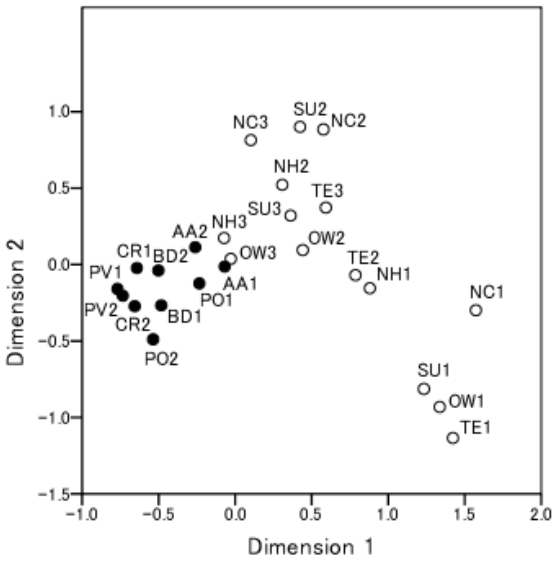



*Figure 6.1.* Correspondence analysis for the 25 textbooks: the row point plot.

Regarding Figure 6.1, although it is difficult to interpret Dimension 2, the axis of Dimension 1 is likely to be labelled the grade levels of the textbooks.[5] This is partly because the textbooks for the first grade of junior high school are found on the right side of the axis, and partly because the points of the English I and II textbooks (i.e., black

coloured points) form a cluster on the left side of the scatter plot. Additionally, the points of textbooks for the second and third grades of junior high school are located in the middle of the dimensional map. Although there some outliers exist, such as NH1, textbooks for each grade are clustered in the same area of the map. In other words, the results indicate that the horizontal dimension separates the higher level textbooks from the lower level ones in terms of DM use.

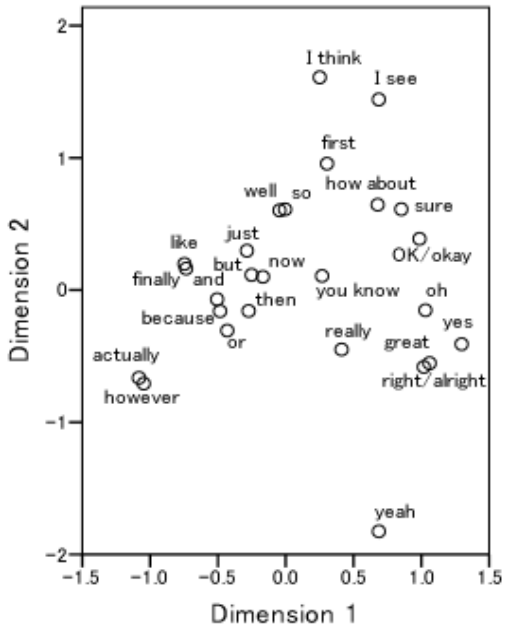



*Figure 6.2.* Correspondence analysis for the 26 DMs: the column point plot.

On the other hand, the interpretation of Dimension 1 can be applied to that of the horizontal dimension of the column point plot. In Figure 6.2, at the axis of Dimension 1, simple interpersonal markers such as *yes*, *I think*, *oh*, *OK/okay*, and *I see* are placed on the right or middle right side, while referential markers such as *and*, *or*, *because*, and *however* are on the left side of the plot. In other words, the results indicate that simple interpersonal markers are associated with textbooks for lower grades, and referential markers are related to those for higher grades.

In sum, the scatter plots can be interpreted as visualizing the results of the frequency analysis described above. Additionally, the correspondence analysis shows that the textbooks for each grade have similarities in the presentation of DMs.

### 6.3.2 RQ4-2: Input of DMs in Textbooks and Japanese Learners' Output in Their Speech

This section explores how the input data of DMs in textbooks can be transferred into learners' output in their speech. As Table 6.8 shown, while certain interpersonal markers such as *yes*, *oh*, and *OK/okay* were much more frequently used in dialogue than in non-dialogue texts, certain referential markers such as *and*, *or*, and *because* were often used in both modes. In other words, although some DMs occur mainly in either mode, others serve the same discourse function in spoken as in written language. DM inputs from non-dialogue texts, as well as from dialogue in textbooks, are likely to contribute to learners' DM use in speech communication. Therefore, the results presented in Table 6.7 were compared with results of the analysis of the use of DMs in Japanese EFL learners' speech.

Table 6.9 provides comparisons of the frequency of DMs. As in the analysis for RQ4-1, if the occurrence rate of DMs was 0.01 per cent or below in each database, the items were not included in the analysis.

The results of chi-square tests revealed that significant differences existed between the textbook database and the learner corpus data in the frequencies of 24 out of 28 DMs with more than 0.01 per cent distribution rate. Additionally, the log-likelihood ratios showed findings similar to those derived from the results of the chi-square tests. To be more specific, the tests of the log-likelihood ratios (i.e., log-likelihood tests or G-tests) revealed that 16 of the 28 DMs were used more frequently in the learners' speech than in the textbooks, and that little difference existed between the two databases in the frequency of four markers: *then*, *just*, *first*, and *finally*.[6]

Table 6.9

*Comparisons of DMs in the NICT JLE Corpus and the Textbook Database*

| | | Frequency per 10,000 words | | | |
|---|---|---|---|---|---|
| DMs | Categories | NICT JLE | Textbooks | LLR | Chi-square value |
| *so* | Ref/Str | 121.96 | 15.29 | 2409.931 | 1599.818*** |
| *yeah* | IP/Str | 76.20 | 2.33 | 2133.273 | 1235.846*** |
| *and* | Ref/Str/Cog | 250.81 | 118.80 | 1364.477 | 1162.227*** |
| *OK/okay* | IP/Str | 47.29 | 11.57 | 620.268 | 453.548*** |
| *yes* | IP | 90.24 | 41.35 | 525.015 | 437.072*** |
| *you know* | IP/Cog | 20.00 | 1.63 | 461.228 | 288.107*** |
| *I think* | IP/Cog | 44.09 | 18.61 | 298.268 | 242.806*** |
| *because/'cause* | Ref | 30.64 | 13.61 | 189.181 | 155.528*** |
| *I mean* | Cog | 6.97 | 0.70 | 150.289 | 95.979*** |
| *actually* | IP | 7.15 | 1.57 | 102.424 | 73.214*** |
| *oh* | IP | 30.92 | 18.84 | 86.126 | 76.164*** |
| *kind of* | IP | 3.02 | 0.12 | 81.867 | 47.831*** |
| *but* | Ref | 70.76 | 52.57 | 80.204 | 74.768*** |
| *or* | Ref | 32.92 | 21.57 | 69.840 | 62.873*** |
| *well* | IP/Str/Cog | 15.61 | 11.46 | 19.049 | 17.608*** |
| *really* | IP | 12.47 | 8.90 | 17.763 | 16.300*** |
| *then* | Str | 15.41 | 14.83 | 0.344 | .341 |
| *just* | IP | 15.58 | 16.28 | −0.483 | .489 |
| *first* | Str | 1.65 | 1.92 | −0.655 | .683 |
| *finally* | Str | 1.50 | 2.09 | −3.221 | 3.529 |
| *right/alright* | IP/Str | 0.34 | 5.29 | −11.911 | 13.329*** |
| *great* | IP | 0.21 | 1.57 | −49.397 | 83.023*** |
| *sure* | IP | 1.97 | 5.52 | −63.049 | 83.741*** |
| *I see* | IP/Cog | 3.88 | 9.19 | −77.816 | 98.737*** |
| *how about* | Str | 5.83 | 15.35 | −158.021 | 206.456*** |
| *like* | IP/Cog | 1.34 | 8.08 | −216.771 | 348.073*** |
| *however* | Ref | 0.46 | 6.05 | −257.850 | 474.754*** |
| *now* | Str | 0.67 | 7.21 | −278.470 | 497.795*** |

*Note*. IP = interpersonal; Ref = referential; Str = structural; Cog = cognitive. LLR = log-likelihood ratio.
***$p < .001$.

In particular, markers such as *so*, *yeah*, *and*, *OK/okay* and *yes* were preferably used by the learners. That is, the results imply that these markers are likely to be accessible to Japanese learners of English because they are relatively simple words which can often

be seen even in textbooks for the first grade of junior high school.

However, the log-likelihood tests also revealed that markers such as *now*, *however*, *like*, *how about*, and *I see* were used less frequently in the Japanese learners' speech than in the textbooks. With reference to the results in the previous section, *however* and *like* were used less frequently in junior high school textbooks than in high school textbooks (see Table 6.7), and moreover, *however* was hardly used in the textbook dialogues (see Table 6.8). Regarding *now*, *how about*, and *I see*, the frequency in the learners' speech may be influenced by the interview tasks employed to collect the data, partly because structural function markers such as *now* and *how about* are often used to organise the conversation, and partly because *I see* is mainly used to show supportive responses. In short, it may be quite difficult for interviewees, who are often questioned in a speaking test, to control their utterance by using *now*, *how about*, and *I see*.

Next, based on the results in Table 6.7 and Study 1[7], focus was placed on how the frequency of these salient items differed according to the grade levels of textbooks and according to the proficiency levels of learners. In other words, the additional analysis

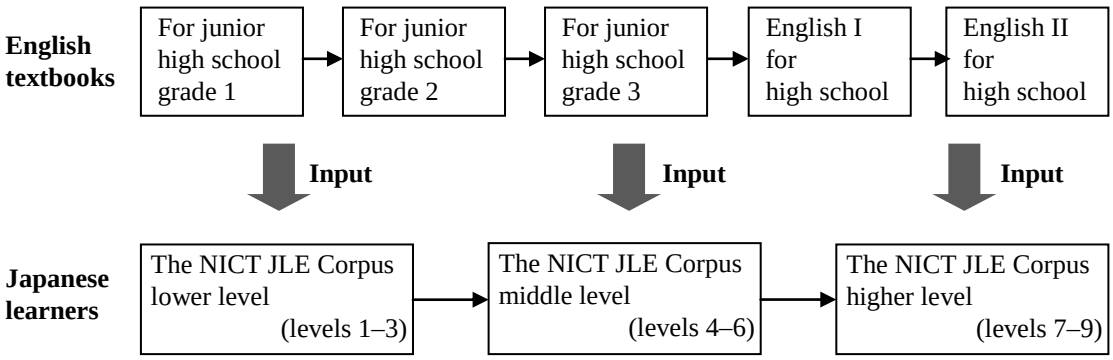



*Figure 6.3.* Design of the additional analysis of the frequency of DMs.

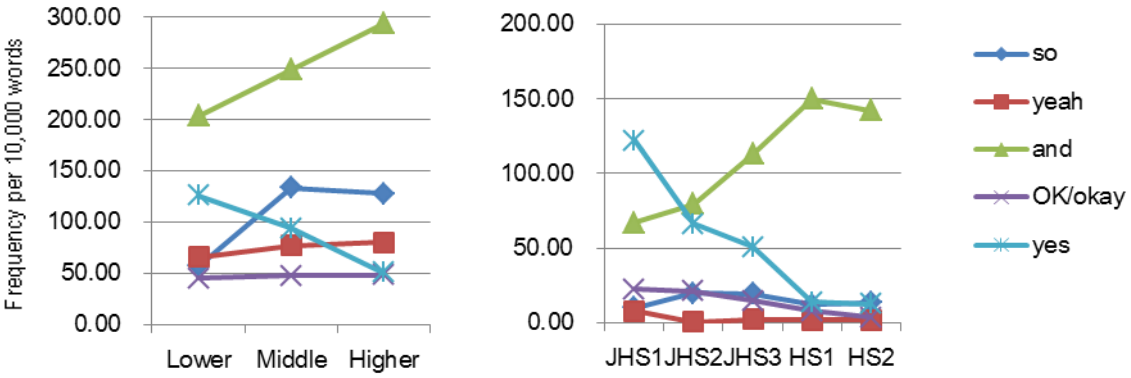



*Figure 6.4.* Changes in the frequency of five of the items which were more often used by the Japanese learners.

examined how the learners acquired the items as the level of the textbooks advanced. Figure 6.3 graphically presents the design of the analysis.

Figure 6.4 represents changes in the frequency of five of the items which were more often used by the Japanese learners. In contrast, Figure 6.5 represents changes in the frequency of five of the items which were less often used by them.

In Figure 6.4 (left), as the learners' language proficiency improved, the frequency of using *and* increased, while the frequency of using *yes* decreased. Additionally, there was almost no variation in the frequency of *yeah* and *OK/okay* regardless of the learners' language proficiency. Interestingly, the tendency of the changes was similar to that shown in the frequency of the four items in the textbooks presented in Figure 6.4 (right). That is, the use of the markers in the learners' speech was proportional to the presentation in the textbooks. However, although there was little change in the frequency of *so* among the five levels of textbooks, the frequency of the marker greatly increased in the speech data as the proficiency level improved from lower to middle. On the other hand, as shown in Figure 6.5, the frequency in the use of *now*, *however*, *like*, *how about*, and *I see* was also proportional to that in the textbooks.

To summarize these frequency analyses, the results indicate that many items

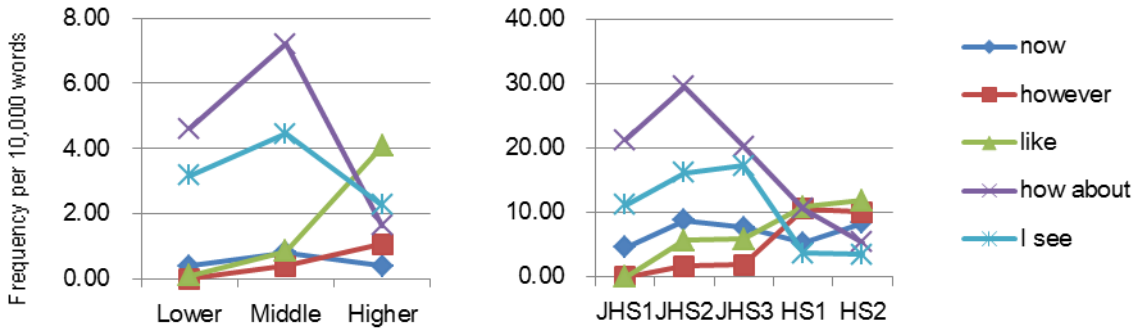



*Figure 6.5.* Changes in the frequency of five of the items which were less often used by the Japanese learners.

which often appear in textbooks may be frequently used by Japanese learners of English, although some items such as *now*, *however*, and *like* may be less common in their speech. Additionally, with the exception of some items such as *so*, the presentation of DMs in textbooks for each grade may be related to the learners' acquisition process in terms of DMs.

In addition to the frequency analyses, some examples of how the marker *so* was used in the NICT JLE Corpus were qualitatively, observed as shown below. The purpose of the observation was to explore why the Japanese learners at middle and higher proficiency levels used the marker more frequently than those at the lower level did.

(4) I go uhm *so* eeto urr ground pitch.

(The NICT JLE Corpus, 00140.stt)

(5) They are always ah chose the same person to elect *so* ah chairperson … or mm mayor man …

(The NICT JLE Corpus, 00340.stt)

(6) … I want to go to many places in the world, *so* I want to see the many kinds of culture.

(The NICT JLE Corpus, 00818.stt)

(7) … this location seems to be very erm peaceful and er not busy at all. *So* the Hiyoshi is more ah Hiyoshi is busi hum more crowd.

(The NICT JLE Corpus, 00842.stt)

While example (4) was extracted from utterances by Japanese learners with a low level proficiency in English, examples (5), (6), and (7) were extracted from utterances by those with a middle level proficiency. In examples (4) and (5), the marker *so* was used as a filler to take some time to find words following each verb, and also, especially in example (4), the learner may have confused the English *so* with the Japanese *so*[8], because the subsequent word *eeto* was a Japanese word equivalent to the marker *well*. As with the English *so*, the Japanese *so* is also used to signal the sequence of discourse segments in the thinking process.

Regarding examples (6) and (7), each learner made an error or mistake in the use of *so*. In other words, *so* in example (6) should be replaced with *and* or *because*, and *so* in example (7) should be changed to a marker suitable for the context: *but* or *however*. Although there were none of these errors or mistakes in speech data at the novice low and middle levels (i.e., SST levels 1 and 2), some learners at the novice high, intermediate, and advanced levels (i.e., SST levels 3–9) made such errors or mistakes.

These observations were drawn from randomly sampled speech data. The statistics are displayed in Table 6.10. In the sample, it was notable that learners at a middle level frequently used *so* as a filler. Thus, the filler use may boost the frequency of *so* at a middle proficiency level. The results also indicate that they may attempt to strategically use the marker to continue their utterance. On the other hand, at a higher proficiency level, the proportion of *so* used as a filler decreased from 40.1% to 27.8%.

This result implies that higher level learners may have a decreased need for using fillers, or they may use other markers such as *well* and *you know* instead of *so*, because they can use a variety of DMs. In addition, as shown in Table 6.10, the more proficient learners became in speech, the more accurately they used the marker. However, some

Table 6.10
*Raw Frequency of* so *in the Random Sample of the NICT JLE Corpus*

| | Correct use | Fillers | E or M | Total |
|---|---|---|---|---|
| Lower (levels 2–3): 8,749 words | 21 (55.3%) | 11 (28.9%) | 6 (15.8%) | 38 (100.0%) |
| Middle (levels 4–6): 18, 694 words | 122 (52.6%) | 93 (40.1%) | 17 (7.3%) | 232 (100.0%) |
| Higher (levels 7–9): 24,729 words | 200 (68.7%) | 81 (27.8%) | 10 (3.4%) | 291 (100.0%) |
| Total: 52,172 words | 343 (61.1%) | 185 (33.0%) | 33 (5.9%) | 561 (100.0%) |

*Note.* In the analysis, five interview data were randomly sampled from each of the subcorpora, except for levels 1 and 2. That was because there was no occurrence of the marker *so* in the learner data at level 1. Thus, I analysed concordance lines of 10 interview data at level 2. E or M = learners' errors or mistakes.

learners at SST levels 3–9 confused the function of *so* with that of other markers such as *because*. This result suggests that an increasing number of new items may be given to learners as their grade in school advances, overloading them with the input data from textbooks.

### 6.4 Summary of Study 4

This study investigated the presentation of DMs in 25 Japanese EFL textbooks for junior high and high school students, and it compared the results with the distribution of DMs in Japanese EFL learners' speech.

Regarding the answer to RQ4-1, the frequency analyses revealed that while

interpersonal function markers such as *yes*, *I think*, *oh*, *OK/okay*, and *I see* were used more frequently in lower level textbooks, referential markers such as *and*, *or*, *because*, and *however* were used more frequently in higher level textbooks. The results indicate that junior high school students can encounter the easy-to-use interpersonal markers which often appear in dialogue, and that high school students can receive much input on the referential markers mainly from reading passages in textbooks, because English I and II[9] textbooks prioritize written language. In addition, these findings were supported by those of the correspondence analysis based on the frequencies of DMs. The correspondence analysis revealed similarities among the textbooks for each grade. Therefore, this result suggests that materials designers or writers may select items depending on the target grades.

With regard to the answer to RQ4-2, the frequency analyses revealed that many items were more frequently used by Japanese learners of English than in the textbook database. Noted among the findings were that relatively simple markers such as *so*, *yeah*, *and*, *OK/okay*, and *yes* were preferably used by the learners, and that the number of less frequent items such as *now*, *however*, and *like* was small. Therefore, given the results of the analyses for RQ4-1, the findings indicate that learners may prefer high frequency items in textbooks and output this input data to communicate with others in speech. In other words, the results from previous studies such as Tono (2002) and Ota et al. (2003) are likely to hold true with the relationship between input and output of DMs in speech (see section 2.4.1 in Chapter 2).

However, some items such as *so* may be influenced by something other than textbooks, partly because the frequency in the use of *so* was not proportional to that in the textbooks, and partly because the qualitative observation of *so* showed that Japanese learners often used the marker as a filler. As found in Study 3, therefore, the filler use is

likely to be influenced by their L1. Among the fillers, as in Study 1 (see section 3.3.2.2), *so* in example (4) may be influenced by mixing up English with Japanese *so* due to the similarities in function and pronunciation. Moreover, the observation revealed that some learners often confused the discourse function of *so* with other items' functions. The finding indicates such errors or mistakes are also likely to occur in the use of other markers. Thus, it would be necessary to examine the L1 transfer and errors of DMs in further extended research. The results could contribute to remarkable findings in the language acquisition of Japanese EFL learners.

Finally, I will point out three major limitations of the study with directions for further research. The first concerns the size of the textbook database. While five of six textbook series for junior high school use were analysed, only five of 36 English I and II textbook series were included in the database. Thus, more high school textbooks should be analysed to obtain a clearer picture of the presentation of DMs in textbooks.

The second limitation relates to the influence of textbooks on DM use by learners. In the present study, it should be noted that a direct causal linkage between the items in textbooks and learners' speech performance cannot be confirmed, because the NICT JLE Corpus users are not given information on course materials learners used in school settings. Additionally, although the corpus mainly includes speech data of young and adult learners aged 15–29, it also includes those of some learners in their 30's and 40's. Among available databases, the choice of the NICT JLE Corpus would be a good compromise to estimate how input from textbooks can be transferred into learners' output because the corpus is comprised of a great deal of speech data at different proficiency levels of English. However, further research is needed to use speech data more suitable for investigating the direct effect of textbooks.

The last limitation involves the influence of different topics of tasks on learners'

DM use. In SST, although all interviewees work on the five tasks, the topics vary according to their proficiency level. For example, while lower level learners are asked to talk about familiar topics such as shopping and seeing movies, higher level learners deal with unfamiliar topics such as car accidents. With the consideration of topic differences, learners' speech data should be analysed from a qualitative perspective.

Despite these limitations, the present study suggests that textbooks may have an important effect on learners' acquisition process in terms of DMs. Especially in speech, during the short period available to think of what to say, learners tend to utilize easy-to-use DMs which can be frequently encountered in textbooks. However, despite obtaining input from textbooks, learners also use some items in error, or infrequently use them in their speech. Thus, based on empirical data such as the results obtained in the present study, materials designers or writers should reconsider what kind of and how many of DMs should be included in textbooks. In other words, it is important to provide learners with appropriate items according to their proficiency level or readiness to accept new input items. Additionally, language instructors should make full use of items which appear in textbooks with careful attention to their students' accuracy in using DMs, and provide the students with rich opportunities to use them in a variety of contexts. As mentioned above, students who take English I or II course may not have many opportunities to encounter DMs in dialogue contexts. Therefore, in order to enable such students to understand the functions and roles of DMs in speech, instructors should supplement textbooks with other teaching materials or communication activities.

**Endnotes**

1. An earlier version of this chapter has been published in Shimada (2013).
2. As reviewed in Chapter 2, Lam (2010) investigated the presentation of DMs in

textbooks for upper-secondary students in Hong Kong and compared it with the distribution of DMs in Hong Kong's advanced English learners' speech data. The results revealed that significant differences existed between the textbook database and the speech data in DM use, but little discussion was conducted on how the textbook dialogue was reflected in learners' speech.

3. The occurrence rate of the marker *I mean* was 0.008 per cent in the textbook database. Therefore, the item was not included in the analysis.
4. In a correspondence analysis, singular values are equivalent to correlation coefficients between the row and column points (Clausen, 1998). In the present study, therefore, the singular value for Dimension 1 (.433) showed that the row points were positively correlated with the column points.
5. While Dimension 1 explained 50.5% of the total inertia, Dimension 2 explained only 11.1%. "In interpreting a dimension, the points with highest values are awarded the greatest importance" (Clausen, 1998, p. 51). In general, therefore, if the percentage of inertia is low, the dimension will be uninterpretable.
6. In the present study, when the ratio was +3.84 or more, the item was considered to be used more frequently in the NICT JLE Corpus than in the textbook database. On the other hand, when the ratio was a negative number with an absolute value 3.84 or more, the item was considered to be used less frequently in the learner corpus.
7. See Table 3.2 in Chapter 3. The table illustrates the frequency of DMs in the NICT JLE Corpus at each level of learner proficiency.
8. Study 1 also pointed out the use of *so* as a filler and confusion with the first language. As for the Japanese *so*, Sadanobu (2002) argues that the item has the discourse function of a filler although it also plays a major role in providing back channel feedback in casual conversations (see endnote 9 in Chapter 3).

9. With the new curriculum introduced by MEXT, instead of English I and II courses, English Communication courses will be phased in by 2015.

# Chapter 7

## Study 5: Discourse Marker Use in EFL Textbooks and Materials Design[1]

### 7.1 Purpose of Study 5 and Research Questions

In the last chapter, Study 4 suggested that Japanese EFL textbooks may have an effect on Japanese junior and high school students' acquisition of DMs. Therefore, the findings of the study support that textbooks are primary resources for learners in the classroom. However, there have been few studies considering the kind and frequency of DMs that should be incorporated in textbooks for Japanese EFL learners.

As reviewed in Chapter 2, many previous studies have revealed that differences between the language of textbooks and native English speech data, but it is controversial whether the language data of NSs are an ideal model for textbook language.

With regard to L2 materials design, many researchers have stressed that the richness of authentic input is beneficial in aspects of affective factors—i.e., motivation, empathy, and emotional involvement—because authentic texts reflect actual situations in the target language speaking community (Mishan, 2005). McCarthy (1991) also argues that naturally occurring structures and vocabulary should be taught in order to prepare students for real-life conversations. The enthusiasm for authentic text usage in teaching materials may suggest that texts written or spoken by NSs should be regarded as the one and only model for textbook language.

However, there has been little empirical research to support the positive effect of using authentic texts in language learning (Gilmore, 2007; Wardman, 2009). Additionally, some researchers (e.g., Feng & Byram, 2002; Kubota, 1997; Richards, 2006) have questioned native-speaker language norms on the ground that English is

used as a lingua franca in international contexts and that there is a burden placed on learners to comprehend unsimplified texts produced for a native-speaking audience.

Therefore, the present study considers the issue of textbook language as a primary input source for learners in terms of DMs. In the previous chapters, the findings of Studies 1, 2, and 3 identified some features of Japanese EFL learners' acquisition of DMs. In addition, the present study addresses the following research questions (RQs) and considers DM input appropriate for Japanese EFL learners.

RQ5-1: How different is the presentation of DMs in EFL textbooks' dialogue from its distribution in NSs' speech data?

RQ5-2: How different is the presentation of DMs in EFL textbooks' dialogue from its distribution in Japanese English learners' speech data?

RQ5-3: What are the features of the presentation of DMs in EFL textbooks' dialogue designed for Japanese English learners?

RQ5-1 and 5-2 are intended to examine empirically the presentation of DMs in EFL textbooks, which account for a significant portion of language learners' input data; RQ5-3 is intended to explore how textbooks should take into consideration the features of Japanese English learners' interlanguage.

## 7.2 Method

### 7.2.1 Databases

In order to conduct an analysis of DMs in textbooks, a database was compiled from nine EFL textbooks available in Japan: five EFL textbooks designed for Japanese college students and four EFL textbooks for sale on the international market (see Table

7.1).

Table 7.1
*EFL Textbook Database*

| Textbook title | Acronym | Publication year and publisher | |
|---|---|---|---|
| *Communication builder* (Rev. ed.) | CBD | 2009 | Nan'un-do |
| *Hear it! say it!* | HIT | 2003 | Kinseido |
| *Everyday talk* | EDT | 2001 | Asahi Press |
| *Passport 2* (2nd ed.) | PAS | 2010 | Oxford University Press |
| *Primary listening* | PRL | 2003 | Kinseido |
| *Join in 2* | JOI | 2009 | Oxford University Press |
| *New American inside out pre-intermediate* | NAM | 2009 | Macmillan |
| *New cutting edge pre-intermediate* | NCE | 2005 | Pearson Education |
| *Touchstone 3* | TOU | 2006 | Cambridge University Press |

*Note.* CBD, HIT, EDT, PAS, and PRL are written exclusively for Japanese college students, whereas JOI, NAM, NCE, and TOU, sold worldwide, are designed to meet the requirements of adult and younger adult learners. The teacher's books for PAS, NAM, and NCE were not examined, because all the transcripts are included at the end of the student's books.

The nine textbooks are new or popular titles since 2001, published by major textbook companies based in Japan, the UK, and the USA.[2] According to the book catalogues and brochures, the approximate levels of these textbooks are set at pre-intermediate, which is equivalent to a TOEIC® score of 380 to 520.

CBD is a textbook designed to improve Japanese English learners' communication skills through pair-work activities. Each chapter provides model dialogues and exercises for pair-work activities. Similarly, HIT and EDT also focus on developing communication skills through spoken English. The former gives both listening exercises and conversation practice in each chapter, while the latter places importance on a balance of listening, speaking, writing, and grammar in each chapter. PRL is a listening textbook with communication practice. It consists of dialogues, news,

speeches, and announcements in various practical situations such as shopping and overseas travel. CBD, HIT, EDT, and PRL are written in American English and published in Tokyo.

PAS is a new title launched by a well-known international publisher based in the UK, but the textbook is a second edition written for Japanese EFL learners. This collection of a two-page spread format, consisting of 20 units, presents listening and communication activities on travelling overseas. The number of units and pages is probably designed to accord with syllabuses for Japanese college classes.[3] In the textbook, both British and American English are used.

Four overseas textbooks were also selected to form part of the database in the present study. JOI, written in American English, focuses on learning communication skills through natural spoken English and provides model dialogues, listening exercises, and communication practice with strategies. NAM and NCE are integrated four-skills courses for adult and younger adult learners. NAM emphasises vocabulary and grammar exercises as well as listening and speaking practice. NCE uses a task-based approach to develop learners' communicative competence. The former is written in American English, while the latter is written in British English.

TOU is probably the most authentic of the four overseas textbooks, since it draws on the analysis of spoken and written texts in the Cambridge International Corpus of North American English.[4] Therefore, the most frequent vocabulary and grammar items in American English are included in the textbook. The corpus-based textbook places emphasis on conversational grammar used in spoken English. Rühlemann (2009) describes it as "a milestone in the history of English textbooks" (p. 421).

The textbook database was made up of dialogue sections extracted from the nine textbooks and the teacher's books or manuals. Thus, narrative passages and spoken

monologues were excluded from the compiled data. All the spoken texts were scanned by the use of optical character recognition and converted to electronic text data.

As in the previous chapters, prior to data analysis for the three research questions, lexical statistical features of the textbook database were confirmed by using WordSmith Tools 5.0. Table 7.2 describes tokens, types, standardized type-token ratio (TTR), and words per sentence in each textbook.

The table shows that there is variability in standardized TTR and words per sentence among the nine textbooks. Although the publishers determined the vocabulary and sentence levels to be of pre-intermediate level, they vary somewhat from textbook to textbook. While secondary school textbooks follow the guidelines of the government, each textbook selected in the present study was edited in accordance with the publisher's policy or the materials designers' views on the language of textbooks. Thus, it was difficult to collect textbooks at the same linguistic level, but the nine textbooks containing more than 20 items of DMs each were selected to create the database in the present study.[5]

In addition to the textbooks, the following three spoken corpora analysed in Study 1 were also used to address the research questions in the present study: (a) the NICT JLE Corpus (Izumi, Uchimoto, & Isahara, 2004), (b) the British National Corpus (BNC)

Table 7.2
*Lexical Statistical Features of the Textbooks*

| | Total | CBD | HIT | EDT | PAS | PRL | JOI | NAM | NCE | TOU |
|---|---|---|---|---|---|---|---|---|---|---|
| Tokens | 66,084 | 3,816 | 5,761 | 8,634 | 6,523 | 5,950 | 12,566 | 7,335 | 4,685 | 10,814 |
| Types | 4,526 | 812 | 881 | 1,633 | 968 | 1,378 | 1,510 | 1,364 | 924 | 1,411 |
| Standardized TTR | 35.52 | 31.47 | 38.47 | 42.04 | 31.88 | 39.62 | 31.55 | 36.96 | 34.75 | 34.52 |
| Words per sentence | 6.96 | 7.28 | 5.37 | 7.87 | 4.66 | 6.85 | 5.44 | 6.25 | 7.25 | 6.63 |

*Note.* Standardized TTR was computed for every 1,000 words. The higher the ratio is, the greater the lexical variety is.

XML Edition, and (c) Child Language Data Exchange System (CHILDES) (MacWhinney, 2000). The NICT JLE Corpus consists of transcriptions (1,536,000 words) of more than 1,200 oral interviews with Japanese EFL learners. With regard to NS data, adult speech data (1,051,215 words), collected from the spoken part of the BNC, were transcribed from dialogues among British and American speakers aged 25–59. Additionally, child speech data (168,135 words) were collected from dialogues uttered by British and American children aged 4–8.

Although the tokens of the nine textbooks were far fewer than those of the spoken corpora, the textbook database was compiled to be representative of a particular linguistic item (i.e., DMs). As mentioned in Chapter 2, corpus size is a controversial issue in corpus linguistics, but what matters is whether the specific features of the observed language are represented in the corpus (Aston, 1997; Römer, 2004). Even such a small database can reveal various features about the presentation of DMs in EFL textbooks for Japanese college students and adult learners.

### 7.2.2 Procedure

The present study also focuses on 57 DMs in Fung and Carter's (2007) framework. As in the previous chapters, the first procedure was to count the number of occurrences of each item for pragmatic function using corpus software[6] while looking at the concordance lines. Some examples are as follows:

Pragmatic function:

*Well*, how about an action movie? (Richards & O'Sullivan, 2009a, p. 6)

I'm really interested in shodo, *you know*, the art of calligraphy.

(Tsuda, 2003a, p. 38)

Other function (i.e., the use of non-DMs):

Yeah, I'm not feeling very *well*. (Cunningham & Moor, 2005, p. 170)

*You know* Brad Pitt? He's my brother-in-law. (Yoshitomi, 2009a, p. 25)

Next, a frequency analysis of the 57 DMs was carried out. Frequency analysis can give materials writers and language teachers information on the most common words (Fox, 1998), and gives them insight into learners' acquisition of the linguistic forms of a second language (Ellis & Barkhuizen, 2005). Therefore, as in the previous chapters, the frequency analysis on DMs is considered an important method to address RQ5-1 and RQ5-2.

In order to provide an answer to RQ5-1, the frequency of each marker in the textbook database, the BNC, and CHILDES was calculated and standardized as a frequency per 10,000 words. Additionally, the log-likelihood ratio and chi-square value were used to find significant differences between the textbook database's presentation of DMs and that of the NS data. As for RQ5-2, the same procedure as for RQ5-1 was followed. That is to say, in comparisons between the textbook database and the Japanese EFL learner corpus for the frequency of each marker, the log-likelihood ratio and chi-square value were calculated.

Finally, the data analysis for RQ5-3 took the form of a correspondence analysis that was used to investigate similarities and differences among the nine textbooks in the presentation of DMs, and to explore the features of materials design in EFL textbooks tailored to the Japanese institutional market. In corpus linguistics studies, the correspondence analysis of vocabulary[7] has been applied to explore similarities and differences in the frequency between categorical variables.

In addition to these quantitative analyses, I also focus attention on qualitative

aspects of the use of DMs in the textbooks. Some qualitative observations can be considered to be a complement of the quantitative analyses.

## 7.3 Results and Discussion

### 7.3.1 RQ5-1: Comparison Between Dialogue in EFL Textbooks and Speech Data of NSs

This section investigates the differences between the dialogue in the nine textbooks and that in the speech data of NSs. First, comparisons of DMs in the textbooks and CHILDES were made as shown in Table 7.3. If the occurrence rate of DMs was 0.01 per cent or below in each database, the items were not included in the analysis.

The results of chi-square tests revealed that there were significant differences between the textbook database and CHILDES in the frequency of DMs with six exceptions including *first* and *now*. Additionally, the log-likelihood ratios were represented as positive number, indicating that most markers were used more frequently in the textbooks than in the spoken data of NS children. In other words, the results suggest that the language level of the textbooks may exceed that of NS children in terms of DMs. As found in Study 1, NS children may have yet to acquire many kinds of DMs because they are still in the acquisition stage of L1 (see section 3.3.2.1 in Chapter 3). Thus, the findings of this analysis confirm that the textbooks are designed for adult learners such as college students.

Table 7.3

*Comparisons of DMs in the Textbooks and CHILDES*

| | | Frequency per 10,000 words | | | |
|---|---|---|---|---|---|
| DMs | Categories | Textbooks | CHILDES | LLR | Chi-square value |
| *really* | IP | 58.26 | 8.68 | 446.512 | 515.446*** |
| *I see* | IP/Cog | 19.07 | 0.12 | 299.585 | 311.813*** |
| *how about* | Str | 19.97 | 0.71 | 259.394 | 286.425*** |
| *OK/okay* | IP/Str | 44.64 | 9.75 | 256.877 | 295.188*** |
| *sure* | IP | 16.34 | 0.30 | 235.671 | 253.273*** |
| *actually* | IP | 14.38 | 0.95 | 159.464 | 180.469*** |
| *oh* | IP | 80.20 | 39.73 | 139.271 | 152.691*** |
| *so* | Ref/Str/ | 39.80 | 15.17 | 116.646 | 130.427*** |
| *well* | IP/Str/Cog | 57.96 | 30.04 | 89.832 | 97.917*** |
| *what about* | Str | 12.11 | 3.45 | 53.119 | 60.365*** |
| *yes* | IP | 67.04 | 42.82 | 52.780 | 56.291*** |
| *great* | IP | 3.93 | 0.12 | 52.713 | 57.775*** |
| *but* | Ref | 71.12 | 47.16 | 47.791 | 50.765*** |
| *exactly* | IP | 4.09 | 0.36 | 41.012 | 46.822*** |
| *I think* | IP/Cog | 52.36 | 34.73 | 35.125 | 37.255*** |
| *right/alright* | IP/Str | 16.49 | 7.61 | 33.676 | 37.016*** |
| *oh great* | IP | 2.42 | 0.12 | 29.259 | 32.719*** |
| *anyway* | Ref | 5.30 | 1.37 | 25.917 | 29.570*** |
| *I mean* | Cog | 17.86 | 9.99 | 22.304 | 24.056*** |
| *or* | Ref | 24.97 | 16.77 | 15.829 | 16.729*** |
| *you know* | IP/Cog | 33.59 | 24.74 | 13.085 | 13.682*** |
| *kind of* | IP | 6.05 | 2.74 | 12.921 | 14.220*** |
| *absolutely* | IP | 1.97 | 0.48 | 10.292 | 11.769** |
| *just* | IP | 37.53 | 34.50 | 1.219 | 1.238 |
| *you see* | IP | 3.93 | 3.45 | 0.305 | 0.311 |
| *sort of* | IP/Cog | 1.21 | 1.07 | 0.082 | 0.084 |
| *finally* | Str | 1.06 | 1.01 | 0.011 | 0.011 |
| *first* | Str | 1.36 | 1.49 | −0.052 | 0.051 |
| *now* | Str | 8.32 | 10.41 | −2.173 | 2.103 |
| *like* | IP/Cog | 11.95 | 19.86 | −18.204 | 16.865*** |
| *yeah* | IP/Str | 44.64 | 101.64 | −204.908 | 181.721*** |
| *then* | Str | 19.82 | 63.88 | −214.955 | 179.878*** |
| *because/'cause* | Ref | 8.17 | 56.32 | −344.128 | 258.473*** |
| *and* | Ref/Str/Cog | 183.71 | 374.04 | −599.350 | 554.262*** |

*Note.* If the occurrence rate of DMs was zero per cent in either corpus, they were excluded from this analysis due to the impossibility of computing the log-likelihood ratio (LLR). IP = interpersonal; Ref = referential; Str = structural; Cog = cognitive. **$p < .01$. ***$p < .001$.

Second, comparisons of DMs in the textbooks and the BNC were made as shown in Table 7.4. The method of analysis was the same as that used in the previous analysis. If the occurrence rate of DMs was 0.01 per cent or below in each database, the items were not included in the analysis.

The results of chi-square tests revealed that there were significant differences between the two databases in the frequency of 31 DMs. Additionally, it was noticeable that a representation of the log-likelihood ratios was clearly divided into two clusters: half the items were calculated as a positive number; half as a negative number. If the log-likelihood ratio applied to the two databases was represented as a positive number, the marker was used more frequently in the textbook database than in the BNC. On the other hand, when the ratio was negative, NSs used the item more frequently. That is to say, the results indicate that there is a considerable discrepancy between textbook dialogue and native adults' speech in the distribution of DMs. Many items that are often used by NSs tend to appear less frequently in textbooks.

In particular, the analysis revealed that native adults used referential or cognitive function markers such as *because/'cause*, *and*, *or*, and *I mean* more frequently than did the textbooks. As for the interpersonal function category, while markers such as *sort of* and *right/alright* were also used more frequently in the BNC than in the textbook database, markers such as *oh*, *really*, *sure*, *I see*, and *yes* were used less frequently in the speech data of adult NSs.

The analyses in this section demonstrated some similarities in DM use between NS children and adults. One of the similarities was that both groups used referential or structural function markers such as *and*, *because/'cause*, *then*, and *now* more frequently than the textbooks. Additionally, interpersonal markers such as *really*, *I see*, and *sure* were used less frequently in the two corpora of NSs than in the textbook database.

Table 7.4

*Comparisons of DMs in the Textbooks and the BNC*

| | | Frequency per 10,000 words | | | |
|---|---|---|---|---|---|
| DMs | Categories | Textbooks | BNC | LLR | Chi-square value |
| *oh* | IP | 80.20 | 19.60 | 623.583 | 987.431*** |
| *really* | IP | 58.26 | 13.18 | 491.821 | 798.213*** |
| *sure* | IP | 16.34 | 0.52 | 409.083 | 1066.735*** |
| *how about* | Str | 19.97 | 1.56 | 359.645 | 796.031*** |
| *I see* | IP/Cog | 19.07 | 3.84 | 180.490 | 303.910*** |
| *what about* | Str | 12.11 | 2.74 | 102.196 | 165.654*** |
| *yes* | IP | 67.04 | 40.73 | 87.194 | 102.143*** |
| *well* | IP/Str/Cog | 57.96 | 34.37 | 82.263 | 97.064*** |
| *great* | IP | 3.93 | 0.48 | 55.491 | 109.356*** |
| *OK/okay* | IP/Str | 44.64 | 27.48 | 55.294 | 64.413*** |
| *oh great* | IP | 2.42 | 0.11 | 53.708 | 132.050*** |
| *to be honest* | IP | 4.09 | 0.71 | 43.947 | 77.458*** |
| *so* | Ref/Str | 39.80 | 25.62 | 41.265 | 47.378*** |
| *first* | Str | 1.36 | 0.04 | 35.339 | 93.654*** |
| *kind of* | IP | 6.05 | 1.90 | 34.337 | 49.870*** |
| *but* | Ref | 71.12 | 59.46 | 13.335 | 14.156*** |
| *exactly* | IP | 4.09 | 2.66 | 4.013 | 4.578* |
| *finally* | Str | 1.06 | 0.49 | 2.952 | 3.753 |
| *anyway* | Ref | 5.30 | 4.06 | 2.114 | 2.292 |
| *actually* | IP | 14.38 | 14.95 | −0.141 | 0.140 |
| *I think* | IP/Cog | 52.36 | 53.88 | −0.270 | 0.270 |
| *absolutely* | IP | 1.97 | 2.52 | −0.822 | 0.766 |
| *just* | IP | 37.53 | 40.48 | −1.372 | 1.347 |
| *yeah* | IP/Str | 44.64 | 48.83 | −2.305 | 2.257 |
| *like* | IP/Cog | 11.95 | 15.36 | −5.118 | 4.772* |
| *however* | Ref | 0.30 | 1.31 | −7.173 | 5.064* |
| *basically* | IP | 0.30 | 1.73 | −11.438 | 7.708** |
| *you see* | IP | 3.93 | 7.80 | −14.750 | 12.283*** |
| *now* | Str | 8.32 | 14.17 | −17.817 | 15.416*** |
| *then* | Str | 19.82 | 29.29 | −21.634 | 19.450*** |
| *obviously* | IP | 0.76 | 4.26 | −27.895 | 18.848*** |
| *you know* | IP/Cog | 33.59 | 48.31 | −31.419 | 28.507*** |
| *or* | Ref | 24.97 | 39.41 | −38.180 | 33.769*** |
| *and* | Ref/Str/Cog | 183.71 | 236.65 | −80.220 | 76.419*** |
| *I mean* | Cog | 17.86 | 37.67 | −81.586 | 67.128*** |
| *because/'cause* | Ref | 8.17 | 25.72 | −103.379 | 77.782*** |

(continued)

Table 7.4 (continued)

| DMs | Categories | Frequency per 10,000 words | | LLR | Chi-square value |
|---|---|---|---|---|---|
| | | Textbooks | BNC | | |
| *right/alright* | IP/Str | 16.49 | 42.47 | −130.811 | 102.883*** |
| *sort of* | IP/Cog | 1.21 | 16.13 | −150.303 | 90.929*** |

*Note.* If the occurrence rate of DMs was zero per cent in either corpus, they were excluded from this analysis due to the impossibility of computing the log-likelihood ratio (LLR). IP = interpersonal; Ref = referential; Str = structural; Cog = cognitive. *$p$ < .05. **$p$ < .01. ***$p$ < .001.

Therefore, as most previous studies have revealed, the quantitative analysis in the present study also indicates that there is a large discrepancy between textbook dialogue and natural conversation. In addition to the analysis, the following examples of how DMs were used in the textbooks and the BNC were qualitatively observed.

(1) … but you you should be *sort of* practising equations all the time.

(The BNC XML Edition, FYA 1817)

(2) But also it would be *kind of* awkward for me to leave before my boss or superiors. (Tsuda, 2003b, p. 22)

(3) Yeah. I *kind of* got tired of parties and clubs and everything.

(McCarthy, McCarten, & Sandiford, 2006b, p. 240)

As shown in (1), the interpersonal function marker *sort of* has "the interactive effect of softening the tone through an element of vagueness" (Fung & Carter, 2007, p. 419), and was ubiquitous in the speech data of NSs. However, *sort of* was used less frequently in the textbook database than in the BNC. Instead, in (2) and (3) extracted from the EFL textbooks, *kind of* serves as the same discourse function as *sort of* does. Accordingly, when different items can fulfil the same function in a discourse, DMs

presented in textbooks may be different from those frequently used by NSs.

Next, the study focuses on aspects of the marking functions of *right/alright*, partly because there was a significant difference between the two databases in the frequency of their occurrences, and partly because *right/alright* has two functions: interpersonal and structural.

(4) A: Carbonate plus an acid?

B: *Right*, like we poured stuff … Salt and water.

(The BNC XML Edition, FMR 1405)

(5) A: Look. Some of them are crying.

B: *Right*. And some of them are covering their ears.

(Richards & O'Sullivan, 2009b, p. 82)

(6) Any questions for me before we finish? *Right*, quarter of an hour coffee break then please. (The BNC XML Edition, JK7 1100)

In the interpersonal domain, *right/alright* is used to provide a speaker's responses to another speaker's utterances. For example, in dialogues (4) and (5), the speaker B gives positive feedback to the speaker A. In the structural domain, *right/alright* serves as a signal to open or close topics in conversation (Fung & Carter, 2007). As shown in (6), the use of *right* indicates the speaker's intention to close the question-and-answer session and let the listeners take a break.

In the frequency analysis on the functional categories, it is particularly worth noting that the structural marker *right/alright* was rarely used in the textbook database. The raw frequency was only one, although the total occurrence of *right/alright* as a DM was 109 in the textbooks (see Table 7.5).

Table 7.5

*Raw Frequency of* right/alright *as a DM in the Textbooks and the Random Sample of the BNC*

| | Raw Frequency of *right/alright* | | |
|---|---|---|---|
| Databases | Interpersonal Marker | Structural Marker | Total |
| Textbooks | 108 (99.1) | 1 (0.9) | 109 (100.0) |
| BNC Sample | 146 (76.4) | 45 (23.6) | 191 (100.0) |

*Note.* The BNC data (39,887 words) were randomly sampled from the speech data analysed in Table 7.4, because concordance lines needed to be viewed to confirm whether the discourse function was interpersonal or structural. Percentages appear in parentheses.

However, in the randomly sampled NS data, the structural marker *right/alright* constituted about a quarter of the total use. Thus, there were large discrepancies between the textbook dialogue and the naturally occurring conversations in the functional distribution of *right/alright*. As Lam (2010) points out, the results suggest that materials writers may pay more attention to a particular role of multi-function markers such as *right/alright* while giving less weight to other discourse functions.

### 7.3.2 RQ5-2: Comparison Between Dialogue in EFL Textbooks and Speech Data of Japanese EFL Learners

In section 7.3.1, the presentation of DMs in the EFL textbook database was compared to its distribution in NSs' speech data. This section, in turn, investigates the differences between dialogue in the nine textbooks and speech data of Japanese EFL learners: the NICT JLE Corpus.

As in the analysis for RQ5-1, the frequency analysis of DMs was carried out. Table 7.6 provides the statistics. If the DMs' occurrence rate was 0.01 per cent or below in each database, the items were not included in the analysis.

Table 7.6

*Comparisons of DMs in the Textbooks and the NICT JLE Corpus*

| | | Frequency per 10,000 words | | | |
|---|---|---|---|---|---|
| DMs | Categories | Textbooks | NICT JLE | LLR | Chi-square value |
| *really* | IP | 58.26 | 12.47 | 538.259 | 926.740*** |
| *well* | IP/Str/Cog | 57.96 | 15.61 | 414.867 | 655.596*** |
| *oh* | IP | 80.20 | 30.92 | 338.067 | 468.558*** |
| *what about* | Str | 12.11 | 0.38 | 327.201 | 1011.833*** |
| *sure* | IP | 16.34 | 1.97 | 240.743 | 510.131*** |
| *I see* | IP/Cog | 19.07 | 3.88 | 185.087 | 324.383*** |
| *like* | IP/Cog | 11.95 | 1.34 | 184.619 | 401.258*** |
| *now* | Str | 8.32 | 0.67 | 155.153 | 376.220*** |
| *right/alright* | IP/Str | 16.49 | 3.51 | 153.287 | 264.226*** |
| *just* | IP | 37.53 | 15.58 | 137.691 | 185.459*** |
| *how about* | Str | 19.97 | 5.83 | 129.000 | 197.582*** |
| *to be honest* | IP | 4.09 | 0.10 | 118.675 | 384.389*** |
| *great* | IP | 3.93 | 0.21 | 88.696 | 242.982*** |
| *you see* | IP | 3.93 | 0.21 | 88.696 | 242.982*** |
| *I mean* | Cog | 17.86 | 6.97 | 73.669 | 101.401*** |
| *anyway* | Ref | 5.30 | 0.80 | 66.423 | 129.847*** |
| *oh great* | IP | 2.42 | 0.04 | 64.788 | 199.019*** |
| *exactly* | IP | 4.09 | 0.49 | 60.582 | 128.795*** |
| *absolutely* | IP | 1.97 | 0.05 | 55.655 | 177.281*** |
| *you know* | IP/Cog | 33.59 | 20.00 | 48.101 | 57.059*** |
| *actually* | IP | 14.38 | 7.15 | 35.175 | 44.357*** |
| *kind of* | IP | 6.05 | 3.02 | 14.708 | 18.521*** |
| *I think* | IP/Cog | 52.36 | 44.09 | 9.241 | 9.794** |
| *then* | Str | 19.82 | 15.41 | 7.309 | 7.926** |
| *sort of* | IP/Cog | 1.21 | 0.57 | 3.351 | 4.300* |
| *but* | Ref | 71.12 | 70.76 | 0.012 | 0.012 |
| *first* | Str | 1.36 | 1.65 | −0.334 | 0.315 |
| *finally* | Str | 1.06 | 1.50 | −0.909 | 0.822 |
| *OK/okay* | IP/Str | 44.64 | 47.29 | −0.961 | 0.949 |
| *or* | Ref | 24.97 | 32.92 | −13.344 | 12.341*** |
| *yes* | IP | 67.04 | 90.24 | −41.696 | 38.556*** |
| *yeah* | IP/Str | 44.64 | 76.20 | −98.036 | 84.883*** |
| *and* | Ref/Str/Cog | 183.71 | 250.81 | −126.021 | 117.959*** |
| *because/'cause* | Ref | 8.17 | 30.64 | −149.667 | 107.956*** |
| *so* | Ref/Str | 39.80 | 121.96 | −481.631 | 365.044*** |

*Note*. IP = interpersonal; Ref = referential; Str = structural; Cog = cognitive. *$p$ < .05. **$p$ < .01. ***$p$ < .001.

The results obtained from chi-square tests revealed that there were significant differences between the two databases in the frequency of DMs with the following exceptions: *but*, *first*, *finally*, and *OK/okay*. The log-likelihood ratios also indicated similar differences. If the log-likelihood ratio applied to the two databases was represented as a positive number, the marker was used more frequently in the textbook database than in the NICT JLE Corpus. On the other hand, when the ratio was negative, Japanese EFL learners used the item more frequently. Thus, the analysis revealed that, although referential function markers such as *so* and *because/'cause* were used more frequently in Japanese EFL learners' speech than in the textbooks, most markers were used more frequently in the textbook database.

In addition to these quantitative analyses, the following examples of how DMs were used in the textbooks and the NICT JLE Corpus were qualitatively observed:

(7) A: It takes me forty-five minutes!

B: *Really?* That's terrible! (Lawrence & Levesque, 2003a, p. 56)

(8) A: What's his new book like?

B: *Well*, it's really long, but I'm enjoying it so far.

(Richards & O'Sullivan, 2009a, p. 66)

(9) A: It's three hundred dollars.

B: *Really?* Um uum I want more uum cheaper one.

(The NICT JLE Corpus, 00712.stt)

(10) A: When you came here.

B: *Well* mhm not not so crowded than urr I expected.

(The NICT JLE Corpus, 00244.stt)

These examples include the markers *really*, and *well*, which were used more frequently in the textbook database. The interpersonal marker *really* in (7) and (9) is used to indicate the speaker B's attitudes towards the speaker A's utterances, whereas the cognitive marker *well* in (8) and (10) has the discourse function of fillers which enable the speaker B to have time to think what to say.

Dialogues (7) and (8), extracted from the EFL textbooks, are denser than dialogues (9) and (10) used in the spoken corpus of Japanese EFL learners. This is because (9) and (10) contain hesitation words, or fillers, such as *uum* and *urr*. Carter (1998) argues that the textbook language "represents a 'can do' society, in which interaction is generally smooth and problem-free" (p. 47). Whether or not this perspective is right, it is certain that the created dialogues are less redundant than the Japanese learners' speech. Thus, the examples indicate that such redundancy in speech data may have an influence on calculating the frequency of DMs. In other words, the ratio of the number of DMs to the total number of words in speech data may be smaller than that in textbook dialogues.

Next, as in section 7.3.1, the frequency of a multi-function marker was quantified by the discourse function. As in the previous chapters, the analysis done here also was focused on *so*, which can be categorized into two discourse functions, because there was a significant difference in the frequency between the textbook database and the NICT JLE Corpus (see Table 7.6). Dialogues (11), (12), (13), and (14) show the use of *so* as a DM in the textbooks:

(11) A: Oh, that's great. I don't know what time it starts.

B: No, neither do I. But my husband has the tickets, *so* he knows.

(Kay & Jones, 2009, p. 151)

(12) A: Yeah. You've got to have a cake on your birthday.

B: What kind is it?

A: It's strawberry cream. I made it myself. *So*, how does it feel to be nineteen? (Lawrence & Levesque, 2003b, p. 18)

(13) A: OK. Is that everything?

B: Yes, I think so.

A: *So* it's just the T-shirt and the earrings. That comes to …

(Buckingham & Lansford, 2010, p. 71)

(14) A: Right, so if you look to your right, you'll see a big cinema ... The Odeon ... got it?

B: Yeah, yeah I can see that.

A: *So* just carry on past the cinema and ... cross the road and you walk along Finlay Street for about 100 metres ….

(Cunningham & Moor, 2005, p. 169)

According to the functional paradigm proposed by Fung and Carter (2007), the referential marker *so* indicates the consequential relationships between utterances. As shown in (11), the speaker B uses *so* in order to introduce the second clause as the result of the preceding one. In the structural domain, Fung and Carter subcategorize the discourse function into three types: topic shifts, summarizing opinions, and continuation of topics. The speaker A in (12) changes the topic from a birthday cake to the speaker B's feelings about being 19 years old. The shift of the topic is marked by *so*. In (13), *so* is used to mark the end of the conversation and summarize the speaker B's utterance about purchasing goods. Additionally, in (14), *so* serves as a signal that the speaker A is continuing with giving directions.

The frequency analysis was done according to the taxonomy of the functional categories and subcategories (see Table 7.7). However, many utterances of *so* by Japanese learners did not fall into the taxonomy. Thus, they were labelled as *others* in the analysis.

Table 7.7
*Raw Frequency of* so *as a DM in the Textbooks and the Random Sample of the NICT JLE Corpus*

| | Raw Frequency of *so* | | | | | |
|---|---|---|---|---|---|---|
| Databases | Ref | Str: TS | Str: SO | Str: CT | Others | Total |
| Textbooks | 118 (44.9) | 58 (22.0) | 30 (11.4) | 57 (21.7) | 0 (0.0) | 263 (100.0) |
| NICT Sample | 329 (48.7) | 23 (3.4) | 21 (3.1) | 53 (7.9) | 249 (36.9) | 675 (100.0) |

*Note.* In calculating the frequency, concordance lines were viewed to confirm the discourse functions. Ref = referential; Str = structural; TS = topic shifts; SO = summarizing opinions; CT = continuation of topics. Percentages appear in parentheses.

As with the analysis in Table 7.5, the speech data of Japanese learners were randomly sampled from the NICT JLE Corpus.[8] In the sampled data (49,253 words), the referential marker *so* constituted about a half of the total use, whereas the proportion of the structural marker was very low. In other words, as found in Study 2 (see section 4.3.3 in Chapter 4), the result implies that many Japanese learners may have difficulty in acquiring the structural marker.

Additionally, it was notable that the proportion of *so* labelled others was high in the NICT sample. This was considered to be mainly due to the use of *so* as a filler and learners' errors. Some examples are shown below.

(15) A: And the movie will begin at, um *so*, six o'clock …

B: Uh-huh.

A: in the afternoon.

(The NICT JLE Corpus, 01145.stt)

(16) A: … I have family's computer. Yes. But I … I can't use it.

B: Why?

A: *So*, I … I don't understand … how to use it.

(The NICT JLE Corpus, 00013.stt)

In dialogue (15), *so* serves the discourse function of a filler, which can lengthen the speaker A's thinking time. Additionally, *so* in dialogue (16) should be replaced with *because*. The utterance of *so* in the form of fillers and errors was ubiquitous in the speech data of Japanese English learners. The findings support those of Study 4 (see section 6.3.2 in Chapter 6). Thus, the analysis suggests that there are large discrepancies between the textbooks and the speech data of Japanese learners in the functional distribution of *so*.

### 7.3.3 RQ5-3: Features of the Presentation of DMs in Dialogue in EFL Textbooks Designed for Japanese English Learners

The purpose of this section is to explore the features of presentation of DMs in EFL textbooks designed for Japanese English learners. In order to serve the purpose, I conducted a correspondence analysis based on the frequency of DMs in the nine EFL textbooks and the NICT JLE Corpus. In other words, the analysis aimed mainly to investigate the differences in DM use between dialogues in the five EFL textbooks for Japanese learners and those in the four EFL textbooks for the worldwide market, and explore how the domestic textbooks reflect DM use in Japanese learners' interlanguage. Additionally, the analysis explores what kinds of DMs are often used in the textbooks.

The analysis was performed on 35 items with more than 0.01 per cent distribution rate in each database (see Appendices 7-A and 7-B). In the analysis, the NICT JLE Corpus was divided into three proficiency groups[9]: higher, middle, and lower. Therefore, the analysis provided a description of the relationship between two categorical variables: the 35 DMs and 12 databases (i.e., the nine EFL textbooks and the three proficiency groups of the NICT JLE Corpus).

In the analysis for categorical data, the two variables were represented as a data matrix of rows and columns (see Appendices 7-C and 7-D). Although the matrix can be displayed as a biplot, i.e., a joint plot of the row and column points, two scatter plots were used in an attempt to visualize the relationship among categories in each variable. In short, the two variables were displayed separately (see Figures 7.1 and 7.2).

In order to output the scatter plots, I selected only the first two dimensions out of 11 dimensions on the basis of the cumulative percentage of inertia, or the proportions of explained variation (see Appendix 7-E). These two dimensions explained 24.6% and 21.9% of the total inertia. Additionally, the singular values were .314 and .286 respectively, and the chi-square value between the two dimensions was significant ($\chi^2$ (374, $N$ = 420) = 4287.611, $p$ < .001). Thus, although the explanation ratio of the variables was not high due to the large number of categories, the correspondence analysis can reveal the relationship among the categories.

First, Figure 7.1 shows similarities and differences among the nine textbooks and three sets of spoken corpus data in the distribution of DMs. Although the distance between the points of categories cannot be defined, the closer the distance is, the more similarities there are between the categories (Clausen, 1998). Therefore, a two-dimensional positioning map can efficiently summarise the relationships among the textbooks.

While it is hard to interpret what Dimension 2 means, the axis of Dimension 1 is likely to indicate target language proficiency, judging from the points of the three groups of the NICT JLE Corpus. That is partly because the higher proficiency group was near the zero point, and partly because the lower and middle proficiency groups were found on the right side of the axis. In other words, the points of the textbooks on the left side are considered to be related to higher language proficiency.

In the light of the interpretation of Dimension 1, TOU may provide DMs which are infrequent or difficult for Japanese learners, and the language level may be above pre-intermediate. On that point, as reviewed in Chapter 2, Miura (2009) also found the gap between TOU and Japanese learner speech in the use of some DMs and pointed out that the textbook may not necessarily meet learners' needs. Since TOU draws on the analysis of the Cambridge International Corpus of North American English, the textbook is likely to reflect natural conversation among NSs the most authentically of the nine textbooks.

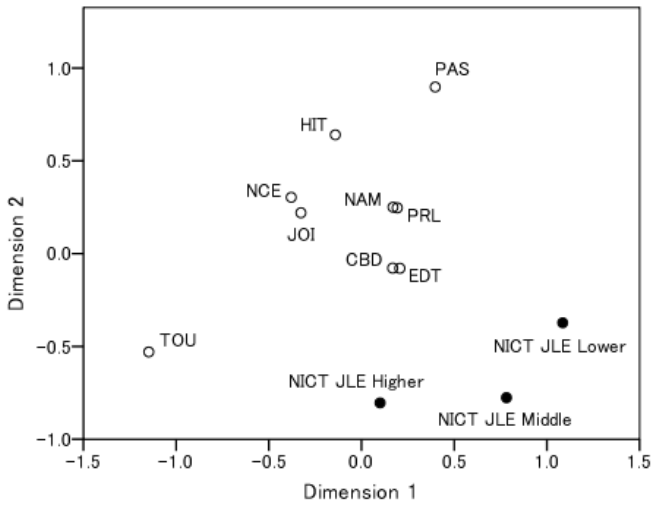



*Figure 7.1.* Correspondence analysis for the textbooks and spoken corpora: the row point plot.

On the other hand, the point of PAS was on the middle left side of the axis of Dimension 1. PAS is a new title launched by a well-known international publisher based in the UK, but the textbook is written exclusively for Japanese learners of English. The result indicates that PAS is the easiest textbook and reflects Japanese learners' proficiency levels as far as the DMs go. As for the lexical statistical features of the textbook, the values are relatively low: standardized TTR is 31.88 and words per sentence is 4.66 (see Table 7.2). Therefore, the vocabulary and sentence levels may also be lower than any other textbook in the present study.

However, the cluster of the other seven textbooks was placed near the intersection of the two axes, or the centre of the map, at which an average pattern of the distribution was represented. The cluster was not separated according to textbooks tailored to the Japanese institutional market and those for sale on the international market. In other words, the correspondence analysis indicates that there is little difference between the two groups in the presentation of DMs, although there are some slight differences among the textbooks.

Thus, Dimension 1 can hardly explain the features about textbooks designed for Japanese learners of English, but there was little difference in the row points to distinguish the five textbooks from the higher or middle group of the NICT JLE Corpus. The result indicates that the textbooks reflect DM use by higher or middle level Japanese learners.

Next, Figure 7.2 shows both similarities and differences among the 35 DMs, the other categorical variables for the analysis. As in Figure 7.1, in this two-dimensional map, the distance between the items represents the degree of the relationship. Additionally, the axis of Dimension 1 can be interpreted as a scale of language proficiency. However, Figure 7.2 shows a remarkably wider distribution than Figure 7.1.

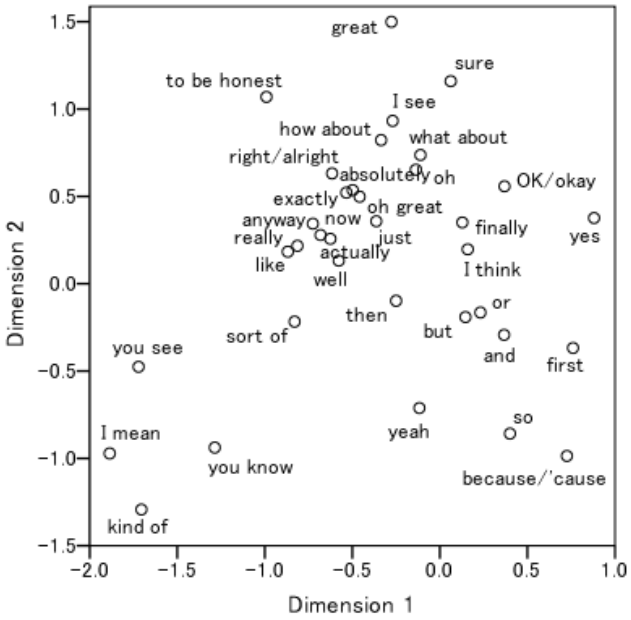



*Figure 7.2.* Correspondence analysis for the 35 DMs: the column point plot.

At the axis of Dimension 1, interpersonal or cognitive function markers such as *I mean*, *kind of*, *you see*, and *you know* were placed on the left side, while referential or structural function markers such as *first*, *because/'cause*, *OK/okay*, *and*, and *so* (but not *yes*) were on the right side of the scatter plot.

From the perspective of language proficiency, the more proficient speakers become in their language skills, the more frequently they can use interpersonal or cognitive function markers in conversation. Therefore, textbooks using the function markers more frequently may be appropriate for not the pre-intermediate but intermediate or upper level of learners. On the other hand, referential or structural function markers may be input data appropriate for the lower level of learners. An example of the presentation of DMs is shown below.

(17) Juan: They really are. Have you ever been to Sequoia National Park?

Kim: No. Have you?

Juan: Yeah. I went last year. The trees there are the tallest in the world.

Kim: Really? I didn't know that.

Juan: Yeah. I had the best time. *I mean*, it's just the greatest place to hike.

(McCarthy, McCarten, & Sandiford, 2006a, p. 26)

(18) Miki: How long does the tour take?

Clerk: About two hours.

Miki: When is the best time to go?

Clerk: You should go early in the morning. It's less crowded, *and* you can enjoy the peaceful gardens.

(Buckingham & Lansford, 2010, p. 78)

Dialogue (17) extracted from TOU includes *I mean*, and dialogue (18) from PAS includes *and*. Schiffrin (1987) claims that *I mean* is used to "preface expansions of speakers' own prior ideas" (p. 296). In (17), therefore, Juan uses *I mean* as a signal to encourage Kim to go to the tourist spot. However, in (18), the clerk uses *and* instead of *I mean* in a very similar situation. The difference in choosing between *I mean* or *and* epitomises the difference between TOU and PAS: two contrastive textbooks at the axis of Dimension 1 in Figure 1.

As for Dimension 2, although it is difficult to interpret the meaning in the row point plot, the dispersion pattern in the column point plot is more conducive to interpretation. In Figure 2, while items such as *great*, *sure*, *I see*, *how about*, and *what about* were found on the upper side of the positioning map, items such as *because/'cause*, *so*, *I mean*, and *you know* were placed near the bottom of the axis. In other words, some markers on the upper side were used more frequently in the

textbooks than in spoken corpora. In contrast, some on the lower side were used less frequently in the textbooks (see Tables 7.3, 7.4, and 7.6). The tendency corresponded to that in Figure 7.1: Most of the textbooks were on the middle upper side; the three proficiency groups of the NICT JLE Corpus were on the lower side. Therefore, Dimension 2 may be taken as an indication of either the features of textbook language or spoken language. In short, the scatter plot may visualize what kinds of DMs are often used in textbooks.

### 7.4 Summary of Study 5

The present study quantitatively investigated the presentation of DMs in EFL textbooks published in Japan and overseas with some qualitative observations. Additionally, the results were compared to the distribution of DMs in spoken corpora.

In regard to the answer to RQ5-1, the two comparative analyses were conducted in order to investigate how speech data of NS children and adults differ from dialogues in the textbook database in terms of DM use. The results revealed that there was a large discrepancy between textbook dialogues and the speech data of NSs in their frequency of DMs. Additionally, the frequency analysis on the functional categories suggests that materials writers may value a particular role of multi-function markers such as *right/alright*, while paying little attention to other discourse functions. If dialogues spoken by NSs are regarded as authentic language, most of the nine textbooks may lack authenticity. Therefore, the finding in the study is consistent with the results of most previous studies on dialogue in EFL/ESL textbooks (see section 2.4.2 in Chapter 2).

Regarding the answer to RQ5-2, the comparison between dialogue in the nine EFL textbooks and the NICT JLE Corpus revealed that, although referential function markers such as *so* and *because/'cause* were used more frequently in the Japanese

learner corpus than in the textbooks, most DMs were used more frequently in the textbook dialogue. Additionally, a qualitative observation on some dialogues in the textbooks and the NICT JLE Corpus indicated that the created dialogues are denser than the Japanese learners' speech. In fact, the latter were found to contain more hesitation words or fillers as well as DMs.

Thus, the results suggest that textbook writers intend to include as many linguistic items as possible in the limited space available. Additionally, from the perspective of interlanguage, Japanese EFL learners are likely to rely on using some simple markers in their speech and have yet to acquire many items presented in textbooks. Interestingly, DM use by NS children shows the same tendency. Moreover, the analysis which focused on the functions of *so* suggests that learners frequently use DMs as a filler or misuse them. These may cause considerable discrepancies between the textbook corpus and the NICT JLE Corpus in the frequency and distribution of DMs.

The answer to RQ5-3 was unclear. The correspondence analysis based on the frequency of DMs revealed that the five titles produced for Japanese college students did not clearly differ from the four international textbooks in DM use. Accordingly, the result indicates that there is little essential distinction between the two groups in the presentation of DMs. However, the analysis also revealed that there was little difference between the five domestic titles and higher or middle groups of the NICT JLE Corpus in terms of language level. This result indicates that they reflect DM use in Japanese learners' interlanguage.

On the other hand, the correspondence analysis provided other findings in the inter-textbook comparison. The analysis revealed that there were small differences among the textbooks in the presentation of DMs. In particular, the result suggests that NS corpus-based textbooks are designed more authentically than non-corpus-based ones.

That is to say, the result implies that the choice and use of DM items depend on textbook writers' or publishers' design policy or views of teaching materials, and generate the features of textbook language. In summary, the exploratory analysis indicates that their design is likely to be influenced by the proficiency level of targeted users.

However, the inhomogeneity of databases may be a limitation of the study's findings. Although the level of the nine textbooks is set at pre-intermediate by the publishers, there is variability in lexical statistical features among the textbooks. Additionally, it is assumed that the selection of topics and situations varies depending on the editing policy of the materials designers. Therefore, these differences are likely to influence the presentation of DMs in textbook dialogue. In other words, textbook writers could purposely select a DM item appropriate to the level of targeted users or for the situation of each unit. For further research, it would be necessary to investigate the presentation of DMs in various topics and situations.

In conclusion, the results in the present study indicated a close relationship between DM use and the proficiency level of target users. From the perspective of second language acquisition, therefore, materials designers or instructors should provide lower level learners with frequent and simple items as language input in order to make DMs accessible to learners. Additionally, at pre-intermediate and intermediate levels, it would be necessary to call learners' attention to the correct usage of DMs as well as to provide them with a variety of items. Given this, the textbook language should not be criticised too strongly, even though the presentation of DMs in the textbooks did not reflect naturally occurring conversation among NSs. Many proponents of authentic language do not take full account of the nature of second language acquisition. In pedagogical settings, there are several important questions with regard to failing to

understand the nature and using only the language produced by native adults as a yardstick.

**Endnotes**

1. An earlier version of this chapter has been published in Shimada (2012).
2. The three publishers based in Japan are affiliated with the Association of English Textbook Publishers in Japan. Cambridge University Press, Macmillan, Oxford University Press, and Pearson Education are considered to be the most significant international publishers in the market of English textbooks in Japan.
3. According to the online book catalogue of the Association of English Textbook Publishers in Japan (http://www.daieikyo.jp/aetp/), most EFL textbooks designed for Japanese college students consist of 12, 15 or 20 units, to suit the number of classes in a college semester.
4. Cambridge International Corpus of North American English is a part of the Cambridge International Corpus, a large-scale database of English texts, built up by Cambridge University Press.
5. While the textbook with the largest number of DM items was TOU (38 items), the textbook with the smallest number was CBD (21 items). In addition to the nine textbooks, some other textbooks were also analysed, but they were excluded from the textbook database due to a small repertoire of DMs.
6. The word frequency count was carried out using WordSmith Tools 5.0 and the NICT JLE Corpus Analysis Tool 1.0.
7. Kobayashi and Yamada (2008) explored the distribution of metadiscourse markers in Japanese learner corpora by using a correspondence analysis. Additionally, Ishikawa (2010) examined Japanese learners' use of *–ly* adverbs in the same way. As

in the last chapter, the analysis in the present study is also based on the data mining technique.

8. The NICT JLE Corpus is made up of nine subcorpora according to learners' proficiency levels (i.e., level 1–9). In the analysis, I randomly sampled five interview data from each of the subcorpora, except the lowest level (level 1) in which no occurrence of the marker *so* was observed.
9. As mentioned in Chapter 3, the NICT JLE Corpus is originally divided into nine proficiency levels, but the present study analysed the learner data at three proficiency levels: lower (level 1–3), middle (level 4–6), and higher (level 7–9) levels.

# Chapter 8

# General Discussion and Conclusion

## 8.1 Overview of the Findings

This research addressed five studies in order to investigate the features of Japanese EFL learners' acquisition of DMs in speech, and to examine the language of textbooks as primary input resources and explore the relationship between input and output of DMs in the Japanese EFL context. In order to attain the two purposes, multiple comparisons were performed using spoken corpora and the language data of textbooks. In this section, I will not refer to the specific research questions of the five studies, but I will recapitulate the main findings of the research.

Study 1 investigated the use of DMs by Japanese EFL learners and native English speakers. The corpus-based analyses revealed that the diversity and quantity of DMs increased as speakers' language proficiency improved, regardless of their L1. These results, in particular, suggest that Japanese learners' language development may have an effect on the frequency of markers such as *you know* and *well*, which are used on an interpersonal level. Although Hays (1992) indicates that Japanese learners use *you know* and *well* less frequently, the current findings imply that higher level learners can acquire *you know* and *well*. Similarly, with increasing age, NSs were found to use a wider variety of markers on the interpersonal and cognitive levels.

However, the results also indicate that there are significant differences in the distribution and development of DMs between Japanese learners and NSs. Overall, Japanese learners were found to use DMs less frequently than native-speaking children and adults, but they overused relatively simple types of DMs such as *OK/okay*, *so*, and *yes*. Thus, the findings of Study 1 support those of the previous studies regarding the

difference in DM use between NSs and NNSs. Additionally, it should be noted that some qualitative observations revealed English *so* and Japanese *so* were mixed up in Japanese learners' utterances.

In Study 2, to explore the features of Japanese learners' DM use, corpora of NNS and NS speech were analysed using CIA. A frequency analysis of DMs revealed significant differences between Japanese learners' and NSs' speech, supporting earlier findings, including those of Study 1.

The quantitative analysis of the learner corpus data suggests that Japanese learners may use *so* more frequently than other non-native English learners with different L1 backgrounds, while also using certain interpersonal or cognitive function markers such as *you know*, *I mean*, and *just* less frequently. In addition, the qualitative analysis indicates that Japanese learners may prefer *so* as a filler when they are thinking about what to say next in the discourse. However, although Study 2 identified these features of Japanese learners' DM use, it did not explain why they overused or underused certain DMs.

To that end, Study 3 explored the influence of L1 transfer on DM use by Japanese EFL learners, partly because some researchers (e.g., Liu, 2013; Müller, 2004; Romero-Trillo, 2002; Sankoff et al., 1997) have pointed out that NNSs may overuse or underuse certain DMs due to the influence of their L1, and partly because the findings of Study 1 indicate that Japanese learners may overuse the English DM *so* as a result of their L1 influence.

The analyses using English-Japanese parallel corpus data revealed a complicated relationship between English DMs and their Japanese equivalent expressions (i.e., Japanese DMs). Most English DMs correspond to a wide variety of Japanese DMs, while some Japanese DMs correspond to different English DMs. Based on these

correspondences, then, both English and Japanese speech data collected from picture description tasks were analysed in terms of DM use. The results suggest that Japanese learners' L1 use may have an influence on the frequency of some referential and structural function markers such as *and*, *so*, and *but*, leading them to overuse these items.

However, Study 3 also did not identify the cause of Japanese learners' underusing some interpersonal or cognitive function markers such as *you know*, *I mean*, and *just*. On the other hand, the findings of Study 1 show that Japanese learners may often use these DMs in speech as their English proficiency improves. Therefore, Study 4 compared Japanese learners' speech data with the language of junior high and high school textbooks which can be regarded as primary resources in the classroom.

The results reveal that the Japanese learners preferred DMs, such as *and*, *OK/okay*, and *yes*, which often appear in the textbooks. In other words, the presentation of DMs in textbooks may have an effect on learners' DM acquisition and use. Additionally, some interpersonal or cognitive function markers such as *you know* and *I mean* were rarely used in the textbooks, although the frequency of *just* was relatively high. Thus, as some previous studies (e.g., Ota et al., 2003; Tono, 2002) point out, Japanese learners' L2 performance is likely to be influenced by input data from textbooks.

The findings of Study 4 confirm the importance of textbook language in the acquisition of DMs. As the final study of this research, therefore, Study 5 examined the presentation of DMs in nine EFL textbooks published in Japan and overseas, and explored how the textbooks take into consideration the features of Japanese EFL learners' DM acquisition.

In order to investigate linguistic features of the textbooks, Study 5 compared the presentation of DMs in the textbooks with their distribution in spoken corpora. The

frequency analysis of DMs in the nine EFL textbooks revealed significant differences between dialogues in the textbooks and the speech data of NSs. Although this finding supports those of previous studies, quantitative and qualitative analyses of the textbook dialogue and speech data of Japanese EFL learners indicate that the design of each textbook is likely to be related to the proficiency level of its target users. In other words, the learners' their proficiency level may be a key factor in the textbook designers' decision about which DMs to present. Additionally, the analyses also revealed that although there is little distinction between the textbooks produced for Japanese learners and international textbooks in terms of DM use; the former reflect DM use in Japanese learners' interlanguage.

In Chapter 2, I clarified two purposes of the current study (see section 2.7.1). Therefore, the concluding remarks attempt to assess whether the findings of the current research contribute to the attainment of these purposes. The purposes are repeated below.

Purpose 1: To identify the features of Japanese EFL learners' DM acquisition

Purpose 2: To examine the language of textbooks and explore the relationship between the input and output of DMs in the Japanese EFL context

With respect to Purpose 1, the findings of Studies 1 and 2 reveal some similarities and differences between Japanese learners and NSs of English in the development of DMs. Additionally, the findings of Study 2 reveal some similarities and differences in DM use by Japanese learners and other NNSs with different L1 backgrounds. Both studies indicate that Japanese learners may acquire and use certain DMs such as *so*, *you know*, *I mean*, and *just* differently from NSs or other NNSs. Moreover, some factors of

the features can be explained by the findings of Studies 3 and 4: Japanese learners' DM acquisition and use may be influenced by L1 transfer and input data from textbooks. Using the framework of the current research (see Figure 2.6 in Chapter 2), Figure 8.1 summarises the four studies relevant to the investigation of Japanese EFL learners' DM acquisition.[1]

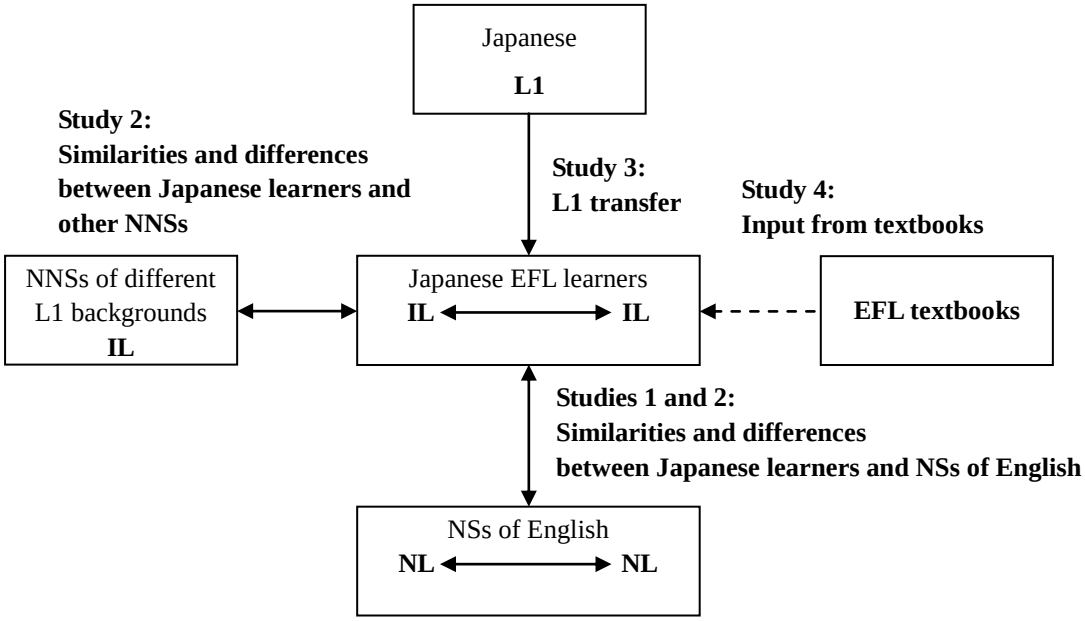



*Figure 8.1.* Findings of the four studies relevant to the investigation of Japanese EFL learners' DM acquisition.

In short, the findings of these four studies largely contribute to the attainment of Purpose 1. However, as mentioned in Chapter 6, Study 4 only estimates the effect of EFL textbooks on Japanese learners' DM acquisition, because the study cannot confirm a direct causal relationship between the two variables. Therefore, the arrow from EFL textbooks in Figure 8.1 has a dotted line.

Concerning Purpose 2, as mentioned above, the findings of Study 4 indicate that

those of Ota et al. (2003) and Tono (2002) are likely to hold true for the relationship between input from textbooks and output of DMs in Japanese learners' speech. Additionally, the findings of Study 5 suggest that the presentation of DMs in textbooks may be related to learners' proficiency levels. In the Japanese EFL context, therefore, textbook design may play an important role in Japanese learners' acquisition of DMs. In other words, although some limitations exist, Purpose 2 is considered to be attained in the current research.

The current study adopted the framework of corpus-based multiple comparisons using learner corpora and the language data of textbooks. The findings indicate the potential usefulness of this framework in identifying features of learners' L2 acquisition and exploring the relationship between the input and output of various linguistic items.

### 8.2 Pedagogical Implications

This research provides several pedagogical implications for classroom instruction and materials design in L2 teaching. Some have already been mentioned earlier in this dissertation, but they will be restated here in order to give a clear picture of the conclusions drawn from the research results.

First, in Study 1, the statistical analysis of Japanese learners' speech data revealed that the frequency of DMs was influenced by their language proficiency levels. Additionally, the tendency held true for the development of L1 English. However, the comparative analysis between Japanese EFL learners' and NSs' speech data suggests that lower level speakers may have difficulty in acquiring cognitive function markers such as *I mean*, *sort of*, and *you know*, regardless of their L1. Therefore, language instructors or material developers should provide beginner and pre-intermediate level learners with referential and structural function markers such as *and*, *because/'cause*,

*then*, *now*, and *first* rather than infrequent or difficult items. In other words, these frequent and easy-to-use items can help Japanese learners control their speech to convey a clear message to others.

Second, Study 2 found that Japanese EFL learners underused certain interpersonal or cognitive markers such as *like*, *really*, *you know*, *kind of*, and *I mean*. Study 4 further revealed that these markers were often not used in the textbooks for junior high and high school students. Therefore, Japanese learners may have more difficulty acquiring certain DMs in these two categories compared to referential or structural markers. Language instructors and materials writers should consider the features of learners' interlanguage and provide infrequent and difficult items in an interpersonal or cognitive category at the intermediate or advanced proficiency level.

Third, Study 4 indicated that Japanese learners may frequently use DMs which often appear in textbooks. This suggests that the presentation of DMs in textbooks may have a substantial effect on Japanese learners' DM acquisition and use. Therefore, it is important for materials designers to carefully include appropriate items in textbooks according to their target proficiency level or the learners' readiness to accept new input items.

Fourth, Studies 1 and 4 revealed that some Japanese EFL learners made errors in their DM use. For example, they confused the function of *so* with that of *because* in speech. The misuse of referential or structural markers such as *so* and *because* may lead to communication failure. Tyler, Jefferies, and Davies (1988) point out that miscues of DMs in connecting discourse segments "can produce a scrambling effect for the listener" (p. 109). Additionally, Study 3 suggested that Japanese learners may overuse some DMs such as *and*, *so*, and *but* because of the influence of their L1. Therefore, language instructors should pay careful attention to their students' accuracy in using

those DMs. The proper use of DMs should be demonstrated in a variety of contexts of classroom instruction.

The last implication for developing teaching materials concerns applications of learner corpora and native child language data. Study 1 adopted Granger's (1996, 1998, 2002) CIA but added a comparison of different NS data (i.e., native child and adult language data) to the learner corpus analysis. Granger's framework is based on multiple comparisons with learners' interlanguage data at the core. She stresses that the findings contribute to the improvement of materials design in L2 teaching. However, I contend that when researchers or materials designers investigate how certain linguistic items are acquired, they should focus on language development in both L1 and L2 speakers, as this perspective can help uncover the similarities and differences between L1 and L2 development. In fact, this study revealed new findings on this issue.

Figure 8.2 summarises the process of designing textbook language based on the comparisons in Studies 1 and 5.

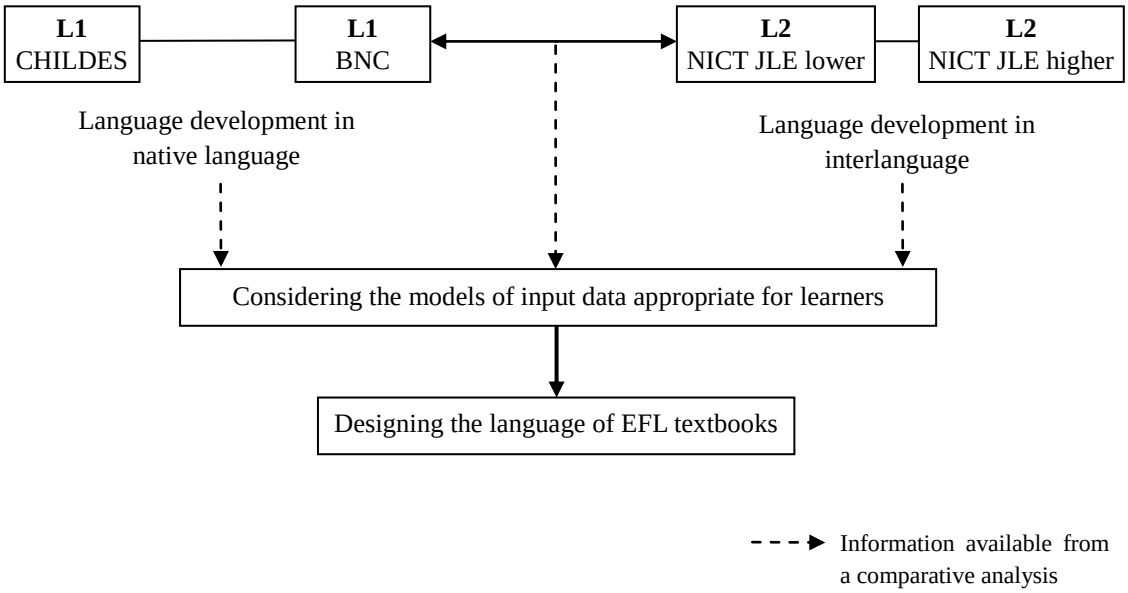



*Figure 8.2.* A model for designing the language of EFL textbooks for Japanese learners.

As mentioned before, there has commonly been no apparent rationale behind the design of textbook language. However, the empirical studies of the current research indicate that corpus-based analyses employing multiple databases can help craft a better model of textbook language.

### 8.3 Limitations of the Research and Suggestions for Further Research

Finally, I will point out four major limitations of this research and suggest directions for further research. As the limitations of each study have been already mentioned, this section will outline general limitations and the potential extension of this dissertation.

The first concerns the definition and taxonomy of DMs in spoken English. As reviewed in Chapter 2, the definition of DMs varies among researchers and lacks clarity. The current research operationally adopted Schiffrin's (1987) definition and Fung and Carter's (2007) functional paradigm of DMs because they illustrate pragmatic features as well as syntactic ones in spoken English. However, it was difficult to draw a distinction between DMs and non-DMs. In spoken corpora, for example, items such as *well* and *so* were often used by the speaker to fill a pause or gain time while searching for what to say next. In such cases, Fung and Carter define *well* as a cognitive function marker, while they regard *so* as being outside of the scope of DMs. On the other hand, Schiffrin accepts these filler expressions as DMs. In the current research, therefore, the filler use of *so* has been regarded as one of the DM functions, but other fillers such as *ah*, *er*, and *uh* have been excluded from the analysis. In short, there has been inconsistency in determining the scope of spoken DMs.

Additionally, some items such as *for example*[2] were not analysed in the current research because they are not included in Fung and Carter's framework. However, many

researchers (e.g., Fraser, 2009; Nattinger & DeCarrico, 1992) accept the exemplifying expression *for example* as a DM or lexical phrase organising discourse. As pointed out by Paquot (2008), NNSs tend to use *for example* frequently in their writing. Therefore, it may be worthwhile to investigate the distribution of *for example* in NNS speech. For further research, the definition and taxonomy of spoken DM should be reconsidered from the viewpoint of learners' use of DMs.

The second limitation involves the reliability and validity of databases used in the current research. In Study 1, it should be noted that there are considerable differences in the method of data collection used for the three spoken corpora. While the NICT JLE Corpus (Izumi, Uchimoto, & Isahara, 2004) is a collection of interviews with Japanese EFL learners in a speaking test, speech data in CHILDES (MacWhinney, 2000) and the BNC are extracted from naturally occurring conversations in daily situations. Additionally, most subcorpora of CHILDES consist of interactions between a child and his or her parents, while the BNC's subcorpora include samples of spoken English in a wide variety of situations. Therefore, as in some of the previous studies, the different types of data collection may imperil the validity of the comparative study because it may elicit different DM uses. However, as Hasselgård and Johansson (2011) point out, it was difficult to collect speech data with high reliability and validity.

Moreover, Study 3 also demonstrates problems in data collection. The speech data collected in the picture description tasks may not be sufficient to investigate certain linguistic features and the monologic nature of the tasks may have elicited a limited number and type of DMs. In this respect, Shimada and Miura (2013) compared data from picture description tasks[3] and question-and-answer tasks to reveal differences in Japanese EFL learners' DM use between the two types of tasks. While the participants often used structural function markers such as *finally*, *next*, *second*, and *then* to describe

the four-frame picture strips, they preferred interpersonal function markers such as *just* and *yeah* to answer the questions. For further research, therefore, speech data should be collected using various tasks, and the analysis should take into account the task differences.

The third limitation concerns the influence of textbooks on learners' DM use. As mentioned in Chapter 6, the learner corpus which was compared with the textbook database did not contain information on the textbooks learners used in junior high or high school. Therefore, Study 4 did not assert that the textbooks' presentation of DMs had a direct effect on the learners' DM use. However, it may be difficult to investigate the direct effect of textbooks on learners' L2 performance because researchers or instructors need to control the input resources available in learning settings in order to avoid other input effects. There have been few studies on this issue due to the difficulty of collecting speech data directly influenced by the language of textbooks.

The last limitation relates to the third one mentioned above. In other words, some input factors may influence L2 learners' DM acquisition and use, but the current research does not deal with all of them. For example, learners are likely to receive various inputs from classroom instruction, including teachers' talk and communication activities, or audio-visual materials used in English classes. Moreover, they can learn DM usage outside the classroom. As Gilquin (2000/2001) points out, researchers should understand this limitation when discussing results obtained in the Integrated Contrastive Model or comparative frameworks based on the model.

Despite these limitations, the current research has shed some light on how Japanese learners acquire and use spoken DMs in the Japanese EFL context. Additionally, based on the findings, the importance of DM inputs for Japanese learners has been confirmed in the dissertation. For further research, it is necessary to attempt to

overcome the limitations and investigate Japanese learners' DM use in various contexts and situations.

**Endnotes**

1. Arrow (G) in Figure 2.6 is deleted here because it is related to Purpose 2.
2. The frequency of *for example* is 10.77 per 10,000 words in LINDSEI-JP, while it is 0.42 per 10,000 words in NICT-NS. In other words, Japanese EFL learners use the exemplifying expression more frequently than interpersonal or cognitive function markers such as *well*, *you know*, and *I mean* (see Table 4.3 in Chapter 4). Additionally, they use the expression more frequently in speech than NSs of English do.
3. As in Study 3B described in Chapter 5, Shimada and Miura (2013) used a comic strip taken from the pre-first grade level of the Eiken Test in Practical English Proficiency.

## References


Aijmer, K. (2004). Pragmatic markers in spoken interlanguage. *Nordic Journal of English Studies*, *3*(1), 173–190.

Aijmer, K. (2011). *Well I'm not sure I think …* The use of *well* by non-native speakers. *International Journal of Corpus Linguistics*, *16*, 231–254. doi: 10.1075/ijcl.16.2.04aij

Aston, G. (1997). Small and large corpora in language learning. In B. Lewandowska-Tomaszczyk & P. J. Melia (Eds.), *PALC '97: Practical applications in language corpora* (pp. 51–62). Lodz, Poland: Lodz University Press. Retrieved from http://www.sslmit.unibo.it/~guy/wudj1.htm

Berman, R., Ragnarsdóttir, H., & Strömqvist, S. (2002). Discourse stance: Written and spoken language. *Written Language & Literacy*, *5*, 253–287. doi: http://dx.doi.org/10.1075/wll.5.2.06ber

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Blakemore, D. (2002). *Relevance and linguistic meaning: The semantics and pragmatics of discourse markers*. Cambridge: Cambridge University Press.

Brinton, L. J. (1996). *Pragmatic markers in English*. Berlin: Mouton de Gruyter.

Buckingham, A., & Lansford, L. (2010). *Passport 2: English for international communication* (Student book, 2nd ed.). New York, NY: Oxford University Press.

Carter, R. (1998). Orders of reality: CANCODE, communication, and culture. *ELT Journal*, *52*, 43–56. doi: 10.1093/elt/52.1.43

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

Chaudron, C., & Richards, J. C. (1986). The effect of discourse markers on comprehension of lectures. *Applied Linguistics*, *7*, 113–127. doi:10.1093/applin/7.2.113

CLAN (Version 25) [Computer software]. Retrieved from http://childes.psy.cmu.edu/clan

Clausen, S.-E. (1998). *Applied correspondence analysis: An introduction*. Thousand Oaks, CA: Sage.

Crystal, D. (2003). *English as a global language* (2nd ed.). Cambridge: Cambridge University Press.

Cullen, R., & Kuo, I.-C. (2007). Spoken grammar and ELT course materials: A missing link? *TESOL Quarterly*, *41*, 361–386. doi: 10.1002/j.1545-7249.2007.tb00063.x

Cunningham, S., & Moor, P. (2005). *New cutting edge pre-intermediate* (Students' book). Harlow: Pearson Education.

De Cock, S. (2011). Preferred patterns of use of positive and negative evaluative adjectives in native and learner speech: An ELT perspective. In A. Frankenberg-Garcia, L. Flowerdew, & G. Aston (Eds.), *New trends in corpora and language learning* (pp. 198–212). London: Continuum.

Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language*. Oxford: Oxford University Press.

Feng, A., & Byram, M. (2002). Authenticity in college English textbooks: An intercultural perspective. *RELC Journal*, *33*, 58–84. doi: 10.1177/003368820203300203

Flowerdew, J., & Tauroza, S. (1995). The effect of discourse markers on second language lecture comprehension. *Studies in Second Language Acquisition*, *17*, 435–458. doi: http://dx.doi.org/10.1017/S0272263100014406

Fox, G. (1998). Using corpus data in the classroom. In B. Tomlinson (Ed.), *Materials development in language teaching* (pp. 25–43). Cambridge: Cambridge University Press.

Fraser, B. (1999). What are discourse markers? *Journal of Pragmatics, 31,* 931–952. doi: http://dx.doi.org/10.1016/S0378-2166(98)00101-5

Fraser, B. (2009). An account of discourse markers. *International Review of Pragmatics, 1,* 293–320. doi: 10.1163/187730909X12538045489818

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

Gilmore, A. (2007). Authentic materials and authenticity in foreign language learning. *Language Teaching, 40,* 97–118. doi: http://dx.doi.org/10.1017/S0261444807004144

Gilquin, G. (2000/2001). The integrated contrastive model: Spicing up your idea. *Languages in Contrast, 3,* 95–123. doi: http://dx.doi.org/10.1075/lic.3.1.05gil

Gilquin, G., De Cock, S., & Granger, S. (Eds.). (2010). *LINDSEI: Louvain international database of spoken English interlanguage*. Louvain-la-Neuve, Belgium: Presses universitaires de Louvain.

Granger, S. (1996). From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. In K. Aijmer, B. Altenberg, & M. Johansson (Eds.), *Languages in contrast* (pp. 37–51). Lund, Sweden: Lund University Press.

Granger, S. (1998). The computer learner corpus: A versatile new source of data for SLA research. In S. Granger (Ed.), *Learner English on computer* (pp. 3–18). New York, NY: Addison Wesley Longman.

Granger, S. (2002). A bird's-eye view of learner corpus research. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition, and foreign language teaching* (pp. 3–33). Amsterdam: John Benjamins.

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

Granger, S., Dagneaux, E., Meunier, F., & Paquot, M. (Eds.). (2009). *International corpus of learner English* (Version 2). Louvain-la-Neuve, Belgium: Presses universitaires de Louvain.

Granger, S., & Tyson, S. (1996). Connector usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes, 15,* 17–27. doi: 10.1111/j.1467-971X.1996.tb00089.x

Hasselgård, H., & Johansson, S. (2011). Learner corpora and contrastive interlanguage analysis. In F. Meunier, S. De Cock, G. Gilquin, & M. Paquot (Eds.), *A taste for corpora: In honour of Sylviane Granger* (pp. 33–61). Amsterdam: John Benjamins.

Hays, P. R. (1992). Discourse markers and L2 acquisition. *Papers in Applied Linguistics-Michigan, 7*, 24–34.

He, A. (2001). On the discourse marker *so*. *Language and Computers, 36,* 41–52.

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. doi: http://dx.doi.org/10.1016/j.pragma.2006.04.008

House, J. (2010). The pragmatics of English as a lingua franca. In A. Trosborg (Ed.), *Pragmatics across languages and cultures* (pp. 363–387). Berlin: Walter de Gruyter.

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. doi: http://dx.doi.org/10.1016/j.pragma.2013.03.001

Hyland, K., & Tse, P. (2004). Metadiscourse in academic writing: A reappraisal. *Applied Linguistics, 25,* 156–177. doi: 10.1093/applin/25.2.156

Ishikawa, S. (2010). Nihonjin eigo gakushusha no *–ly* hukushi shiyou: Gakushusha kopasu CEEAUS ni motozuku keiryouteki kousatsu [Use of *–ly* adverbs by Japanese learners of English: A quantitative analysis of the learners' corpus,

CEEAUS]. *CELES (Chubu English Language Education Society) Bulletin*, *39*, 181–188.

Izumi, E., Uchimoto, K., & Isahara, H. (Eds.). (2004). *Nihonjin 1200 nin no eigo supikingu kopasu* [A spoken corpus of 1200 Japanese learners of English]. Tokyo: ALC Press.

Jiji Press. (2011). Jyouhou A no genshou keikou ni hadome: 2011 nendo koukou kyoukasho saitaku jyoukyo—Monkashou matome [The downward trend in adopting Information A textbooks has been stemmed: The adoption of the textbooks for Japanese high schools in 2011—A report compiled by the Ministry of Education, Culture, Sports, Science and Technology in Japan]. *Naigai Kyouiku*, *6054*, 10–17.

Kaneko, T. (2003). The use of negative emotional expression in English by non-native speakers: A corpus-based comparative study. *Gakuen*, *752*, 1–24.

Kay, S., & Jones, V. (2009). *New American inside out pre-intermediate* (Student's book). Oxford: Macmillan.

KH Coder (Version 2.0) [Computer software]. Retrieved from http://khc.sourceforge.net/dl.html

Kobayashi, Y. (2009). Profiling metadiscourse markers in native and non-native English. *Lexicon*, *39*, 1–17.

Kobayashi, Y., & Yamada, H. (2008). Nihonjin eigo gakushusha kopasu ni miru metadiscourse markers no shiyou keikou [The use of metadiscourse markers in Japanese EFL learners' English]. *English Corpus Studies*, *15*, 161–173.

Kubota, A. (1997). Gakushu no kontekusuto to kyouzai ['Context of learning' and ELT materials]. *Studies in Languages and Cultures*, *44*, 69–87.

Lam, P. W. Y. (2010). Discourse particles in corpus data and textbooks: The case of *well*. *Applied Linguistics*. doi: 10.1093/applin/amp026

Lawrence, N., & Levesque, G. (2003a). *Hear it! say it!: An English speaking and*

*listening text for Japanese students*. Tokyo: Kinseido.

Lawrence, N., & Levesque, G. (2003b). *Hear it! say it!: An English speaking and listening text for Japanese students* (Teacher's manual). Tokyo: Kinseido.

Lee, W. Y. (1995). Authenticity revisited: Text authenticity and learner authenticity. *ELT Journal*, *49*, 323–328. doi: 10.1093/elt/49.4.323

Leech, G. (1991). The state of the art in corpus linguistics. In K. Aijmer & B. Altenberg (Eds.), *English corpus linguistics: Studies in honour of Jan Svartvik* (pp. 8–29). London: Longman.

Leech, G. (1998). Preface. In S. Granger (Ed.), *Learner English on computer* (pp. xiv–xx). New York, NY: Addison Wesley Longman.

Leech, G. (2011). Frequency, corpora and language learning. In F. Meunier, S. De Cock, G. Gilquin & M. Paquot (Eds.), *A taste for corpora: In honour of Sylviane Granger* (pp. 7–31). Amsterdam: John Benjamins.

Lenk, U. (1998). Discourse markers and global coherence in conversation. *Journal of Pragmatics*, *30*, 245–257. doi: 10.1016/S0378-2166(98)00027-7

Liu, B. (2013). Effect of first language on the use of English discourse markers by L1 Chinese speakers of English. *Journal of Pragmatics*, *45*, 149–172. doi: org/10.1016/j.pragma.2012.11.002

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

McCarthy, M. (1991). *Discourse analysis for language teachers*. Cambridge: Cambridge University Press.

McCarthy, M., & Handford, M. (2004). "Invisible to us": A preliminary corpus-based study of spoken business English. In U. Connor & T. A. Upton (Eds.), *Discourse in the professions: Perspectives from corpus linguistics* (pp. 167–201). Amsterdam: John Benjamins.

McCarthy, M., McCarten, J., & Sandiford, H. (2006a). *Touchstone 3* (Student's book). New York, NY: Cambridge University Press.

McCarthy, M., McCarten, J., & Sandiford, H. (2006b). *Touchstone 3* (Teacher's book). New York, NY: Cambridge University Press.

McDonough, J., & Shaw, C. (2003). *Materials and methods in ELT: A teacher's guide* (2nd ed.). Oxford: Blackwell.

Mishan, F. (2005). *Designing authenticity into language learning materials.* Bristol: Intellect.

Miura, A. (2009). A study of L2 spoken corpora: The gap between productive skills of Japanese EFL learners and the corpus-based ELT textbooks. *OnCUE Journal*, *3*, 136–159.

Miura, A. (2011, September). *Discourse markers in spoken corpora of Japanese EFL learners.* Paper presented at Learner Corpus Research 2011, Louvain-la-Neuve, Belgium.

Mizumoto, A. (2009). Effect size calculation sheet. Retrieved from http://www.mizumot.com/stats/effectsize.xls

Mizutani, N. (2001). *Zoku niche-ei hikaku hanashi kotoba no bunpo* [Grammar of spoken language: Comparison between Japanese and English]. Tokyo: Kurosio.

Müller, S. (2004). 'Well you know that type of person': Functions of *well* in the speech of American and German students. *Journal of Pragmatics*, *36*, 1157–1182. doi: http://dx.doi.org/10.1016/j.pragma.2004.01.008

Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.

Onodera, N. (2001). *Japanese discourse markers: Synchronic and diachronic discourse analysis*. Amsterdam: John Benjamins.

Osborne, J. (2008). Phraseology effects as a trigger for errors in L2 English: The case of

more advanced learners. In F. Meunier & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp. 67–83). Amsterdam: John Benjamins.

Ota, H., Kanatani, K., Kosuge, A., & Hidai, S. (2003). *Eigo ryoku wa donoyouni nobite yukuka: Chuugakusei no eigo shuutoku katei wo ou* [How English ability is developed: Exploring English acquisition process of Japanese junior high school students]. Tokyo: Taishukan.

Paquot, M. (2008). Exemplification in learner writing: A cross-linguistic perspective. In F. Meunier & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp. 101–119). Amsterdam: John Benjamins.

ParaConc (Version 1.0, Build 269) [Computer software]. Houston, TX: Athelstan.

Pearson, J. (1998). *Terms in context*. Amsterdam: John Benjamins.

Petch-Tyson, S. (1998). Writer/reader visibility in EFL written discourse. In S. Granger (Ed.), *Learner English on Computer* (pp. 107–181). New York, NY: Addison Wesley Longman.

Polat, B. (2011). Investigating acquisition of discourse markers through a developmental learner corpus. *Journal of Pragmatics, 43,* 3745–3756. doi: 10.1016/j.pragma.2011.09.009

Porter, D., & Roberts, J. (1981). Authentic listening activities. *ELT Journal, 36,* 37–47. doi: 10.1093/elt/36.1.37

Pritchard, R. M. O. (1995). *Amae* and the Japanese learner of English: An action research study. *Language, Culture and Curriculum, 8,* 249–264. doi: 10.1080/07908319509525207

Rayson, P., Berridge, D., & Francis, B. (2004). Extending the Cochran rule for the comparison of word frequencies between corpora. In G. Purnelle, C. Fairon & A. Dister (Eds.), *Le poids des mots: Proceeding of the 7th International Conference on Statistical Analysis of Textual Data* (pp. 926–936). Louvain-la-Neuve,

Belgium: Presses universitaires de Louvain.

Rayson, P., & Garside, R. (2000). Comparing corpora using frequency profiling. In A. Kilgarriff & T. B. Sardinha (Eds.), *Proceedings of the Workshop on Comparing Corpora: Held in Conjunction with the 38th Annual Meeting of the Association for Computational Linguistics* (pp. 1–6). Hong Kong: Hong Kong University of Science and Technology.

Redeker, G. (2006). Discourse markers as attentional cues at discourse transitions. In K. Fischer (Ed.), *Approaches to discourse particles* (pp. 339–358). Amsterdam: Elsevier Science.

Richards, J. C. (2006). Materials development and research: Making the connection. *RELC Journal*, *37*, 5–26. doi: 10.1177/0033688206063470

Richards, J. C., & O'Sullivan, K. (2009a). *Join in 2: Developing conversation strategies* (Student's book). New York, NY: Oxford University Press.

Richards, J. C., & O'Sullivan, K. (2009b). *Join in 2: Developing conversation strategies* (Teacher's book). New York, NY: Oxford University Press.

Römer, U. (2004). Comparing real and ideal language learner input: The use of an EFL textbook corpus in corpus linguistics and language teaching. In G. Aston, S. Bernardini, & D. Stewart (Eds.), *Corpora and language learners* (pp. 151–168). Amsterdam: John Benjamins.

Romero-Trillo, J. (2002). The pragmatic fossilization of discourse markers in non-native speakers of English. *Journal of Pragmatics*, *34*, 769–784. doi: http://dx.doi.org/10.1016/S0378-2166(02)00022-X

Rühlemann, C. (2009). Discourse presentation in EFL textbooks: A BNC-based study. *Language and Computers*, *69*, 415–435.

Sadanobu, T. (Ed). (2002). *'Un' to 'so' no gengogaku* [Linguistics on 'un' and 'so']. Tokyo: Hituzi Syobo.

Sankoff, G., Thibault, P., Nagy, N., Blondeau, H., Fonollosa, M.-O., & Gagnon, L. (1997). Variation in the use of discourse markers in a language contact situation. *Language Variation and Change*, *9*, 191–217. doi: http://dx.doi.org/10.1017/S0954394500001873

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

Scotton, C. M., & Bernsten, J. (1988). Natural conversations as a model for textbook dialogue. *Applied Linguistics, 9*, 372–384. doi: 10.1093/applin/9.4.372

Shimada, K. (2011). Discourse marker use in learners and native speakers: A corpus-based analysis of spoken English. *Annual Review of English Language Education in Japan (ARELE), 22*, 377–392.

Shimada, K. (2012). Discourse markers in EFL textbooks and spoken corpora: Materials design and authenticity. *Language Education & Technology*, *49*, 215–244.

Shimada, K. (2013). A comparative analysis of discourse markers in Japanese EFL textbooks and learner corpus data. *Language Education & Technology*, *50*, 69–91.

Shimada, K. (2014). Contrastive interlanguage analysis of discourse markers used by nonnative and native English speakers. *JALT Journal*, *36*, 47–67.

Shimada, K., & Miura, A. (2013, December). *Japanese EFL learners' use of discourse markers in their speech and writing*. Paper presented at the 131st Kanto Chapter Conference of the Japan Association for Language Education and Technology, Tokyo.

Someya, Y. (2001, April). *Bizinesu eigo gakushusha kopasu ni mirareru metadiscourse markers no bunpu to shiyou keikou* [Use of metadiscourse markers in business letters written by Japanese learners of EBP]. Paper presented at FLT Association, University of Tokyo. Abstract retrieved from http://gamp.c.u-tokyo.ac.jp/~flta/kojin/someya2.doc

Svartvik, J. (1980). 'Well' in conversation. In S. Greenbaum, G. Leech & J. Svartvik (Eds.), *Studies in English linguistics for Randolph Quirk* (pp. 167–177). London:

Longman.

Tono, Y. (2002). *The role of learner corpora in SLA research and foreign language teaching: The multiple comparison approach* (Unpublished doctoral dissertation). Lancaster University, Lancaster.

Tono, Y. (2004). Multiple comparisons of IL, L1 and TL corpora: The case of L2 acquisition of verb subcategorization patterns by Japanese learners of English. In G. Aston, S. Bernardini, & D. Stewart (Eds.), *Corpora and language learners* (pp. 45–66). Amsterdam: John Benjamins.

Tono, Y. (Ed.). (2007). *Nihonjin chukousei ichimannin no eigo kopasu: Chukousei ga kaku eibun no jittai to sono bunseki* [The JEFLL Corpus: A written corpus of 10,000 secondary school students]. Tokyo: Shogakukan.

Tono, Y. (2011). TaLC in action: Recent innovations in corpus-based English language teaching in Japan. In A. Frankenberg-Garcia, L. Flowerdew, & G. Aston (Eds.), *New trends in corpora and language learning* (pp. 3–25). London: Continuum.

Tsuda, A. (2003a). *Primary listening: 20 tips for better communication*. Tokyo: Kinseido.

Tsuda, A. (2003b). *Primary listening: 20 tips for better communication* (Teacher's manual). Tokyo: Kinseido.

Tyler, A. E., Jefferies, A. A., & Davies, C. E. (1988). The effect of discourse structuring devices on listener perceptions of coherence in non-native university teacher's spoken discourse. *World Englishes, 7*, 101–110. doi: 10.1111/j.1467-971X.1988.tb00223.x

Uesugi, A., Itoh, N., & Powell, R. (2001a). *Everyday talk: A total approach to listening, speaking, writing and grammar*. Tokyo: Asahi Press.

Uesugi, A., Itoh, N., & Powell, R. (2001b). *Everyday talk: A total approach to listening, speaking, writing and grammar* (Teacher's manual). Tokyo: Asahi Press.

Wardman, C. (2009). Authentic materials in English language teaching: Are they as

important as we think? *Folio, 13*(2), 14–17.

Watanabe, A. (2009). Zennendo hi 0.2% gen no 3533 man satsu ni: 2010 nendo chuugakko kyoukasho saitaku jyoukyo—Monkashou matome [35.33 million copies were adopted, a 0.2 per cent decrease over the previous year: The adoption of the textbooks for Japanese junior high schools in 2010—A report compiled by the Ministry of Education, Culture, Sports, Science and Technology in Japan]. *Naigai Kyouiku, 5960,* 2–4.

Watts, R. J. (1989). Taking the pitcher to the 'well': Native speakers' perception of their use of discourse markers in conversation. *Journal of Pragmatics, 13,* 203–237. doi: http://dx.doi.org/10.1016/0378-2166(89)90092-1

Williams, J. (1992). Planning, discourse marking, and the comprehensibility of international teaching assistants. *TESOL Quarterly, 26,* 693–711. doi: 10.2307/3586869

Wolfe-Quintero, K., Inagaki, S., Kim, H.-Y. (1998). *Second language development in writing: Measures of fluency, accuracy & complexity*. Honolulu, HI: University of Hawai'i Press.

WordSmith Tools (Version 5.0) [Computer software]. Liverpool: Lexical Analysis Software.

Yamamoto, A. (2010). Danwa hyoushiki *so* no kinou no shuutoku: Kaigai ryuugaku kenshuu ni sanka shita nihonjin eigo gakushusya no baai [The acquisition of the discourse marker *so* by Japanese learners of English: A comparison of learners' discourse before and after study abroad]. *Journal of Ochanomizu University English Society, 1,* 68–76.

Yamamoto, A. (2012). Learning how to provide reasons: A comparison of learners' usage of *because* before and after study abroad. *The Lark Hill: Bulletin of the School of Humanities and Social Engineering, 34,* 13–25.

Yoshitomi, N. S. (2009a). *Communication builder: Practical English for beginners and*

*intermediate students* (Rev. ed.). Tokyo: Nan'un-do.

Yoshitomi, N. S. (2009b). *Communication builder: Practical English for beginners and intermediate students* (Teacher's manual). Tokyo: Nan'un-do.

## Appendices

Appendix 4-A

*Raw Frequency of DMs in LINDSEI Subcorpora (422,476 words)*

| DMs | Categories | LINDSEI subcorpora | | | | | |
|---|---|---|---|---|---|---|---|
| | | JP 37,126 words | CH 63,542 words | DU 79,652 words | GE 85,950 words | FR 91,402 words | SP 64,804 words |
| *actually* | IP | 18 | 102 | 238 | 136 | 56 | 10 |
| *and* | Ref/Str/Cog | 1561 | 1900 | 3424 | 3610 | 3533 | 2721 |
| *anyway* | Ref | 4 | 11 | 17 | 14 | 31 | 18 |
| *basically* | IP | 1 | 0 | 12 | 24 | 8 | 7 |
| *because/'cause* | Ref | 177 | 388 | 587 | 600 | 627 | 621 |
| *but* | Ref | 541 | 553 | 1113 | 896 | 1241 | 800 |
| *cos* | Ref | 16 | 16 | 89 | 67 | 43 | 51 |
| *exactly* | IP | 8 | 16 | 7 | 31 | 16 | 12 |
| *finally* | Str | 11 | 3 | 2 | 14 | 7 | 16 |
| *first* | Str | 11 | 4 | 18 | 13 | 6 | 4 |
| *I mean* | Cog | 4 | 123 | 56 | 163 | 156 | 105 |
| *I think* | IP/Cog | 164 | 404 | 385 | 461 | 429 | 424 |
| *just* | IP | 40 | 412 | 493 | 468 | 296 | 170 |
| *kind of* | IP | 10 | 0 | 17 | 54 | 11 | 12 |
| *like* | IP/Cog | 107 | 136 | 408 | 447 | 214 | 512 |
| *now* | Str | 50 | 45 | 92 | 47 | 82 | 48 |
| *oh* | IP | 28 | 143 | 81 | 169 | 222 | 95 |
| *OK/okay* | IP/Str | 85 | 175 | 85 | 191 | 74 | 209 |
| *or* | Ref | 186 | 198 | 406 | 621 | 409 | 499 |
| *really* | IP | 32 | 114 | 710 | 642 | 495 | 256 |
| *right/alright* | IP/Str | 1 | 23 | 24 | 24 | 5 | 7 |
| *so* | Ref/Str | 768 | 634 | 939 | 711 | 948 | 469 |
| *then* | Str | 57 | 168 | 366 | 314 | 264 | 183 |
| *well* | IP/Str/Cog | 20 | 49 | 774 | 471 | 1012 | 392 |
| *yeah* | IP/Str | 321 | 635 | 871 | 1359 | 727 | 704 |
| *yes* | IP | 267 | 343 | 274 | 266 | 1932 | 444 |
| *you know* | IP/Cog | 8 | 120 | 196 | 77 | 192 | 184 |

*Note.* Raw frequencies were counted on 27 items with more than 0.01 per cent distribution rate in the subcorpora. IP = interpersonal; Ref = referential; Str = structural; Cog = cognitive.

Appendix 5-A

*Comic Strip A for the Picture Description Task*

You have **one minute** to prepare.

This is a story about a man who lived with his family in a small house.
You have **two minutes** to narrate the story.

Your story should begin with the following sentence:
**One day, a man was at home with his family.**


Reprinted from the second stage test of the pre-first grade level of the Eiken Test in Practical English Proficiency, conducted in the fall session 2010.

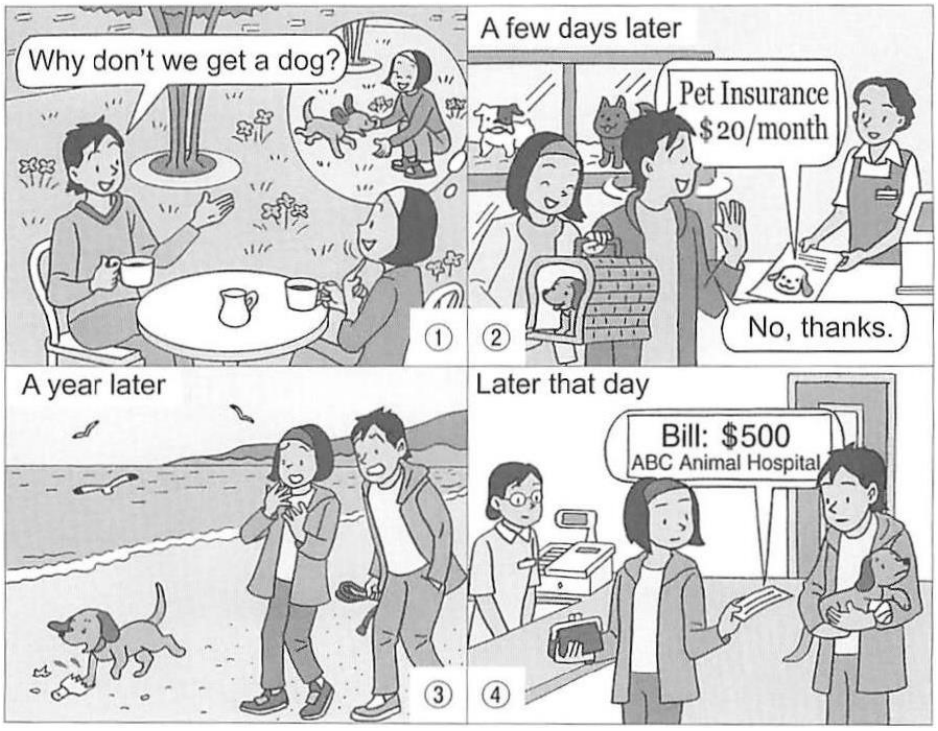
Appendix 5-B

*Comic Strip B for the Picture Description Task*

You have **one minute** to prepare.

This is a story about a couple who moved to a house with a yard.
You have **two minutes** to narrate the story.

Your story should begin with the following sentence:
**One day, a couple was sitting outside in their new yard.**




Reprinted from the second stage test of the pre-first grade level of the Eiken Test in Practical English Proficiency, conducted in the fall session 2011.

Appendix 6-A

*Raw Frequency of DMs in Japanese Junior High School English Textbooks (81,963 words)*

| DMs | Categories | NH1 3,324 words | NC1 3,799 words | SU1 5,517 words | TE1 2,277 words | OW1 4,785 words |
|---|---|---|---|---|---|---|
| *actually* | IP | 0 | 0 | 0 | 0 | 0 |
| *and* | Ref/Str/Cog | 27 | 18 | 44 | 11 | 32 |
| *because/'cause* | Ref | 0 | 0 | 2 | 0 | 0 |
| *but* | Ref | 15 | 6 | 20 | 13 | 13 |
| *finally* | Str | 0 | 0 | 0 | 0 | 0 |
| *first* | Str | 2 | 0 | 1 | 0 | 1 |
| *great* | IP | 2 | 0 | 5 | 0 | 3 |
| *how about* | Str | 6 | 1 | 9 | 0 | 5 |
| *I see* | IP/Cog | 3 | 4 | 4 | 0 | 0 |
| *I think* | IP/Cog | 1 | 0 | 0 | 0 | 1 |
| *just* | IP | 1 | 1 | 5 | 0 | 5 |
| *like* | IP/Cog | 0 | 0 | 0 | 0 | 0 |
| *now* | Str | 0 | 2 | 0 | 0 | 7 |
| *oh* | IP | 21 | 14 | 28 | 7 | 27 |
| *OK/okay* | IP/Str | 6 | 18 | 8 | 0 | 12 |
| *or* | Ref | 9 | 2 | 4 | 4 | 11 |
| *really* | IP | 3 | 3 | 8 | 8 | 4 |
| *right/alright* | IP/Str | 6 | 6 | 7 | 5 | 9 |
| *so* | Ref/Str | 6 | 4 | 5 | 2 | 2 |
| *sure* | IP | 4 | 2 | 2 | 2 | 7 |
| *then* | Str | 9 | 1 | 5 | 2 | 4 |
| *well* | IP/Str/Cog | 5 | 1 | 1 | 2 | 3 |
| *yeah* | IP/Str | 1 | 0 | 6 | 0 | 8 |
| *yes* | IP | 26 | 39 | 68 | 41 | 66 |
| *you know* | IP/Cog | 0 | 1 | 0 | 0 | 1 |

(continued)

Appendix 6-A (continued)

| DMs | Categories | NH2 6,313 words | NC2 5,540 words | SU2 7,853 words | TE2 3,908 words | OW2 6,146 words |
|---|---|---|---|---|---|---|
| *actually* | IP | 0 | 0 | 0 | 0 | 3 |
| *and* | Ref/Str/Cog | 72 | 43 | 53 | 30 | 38 |
| *because/'cause* | Ref | 5 | 2 | 6 | 11 | 14 |
| *but* | Ref | 41 | 21 | 56 | 20 | 30 |
| *finally* | Str | 0 | 0 | 3 | 0 | 2 |
| *first* | Str | 0 | 6 | 1 | 0 | 1 |
| *great* | IP | 1 | 1 | 1 | 1 | 0 |
| *how about* | Str | 4 | 8 | 18 | 5 | 9 |
| *I see* | IP/Cog | 4 | 10 | 5 | 4 | 1 |
| *I think* | IP/Cog | 16 | 13 | 25 | 4 | 9 |
| *just* | IP | 12 | 8 | 11 | 3 | 10 |
| *like* | IP/Cog | 4 | 5 | 3 | 1 | 4 |
| *now* | Str | 6 | 5 | 5 | 1 | 9 |
| *oh* | IP | 25 | 12 | 24 | 13 | 15 |
| *OK/okay* | IP/Str | 13 | 12 | 8 | 9 | 20 |
| *or* | Ref | 8 | 8 | 18 | 11 | 8 |
| *really* | IP | 3 | 2 | 4 | 10 | 13 |
| *right/alright* | IP/Str | 7 | 5 | 3 | 2 | 2 |
| *so* | Ref/Str | 18 | 10 | 21 | 5 | 12 |
| *sure* | IP | 4 | 5 | 8 | 8 | 4 |
| *then* | Str | 2 | 7 | 5 | 10 | 13 |
| *well* | IP/Str/Cog | 14 | 5 | 17 | 7 | 6 |
| *yeah* | IP/Str | 1 | 0 | 0 | 0 | 3 |
| *yes* | IP | 30 | 32 | 43 | 46 | 39 |
| *you know* | IP/Cog | 1 | 1 | 0 | 1 | 0 |

(continued)

Appendix 6-A (continued)

| DMs | Categories | NH3 5,626 words | NC3 6,294 words | SU3 8,577 words | TE3 4,736 words | OW3 7,268 words |
|---|---|---|---|---|---|---|
| *actually* | IP | 0 | 0 | 0 | 0 | 1 |
| *and* | Ref/Str/Cog | 83 | 63 | 90 | 41 | 90 |
| *because/'cause* | Ref | 2 | 7 | 11 | 6 | 20 |
| *but* | Ref | 44 | 30 | 46 | 24 | 30 |
| *finally* | Str | 0 | 3 | 0 | 0 | 4 |
| *first* | Str | 0 | 4 | 2 | 0 | 4 |
| *great* | IP | 1 | 2 | 2 | 0 | 1 |
| *how about* | Str | 3 | 3 | 10 | 7 | 10 |
| *I see* | IP/Cog | 1 | 14 | 5 | 2 | 6 |
| *I think* | IP/Cog | 5 | 9 | 12 | 8 | 11 |
| *just* | IP | 6 | 14 | 32 | 11 | 7 |
| *like* | IP/Cog | 4 | 8 | 2 | 1 | 4 |
| *now* | Str | 7 | 6 | 7 | 1 | 4 |
| *oh* | IP | 6 | 6 | 16 | 17 | 13 |
| *OK/okay* | IP/Str | 5 | 9 | 14 | 13 | 6 |
| *or* | Ref | 4 | 10 | 7 | 7 | 11 |
| *really* | IP | 2 | 1 | 5 | 10 | 8 |
| *right/alright* | IP/Str | 4 | 4 | 4 | 2 | 3 |
| *so* | Ref/Str | 18 | 11 | 20 | 5 | 9 |
| *sure* | IP | 6 | 7 | 6 | 10 | 1 |
| *then* | Str | 6 | 8 | 17 | 5 | 10 |
| *well* | IP/Str/Cog | 4 | 9 | 15 | 8 | 7 |
| *yeah* | IP/Str | 1 | 0 | 0 | 0 | 6 |
| *yes* | IP | 21 | 23 | 66 | 29 | 26 |
| *you know* | IP/Cog | 1 | 1 | 0 | 1 | 0 |

*Note.* Raw frequencies were counted on 25 items with more than 0.01 per cent distribution rate in the textbook database. IP = interpersonal; Ref = referential; Str = structural; Cog = cognitive.

Appendix 6-B

*Raw Frequency of DMs in English I Textbooks for Japanese High School Students (37,696 words)*

| DMs | Categories | CR1 12,405 words | PO1 6,000 words | PV1 9,486 words | AA1 3,694 words | BD1 6,111 words |
|---|---|---|---|---|---|---|
| *actually* | IP | 3 | 0 | 0 | 0 | 0 |
| *and* | Ref/Str/Cog | 235 | 90 | 141 | 31 | 68 |
| *because/'cause* | Ref | 17 | 14 | 25 | 9 | 13 |
| *but* | Ref | 72 | 23 | 52 | 16 | 44 |
| *finally* | Str | 2 | 2 | 3 | 0 | 4 |
| *first* | Str | 2 | 2 | 1 | 0 | 2 |
| *great* | IP | 2 | 1 | 0 | 2 | 0 |
| *how about* | Str | 4 | 6 | 1 | 4 | 5 |
| *I see* | IP/Cog | 2 | 3 | 1 | 0 | 1 |
| *I think* | IP/Cog | 7 | 3 | 4 | 2 | 1 |
| *just* | IP | 28 | 18 | 18 | 5 | 9 |
| *like* | IP/Cog | 16 | 3 | 9 | 11 | 2 |
| *now* | Str | 10 | 1 | 0 | 3 | 6 |
| *oh* | IP | 2 | 13 | 1 | 5 | 8 |
| *OK/okay* | IP/Str | 7 | 9 | 1 | 6 | 6 |
| *or* | Ref | 27 | 21 | 22 | 6 | 15 |
| *really* | IP | 7 | 9 | 5 | 3 | 7 |
| *right/alright* | IP/Str | 1 | 2 | 2 | 1 | 1 |
| *so* | Ref/Str | 17 | 7 | 14 | 6 | 1 |
| *sure* | IP | 6 | 2 | 1 | 0 | 4 |
| *then* | Str | 21 | 8 | 16 | 4 | 13 |
| *well* | IP/Str/Cog | 23 | 6 | 8 | 3 | 2 |
| *yeah* | IP/Str | 3 | 1 | 2 | 0 | 0 |
| *yes* | IP | 12 | 14 | 5 | 14 | 7 |
| *you know* | IP/Cog | 1 | 1 | 0 | 0 | 0 |

*Note.* Raw frequencies were counted on 25 items with more than 0.01 per cent distribution rate in the textbook database. IP = interpersonal; Ref = referential; Str = structural; Cog = cognitive.

Appendix 6-C

*Raw Frequency of DMs in English II Textbooks for Japanese High School Students (52,304 words)*

| DMs | Categories | CR2 13,045 words | PO2 8,960 words | PV2 15,856 words | AA2 5,262 words | BD2 9,181 words |
|---|---|---|---|---|---|---|
| *actually* | IP | 7 | 2 | 7 | 0 | 3 |
| *and* | Ref/Str/Cog | 202 | 131 | 218 | 74 | 118 |
| *because/'cause* | Ref | 20 | 13 | 17 | 7 | 13 |
| *but* | Ref | 70 | 33 | 102 | 22 | 61 |
| *finally* | Str | 5 | 1 | 3 | 1 | 3 |
| *first* | Str | 0 | 2 | 0 | 1 | 1 |
| *great* | IP | 2 | 0 | 0 | 0 | 0 |
| *how about* | Str | 7 | 1 | 0 | 3 | 3 |
| *I see* | IP/Cog | 0 | 1 | 2 | 4 | 2 |
| *I think* | IP/Cog | 7 | 2 | 4 | 5 | 11 |
| *just* | IP | 28 | 8 | 27 | 5 | 8 |
| *like* | IP/Cog | 20 | 4 | 13 | 10 | 15 |
| *now* | Str | 18 | 7 | 9 | 3 | 7 |
| *oh* | IP | 8 | 9 | 7 | 17 | 10 |
| *OK/okay* | IP/Str | 1 | 2 | 1 | 7 | 6 |
| *or* | Ref | 42 | 30 | 39 | 13 | 34 |
| *really* | IP | 4 | 6 | 3 | 13 | 12 |
| *right/alright* | IP/Str | 4 | 4 | 3 | 0 | 4 |
| *so* | Ref/Str | 14 | 11 | 17 | 12 | 16 |
| *sure* | IP | 1 | 1 | 1 | 1 | 2 |
| *then* | Str | 18 | 17 | 17 | 11 | 26 |
| *well* | IP/Str/Cog | 13 | 9 | 10 | 11 | 8 |
| *yeah* | IP/Str | 4 | 4 | 0 | 0 | 0 |
| *yes* | IP | 18 | 13 | 14 | 7 | 12 |
| *you know* | IP/Cog | 3 | 0 | 1 | 0 | 0 |

*Note.* Raw frequencies were counted on 25 items with more than 0.01 per cent distribution rate in the textbook database. IP = interpersonal; Ref = referential; Str = structural; Cog = cognitive.

Appendix 6-D

*Overview of the Row Points*

| | Scores | | Contribution to the inertia of the dimension | |
|---|---|---|---|---|
| Row categories | Dimension 1 | Dimension 2 | Dimension 1 | Dimension 2 |
| NH1 | .880 | −.155 | .039 | .003 |
| NC1 | 1.573 | −.299 | .100 | .008 |
| SU1 | 1.233 | −.813 | .116 | .107 |
| TE1 | 1.423 | −1.135 | .065 | .087 |
| OW1 | 1.337 | −.932 | .130 | .134 |
| NH2 | .307 | .522 | .009 | .056 |
| NC2 | .576 | .884 | .024 | .121 |
| SU2 | .424 | .901 | .020 | .192 |
| TE2 | .786 | −.069 | .041 | .001 |
| OW2 | .442 | .095 | .017 | .002 |
| NH3 | −.073 | .174 | .000 | .005 |
| NC3 | .102 | .814 | .001 | .117 |
| SU3 | .362 | .320 | .017 | .028 |
| TE3 | .592 | .373 | .024 | .020 |
| OW3 | −.029 | .038 | .000 | .000 |
| CR1 | −.642 | −.022 | .072 | .000 |
| PO1 | −.234 | −.123 | .005 | .003 |
| PV1 | −.770 | −.159 | .066 | .006 |
| AA1 | −.069 | −.013 | .000 | .000 |
| BD1 | −.483 | −.267 | .018 | .012 |
| CR2 | −.657 | −.272 | .075 | .027 |
| PO2 | −.537 | −.489 | .030 | .054 |
| PV2 | −.735 | −.204 | .093 | .015 |
| AA2 | −.262 | .113 | .005 | .002 |
| BD2 | −.502 | −.038 | .032 | .000 |
| Total | | | 1.000 | 1.000 |

Appendix 6-E

*Overview of the Column Points*

| | Scores | | Contribution to the inertia of the dimension | |
|---|---|---|---|---|
| Column categories | Dimension 1 | Dimension 2 | Dimension 1 | Dimension 2 |
| *actually* | −.1.082 | −.664 | .010 | .008 |
| *and* | −.484 | −.160 | .157 | .037 |
| *because/'cause* | −.506 | −.070 | .020 | .001 |
| *but* | −.252 | .115 | .019 | .008 |
| *finally* | −.735 | .162 | .006 | .001 |
| *first* | .304 | .956 | .001 | .021 |
| *great* | 1.018 | −.582 | .009 | .006 |
| *how about* | .678 | .645 | .020 | .038 |
| *however* | −1.047 | −.708 | .037 | .036 |
| *I see* | .686 | 1.441 | .012 | .115 |
| *I think* | .250 | 1.609 | .003 | .290 |
| *just* | −286 | .298 | .008 | .017 |
| *like* | −.747 | .198 | .026 | .004 |
| *now* | −.168 | .100 | .001 | .001 |
| *oh* | 1.030 | −.153 | .113 | .005 |
| *OK/okay* | .986 | .387 | .064 | .021 |
| *or* | −.432 | −.306 | .023 | .024 |
| *really* | .412 | −.450 | .009 | .022 |
| *right/alright* | 1.060 | −.553 | .034 | .019 |
| *so* | −.005 | .610 | .000 | .069 |
| *sure* | .852 | .612 | .023 | .025 |
| *then* | −.273 | −.158 | .006 | .004 |
| *well* | −.046 | .603 | .000 | .050 |
| *yeah* | .687 | −.1.825 | .006 | .093 |
| *yes* | 1.297 | −.409 | .393 | .083 |
| *you know* | .267 | .106 | .000 | .000 |
| Total | | | 1.000 | 1.000 |

Appendix 6-F

*Contribution to the Inertia*

| | Contribution to the inertia | |
|---|---|---|
| Dimension | Proportion | Cumulative proportion |
| 1 | .505 | .505 |
| 2 | .111 | .616 |
| 3 | .068 | .684 |
| 4 | .049 | .733 |
| 5 | .041 | .775 |
| 6 | .035 | .810 |
| 7 | .034 | .844 |
| 8 | .028 | .873 |
| 9 | .026 | .899 |
| 10 | .018 | .917 |
| 11 | .017 | .934 |
| 12 | .014 | .948 |
| 13 | .012 | .961 |
| 14 | .010 | .971 |
| 15 | .008 | .979 |
| 16 | .006 | .985 |
| 17 | .004 | .989 |
| 18 | .003 | .992 |
| 19 | .003 | .995 |
| 20 | .002 | .997 |
| 21 | .002 | .999 |
| 22 | .001 | 1.000 |
| 23 | .000 | 1.000 |
| 24 | .000 | 1.000 |
| Total | 1.000 | 1.000 |

Appendix 7-A

*Raw Frequency of DMs in EFL Textbooks for Japanese Learners (30,684 words)*

| DMs | Categories | CBD 3,816 words | HIT 5,761 words | EDT 8,634 words | PAS 6,523 words | PRL 5,950 words |
|---|---|---|---|---|---|---|
| *absolutely* | IP | 0 | 0 | 0 | 1 | 1 |
| *actually* | IP | 2 | 14 | 4 | 4 | 8 |
| *and* | Ref/Str/Cog | 89 | 64 | 93 | 95 | 136 |
| *anyway* | Ref | 0 | 2 | 6 | 2 | 2 |
| *because/'cause* | Ref | 2 | 5 | 11 | 2 | 8 |
| *but* | Ref | 32 | 39 | 73 | 36 | 33 |
| *exactly* | IP | 0 | 0 | 0 | 0 | 3 |
| *finally* | Str | 0 | 2 | 0 | 0 | 0 |
| *first* | Str | 4 | 0 | 0 | 1 | 1 |
| *great* | IP | 0 | 4 | 0 | 6 | 0 |
| *how about* | Str | 5 | 17 | 5 | 4 | 2 |
| *I mean* | Cog | 0 | 0 | 0 | 0 | 1 |
| *I see* | IP/Cog | 0 | 4 | 0 | 10 | 10 |
| *I think* | IP/Cog | 9 | 17 | 12 | 21 | 15 |
| *just* | IP | 4 | 24 | 37 | 25 | 11 |
| *kind of* | IP | 0 | 0 | 0 | 0 | 1 |
| *like* | IP/Cog | 1 | 2 | 4 | 3 | 6 |
| *now* | Str | 3 | 0 | 2 | 5 | 0 |
| *oh* | IP | 13 | 63 | 12 | 68 | 33 |
| *oh great* | IP | 0 | 0 | 1 | 1 | 0 |
| *OK/okay* | IP/Str | 0 | 35 | 3 | 80 | 12 |
| *or* | Ref | 2 | 6 | 11 | 14 | 37 |
| *really* | IP | 4 | 23 | 15 | 35 | 24 |
| *right/alright* | IP/Str | 0 | 5 | 1 | 9 | 22 |
| *so* | Ref/Str | 8 | 23 | 21 | 23 | 6 |
| *sort of* | IP/Cog | 0 | 0 | 0 | 0 | 0 |
| *sure* | IP | 10 | 18 | 3 | 26 | 16 |
| *then* | Str | 9 | 3 | 8 | 11 | 6 |
| *to be honest* | IP | 0 | 1 | 0 | 0 | 0 |
| *well* | IP/Str/Cog | 29 | 40 | 10 | 34 | 10 |
| *what about* | Str | 1 | 6 | 9 | 4 | 1 |
| *yeah* | IP/Str | 7 | 43 | 4 | 6 | 13 |
| *yes* | IP | 6 | 16 | 22 | 115 | 53 |
| *you know* | IP/Cog | 2 | 0 | 3 | 2 | 5 |
| *you see* | IP | 0 | 0 | 0 | 0 | 3 |

*Note.* Raw frequencies were counted on 35 items with more than 0.01 per cent distribution rate in the textbook database. IP = interpersonal; Ref = referential; Str = structural; Cog = cognitive.

Appendix 7-B

*Raw Frequency of DMs in EFL Textbooks for the International Market (35,400 words)*

| DMs | Categories | JOI 12,566 words | NAM 7,335 words | NCE 4,685 words | TOU 10,814 words |
|---|---|---|---|---|---|
| *absolutely* | IP | 4 | 4 | 0 | 3 |
| *actually* | IP | 20 | 9 | 9 | 25 |
| *and* | Ref/Str/Cog | 245 | 192 | 96 | 204 |
| *anyway* | Ref | 1 | 5 | 7 | 10 |
| *because/'cause* | Ref | 2 | 11 | 7 | 6 |
| *but* | Ref | 71 | 70 | 27 | 89 |
| *exactly* | IP | 10 | 7 | 4 | 3 |
| *finally* | Str | 0 | 3 | 2 | 0 |
| *first* | Str | 1 | 1 | 0 | 1 |
| *great* | IP | 10 | 2 | 1 | 3 |
| *how about* | Str | 12 | 3 | 10 | 8 |
| *I mean* | Cog | 17 | 0 | 5 | 36 |
| *I see* | IP/Cog | 15 | 6 | 8 | 10 |
| *I think* | IP/Cog | 50 | 22 | 8 | 19 |
| *just* | IP | 57 | 13 | 37 | 40 |
| *kind of* | IP | 7 | 0 | 0 | 12 |
| *like* | IP/Cog | 18 | 13 | 8 | 24 |
| *now* | Str | 7 | 15 | 5 | 18 |
| *oh* | IP | 127 | 78 | 48 | 88 |
| *oh great* | IP | 2 | 2 | 0 | 2 |
| *OK/okay* | IP/Str | 51 | 42 | 34 | 38 |
| *or* | Ref | 45 | 11 | 16 | 23 |
| *really* | IP | 89 | 38 | 34 | 123 |
| *right/alright* | IP/Str | 26 | 0 | 27 | 19 |
| *so* | Ref | 47 | 25 | 39 | 71 |
| *sort of* | IP/Cog | 1 | 1 | 1 | 1 |
| *sure* | IP | 15 | 2 | 7 | 11 |
| *then* | Str | 23 | 18 | 24 | 29 |
| *to be honest* | IP | 6 | 0 | 2 | 0 |
| *well* | IP/Str/Cog | 72 | 35 | 48 | 105 |
| *what about* | Str | 9 | 5 | 1 | 4 |
| *yeah* | IP/Str | 71 | 8 | 23 | 120 |
| *yes* | IP | 134 | 48 | 38 | 11 |
| *you know* | IP/Cog | 29 | 3 | 8 | 59 |
| *you see* | IP | 1 | 0 | 1 | 8 |

*Note.* Raw frequencies were counted on 35 items with more than 0.01 per cent distribution rate in the textbook database. IP = interpersonal; Ref = referential; Str = structural; Cog = cognitive.

Appendix 7-C

*Overview of the Row Points*

| | Scores | | Contribution to the inertia of the dimension | |
|---|---|---|---|---|
| Row categories | Dimension 1 | Dimension 2 | Dimension 1 | Dimension 2 |
| CBD | .168 | −.078 | .005 | .001 |
| HIT | −.141 | .641 | .005 | .113 |
| EDT | .207 | −.079 | .005 | .001 |
| PAS | .398 | .898 | .046 | .259 |
| PRL | .193 | .247 | .009 | .016 |
| JOI | −.327 | .220 | .034 | .017 |
| NAM | .170 | .251 | .008 | .019 |
| NCE | −.379 | .304 | .054 | .038 |
| TOU | −1.148 | −.529 | .467 | .109 |
| NICT JLE Lower | 1.085 | −.373 | .210 | .027 |
| NICT JLE Middle | .782 | −.776 | .152 | .165 |
| NICT JLE Higher | .100 | −.804 | .003 | .233 |
| Total | | | 1.000 | 1.000 |

Appendix 7-D

*Overview of the Column Points*

| | Scores | | Contribution to the inertia of the dimension | |
|---|---|---|---|---|
| Column categories | Dimension 1 | Dimension 2 | Dimension 1 | Dimension 2 |
| *absolutely* | −.498 | .536 | .001 | .001 |
| *actually* | −.578 | .132 | .014 | .001 |
| *and* | .367 | −.292 | .090 | .063 |
| *anyway* | −.726 | .344 | .007 | .002 |
| *because/'cause* | .725 | −.986 | .024 | .049 |
| *but* | .146 | −.191 | .005 | .009 |
| *exactly* | −.534 | .521 | .003 | .003 |
| *finally* | .128 | .351 | .000 | .001 |
| *first* | .760 | −.368 | .004 | .001 |
| *great* | −.277 | 1.499 | .001 | .022 |
| *how about* | −.336 | .822 | .007 | .044 |
| *I mean* | −1.886 | −.972 | .142 | .041 |
| *I see* | −.270 | .932 | .004 | .048 |
| *I think* | .159 | .197 | .004 | .007 |
| *just* | −.363 | .358 | .014 | .015 |
| *kind of* | −1.705 | −1.292 | .039 | .025 |
| *like* | −.869 | .184 | .021 | .001 |
| *now* | −.682 | .279 | .009 | .002 |
| *oh* | −.138 | .653 | .004 | .105 |
| *oh great* | −.458 | .498 | .001 | .001 |
| *OK/okay* | .369 | .557 | .021 | .052 |
| *or* | .231 | −.165 | .005 | .003 |
| *really* | −.814 | .217 | .096 | .007 |
| *right/alright* | −.616 | .631 | .018 | .021 |
| *so* | .400 | −.858 | .030 | .150 |
| *sort of* | −.830 | −.217 | .003 | .000 |
| *sure* | .062 | 1.159 | .000 | .071 |
| *then* | −.251 | −.098 | .004 | .001 |

(continued)

Appendix 7-D (continued)

| | Scores | | Contribution to the inertia of the dimension | |
|---|---|---|---|---|
| Column categories | Dimension 1 | Dimension 2 | Dimension 1 | Dimension 2 |
| *to be honest* | −.991 | 1.069 | .009 | .011 |
| *well* | −.626 | .257 | .064 | .012 |
| *what about* | −.111 | .737 | .000 | .018 |
| *yeah* | −.117 | −.712 | .002 | .089 |
| *yes* | .880 | .376 | .187 | .038 |
| *you know* | −1.286 | −.938 | .142 | .083 |
| *you see* | −1.721 | −.475 | .026 | .002 |
| Total | | | 1.000 | 1.000 |

Appendix 7-E

*Contribution to the Inertia*

| | Contribution to the inertia | |
|---|---|---|
| Dimension | Proportion | Cumulative proportion |
| 1 | .264 | .264 |
| 2 | .219 | .483 |
| 3 | .149 | .631 |
| 4 | .098 | .729 |
| 5 | .072 | .802 |
| 6 | .066 | .867 |
| 7 | .049 | .916 |
| 8 | .034 | .950 |
| 9 | .025 | .975 |
| 10 | .021 | .996 |
| 11 | .004 | 1.000 |
| Total | 1.000 | 1.000 |