

**UCLA**

**UCLA Electronic Theses and Dissertations**

**Title**

From Frequency to Formulaicity: Morphemic Bundles and Semi-Fixed Constructions in Japanese Spoken Discourse

**Permalink**

<https://escholarship.org/uc/item/1zp613xj>

**Author**

Kaneyasu, Michiko

**Publication Date**

2012

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

From Frequency to Formulaicity:

Morphemic Bundles and Semi-Fixed Constructions in Japanese Spoken Discourse

A dissertation submitted in partial satisfaction of  
the requirements for the degree Doctor of Philosophy  
in Asian Languages and Cultures

by

Michiko Kaneyasu

2012

© Copyright by

Michiko Kaneyasu

2012

## ABSTRACT OF THE DISSERTATION

From Frequency to Formulaicity:

Morphemic Bundles and Semi-Fixed Constructions in Japanese Spoken Discourse

by

Michiko Kaneyasu

Doctor of Philosophy in Asian Languages and Cultures

University of California, Los Angeles, 2012

Professor Shoichi Iwasaki, Chair

This dissertation investigates patterns and functions of fixedness, or formulaicity, in Japanese spoken discourse. A growing number of language researchers have pointed to the centrality of formulaicity and idiomaticity in understanding what native speakers know and do with the language. However, an overwhelming majority of evidence comes from English, which differs from Japanese typologically. In particular, agglutinating morphology poses methodological and conceptual challenges for examining formulaicity in Japanese language. As a result, not much is known about Japanese formulaicity, except for fixed idioms and phrases, such as *te o yaku* ‘have difficulty with’ (lit. ‘burn one’s hand’) and *o-tsukare-sama deshita* ‘thank you for your hard work’ (lit. ‘(you) were honorably tired’ or ‘(you) seemed tired’).

The present study is an attempt to explore formulaicity in Japanese spoken language by employing a morpheme-based, frequency-driven approach with large corpora. It identifies the most frequently recurring multi-morphemic sequences, or morphemic bundles, in three spoken registers, ordinary conversation, formal interview, and academic speech. Detailed analysis of the morphemic bundles reveals structural and functional characteristics of fixedness in spoken language that are linked to particular communicative needs of different registers and contexts. Findings in this study, however, also indicate that the morphemic bundles do not necessarily represent form-function units. Most significantly, about half of the morphemic bundles are encompassed by three partially-filled, or semi-fixed, constructions, with open slots and variations in some forms. Rather than adding to the conceptual meaning of utterances, the three semi-fixed constructions serve as discourse frames for fluent, socio-pragmatically appropriate, and idiomatically natural utterances, thereby contributing to the smooth on-line production and comprehension of spoken language.

Findings and discussions in this study provide new insight into the formulaic and socio-interactive nature of native speakers' language use in real context. The new insight also has important implications for Japanese as a foreign language learning and teaching, as a significant part of language acquisition takes place in learning idiomatic ways of saying or doing things in the target socio-linguistic community.

The dissertation of Michiko Kaneyasu is approved.

John H. Schumann

Sung-Ock Sohn

Hongyin Tao

Shoichi Iwasaki, Committee Chair

University of California, Los Angeles

2012

*To Mayu*

## TABLE OF CONTENTS

LIST OF FIGURES .....	ix
LIST OF TABLES .....	ix
LIST OF GLOSS ABBREVIATIONS .....	x
ACKNOWLEDGEMENTS .....	xi
VITA .....	xiii
<b>Chapter 1: Introduction</b> .....	<b>1</b>
1.1. Purpose of the study .....	1
1.2. Basic unit of analysis .....	5
1.3. Definition of formulaic language .....	6
1.4. Scope of the research .....	9
1.5. Research objectives .....	10
1.6. Outline of the dissertation .....	11
<b>Chapter 2: Literature Review</b> .....	<b>13</b>
2.1. Overview .....	13
2.2. Formal-semantic perspectives on formulaic language .....	14
2.3. Pragmatic/socio-interactive aspect of utterances .....	21
2.4. Cognitive-oriented approaches to formulaic language .....	24
2.4.1. Grammar as constructions .....	24
2.4.2. Cognitive representation model .....	26
2.4.3. Cognitive models for first language acquisition .....	27
2.4.4. Intersection between formulaicity and creativity .....	31
2.4.5. Problems for second/foreign language learners .....	33
2.5. Usage-based approaches to Japanese discourse .....	34
2.5.1. Introduction .....	34
2.5.2. Patterns of fixedness in Japanese discourse .....	35
2.5.2.1. Varied word order .....	36
2.5.2.2. Ellipsis and syntactically incomplete utterances .....	38
2.5.2.3. Particles .....	41
2.5.2.4. Discourse markers .....	42
2.6. Lexical bundles and register variations .....	43
2.7. Summary .....	47

<b>Chapter 3: Data and Methodology</b>	<b>49</b>
3.1. Introduction	49
3.2. The data	49
3.3. Methodology	51
3.4. Procedures	52
3.4.1. Tokenization of texts	53
3.4.2. Standardization of tokenized texts	54
3.4.3. N-gram analysis	55
<b>Chapter 4: Morphemic Bundles in Conversation, Interview, and Speech</b>	<b>58</b>
4.1. Frequency of morphemic bundles among three registers	58
4.2. Structural characteristics of the morphemic bundles	60
4.2.1. VP-related bundles	64
4.2.2. NP-related bundles	68
4.3. Functional characteristics of the morphemic bundles	69
4.3.1. Stance expressions	77
4.3.1.1. Epistemic stance bundles	79
4.3.1.2. Attitudinal stance bundles	87
4.3.1.3. Summary of stance expressions	94
4.3.2. Discourse organizers	94
4.3.2.1. Topic introduction/elaboration bundles	95
4.3.2.2. Cohesion bundles	103
4.3.2.3. Background bundles	113
4.3.2.4. Summary of discourse organizers	131
4.3.3. Referential expressions	132
4.3.3.1. Conceptualization bundles	132
4.3.3.2. Equivocation bundles	138
4.3.3.3. Summary of referential expressions	152
4.3.4. Socio-interactive expressions	153
4.3.4.1. Reporting bundles	154
4.3.4.2. Interpersonal bundles	164
4.3.4.3. Reactive bundles	167
4.3.4.4. Word/expression search bundles	175
4.3.4.5. Inquiry bundles	187
4.3.4.6. Other socio-interactive bundles	191
4.3.4.7. Summary of socio-interactive expressions	201
4.3.5. Objective statements	202
4.4. Register variations in the use of the morphemic bundles	210
4.5. Cross-linguistic perspectives on the bundles	215
4.6. Formulaic status of the morphemic bundles	217

<b>Chapter 5: Semi-fixed Constructions in Spoken Japanese</b>	220
5.1. Introduction .....	220
5.2. Thought-reporting construction .....	222
5.3. Utterance-expansion construction .....	233
5.4. Conceptualization construction .....	242
5.5. Motivations for the use of the constructions .....	248
<b>Chapter 6: Conclusions</b>	254
APPENDIX .....	259
REFERENCES .....	284
DATA SOURCES .....	296
SOFTWARE PROGRAMS .....	296

## LIST OF FIGURES

Figure 2.1. Balance of holistic and analytic processing from birth to adulthood .....	30
Figure 4.1. Number of different four-morphemic bundles across registers .....	58
Figure 4.2. Overall frequency of four-morphemic bundles across registers .....	59
Figure 4.3. Distribution of morphemic bundles across functional categories .....	210
Figure 4.4. Register distribution of the bundle functional types .....	213

## LIST OF TABLES

Table 3.1. Description of the dataset .....	50
Table 4.1. Distribution of morphemic bundle types across structural types .....	63
Table 4.2. Distribution of VP-related morphemic bundle types .....	64
Table 4.3. Distribution of morphemic bundle types across functional types .....	70
Table 4.4. Functional classification of common morphemic bundles across registers .....	71
Table 4.5. Bare <i>ka + na</i> [Q + PP] vs. quoted <i>ka + na</i> [Q + PP] in three registers .....	78
Table 4.6. Socio-interactional expression sub-types across three registers .....	154
Table 5.1. Collocates of <i>n+da</i> (F(n) = 2941) in Conversation (span +1) .....	235
Table 5.2. Collocates of <i>n+desu</i> (F(n) = 1773) in Interview (span +1) .....	235
Table 5.3. Collocates of <i>n+desu</i> (F(n) = 1249) in Speech (span +1) .....	236

## LIST OF GLOSS ABBREVIATIONS

ACC	accusative	NEG	negative
ADV	adverbial form	NML	nominalizer
ALL	allative	NOM	nominative
ASP	aspect	NONPST	non-past tense
ATT	attributive form	ONM	onomatopoeia
AUX	auxiliary	PASS	passive
CAUS	causative	PL	plural
CLS	classifier	POL	polite form
COND	conditional	POT	potential
COP	copula	PP	pragmatic particle
DAT	dative	PRF	prefix
DM	discourse marker	PST	past tense
GEN	genitive	PT	particle
HON	honorific	Q	question
HUM	humble	QT	quotative particle
IMPF	imperfect	SFX	suffix
INJ	interjection	TE	te-form
INS	instrumental	TOP	topic particle
LOC	locative	VOL	volitional

## ACKNOWLEDGEMENTS

The completion of this project owes much to the inspiration and support of many people. My interest in formulaicity of language was generated during years of teaching Japanese as a foreign language to college students. I would like to thank my students at UCLA and Santa Monica College for inspiring me to pursue this topic as my doctoral research.

Professor Shoichi Iwasaki's seminar on formulaic language provided an opportunity to learn and think more about the topic with fellow graduate students. The Japanese formulaic language research group, which emerged from this seminar, continued to provide a forum for further discussions, which led to some collaborative research projects. I thank the group members for many interesting discussions and all the good laughs we had together.

Professor Hongyin Tao's courses on corpus linguistics gave me invaluable training in corpus-based methodologies and tools. As I decided to adopt a corpus-based, frequency-driven approach to formulaic language, Professor Sung-Ock Sohn kindly shared her yet-to-be-published paper she had written with Professor Kilim Nam on a corpus-based study of Korean multi-morphemic sequences involving bound nouns *kes* and *ke* 'thing.' This paper gave me a practical insight into conducting a corpus-based formulaic research on agglutinating languages like Korean and Japanese.

The data for the present study was compiled from seven different sources. I would like to thank the following people and organizations for allowing me to use their publicly available corpora: Mayumi Usami at Tokyo University of Foreign Studies, Gendai Nihongo Kenkyukai, TalkBank, Linguistic Data Consortium, National Institute for Japanese Language and National Institute of Information and Communications Technology, and Ryuichi Uemura at Fukuoka Institute of Technology. I would also like to thank Tsuyoshi Ono and Yuki Taylor for sharing

their data with me. In addition, I would like to express my gratitude to Laurence Anthony and Nara Institute of Science and Technology (NAIST) for developing the user-friendly software programs, AntConc and ChaSen, and making them available to public free of charge.

Portions of this dissertation were presented at the 18<sup>th</sup> Workshop on East Asian Languages (WEAL) at University of California at Santa Barbara and at the 5<sup>th</sup> Meeting of the Formulaic Language Research Network (FLaRN 5) at Tilburg University in the Netherlands. I would like to acknowledge helpful comments and feedback from participants at these meetings.

I am indebted to my dissertation committee members — Shoichi Iwasaki, Hongyin Tao, Sung-Ock Sohn, and John Schumann — for sharing their valuable insights with me and providing much guidance and support throughout my graduate career and dissertation research.

The completion of this dissertation was not possible without the financial support I received from University of California at Los Angeles and Department of Asian Languages and Cultures during my graduate studies. I am also grateful to the Terasaki Center for Japanese Studies for awarding me the Sasakawa and Aratani Fellowships for dissertation research and writing.

In addition, I wish to express my appreciation to the following individuals for their friendship and support, and for keeping me sane through graduate school: Jihyeon Cha, Mayumi Ajioka, Kotoka Nakamura, Yoshiko Fukuyasu, Hee Ju, and Noriko Day.

Lastly, I would like to thank my parents for being my role models and letting me be me, my sister for her generosity and kindness, and my daughter for bringing so much love and laughter, being my *true* inspiration, and for her patience.

## VITA

- 2005  
B.A., Linguistics, TESL minor  
University of California, Los Angeles  
Los Angeles, California
- 2005-2007  
Teaching Assistant  
Asian Languages and Cultures Department  
University of California, Los Angeles
- 2007  
M.A., East Asian Linguistics  
University of California, Los Angeles  
Los Angeles, California
- 2007-2008  
Adjunct Faculty  
Modern Languages Department  
Santa Monica College, Santa Monica, California
- 2009-2010  
Teaching Associate  
General Education Freshman Cluster Program  
University of California, Los Angeles
- 2010-2012  
Teaching Fellow  
Asian Languages and Cultures Department  
University of California, Los Angeles

## Publications and Presentations

- Kaneyasu, M. (2007, April). *Distinguishing artificial vs. spontaneous spoken discourse: A corpus-based study of Japanese final particles ne, yo, and yone in drama and conversation*. Paper presented at the American Association for Applied Linguistics (AAAL) Conference, Costa Mesa, CA.
- Kaneyasu, M. (2011, March). *Context-specific form-function links: The case of no-wa constructions in Japanese newspaper discourse*. Paper presented at the 17<sup>th</sup> Workshop on East Asian Languages, Los Angeles, CA.
- Kaneyasu, M. (2011, April). *The role of the particle GA in the local discourse organization of conversational Japanese*. Paper presented at the Association of Teachers of Japanese Annual Conference, Honolulu, HI.
- Kaneyasu, M. (2011, July). *Stance taking in Japanese newspaper discourse: The use and non-use of copulas da and dearu*. Paper presented at the Stance Workshop: Stance Marking Across Languages: Typological, Diachronic & Discourse Perspectives, Hong Kong, China.
- Kaneyasu, M. (2012, March). *Frequency-based approach to formulaicity: Morphemic bundles in Japanese spoken discourse*. Paper presented at the 18<sup>th</sup> Workshop on East Asian Languages, Santa Barbara, CA.
- Kaneyasu, M. (2012, March). *Semi-schematic constructions in Japanese spoken discourse: A corpus-driven approach*. Paper to be presented at the 5<sup>th</sup> Meeting of the Formulaic Language Research Network (FLaRN 5), Tilburg, the Netherlands.
- Ajioka, M., Kaneyasu, M., Kawanishi, Y., & Iwasaki, S. (to appear). *Mikan yo mikan: Formulaic constructions and their implicature in conversation*. In Nam, S., Ko, H., and Jun, J. (Eds.), *Japanese/Korean Linguistics, Volume 21*. Stanford, CA: CSLI Publication.
- Kaneyasu, M. (to appear). *Grammar and interaction: Marking non-topical subject in Japanese conversation*. In Kondo-Brown, K., Saito-Abbot, Y., Satsutani, S., Tsutsui, M., and Wehmeyer, A. K. (Eds.), *New Perspectives on Japanese Language Learning and Linguistics: The Classroom and Beyond* [Tentative title]. Honolulu, HI: National Foreign Languages Resource Center.

## Chapter 1

### Introduction

#### 1.1. Purpose of the study

This dissertation investigates patterns and functions of fixedness or formulaicity in Japanese spoken language. A growing number of researchers, whose interests span a wide range of language-related fields, (e.g., lexicography, corpus linguistics, cognitive linguistics, usage-based linguistics, first/second language acquisition, inter alia), have argued for the centrality of the notion of formulaicity/idiomaticity in understanding what native speakers know and do with the language. However, an overwhelming majority of evidence comes from English, which differs from Japanese typologically. In particular, agglutinating morphology poses methodological and conceptual challenges for examining formulaicity in Japanese language. As a result, not much is known about Japanese formulaicity, except for fixed idioms and phrases, such as *te o yaku* ‘have difficulty with’ (lit. ‘burn one’s hand’) and *o-tsukare-sama deshita* ‘thank you for your hard work’ (lit. ‘(you) were honorably tired’ or ‘(you) seemed tired’). The present study employs a corpus-based, frequency-driven approach to empirically explore formulaicity in Japanese spoken language.

Conducting research into formulaic language entails taking a particular perspective on what language is and how linguists should approach their object of study. One perspective that is shared by many linguists working extensively on formulaicity/idiomaticity of language and related issues is distrust in an intuition-based analysis. Meanings of individual words may be more accessible to our intuition because they represent static, decontextualized meanings. On the contrary, functional and structural patterning of discourse elements emerge from language

use (e.g., Bybee 2001, 2006; Hopper 1987, 1998), and formulaic strings may also perform highly contextualized socio-interactional functions. Sinclair (1991) states that “human intuition about language is highly specific, and not at all a good guide to what actually happens when the same people actually use the language” (p. 4). A number of usage-based studies have provided strong evidence supporting this statement (see Thompson & Ono 2010; see also Clancy 2010).

Another perspective, commonly taken by formulaic language researchers, speaks directly to the issue of formulaicity, which is summarized in Sinclair’s (1991) “idiom principle” stated as follows:

The principle of idiom is that a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments. (p. 110)

Wray (2002) conveys a similar idea in her ‘needs-only analysis’ account for first language acquisition. In this account, children learn and use word strings as chunks unless there is need to break them down. What is crucial is that the need does not derive from an innate grammatical knowledge (see Hopper 1998) but arises from socio-interactional consciousness. For example, the sequence “thank you for having me” is linked to a particular meaning/function within a specific social situation, allowing it to remain unanalyzed into adulthood. For many multiword items, nevertheless, children first learn them holistically and later analyze and abstract the meaning of component parts as well as general syntactic rules based on input evidence from other strings and/or communicative necessity (Bolinger 1976: 1-9; Nattinger & DeCarrico 1992: 12). Thus, adult native speakers have stored and available in their mental lexicon linguistic chunks of various sizes from morphemes to sentential units with overlapping elements (see, for example, heteromorphic distributed lexicon model,<sup>1</sup> Wray 2002, 2008). While adult native

---

<sup>1</sup> Heteromorphic distributed lexicon model permits multiple part-mappings of the same information in the lexicon (cf. continuum models discussed earlier). In this model, linguistic material is stored in chunks of different sizes,

speakers have both creative and formulaic potentials, they are not well balanced in how often they are utilized in naturally occurring language production and comprehension.

[N]ative speakers do not exercise the creative potential of syntactic rules to anything like their full extent, and ... if they did so they would not be accepted as exhibiting nativelike control of the language. The fact is that only a small portion of the total set of grammatical sentences are nativelike in form — in the sense of being readily acceptable to native informants as ordinary, natural forms of expression, in contrast to expressions that are grammatical but are judged to be ‘unidiomatic’, ‘odd’, or ‘foreignisms.’

(Pawley & Syder 1983: 193)

As anyone who has learned or taught a foreign or second language as/to an adult may have also experienced, I have noticed on many occasions a serious gap between grammaticality and nativelikeness in sentences/utterances produced by non-native speakers. For instance, in the intermediate and advanced Japanese classes, my students repeatedly produced grammatically accurate, and yet, idiomatically unnatural sentences/utterances. It was not just my students who struggled with nativelikeness of utterances. On some occasions, I, as a teacher, found it extremely difficult to find adequate explanations for preferring to say something a certain way rather than some other ways. One of the main objectives of the current research is to find some patterns of formulaicity in native speakers’ speech and provide functional and socio-cultural explanations for the way they speak.

Uses of linguistic expressions can also be highly restricted to particular contexts.

Fillmore (1979) emphasizes that formulaic expressions must be learned in association with the context in which their use is appropriate. At least some of the routinized language uses seem to be inseparable from the associated social and communicative situations. As an illustration, I list below some expressions of prohibition whose uses are linked to specific contexts.

---

including morphemes, words, multi-word strings, and partially lexically filled constructions with slots for variable items, all of which are treated as if they are single morphemes (or morpheme equivalent units). “The determination of which components are separately stored and, thus, how much flexibility there is for varying the form of the complete string, depends on input evidence and use need (i.e., needs only analysis)” (Wray 2008: 15).

<u>Prohibition expression</u>	<u>Context of use</u>
(1-1) [Verb] <i>cha dame.</i> <i>tatai cha dame.</i> hit TE.TOP no.good '(You) can't hit.'	caregiver to child (or child to caregiver)
(1-2) [Verb] <i>nai.</i> <i>oshaberi shi nai.</i> talk do NEG 'Don't chatter.'	teacher to student
(1-3) [Noun] <i>wa</i> [Verb] <i>nai koto.</i> <i>rooka wa hashira nai koto.</i> hallway TOP run NEG NML 'Don't run the hallway.'	school hallway
(1-4) [Verb] <i>bekarazu.</i> <i>shibafu ni hairu bekarazu.</i> grass DAT enter AUX:NEG 'Keep off the grass.'	park grass area
(1-5) [Noun] <i>kinshi.</i> <i>kootsuu kinshi. / tachiiri kinshi.</i> traffic prohibition / entry prohibition 'No thru traffic.' / 'No entry.'	construction site / traffic sign

Investigation into formulaic language should not be limited to a small set of fixed linguistic expressions represented by proverbs and idioms, which are considered rather peripheral to the natural language use. Rather than relying on our perceptual salience, the study adopts a frequency-driven approach to identify and analyze most commonly used multi-morphemic sequences, or “morphemic bundles,” in Japanese spoken discourse. The identified morphemic bundles are further scrutinized from functional and socio-interactional perspectives. The overall aim of the present study is to provide an empirically adequate account of the observed patterns of formulaicity as well as a further insight into cognitive and social foundations for the native speakers' formulaic speech.

## 1.2. Basic unit of analysis

The status of “word,” which is the basic unit for analyzing formulaicity in English data, needs to be reexamined before investigating Japanese data. Dixon and Aikhenvald (2002) state that “the idea of ‘word’ as a unit of language was developed for the familiar languages of Europe” (p. 3). They further note that only some languages have a lexicon with the meaning “word.” In Japanese, the native word *kotoba* does not carry a specific meaning of “word,” and it may refer to morphemes, words, phrases, sentences, speech, or even language itself. Another Japanese term *tango* refers to smallest grammatically meaningful units, and it may be regarded as more equivalent to the concept of “word” in English. However, in ordinary sense of the term, *tango* only refers to unconjugated dictionary forms of words, which are difficult to find in natural spoken data.

In addition to the lack of a term for “word,” the agglutinating nature of Japanese language, especially a rich array of verbal suffixes, gives another reason for not adopting word as the basic unit of analysis in Japanese language. For instance, the following utterance, excerpted from naturally occurring conversation, the verb, *iu* ‘say,’ is followed by five suffixes (i.e., *n-ja-nai-n-da*) and two utterance final pragmatic particles (i.e., *yo-na*).

(1-6) Verbal suffixes and pragmatic particles (Conversation: Ono\_Ryokoo)

好き っ て いう -ん- じゃ -ない -ん- だ -よ -な 。

*suki tte iu-n-ja-nai-n-da-yo-na.*

like QT say-NML-COP-NEG-NML-COP-PP-PP

‘It’s not that (he) likes (his job).’

Identifying word boundaries in verbal phrases like the one above would be very difficult, and can be done only arbitrarily at best. Although such arbitrariness might not be a major issue in a

small-scale qualitative study, it would be a major obstacle to achieving consistency and objectivity in a large-scale corpus-based study like this one.

The agglutinating nature of the language and the uncertainty surrounding the status of word are critical reasons for not adopting word as the basic unit of analysis in the current investigation of formulaic sequences in Japanese spoken discourse. Instead, the morpheme will be adopted as the basic unit of analysis in the present study.

### **1.3. Definition of formulaic language**

A great number of definitions have been proposed for formulaic language. These definitions can be classified into four types, according to aspects of formulaic expressions they consider most important. At one end, researchers focus on formal characteristics of formulaic language; formulaic expressions are defined as frequent and constrained multi-word sequences. At the other end, rather than formal properties, conventionalized functions of linguistic expressions are recognized as the central characteristics of formulaic language. In this type of definition, whether an expression is multiword or single word is not an important factor. The third type of definition is a psychologically oriented one as it focuses on holistic storage and retrieval of linguistic units. The fourth type of definition is probably the most inclusive and, at the same time, more obscure than the other three, as it is solely based on the fuzzy concept of native speakers' preference.

Below, I list a few examples of the four types of definitions of formulaic language:

#### **Form-based definitions**

“any sequence of two or more words that are perceived to be more constraint than usual in their co-occurrence” (Hudson & Wiktorsson 2009: 81)

“a multiword piece of language that occurs a lot” (Bannard & Lieven 2009: 299)

### **Function-based definitions**

“all conventional multiword expressions and ... single word expressions that serve speech act functions, such as *Hello!* and *Thanks!*” (Pawley 2007: 3)

“conventional ways of expressing common functional messages” (Wray & Grace 2007: 556)

### **Cognitive-oriented definitions**

“a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar” (Wray 2002: 9)

“various kinds of strings of linguistic items which seem to be holistically stored and retrieved from memory” (Kerz & Haas 2009: 97)

### **Preference-based definitions**

“a combination of at least two words favored by native speakers in preference to an alternative combination which could have been equivalent had there been no conventionalization” (Erman & Warren 2000: 31)

“native speakers’ preferred combinations of words” (Erman 2009: 324)

One way to evaluate the usefulness of these definitions is to determine whether or not these definitions are empirically observable or testable. From this perspective, formal and functional variables in the form-based and function-based definitions are empirically identifiable and measurable. On the other hand, the cognitively oriented definition and the preference-based definition can only be operationalized if we assign measurable variables, such as frequency counts or collocational strength, to the concept of holistic processing or native speakers’ preference.

Although methodological applicability is a practical way of looking at various definitions, it does not capture the dynamics of having diverse views on formulaic language. The four types of definitions presented above capture different aspects of essentially the same or related phenomenon. For instance, in describing a greeting expression “What’s up?” all four types of definitions can be applied.

**Form:** “What’s up?” is a constrained multiword sequence (i.e., it is non-compositional; \*what+is+up) which appears frequently in the language use.

**Function:** “What’s up?” carries a conventionalized functional message tied to a particular socio-interactive setting.

**Cognitive:** “What’s up?” is stored and retrieved from memory holistically rather than analyzed and generated by the language grammar.

**Preference:** “What’s up?” is a nativelike, idiomatic choice by native speakers in a particular socio-interactive setting, in preference to alternative expressions such as “How are you?” and “Hi.”

There is a stronger connection between the form-based and the cognitive-oriented perspectives on the one hand, and between the function-based and the preference perspectives on the other. That is, the former pair focuses on non-compositionality of formulaic language, whether in form or cognitive representation, while the latter pair highlights the socio-interactive appropriateness and idiomaticity of formulaic language.

In the present study, I begin with a form-based definition of the formulaic language as it allows us to empirically identify a set of linguistic (morphemic) chunks that occur a lot in language use. The most frequently occurring strings of words in a given register have been identified as lexical bundles (e.g., Biber, Johansson, Leech, Conrad & Finegan 1999). This study adopts the basic methodology for identifying bundles in a corpus (see Section 2.6 for the review of lexical bundle studies and Chapter 3 for the methodology used in this study). We can use the

identified morphemic sequences as a starting point for further investigating other aspects of formulaic language, such as their form-function links, cognitive representation, and preference, with an ultimate objective of gaining a realistic account of formulaic language in Japanese spoken discourse.

#### **1.4. Scope of the research**

In accordance with the usage-based view of linguistic patterns as the cognitive organization of one's experience with language (Bybee 2006), the current study takes different social/communicative situations as constituting different language experiences for the speaker. From a cognitive perspective, even within a single linguistic community, native speakers are likely to have different sets of familiar formulaic language depending on their individual experiences with language in various situations (Biber & Finegan 1994; Biber, Reppen & Conrad 1998). "Verbal repertoire" (Hymes 1984) of individual speakers consists of both dialectal and register variations. An examination into formulaic language must take into account the linguistic variation; in this sense, the formulaic language analysis can be considered a kind of register analysis which "explore[s] the link between linguistic expression and social situation, with a view toward explanation" (Biber & Finegan 1994: 7).

The linguistic registers can be highly specialized, for example, sports announcer talk (see Ferguson 1983) and academic discourse (e.g., Kerz & Haas 2009; Dorgeloh & Wanner 2009), in which only a small number of speakers within a given linguistic community would ever develop fluency. In this study, three spoken registers are considered: informal conversation, formal interview, and academic speech. Ordinary informal conversation among friends and family members is arguably the primordial form of language use (Schegloff 1979) which a child native

speaker of any language encounters and acquires before other more sophisticated or specialized language use. It is placed at the most central and fundamental part of the individual's verbal repertoire. Formal interview and academic speech are much more specialized than conversation, and among the three, academic speech is the most specialized register.<sup>2</sup> Examining the three spoken registers allows us to compare the types and functions of multi-morphemic sequences found in different registers, which helps us gain a deeper insight into the interrelationship among form, function, and context.

### **1.5. Research objectives**

The ultimate goal of this dissertation is to gain a realistic account of formulaicity, in the sense of ubiquity in actual language use, in Japanese spoken discourse. To accomplish this goal, first, a corpus-based, frequency-driven approach is taken to empirically identify morphemic chunks that are used most commonly in the three spoken registers. These morphemic strings, identified as “morphemic bundles,” are scrutinized under structural and functional analyses based on a detailed examination of actual contexts of use in three registers. The structural and functional analyses of the morphemic bundles reveal much about language specific characteristics of fixedness in spoken language as well as register variations in the use of different types of morphemic bundles, which are linked to particular communicative needs of the different registers.

At the same time, the morphemic bundle analysis exhibits some limitations. It is found that about three fourths of the bundles do not represent form-function units on their own. That is, there are shorter and longer formulaic sequences with respect to form-function link (see Section 4.6). Many of these non-form-function unit bundles are found to be part of partially-filled, or

---

<sup>2</sup> The three registers have other difference as described in section 3.3.

semi-fixed, constructions. The present study argues that these constructions, with open slots and variations in some forms, represent an important type of formulaicity in Japanese native speaker's language use and knowledge. The study aims to show how this type of formulaicity is multifariously linked to functional, cognitive, and socio-interactional motivations, serving to fulfill various needs of the speakers in the Japanese socio-linguistic community.

## **1.6. Outline of the dissertation**

In Chapter 1, I have introduced the context and scope of the study, defined the concept of formulaic language and the basic unit of analysis, and described the main aims as well as specific objectives to be pursued in the present study. The investigation in the present study begins with the form-based definition of formulaic language as the most frequently occurring strings of morphemes in a given register, and later explores other aspects of the multi-morphemic sequences. Chapter 2 reviews relevant studies on formulaic language and other related issues, and presents a theoretical background for the present study. Chapter 3 describes and provides a rationale for the data and methodologies. Chapter 4 describes and analyzes the most frequently occurring morphemic sequences in detail, and discuss register variations in the use of the multi-morphemic units. In Chapter 5, the study looks beyond the fixed strings of morphemes, and presents a more dynamic view on formulaicity by describing and analyzing partially-filled constructions, or semi-fixed constructions, which encompass many of the specific morphemic strings. The analysis of the semi-fixed constructions help us gain a deeper insight into the patterns of formulaicity in native speakers' speech and the cognitive and socio-cultural basis for preferring to speak in a certain way. Finally, Chapter 6 summarizes the major findings in the

present study and briefly discusses pedagogical implications for Japanese as a foreign language learning and teaching.

## Chapter 2

### Literature Review

#### 2.1. Overview

This chapter presents theoretical background for the present study. Studies reviewed in this section fall under a wide range of language-related disciplines, including phraseology, lexicography, speech act theories, corpus linguistics, cognitive linguistics, usage-based linguistics, conversation analysis, language acquisition research, and formulaic language research. One common thread among these diverse fields is that they take as their starting point the actual use of language. Language use refers to both spoken and written production and comprehension of linguistic discourse in natural settings. The shared commitment to the naturally occurring language as the target of linguistic inquires and the data source makes it highly significant and unlikely a coincidence that researchers in these diverse fields have arrived at similar conclusions regarding the nature of language. They all emphasize the emergent nature of language as well as the central role of holistic processing in the use of language.

The usage-based view of language is also widespread among researchers of conversation and grammar, cognitive linguistics, conversation analysis, and other related fields, working on Japanese data (e.g., Ono & Jones 2008; Ono & Thompson 2003; Ohori 1997; Mori 1999; Maynard 1993, 2002). However, as I have mentioned in the introduction, there is a lack of evidence and studies in general concerning holistic processing of linguistic elements in Japanese spoken discourse. This is partially due to the agglutinating and context-oriented nature of the language and the entailing methodological challenges. One of the main goals in this section is to review previous Japanese usage-based studies and conversation analytic studies from a

formulaicity perspective, in an attempt to uncover possible links between characteristics of Japanese language-in-use and formulaicity.

Fixedness in the language use in natural settings has been examined from formal-semantic, socio-interactional, and cognitive aspects, with a limited literature also looking into the interrelationships among those aspects. This chapter will unfold in four main sections to highlight each of these three aspects and present a new perspective on findings from previous usage-based studies on Japanese discourse. First, I will review relevant studies with a particular focus on the formal-semantic aspect of formulaicity. Secondly, I will turn to the socio-interactional aspect of formulaic language, reviewing studies related to speech act theories and other pragmatic concerns. The following section will highlight the cognitive aspect, that is, the holistic storing and retrieving of the formulaic language. This section also touches on the ontogenetic and phylogenetic consequences of the way our minds handle/respond to linguistic activities. I will also survey first language acquisition and second language learning studies that deal with the notion of native-like idiomaticity. The next section will survey usage-based approaches to Japanese language which offer much insight into the interrelation between typological characteristics of the language and formulaicity. The last section focuses on studies of lexical bundles which are directly relevant to the corpus-based, frequency-driven approach taken in the present study.

## **2.2. Formal-semantic perspectives on formulaic language**

One of the earliest language-related fields to recognize and describe noncompositional features of the language use by native speakers was lexicography (Cowie 1998; Pawley 2007). Cowie (1998) names H.E. Palmer and A.S. Hornby as two “founding fathers of EFL [(English as

a foreign language)] lexicography” (p. 210) who began collecting and classifying multiword units in the late 1920s, published a classification of word combinations (Palmer 1933), and an idiomatic dictionary (Hornby, Gatenby & Wakefield 1942). According to Palmer (1933: 5), “a collocation<sup>3</sup> is a succession of two or more words that must be learned as an integral whole, and not pieced together from its component parts.” His definition resonates with the views of many usage-based linguists today (e.g., Bybee 2001; Hopper 1998; Altenberg 1991; Wray 2002). Palmer (1933: 8) argues, these collocations (e.g., *to ask a question, to do a favor, to give trouble, or to have patience*) present challenges to EFL learners who must rely on guess work and analogy of their native languages. Thus, the learners may produce unnatural expressions such as *to make a question, to perform a favor, to do trouble, and to keep patience*.

The kind of multiword strings extensively studied by Palmer and Hornby are not classical idioms (e.g., *spill the beans, bite the dust*) that traditional lexicographers often include in their descriptions, but rather “restricted collocations” (Cowie 1981, 1998; Gläser 1988; Howarth 1996) in which “at least one element has a specialized sense that occurs only in combination with the other elements” (Pawley 2007: 25). More recent corpus-based studies of natural language texts have revealed that the restricted collocations are much more commonly used than true idioms (Cowie 1998; Howarth 1996, 1998; Moon 1998). Two volumes of large-scale collocational dictionary, Oxford Dictionary of Current Idiomatic English (ODCIE) (1975/1983), were “the first English dictionary to distinguish between pure idioms, figurative idioms, and restricted collocations” (Pawley 2007: 25).

The three types of multiword units are considered prototypes along a continuum from most compositional/least formulaic to least compositional/most formulaic. The continuum model basically promotes the idea that “formulaic sequences can best be described as lying along

---

<sup>3</sup> Palmer (1933) uses the term ‘collocation’ in a general sense to cover any multiword unit.

a continuum” (Wray 2002: 62), and it is widely accepted (e.g., Cowie 1981; Pawley & Syder 1983; Peters 1983; Gläser 1986; Van Lancker 1987; Givón 1989; Altenberg 1991, 1998; Howarth 1998). The continuum is broadly divided into three major categories: free combinations, restricted collocations, and idioms, which are further classified into figurative idioms and pure idioms (Howarth 1998: 164):

<i>blow a trumpet</i>	(free combination)
<i>blow a fuse</i>	(restricted collocation)
<i>blow your own trumpet</i>	(figurative idiom)
<i>blow the gaff</i>	(pure idiom)

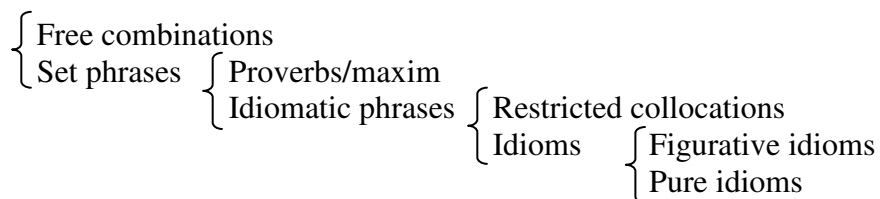
Altenberg (1998) examined recurrent word-combinations in the London-Lund Corpus of Spoken English, and concluded that “the great majority of the examples occupy a position along the cline between fully lexicalized units and more or less conventionalized expressions, and they illustrate very clearly the difficulty — or impossibility — of making a sharp distinction between lexicon and grammar” (p. 121). Altenberg (1998) further argues:

Speakers engaged in spontaneous interaction are in constant need of easily retrieved expressions to convey their intentions and reactions in discourse. At their disposal, they have a large stock of recurrent word-combinations that are seldom completely fixed but can be described as ‘preferred’ ways of saying things—more or less conventionalized building blocks that are used as convenient routines in language production.  
(pp. 121-122)

As Wray (2002) points out, many of the continuum models draw on multidimensional modes of description. For example, Van Lancker’s continuum (1987) lists terms that lie within three different domains of linguistic description: forms (*sentence stems, collocations, lists, etc.*), functions (*indirect requests, song lyrics and titles, pause fillers, etc.*), and meanings (*metaphors, idioms, proverbs, etc.*) (Wray 2002: 63). The classical continuum model illustrated by Howarth (1998) (see above), also, incorporates both formal and semantic aspects. These continuum

models underscore the difficulty of achieving a comprehensive and realistic account of formulaic language based on a single dimension of description.

The fuzzy boundary between free combinations and restricted collocations has also been recognized by Japanese linguists (e.g., Miyaji 1982, 1985; Kunihiro 1985). Miyaji (1985) defines formulaic language (*kanyooku*) as the combination of two or more words whose connection is rather fixed, and whose meaning cannot be obtained from the meanings of its components. Miyaji (1985: 63) classifies phrases (*ku*) into the following categories:



Free combinations are phrases whose component words can be freely combined, and the meanings of the phrases are transparent, that is, the sum of their parts. According to Miyaji, phrases such as *kumo ga nagareru* ‘clouds drift,’ *hana o nagameru* ‘gaze at flowers,’ and *sora ga aoi* ‘the sky is blue’ belong to this category. Set phrases, on the other hand, are restricted in how individual words combine. Examples of this type are proverbs and idiomatic phrases such as *ase o kaku* ‘sweat,’ *uso o tsuku* ‘lie,’ and *kaze o hiku* ‘catch a cold.’ Miyaji notes that some phrases can be both free combinations and set phrases. For example, a phrase, *me o tsuburu* (eye ACC close), has two different meanings, one as a free combination ‘close one’s eyes,’ and the other as a set phrase ‘pretend not to see something.’ Based on this observation, the author emphasizes that there are no clear-cut boundaries between the categories of phrases.

With respect to the type of formulaic language, it is the idiomatic expressions, as listed and categorized by Miyaji (1985) above, that has been given much attention in Japanese

phraseological studies There is no doubt that these idiomatic expressions should be taken as formulaic language in the sense that they constitute “single choices” (Sinclair 1991) for the speakers. However, we have yet to find out how much of our natural language use is based on set phrases in contrast to free combinations. In addition, Japanese phraseological tradition does not discuss other types of formulaic language, including native speakers’ preferred ways of conveying common messages and most frequently occurring strings of morphemes. This unexplored area of formulaicity is the target of the current study.

Advancement of computer-assisted technology and availability of large corpora of spoken and written discourse have made the tasks of identifying formulaic sequences and describing them more scientific and practical in one respect, but even more challenging in another. Corpus linguistics “is a research approach developed for empirical investigation of language use” (Biber, Kim & Tracy-Ventura 2010: 76). It is considered one of the most useful methodologies in discovering the association patterns of linguistic features with other linguistic as well as non-linguistic factors (Biber et al. 1998). Biber et al. (1998: 4) lists four central characteristics of corpus-based analysis:

1. It is empirical, analyzing the actual patterns of use in natural texts.
2. It utilizes a large and principled collection of natural texts, known as a “corpus,” as the basis of analysis.
3. It makes extensive use of computers for analysis, using both automatic and interactive techniques.
4. It depends on both quantitative and qualitative analytical techniques.

McEnery, Xiao, and Tono (2006) emphasize the methodological role of corpus linguistics, which provides a new facet for functional linguists in implementing usage-based approaches

using large corpora. Among a variety of linguistic issues examined based on this computer-mediated empirical methodology, lexical bundles (e.g., Biber et al. 1999), collocations<sup>4</sup> (e.g., Sinclair 1987, 1991; Stubbs 1995), and other co-occurring patterns of grammatical and lexical items (e.g., Tao 2001, 2003) are particularly relevant to the present investigation of formulaic language.

Tao (2001), using corpus linguistic tools, reports the co-occurring patterns of the verb *remember* and types of subject, tenses, utterance types, modal words, post-verbal elements, syntactic positions, and prosodic properties in spoken discourse. In particular, he has found that the post-verbal elements of *remember* tend to be simple forms (i.e., a zero form, simple NPs, or simple relative clauses), rather than complex complement clauses with *that*, *-ing*, and *to + infinitive*, which had been the most studied types of post-verbal elements in traditional non-empirical syntactic analyses. He further illustrates, from a discourse-pragmatic perspective, that the recurrent patterns of *remember* in spoken discourse are motivated by discourse functions they perform, namely marking epistemic stance and regulating interaction.

Based on an extensive corpus data, Sinclair (1991) generalizes that “most everyday words do not have an independent meaning, or meanings, but are components of a rich repertoire of multi-word patterns that make up text” (p. 108). This tendency is stronger for frequent words or frequent senses of words than less frequent words or senses. For instance, words that syntactically constitute the phrase *of course*, that is, *of* and *course*, do not carry independent meanings in the phrase; the meaning of the phrase is the property of the phrase itself. While Tao (2001) focuses on a relationship between recurrent formal patterns and discourse functions,

---

<sup>4</sup> Collocation is a corpus-derived concept, and refers to “characteristic co-occurrence of patterns of lexical or grammatical words” (McEnery et al. 2006: 56), or “patterns of preferred co-occurrence of particular words” (Ellis & Frey 2009: 474).

Sinclair draws a link between the evidence of collocation and the cognitive processing of language. He argues that the presence of collocation is the result of a single choice, or the idiom principle:

The principle of idiom is that a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments. (Sinclair 1991: 110)

Biber et al. (2010) investigate, cross-linguistically, types of lexical bundles, which refer to “multi-word sequences that recur most commonly in a given register” (p. 77), in English, Spanish, and Korean corpus data. Three major functional categories of lexical bundles are identified as (1) stance expressions, (2) discourse organizers, and (3) referential expressions (Biber, Conrad & Cortes 2003; Biber, Conrad & Cortes 2004). The ultimate goal of the comparative corpus-driven study is “to investigate the extent to which lexical bundles can be considered a universal building block of discourse across languages” (Biber et al. 2010: 78). They found the overall rarity of lexical bundles in Korean data,<sup>5</sup> which they explain has to do with the status of word, various typological factors, including the flexibility of word order, the phenomenon of ellipsis, and morphological inflections. The study “suggests the possibility that formulaic language may not be universally important in the same way or to the same extent across languages” (2010: 92).

There are two significant advantages to the corpus-based, computer-assisted studies. One is that they allow a linguist to handle a large amount of data otherwise impossible. The other advantage is frequency counts and statistical analyses of various co-occurring patterns. As I mentioned earlier, however, these advantages are by no means flawless. First, we need to recognize that no matter how large a collection of corpora is, it is only a small sample of

---

<sup>5</sup> Sohn and Nam (2011) point out that the number of Korean lexical bundles in this study “is based on orthographic words and does not represent morphological characteristics of Korean.”

naturally occurring languages in our daily lives. Secondly, a corpus may be biased with respect to types of speakers, settings, and the like. With respect to the power of computer processing in detecting co-occurring patterns and counting their frequencies, there is a significant limit to what the computer can recognize as patterns, and there are also orthographic difficulties in the case of Japanese corpora.

Despite these limitations and challenges, the corpus-based approach gives us much insight into the frequently occurring formal patterning of the language, which we can use as a point of departure for more detailed qualitative analyses of the nature of fixedness in the use and the mental representation of language. I will come back to the discussion of lexical bundles and register variations in Section 2.6.

### **2.3. Pragmatic/socio-interactional aspect of utterances**

In both Western and Japanese linguistic traditions, the pragmatic nature of language has long been recognized. A Japanese linguist, Jimbo (1948), asserts that learning a language means learning common practices or idiomatic ways of doing things in a given speech community. Language is based on the common practice of the community. Shiraishi (1950) argues that idiomaticity should be the basis for determining whether or not a certain linguistic use is correct. Idiomatic expressions (*kanyooku*) refer to any expressions, whether they are everyday ordinary ones or specialized academic ones, that are commonly used, and have become fixed and specialized in their meanings or pragmatic functions. In certain cases, they may also refer to coined terms or words that certain individuals use repeatedly. He also claims that we should not think of idioms and idiomatic expressions as something special or unordinary. Only when we

compare them to grammatical facts, do they seem to be rather specialized and exceptional cases. But they themselves are quite natural ordinary speech.

Speech act theory proposed by Austin (1962) also advocates the view that we do much more than just conveying information with language. Austin argues that “if we are to see the parallel between statements and performative utterances, ... we must consider the total situation in which the utterance is issued—the total speech act” (p. 52). There are three levels of action identified by Austin: a locutionary act, an illocutionary act, and a perlocutionary act. Among these, one that is particularly relevant to formulaic language is the illocutionary act, that is, social and interactional actions performed by utterances, including greeting, apologizing, requesting, promising, and so on.

Pawley (2007), following Lyons (1968), calls conventional expressions for performing illocutionary speech acts “situation-bound expressions” (p. 5). Expressions “such as *Hello!*, *Pleased meet you*, *Dear X*, *Here’s to X!*, *I declare this (meeting, etc.) closed*, *If it’s good enough for X, it’s good enough for Y*, or *Give me a break, will you!*” are “tried and true ways of doing things, standard recipes for achieving social purposes” (Pawley 2007: 17). However, as Strawson (1964) points out, most illocutionary acts do not conform to convention, for instance, a convention for using performative verbs (Searle 1969), but they are essentially achieved by recognition of the speaker’s intention by the addressee. Bach (1998) illustrates this point, using an illocutionary act of apologizing. Use of a conventional performative, *I apologize*, may facilitate the addressee to recognize the speaker’s communicative intention of expressing regret, but other unconventional utterances, such as *oops*, may work just as effectively when uttered in a right way in a right situation.

Although the debates over the conventionality of speech act expressions mainly concern explicitness and formal fixedness of utterances, it is also important to consider the conventionality in the sense of a link between form, function, and context. In the case of apology, *oops* may be more conventional than *I apologize* in a particular context. Its conventional significance may be applicable to an entire speech community, to a group within the speech community, or just to an individual (Hymes 1968). Whatever the case, formulaic expressions are learned in association with the context in which their use is appropriate (Fillmore 1979). The routinized language use and the associated social and communicative situations are inseparable.

Genre-specific formulaicity is one of the conventional links that have been paid much attention (e.g., Kerz & Haas 2009; Sams 2009; Dorgeloh & Wanner 2009). For example, Kerz and Haas (2009) demonstrate that there is a range of register-specific partially lexically filled constructions around “research predicates,” such as *analyze*, *study*, *examine*, and *investigate*, in academic writing. One of the examples they present is “[det *aim/objective/purpose* (of NP1)] [*be to V<sub>research verb</sub> NP2*], whereby the N slot of NP1 is usually filled by the research noun *study* or *analysis* and NP2 denotes a phenomenon under investigation” (p. 109). The construction is semantically predictable and involves lexically constrained slots. According to the authors, these constructions are prefabricated in the sense that they are registered in linguistic memory as chunks, indexed with their implications and contexts of use. The study provides empirical evidence that some constructions are frequently used only in a certain domain of linguistic activity, and these constructions become entrenched and conventionalized for the users of the specific register. Formulaic sequences emerge from frequent use in a specific communicative

context, and as they become more and more conventionalized, the formulas in turn contribute to shaping the linguistic distinctiveness of a particular register.

In this section, I have reviewed some studies with particular focus on pragmatic aspect of utterances, which are found to be strongly tied to the contexts of their uses. In the next section, I will survey cognitive-oriented views of formulaic language that seek to provide a missing link between language use and linguistic knowledge of native speakers.

## **2.4. Cognitive-oriented approaches to formulaic language**

### **2.4.1. Grammar as constructions**

Much work has been done on the cognitive aspect of formulaicity within the paradigm of cognitive linguistics. One of the foundational goals of cognitive linguistics is to provide “a psychologically plausible account of language” (Segalowitz 2001: 11), which is grounded on the view of language as “an integral facet of cognition, ([and] not as a separate “module” or “mental faculty”)” (Langacker 1998: 1) and grammar as a bottom-up structured inventory of conventional linguistic units, that is, form-meaning pairings (Langacker 1987, 1991). On this view, traditionally distinguished domains of linguistic structures, namely, phonemes, morphemes, lexicon, phrase-level constructions, and sentence-level constructions, are all part of the same phenomenon of form-meaning mappings.

In particular, a sub-field called Construction Grammar (Fillmore, Kay & O’Connor 1988; Fillmore 1988; Kay 1997) has presented evidence that many linguistic constructions are mixtures of more regular and more idiomatic sub-constructions. A construction can be a single word or a longer sequence and it can be more concrete (lexically specified) or abstract (lexically

unspecified). Goldberg (1995) argues that what she calls argument structure constructions<sup>6</sup> carry their own meanings, independently of the constituent words consisting the constructions. For instance, she illustrates that, in the caused-motion construction, the meaning of causer directly causing the theme to move along a path described by the directional phrase (X causes Y to move Z) cannot be attributed to the main verb. Observe the following sentences.

(2-1) Argument structure construction (Goldberg 1995: 152)

- (a) Frank pushed the issue off the table.
- (b) Frank sneezed the tissue off the table.

In (a), it is possible to attribute the meaning of “causing something to move” to the verb *push*. In (b), by contrast, the verb *sneeze* cannot contribute the meaning of “causing something to move” to the construction since the verb itself is an intransitive verb. Therefore, the meaning of the caused-motion must come from the construction itself. The fact that one can use a non-prototypical intransitive verb to convey the meaning of caused-motion proves the conventionalization of construction-meaning pair.

From a formulaic perspective, this is very much similar to the form-meaning pairings seen in fixed idiomatic expressions, such as *kick the bucket*, in that the meaning of the whole is not predictable from the sum of its parts (Fillmore 1988). The study of argument structure constructions suggest that conventionalization of form-meaning pairings can occur not only at a concrete level of lexical co-recurrences, but also at a more abstract level, that is, at the level of grammatical constructions.

---

<sup>6</sup> Argument structure constructions are defined as “a special subclass of constructions that provides the basic means of clausal expression in a language” (Goldberg 1995: 3).

#### 2.4.2. Cognitive representation model

Grounded on the usage-based view of grammar as “the cognitive organization of one’s experience with language” (Bybee 2006: 711), Bybee (2006) explores implications of the fact that natural language use is highly lexically particular and conventionalized (e.g., idioms, prefabs) for cognitive representation of language.

In her search for a model that allows particular instances of use to affect representation, she adopts a version of exemplar theory (see Johnson 1997; Pierrehumbert 2001, 2002). In the exemplar model, each category is represented in memory by a large cluster of remembered tokens of that category. Each individual token of linguistic experience is categorized and mapped onto an identical, or most similar, exemplar, thereby strengthening its representation. If there is no similar exemplar, it is analogically classified according to perceived similarity to other existing ones. Exemplars with high token frequency serve as central members of these categories and they can be regarded as prototypes. Some members are more central than others (i.e., graded membership), but the boundaries are not discrete. Marginal members may share characteristics with the central members (i.e., family resemblance).

With empirical evidence from grammaticization studies (e.g., Haiman 1994; Bybee 2003) and construction studies (e.g., Kay & Fillmore 1999), Bybee (2006) argues that in order for a new construction to arise, whether with or without the hallmark of grammaticization, the speakers/hearers must register in memory the particular instances of constructions indexed with their implications and context of use. “They could not wait until they had heard the expression frequently in a certain context to register this in memory, because if they did not remember each time, they would not know that they had heard it before” (p. 722). So, for example, an expression *that drives me crazy* may be represented as a chunk in memory storage, but it can still

be related in representation to the individual units that comprise them (*that / drives / me / crazy*) as well as to the more general construction (*drives someone + Adj.*)

It is important to realize that the exemplar model was adopted to account for the diachronic language change including grammaticization. It makes sense, for instance, that throughout the development of *be going to* from a marker of intention to a marker future tense, specific instances of the purposive construction somehow had to be registered in cognitive representation, since otherwise “the construction could not be subject to the processes that comprise grammaticization” (Bybee 2006: 721).

But what about linguistic sequences that have been conventionalized to a specific meaning or function? Are the instances of formulaic sequences also registered in memory with representation of all the smaller component units and the general construction as well? The exemplar model seems to be in favor of this view. In ontogenetic experience with language, however, one is likely to experience the conventionalized use of the construction repeatedly without maybe ever experiencing the original forms or constructions. This seems especially relevant in how children acquire their first language. Since the present study is concerned with synchronic patterns and functions of formulaic language used by native Japanese speakers, it is essential to review and discuss cognitive models for the first language acquisition and language processing that attempt to explain the uses and patterns of formulaic expressions in ontogenetic language use.

### **2.4.3. Cognitive models for first language acquisition**

In usage-based first language acquisition research, children’s early syntactic development is found to be built around concrete individual words and phrases, and not based on some

abstract rules and principles (Peters 1983; Tomasello 2003). Even infants at one year of age are learning to identify their caregivers' communicative intentions expressed in utterances and to “use a single linguistic symbol to express their own communicative intentions about an entire experiential scene” (Tomasello 2003: 139). When children experience fixed utterances recurrently in their daily lives, and as they learn socio-communicative significance of these utterances in specific contexts, they are likely to store and access these expressions holistically to convey the conventional meaning or function. Therefore, the children's mental storage of linguistic material must contain linguistic chunks of various sizes from single morphemes and words to multi-word strings (Wray 2002, 2008).

In contrast to Bybee's exemplar model in which multi-word expressions are not only registered in memory as chunks, but also broken down into smaller components and mapped onto general constructions in mental representation, Wray's (2002, 2008) Needs Only Analysis model proposes that:

[t]he process of analysis which the native speaker child engages in is not that of breaking down as much linguistic material as possible into its smallest components. Rather, nothing is broken down unless there is a specific reason. (Wray 2002: 130)

The holistic storage of multi-word units advocated by Wray parallels the initial stage of a language acquisition model proposed by Peters (1983, 2009). In her model, children first extract and remember chunks from ambient speech. The chunks include greeting and farewell expressions as well as prefabs such as pronouns plus modals (e.g., *didja, do'ya, lemme, whatcha*), modals plus *to* (e.g., *wanna, gonna, hafta*), and locatives or demonstratives with forms of *be* (e.g., *here's, there's, that's*). In the process of extraction, Peters considers pragmatics to be the most influential driving force because it enables the child to interact with adults in desired ways to meet their physical and emotional needs. In Wray's words “[these chunks] are a local linguistic

response to a local nonlinguistic problem” (2002: 131). One kind of evidence Peters presents is the children’s association of *thankyou* with the act of both giving and receiving because they hear this expression when they give something to their caregivers. Another example she gives is the use of *piys* (*please*) by a child to mean *yes* as his caregiver teaches the child to respond with *please* to questions such as *do you want X?*<sup>7</sup>

All three researchers cited above, namely, Tomasello, Wray, and Peters, recognize the highly analytic phase in the children’s language development, following the initial holistic phase just described. At around 18 months (Tomasello 2003) to 20 months (Wray 2002), children begin to produce multi-word units, which can be classified into two types: word combination (e.g., *ball table*) and pivot schema (*more X*). Tomasello points out that although pivot schema involves productive use of linguistic symbols, it is not until children reach 24 months or so that syntactic symbols come into play. He thus characterizes these syntactically marked utterances as item-based constructions because they revolve around specific linguistic items in specific contexts.

In a similar vein, Wray and Perkins (2000) and Wray (2002), following Locke’s (1993, 1995, 1997) work, present a language development model in which “the relative proportions of holistic and analytic involvement in language processing alter, first in one direction and then the other” (Wray 2002: 132-133). In Phase 1, from birth to around 20 months, the language processing is governed by holistic strategies as described above. In Phase 2, beginning at 20 to 30 months, as children have undergone the vocabulary spurt, and lasting until about age 8, the initial holistic processing is taken over by more analytic processing of inputs. The adoption of the analytic strategy may also be affected by the development of literacy and formal education which focus on expression of more complicated messages and processing of more complex and decontextualized linguistic input. Then from about age 8 to 18, their language processing again

---

<sup>7</sup> This child eventually learned other ways to mean *yes* such as *mhm*, *yeah*, and *uh-huh* (Peters 2009: 393).

becomes increasingly holistic (Phase 3). Perkins (1999) and Wray (2002) provide functional explanation for this shift:

[The shift to holistic processing] occurs as it becomes increasingly inefficient to generate from scratch the now very wide range of utterances which are needed regularly. If the same, or similar, groups of elements are being continually encountered and/or produced, it will make good economical sense to store them as separate holistic items.  
(Wray 2002: 135)

By the time the first language learners reach Phase 4, the balance between holistic and analytic processing has been fixed. The relative proportion of holistic and analytic involvement in language processing from birth to adulthood is summarized in the figure below (adopted from Wray 2002: 133):

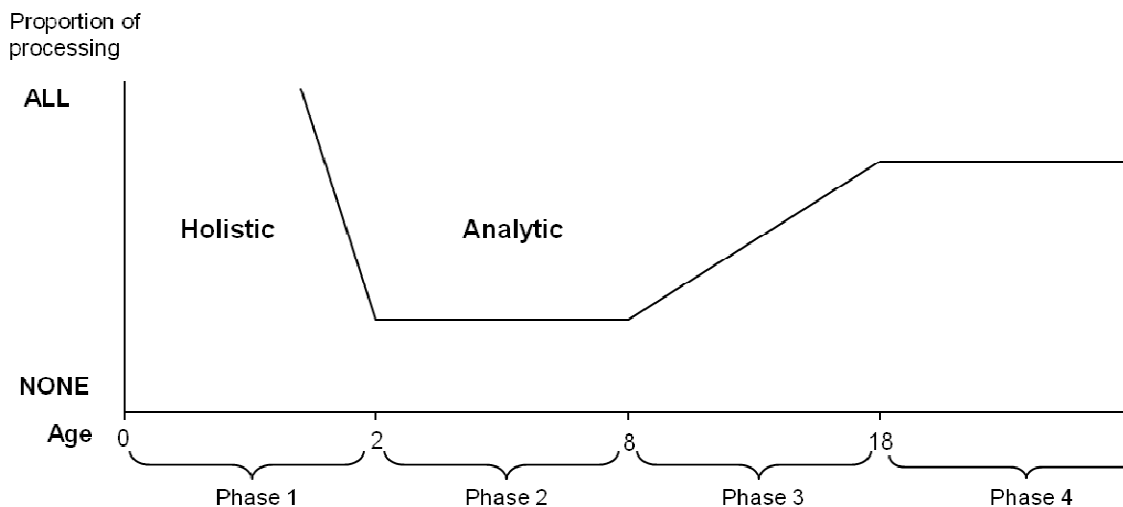


Figure 2.1. Balance of holistic and analytic processing from birth to adulthood

Although from age 2 to 8 the language is processed more analytically, at least some of the formulaic sequences observed in adult language, including the ones with obsolete vocabulary and structures, are not subjected to analysis because their usage is highly conventionalized. As we can see in Figure 2.1, adult language processing is characterized by preferred use of holistic

strategies, that is, storing of and accessing to a large number of prefabricated chunks in their daily communication.

The first language acquisition model described above also allows for the possibility that some types of holistically stored/accessed language in adulthood are newly acquired during and/or after the phrase 3 (8-18 years old). For example, languages like Japanese have different speech levels and styles (e.g., polite and honorific languages) whose use is required or encouraged in certain social situations. Socially-complex language use can only be learned by the first language learners as their area of activity and socialization broadens from the family to public sphere. A majority of the formulaic sequences identified in the present study belong to this type of adulthood formulaic language.

#### **2.4.4. Intersection between formulaicity and creativity**

Besides the balance between holistic and analytic processing, there is another issue that we need to consider in relation to the adult language processing: the question of creativity. This issue is discussed by some researchers working on formulaic-related phenomena (e.g., Pawley & Syder 1983; Tomasello 2003; Wray 2002). Pawley and Syder (1983) make a point that a reliance on ready-made expressions does not necessarily detract from the creativity of spoken discourse. Because the memorized sequences come ready-made, the speaker is freed from the task of composing such sequences word-by-word, and therefore can attend to other conversational activities. They believe that memorized sequences and phrases are the normal building blocks of fluent spoken discourse.

With respect to construction grammar, Tomasello (2003) states that “much of the creativity of language comes from fitting specific words into linguistic constructions that are

non-prototypical for that word on a specific occasion of use, with no implication that this requires a corresponding permanent lexical entry for the verb involved” (pp. 160-161). One example is Goldberg’s (1995) caused-motion construction presented earlier. Employing or interpreting the verb *sneeze* in *Frank sneezed the tissue off the table* as part of the caused motion meaning of the construction requires some imagination. Another example Tomasello gives is *He smiled her answer* which requires even more imaginative interpretation to be taken as the transfer of possession meaning of the construction.

Another type of creativity found to intersect with formulaicity is substitution of one lexical item in idiomatic expressions (Bybee & Cacoullos 2009). For instance, Bybee and Cacoullos report that a radio news reporter was heard to say *all chaos broke loose* instead of the more canonical *all hell broke loose*. Yet another type of creativity is extraction of one or more lexical items from fixed idiomatic expressions. Many films have titles of this type, such as *The Bucket List* (as in *kick the bucket*)<sup>8</sup>. These two types of creativity are not difficult to find especially within linguistic genres that value originality.

While the kinds of creativity mentioned so far represent somewhat specialized cases, the intersection between native speakers’ creative capabilities and reliance on formulaicity can be observed in the use of ordinary words in ordinary situations as well. For example, a Japanese expression *ii hito*, which is a combination of *ii* ‘good’ and *hito* ‘person,’ is considered to be a fully lexicalized unit, meaning ‘partner’ or ‘lover’ (Thompson & Ono 2009: 126), but it is also possible to use *ii hito* more or less compositionally (*ii + hito*) to mean a ‘good person.’ Thus, the unit *ii hito* requires two separate cognitive representations, one as a single idiomatic unit and the other as individual entries of lexical items despite the fact that their forms are identical.

---

<sup>8</sup> On a side note, although it is quite obvious, when this film was released in Japan, it was given a totally different title *Saikoo no jinsee no mitsukekata*, which can be translated as *How to find the best life*. Sense of creativity or originality is contingent upon the conventional norm that everyone is expected to share.

A kind of creativity-formulaicity interface that is relevant to the present study is a discourse-oriented one. Morphemic sequences identified in this study serve as various kinds of discourse frames anchoring novel or new information which requires more creative ability in composing.

#### **2.4.5. Problems for second/foreign language learners**

Second/foreign language teachers and researchers often comment on the difficulty formulaic language poses for adult learners of second or foreign languages. As an example, I cite from a preface to a book titled “Japanese core words and phrases: things you can’t find in a dictionary” (Shoji 1999):

At the intermediate and advanced level, what is required are means of expression that are culturally textured and contextually interwoven... At this stage, not only must students increase their general vocabulary and store of idiomatic expressions, but also realize that previously learned words and phrases must be transformed from locutions that are simple and fixed in meaning into ones that are multidimensional and many-layered. (p. 8)

Knowing the grammar and the vocabulary of the target language does not help the learners predict the meanings and/or functions of formulaic expressions. And even when they somehow figure out the meaning of a formulaic expression, “knowing that” does not always translate into “knowing how” (Fillmore 1979: 92). Because formulaic expressions are learned and used in close association with real time contexts or situations, they are particularly problematic for most adult learners who must learn these socially-particular and context-sensitive expressions in classroom settings.

The kind of formulaic expressions that have been identified to be most problematic for the learners are not pure idioms, but restricted collocations and discourse markers (Cowie 1981, 1998; Gläser 1988; Howarth 1996; Moon 1998). These collocations and discourse markers are

much more frequently used by native speakers than idioms, but they have not been the focus of ESL teaching material. With the lack of awareness on the part of second/foreign language researchers and teachers regarding the types and functions of most commonly used formulaic language in natural discourse, the intermediate and advanced learners tend to think that lexical resources are simply split into two extremes: free generative combinations and completely fixed idioms. Because the completely fixed idioms are much fewer than expressions that can be constructed from parts, many learners are unaware that formulaicity is a much more prevalent and essential part of the ordinary language use.

If we are to help second/foreign language learners get a real sense of fixedness involved in the target language use, we must first understand patterns and functions of formulaicity in ordinary language use ourselves. This study is to be taken as a first step into explicating the actual patterns and functions of formulaic language in Japanese conversational discourse.

## **2.5. Usage-based approaches to Japanese discourse**

### **2.5.1. Introduction**

As mentioned earlier, a central tenet in the usage-based paradigm is a methodological principle that entails naturally occurring discourse as one's primary data and discourse structures as the explaining ground for observable recurrent structures (Ochs, Schegloff & Thompson 1996). Clancy (2010) states:

Not only is much of the information arising from usage-based research new, it is also more accurate. When linguistic analysis is based on language use, the data are empirical and objective, and the findings are readily replicable, meeting the basic standards for scientific inquiry found in other disciplines. ... Data of language use thus avoids many pitfalls of intuition-based research, increasing the reliability of grammatical findings and, consequently, their descriptive adequacy. (p. 140)

Proponents of this principle have studied topics such as discourse structure (e.g., Du Bois 1985, 1987), child language (e.g., Clancy 1986), discourse markers (e.g., Schiffrin 1987), comparisons between written and spoken texts (e.g., Biber 1988; Chafe 1982; Chafe & Danielewicz 1987), cognitive processing (e.g., Chafe 1994), interactional units (e.g., Ford & Thompson 1996), and prosody (e.g., Fox 2001).

Investigating natural conversation has also been a long tradition in branches of sociology and anthropology, most notably sociolinguistics, conversational analysis, and linguistic anthropology. The basic tenet of conversation analytic works is to ask the question: What conversational participants are doing by talking in a certain way in terms of both its position and composition in “talk-in-interaction” (Sacks, Schegloff & Jefferson 1974; Schegloff 1987). This principle parallels, to a great degree, the view of linguistic anthropologists on conversation as “situated action” (Ochs et al. 1996).

With respect to formulaic language, many studies within the usage-based paradigm take the view that “routinized patterns ... exist because speakers need routinized ways to implement actions” tied to social contexts (Thompson & Couper-Kuhlen 2005: 482).

### **2.5.2. Patterns of fixedness in Japanese discourse**

In this section, I review previous Japanese usage-based studies and conversation analytic studies from a formulaicity perspective, in an attempt to uncover possible links between characteristics of Japanese language-in-use and formulaicity. In particular, I focus my review on varied word order, ellipsis and syntactically incomplete utterances, particles, and discourse markers. The brief survey of previous usage-based literature sheds new light on the ongoing

conventionalization of linguistic patterns and social actions implemented by conversational participants in real-time contexts.

### 2.5.2.1. Varied word order

In traditional grammatical studies and textbooks, it is generally taken for granted that Japanese is a predicate final, SOV language (e.g., Kuno 1973; Shibatani 1990). However, inspection of actual conversational data indicates that SOV word order is rarely used. This is generally attributed to the ellipsis of subject and other constituents as well as deviations from predicate-final word order (e.g., Fujii 1991; Ono & Suzuki 1992; Tanaka 2005).

Within a conversation analytic framework, Tanaka (2005) illustrates that, in real interaction, so-called canonical word order (i.e., SOV) is strongly correlated with performing dispreferred second actions of adjacency pairs, such as requests, suggestions, invitations, and offers. On the other hand, non-canonical word order, along with ellipsis, is closely related to performing preferred second actions. Since dispreferred actions are marked social actions and are much less frequently observed than unmarked preferred actions, Tanaka's analysis provides explanation for the rarity of the SOV word order in Japanese conversation. For instance, she demonstrates how a preferred second action to a request, namely, an acceptance, is done with a turn-initial verbal predicate *wakatta* 'understood,' which has a non-canonical word order with ellipsis of subject and object (p. 403, example 11). On the contrary, a dispreferred action of declination is done in a canonical word order with informationally redundant topic noun phrases *NP tte no wa* 'when it comes to NP' and *kochira de wa* 'on our side' (p. 412, example 19). Inclusion of hesitation markers *e:to* and *ano::*, and an adverb *chotto* 'a bit' further indicates the dispreferred status of the utterance. This study elucidates the interrelation between grammar and

socio-interactional action; “word order and ellipsis are vital resources available to Japanese participants for managing the ‘timing’ of social actions in the realization of preferred and dispreferred responses” (p. 425). It remains to be determined whether or not such correlation between word order and social action is quantitatively significant, and whether or not the association is in any way conventionalized for the speakers.

In a series of studies by Ono and his colleagues, it has also been suggested that a non-canonical, post-predicate constituent order may be tied to a function of expressing strong emotion of the speaker (Ono & Suzuki 1992; Ono & Thompson 2003; Ono 2006). For instance, Ono and Thompson (2003) report that when first person pronouns (e.g., *atashi* ‘I,’ *watashi* ‘I’) are used for emotive function, they tend to appear after the co-occurring predicates. In the following two examples, first person pronouns *watashi* and *atashi* occur after the predicates *warukute* and *aji o shimeten no*, respectively.

(2-2) Post-predicate utterances (Ono & Thompson 2003: 330-331)

- (a) *sugoi warukute watashi*  
very bad I  
‘I (feel) terrible.’
- (b) *moo sapuraizu no aji o shimeten no atashi*  
already surprise GEN taste ACC get PP I  
‘I (know how fun) a surprise party is.’

Prosodic and distributional evidence indicate that the post-predicate utterances are in fact planned and produced as non-predicate-final utterances from the onset (Ono 2006). The claim for the conventionalized link between the post-predicate construction and the function of emotive expression is similar to the links between various constructions and their functions in construction grammar studies (e.g., Goldberg 1995).

Both Tanaka and Ono found evidence in natural conversational data that so-called non-canonical word orders are associated with very common conversational practices (i.e., delivering preferred responses and achieving an emotive effect). Their findings raise fundamental questions regarding the appropriateness of labeling spoken Japanese predicate-final SOV language, and of conceptualizing word-order as a mere syntactic phenomenon.

### 2.5.2.2. Ellipsis and syntactically incomplete utterances

The Japanese language is well known for its elliptic nature especially in informal spoken discourse. However, qualitative micro analyses of partially expressed utterances suggest that what seems to be formal incompleteness of utterances may in fact be the result of different types of discourse-oriented phenomena. In this section, I review, from a formulaicity perspective, two studies that inspect the use of sentence-final inferential evidential auxiliaries in pre-nominal forms: *yoona* and *mitaina*.

Cook (2008) examines style shifting between polite (*-masu*) and plain verbal forms in academic consultation sessions between college students and their professors. She illustrates how a student avoids marking a particular speech style, and therefore avoids creating a hierarchical relationship with his professor, by not completing a sentence. Below is the example she provides:

(2-3) Utterance-final *yoona* by student (S) to professor (P) (Cook 2008: 29-30)

5. P: *hyaku nijuu hachi peeji made yonda?*  
hundred twenty eight page until read:PST  
'Have you read up to page 128?'
6. S: *hai*  
yes  
'Yes.'

7. P: *hu:n. muzukashikatta desho. muzukashiku nakatta?*  
 INJ difficult:PST AUX / difficult NEG:PST  
 ‘Um: it was difficult, wasn’t it? Wasn’t it difficult?’
8. S: *muzukashii iya: kotchi yori wa*  
 difficult no this than TOP  
 ‘Difficult. No, compared to this one,’
9. → *yomiyasukatta yoona*  
 readable:PST AUX:ATT  
 ‘(it) seemed readable.’
10. P: *u::n un un*  
 INJ yes yes  
 ‘Um um um.’

In line 9, the student responds to his professor’s question with an incomplete sentence *yomiyasukatta yoona* ‘(it) seemed readable’ without a predicate. Even though the sentence is incomplete, a native Japanese speaker would be able to infer the complete expression to be something like *yomiyasukatta yoona ki ga suru/shimasu* ‘(I) feel that it seemed easier to read.’ Cook argues that “by not completing the predicate, the student does not define the relationship with the professor as either a hierarchical or an equal relationship” (p. 30). The point of interest for our purpose is that there are extremely limited types of predicates that can follow the prenominal form of the auxiliary *yoona* in line 9. *Yoona* can be followed by *ki ga suru/shimasu* or *kanji ga suru/shimasu*, both of which can be translated as ‘(I) feel.’ I take this to be an example of the use of formulaic language as a resource for achieving socio-interactional actions in conversation. That is, the idiomatic status of the predicate allowed the speaker to leave it unsaid in his expectation that the addressee would be able to complete the incomplete utterance.

The use of incomplete sentences by conversational participant also appears in Maynard (2005, 2007) who examines inserted speech and utterance final *mitaina*. Despite the fact that

*mitaina* is a pre-nominal auxiliary form and thus requires a nominal, Maynard (2005)

demonstrates that it appears utterance finally. She gives the following example:

(2-4) Utterance-final *mitaina* (Maynard 2005: 838, 850)

1. H: *onna no ko ni iwareru hoo ga ureshii?*  
female GEN person DAT say:PASS way NOM happy  
'Are you happier when you are told so by girls?'
2. M: *ureshii shi "arigatoo" tte shizenni kaeseru n desu.*  
happy and thank.you QT naturally return NML COP  
'I'm happy, and I can naturally say "Thank you."'
3. *otoko no hito ni iwareru to, etto watashi, doo sureba ii n daroo to yuu kanji de.*  
male GEN person DAT say:PASS COND INJ I how do:COND good NML AUX QT say feeling COP  
'When I am told by men, I feel like, what, uh, what should I do?'
4. H: *"akushu shite kudasai" toka iwareru to, hiichaimasu?*  
hand.shake do:TE please such say:PASS COND hesitate:POL  
'Do you hesitate when you are told "Please shake hands with me"?'
5. → M: *watashi, te o arawanakute ii no ka na, mitaina* ((laugh))  
I hand ACC wash:NEG:TE good NML Q PP AUX:ATT  
'It's like, shouldn't I wash my hands?'

According to Maynard, *mitaina* in general has the effect of distancing the speaker himself or herself from the content of the inserted speech (inner speech in this case), and hence "indicates the speaker's separation from the just expressed attitude (in the inserted speech)" (p. 850). In contrast to the case of *yoona* in Cook (2008), it is more difficult to think of a single predicate that is likely to follow *mitaina* if it were a complete sentence. Maynard mentions three possibilities: *mitaina koto o yuu* 'to say things like,' *mitaina kanji da* 'to have a feeling like,' and *mitaina jookyoo da* 'to be in a situation like.' The utterance final *mitaina* seems to be becoming an independent quotative or discourse marker, and is employed as such in the above example.<sup>9</sup>

---

<sup>9</sup> The independent status of *mitaina* can also be seen in the attachment of sentence-final particle directly to *mitaina* (S. Suzuki 1995: 70).

Although *yoona* and *mitaina* share similar semantic and syntactic properties, the examples and analyses above suggest that the kinds of discourse elements that precede or follow *yoona/mitaina* as well as the degree of association between the preceding/following elements and *yoona/mitaina* may differ. Although it is beyond the scope of this paper, examining these items and phenomena from a formulaic perspective would deepen our understanding of how apparently similar linguistic items behave in actual language use.

### **2.5.2.3. Particles**

Japanese particles are all postpositional, and they supply various kinds of information regarding the relations of the preceding elements to the following components. In the case of final or pragmatic particles appearing at the end of a sentence, they express the speaker's stance towards the proposition. Particles are considered vital components in the construction of Japanese sentences.

Hayashi (2004), in his study of postpositional particles, illustrates that a sentence is constructed as an “interactional achievement — that is, as an outcome of intricate interactional work performed by speaker and hearer in response to local exigencies of the moment of its production” (p. 344). In this interactional practice, case or adverbial postpositional particles are employed by conversational participants as important grammatical resource for projecting further components of an unfolding utterance or for retroactively transforming turn-shapes by creating a grammatical link between two discourse elements (Tanaka 1999).

While Tanaka (1999) and Hayashi (2004) emphasize the grammatical function of postpositional particles, Ono, Thompson & R. Suzuki (2000) focus on the pragmatic aspect of one case particle *ga*. The authors challenge the traditional characterization of the particle *ga* as a

pure grammatical marker of subject. Based on examination of conversational data, they argue that the use of *ga* is highly infrequent in naturally occurring conversation, and when it occurs, it does with a limited number of intransitive predicates with low semantic content. Furthermore, the utterances containing *ga* are found to be associated with various degrees of conventionalization. Among the most conventionalized cases, they list ... *hazu ga nai* ‘there is no way that ...’ and ... *hoo ga + Adjective* ‘it is more Adjective to ...’. These examples provide an empirical evidence for the role of frequency as well as low semantic content of component in the path for conventionalization of multi-word sequences.

#### **2.5.2.4. Discourse markers**

There are a large number of usage-based studies on Japanese discourse markers (e.g., Maynard 1989; Matsui 2002; Onodera 2004; R. Suzuki 2006). Discourse markers are broadly defined as “sequentially dependent elements which bracket units of talk” (Schiffrin 1987: 31). Although sequentially dependent, they are “independent of sentential structure, ... [and] several markers — *y’know, I mean, oh, like* — can occur quite freely within a sentence” (Schiffrin 1987: 31). Other features include grammatical optionality, lack of semantic content, orality, and multifunctionality (see Müller 2005). Discourse markers may serve either textual or interactional functions (Lenk 1997; Müller 2005). Due to the formal fixedness and the lack of semantic content, some researchers regard discourse markers as formulaic language (e.g., Raupach 1984; Sidtis 2009).

One of the most extensive investigations of Japanese discourse markers is conducted by Onodera (2004) who inquired into the pragmaticalization of two conjunctions *demo* and *dakedo* from synchronic and diachronic perspectives. Diachronic change occurred from a clause-final V

(copula *de/da*) + a conjunctive particle *mo/kedo* into an utterance-initial discourse marker. The shift to the sentence-initial location is explained by the effectiveness of the initial location in serving various kinds of signaling functions. Of particular interest to the present study are discussions on the role of typological characteristics of Japanese language in the process of discourse marker formation and the conventionalization of conversational implicature. The author argues that “at least two typological feature of Japanese, i.e. postpositional and agglutinative, ... affect the pragmaticalization” (p. 205). That is, agglutination directly connects *mo/kedo* to the preceding sentence ending with copula *de/da*; in addition, the shift from unit-final to initial position is attributed to the postpositional structure of the language.

The significance of this study to the present investigation of formulaicity is that it describes the diachronic process in which recurrently co-occurring linguistic items (e.g., *de + mo* and *da + kedo*) become holistically processed chunks as they gain new functions based on conversational implicature. Moreover, the study shows that typological features of the language play a major role in the pragmaticalization process.

## **2.6. Lexical bundles and register variations**

Lexical bundles are empirically identified recurrent strings of words in a register of natural language use. The term was first used in a corpus-based reference grammar of English, *Longman grammar of spoken and written English* (Biber et al. 1999). Instead of relying on pre-determined sets of linguistic items and chunks, the authors took a purely empirical approach to extract “sequences of word forms that commonly go together in natural discourse” (p. 990). Based on the computer-assisted frequency-based analysis, most frequently recurring multi-word sequences in given registers are identified. Biber et al.’s (1999) and subsequent lexical bundle

studies (e.g., Biber & Conrad 1999; Biber et al. 2004; Biber & Barbieri 2007) are not only quantitative but seek a functional explanation for the use of the identified lexical bundles in particular registers. Biber et al. (2004: 400) state that “high frequency patterns are not accidental, nor are they explanatory;” in other words, “corpus-based frequency evidence provides descriptive facts that require explanation.”

Lexical bundles naturally fit the form-based definition of formulaic language; at the same time, on the basis of the qualitative findings in the lexical bundles studies cited above, many of the lexical bundles also fit the function-based and preference-based (and possibly cognitive-oriented) definitions of formulaic language presented in Chapter 1. This shows that frequency, at least to certain extent, is a useful indicator of formulaic language. The present study adopts the basic theoretical as well as methodological principles of the lexical bundle studies. In the rest of this section, I review some of the major findings from the studies within this framework, and suggest some modifications for investigating Japanese spoken data.<sup>10</sup>

Most English lexical bundle studies that include in their data the corpus of spoken language compare the patterns and functions of the bundles in spoken and written registers (e.g., Biber et al. 1999; Biber et al. 2004; Biber & Barbieri 2007). Written registers include academic research articles and books (Biber et al. 1999; Biber et al. 2004), textbooks (Biber et al. 1999; Biber et al. 2004; Biber & Barbieri 2007), university course syllabi and assignments (Biber & Barbieri 2007), and campus-related written texts<sup>11</sup> (Biber & Barbieri 2007). In some ESL studies, published academic texts are compared to second language (L2) students’ writings (e.g., Allen 2009; Chen & Baker 2010). Spoken registers include conversation (Biber et al. 1999; Biber et al.

---

<sup>10</sup> Since the present study only examines spoken language, I limit the review to studies that deal with spoken corpus.

<sup>11</sup> They include academic program brochures, student handbooks, and academic policy statements and descriptions in university catalogs (Biber & Barbieri 2007: 267).

2004; Biber & Barbieri 2007), classroom teaching (Biber et al. 2004), teacher-student or student-student interaction outside class (Biber & Barbieri 2007).

In the first study of lexical bundles in conversation and academic prose, Biber et al. (1999) found the following three characteristics of the lexical bundles.

1. Lexical bundles are by definition extremely common.
2. Lexical bundles do not usually represent a complete structural unit.
3. Most lexical bundles are not idiomatic in meaning.

They reported that there were more lexical bundles in conversation than in academic prose. In speech, bundles appear at a clause boundary and link two clauses (e.g., *I don't know what, I mean you know*), while in academic prose, bundles tend to appear at a phrase boundary and connect two phrases (e.g., *in the case of, the base of the*).

In Biber et al. (2003) and Biber et al. (2004), the functional taxonomy of the lexical bundles are developed. Although the initial identification of the bundles is done automatically using computational technology, the functional analyses are done manually by checking the concordance lines and analyzing them in their discourse context. Biber et al. (2004) recognize three primary functions of the bundles: (1) stance expressions, (2) discourse organizers, (3) referential expressions.<sup>12</sup> They also demonstrate the strong association between structure and function of the bundles. VP-based bundles are mostly used for stance expressions and discourse-organizing functions, whereas NP-based bundles are largely used for referential expressions.

There have been attempts to apply the methodological principles of the English lexical bundle studies to examine other languages (e.g., Tracy-Ventura et al. 2007; Kim 2009). The findings from these studies point to interesting similarities and differences in the patterns and uses of the lexical bundles. In Spanish (see Tracy-Ventura et al. 2007), major structural and

---

<sup>12</sup> The definitions of these functions are presented and discussed in association with the data and analyses in the present study in Chapter 4.

functional types of the bundles are similar to English. However, distributional patterns differ significantly. First, the overall number of the lexical bundles is greater in written corpus (academic prose) than in spoken corpus (interviews). Moreover, in Spanish, referential bundles (which are mostly NP-bundles) are the most common type in both spoken and written registers. Biber et al. (2010) explain this difference by the structural and typological difference between English and Spanish. In particular, gender and number markings as well as the strong dependence on *de*-phrase for phrasal modification in Spanish contribute to the larger number of NP-based and referential bundles. Unlike both English and Spanish, lexical bundles in Korean (see Kim 2009) are extremely rare. This is due to the agglutinating nature of the language. In the analysis of shorter sequences, it is found that NP-based bundles are much more common than VP-based bundles in both spoken (conversation) and written (academic prose) registers. Functionally, stance expressions are most common in both spoken and written corpora. As the authors admit, the approach that was designed for languages with isolating morphology “penalizes a language like Korean, where grammatical functions are typically realized as morphological inflections attached to a content word” (Biber et al. 2010: 92).

As Sohn and Nam (2011) state, in order to incorporate grammatical morphemes into the bundles, a morpheme-based analysis is more appropriate. According to them, since the Korean bundles in Kim’s study (2009) are based on orthographic words, many grammatical suffixes are excluded. They cite the following example:

<i>swu</i>	<i>iss-ul</i>	<i>kes-i-ta</i>	‘it would be probably’
possibility	be -prospective	thing-be-declarative	

The formulaic sequence is identified as a three-word lexical bundle. However, the sequence can be further broken down into six morphemes. Although there are methodological challenges in

conducting a morpheme-based corpus study both in Korean and in Japanese, in order to provide a more adequate and realistic account for the patterns and functions of multi-linguistic sequences in the agglutinating languages, it is crucial to take morphemes as the basic units of analysis.

## **2.7. Summary**

This chapter presented theoretical background for the present study. I have presented currently available evidence for the centrality of formulaicity in discourse. Studies within lexicography and phraseology have categorized formulaic language into three major categories: free combinations, restricted collocations, and idioms, based on formal-semantic factors. Among them, empirical corpus-based studies have found that restricted collocations are the most commonly observed type of formulaic language in naturally occurring discourse. Although the corpus-based approach is not without limitations and challenges, it gives us much insight into the frequently occurring formal patterning of the language, which we can use as a reference point for more detailed qualitative analyses or psycholinguistic speculations. Studies with a particular focus on socio-interactional aspect of utterances revealed inseparable connection between formulaic language and contexts of their use. I have also reviewed cognitive-oriented views of formulaic language that sought to provide a missing link between language use and linguistic knowledge of native speakers. Because formulaic expressions are learned and used in close association with real time contexts or situations, they are particularly problematic for most adult learners who must learn these socially-particular and context-sensitive expressions in classroom settings. If we are to help second/foreign language learners get a real sense of fixedness involved in the target language use, we must first understand patterns and functions of formulaicity in ordinary language use ourselves.

In Japanese linguistics, there is a lack of evidence and studies in general concerning holistic processing of linguistic elements in spoken discourse partially due to the agglutinating and context-dependent nature of the language and the entailing methodological difficulties. However, a survey of previous Japanese usage-based studies and conversation analytic studies from a formulaicity perspective suggested possible links between Japanese language-in-use and formulaicity.

Finally, I reviewed lexical/morphemic bundle research which is directly relevant to the present study. Unlike many other types of formulaic language, lexical/morphemic bundles are empirically identified based solely on their frequency. The previous studies have demonstrated that the high frequency patterns are not accidental and that further qualitative analysis can reveal much about the nature of these units as well as their link to the functional needs of the particular registers as well as to the typological characteristics of the particular languages. Following this line of research, the present study investigates multi-morphemic units in Japanese spoken language, using both quantitative and qualitative analyses. The examination of morphemic bundles in Japanese not only explicates the functional characteristics of the bundles and their association with particular registers, but it also sheds new light on the socio-cultural basis for the most frequently used morpheme-based sequences (Chapter 4) and semi-fixed constructions (Chapter 5).

## Chapter 3

### Data and Methodology

#### 3.1. Introduction

In this chapter, I will describe and provide a rationale for the data and the methodologies to be employed in the present study. Adopting the emergent view of language (i.e., view that patterns of linguistic forms and functions emerge from language use in natural contexts) requires a methodological framework that acknowledges the importance of empirical investigations of language use and one that regards genres/registers as constituting different kinds of contexts for language users. The inductive approach to data analysis and theory building is crucial in ensuring that explanations are sought to “fit the evidence, rather than adjusting the evidence to fit a pre-set explanation” (Sinclair 1991: 36).

#### 3.2. The data

The present study employs a corpus-based approach, and analyzes three spoken registers: informal conversation, formal interviews, and academic speech. The three registers spread across two different sociolinguistic dimensions: interactive (conversation and interview) vs. non-interactive (speech) on the one hand, and informal (conversation) vs. formal (interview and speech) on the other. Another dimension along which three registers differ is whether the language use is ubiquitous or specialized. Informal conversation is considered the “primordial” form of language use (Schegloff 1996, 1999) which a native speaker encounters and acquires before other more sophisticated (e.g., formal interview), and/or specialized (e.g., academic speech) language use. The dataset for this study consists of 138 conversations, 86 interviews,

and 108 speeches as shown in Table 3.1 below. These nine sub-corpora together contain more than one million morphemes.

Table 3.1. Description of the dataset

Register	Corpus	No. of texts	No. of morphemes
Conversation	BTS corpus <sup>13</sup> (26)	138	439,270
	Japan corpus <sup>14</sup> (13)		
	Workplace corpus <sup>15</sup> (14)		
	Sakura corpus <sup>16</sup> (14)		
	CallHome corpus <sup>17</sup> (71)		
Interview	CSJ corpus <sup>18</sup> (26)	86	228,917
	Hypermedia corpus <sup>19</sup> (50)		
	Tetsuko's room corpus <sup>20</sup> (10)		
Speech	CSJ corpus <sup>21</sup> (108)	108	354,595
TOTAL		332	1,022,782

Since there was no pre-existing corpus that included all three spoken registers, I compiled the dataset from eight different corpora that were made available to me. Several criteria were taken into consideration in selecting corpora as well as texts within individual corpora. For informal conversations, two main criteria were used in choosing texts from five corpora listed above. First, speakers of Tokyo dialects were selected to achieve homogeneity. Second, especially for Workplace corpus, the relationships between the interlocutors were limited to

<sup>13</sup> Usami (ed.) (2007)

<sup>14</sup> Aoki et al. (to appear)

<sup>15</sup> Gendai Nihongo Kenkyukai (1997); Gendai Nihongo Kenkyukai (2002)

<sup>16</sup> Miyata et al. (2010)

<sup>17</sup> Wheatley et al. (1996); Canavan and Zipperlen (1996)

<sup>18</sup> Furui et al. (2004)

<sup>19</sup> Uemura (1997)

<sup>20</sup> Fujii, T. (General Producer). *Tetsuko no heya* [Tetsuko's room].

<sup>21</sup> Furui et al. (2004)

friends, family, and couples.<sup>22</sup> For formal interviews, particular attention was paid to include as many interviewers as possible to avoid idiosyncratic uses of language by individual interviewers.<sup>23</sup> For academic speech, CSJ corpus was used exclusively. The CSJ corpus contains more than one thousand academic speeches at nine different conferences in three different disciplines (i.e., engineering, humanities, and social and behavioral sciences). Among them, a smaller balanced corpus was compiled with respect to the gender of the speakers and academic disciplines as well as the number of texts and morphemes comparable to other two registers.

### **3.3. Methodology**

The fundamental methodological framework for this study is in line with Tao (2001) and Thompson and Ono (2010), who call for an integrated approach to the study of language use, comprising both quantitative/macro and qualitative/micro analyses (Thompson & Ono 2010: 110), or more specifically, “corpus linguistic tools and sociocultural linguistic insights” (Tao 2001: 116).

The methodological framework is based on a frequency-driven approach for indentifying uninterrupted sequences of linguistic elements. As described in Chapter 2, Biber and his colleagues have done much work on multi-word sequences which they have termed “lexical bundles” in English (e.g., Biber et al. 1999; Biber et al. 2004; Biber & Barbieri 2007), and, to a lesser extent, in Spanish (e.g., Tracy-Ventura et al. 2007). To reiterate, lexical bundles are defined as recurring sequences of words that occur most frequently in a given register. The sequences must be continuous, that is, strings that cross a turn boundary or a punctuation mark

---

<sup>22</sup> Both kinds of selections were made possible by the information provided by the transcribers of each text.

<sup>23</sup> There were two interviewers in CSJ corpus, four interviewers in Hypermedia corpus, and one interviewer in Tetsuko’s room corpus.

are excluded. For the most part, lexical bundles neither represent syntactically complete units nor are they idiomatic in their meanings. Generally, they join two structural units, thereby serving as important discourse building blocks for verbal and clausal units in conversation and for extended noun phrases and prepositional phrases in academic prose (Biber et al. 2004: 380-381).

It is in this sense that lexical bundles are regarded as being part of grammar (Biber et al. 1999). According to the authors, grammar concerns not only abstract classes and structures but also particular words and their particular functions (pp. 989-999). This perspective is essential in addressing nativelikeness of only some grammatically well formed utterances. As I briefly discuss in Chapter 6, it also has pedagogical implications for second/foreign language teaching. Learners must be made aware of the lexical end of the grammar and lexical/morphemic bundles can serve as concrete examples.

All the previous lexical bundle studies stress the functional basis of lexical bundles; they are strongly associated with discourse functions and communicative purposes of individual registers. However, as the present study on Japanese morphemic bundles demonstrates, functional motivation is not limited to communicative needs and situational characteristics of registers. The findings in the current study point to a wider socio-pragmatic aspect of linguistic community as closely tied to the use of recurring morphemic strings and semi-fixed construction of which specific morphemic bundles are part.

### **3.4. Procedures**

Since the current dissertation is a first attempt to analyze frequently used linguistic strings in Japanese spoken discourse using a large corpus, it is important to document procedures

involved in designing the corpus which can be used in computer-assisted programs. Some of the steps are specific to Japanese data due to the characteristics of Japanese orthography.

Justifications for specific decisions are also provided.

### 3.4.1. Tokenization of texts

The first step is to put spaces between morphemes, namely, tokenization, which makes automatic searching for sequences of morphemes possible. Initial tokenization was performed by a Japanese morphological analyzer called ChaSen (version 2.0; Nara Institute of Science and Technology [NAIST] 1999). Before the texts were tokenized by ChaSen, two modifications were made on the texts. First, notations for non-linguistic (e.g., laugh, cough) and non-verbal (e.g., physical action, facial expression) were omitted from the texts. Although these features play important role in spoken discourse, they are irrelevant to the identification of multi-morpheme strings. Tagging of these elements (instead of complete elimination) was also an option, but the tagging markers intervening morpheme strings can cause problems when searching is performed for recurrent morpheme units. In addition to the complete omission of non-linguistic features, all the *tooten* (commas) were eliminated. Compared to *kuten* (periods), which are mostly used to mark the end of sentences or grammatically complete units, the use of *tooten* is not systematic but rather idiosyncratic. Again, the complete omission of *tooten* was chosen over the tagging to avoid a potential disruption in the morpheme sequence search.

After these modifications,<sup>24</sup> ChaSen was used to segment the texts into morpheme-based units (with spaces between morphemes). The morpheme-based tokenization requires a morphological analysis and ChaSen is equipped with a morpheme dictionary, containing both

---

<sup>24</sup> It is important to save these modified texts separately and keep the original texts without any modification. This allows the researcher to come back to the original texts when necessary.

morphemes and general morphemic collocational patterns. Even though the automatic tokenization is a convenient and inevitable step in the study of multi-linguistic bundles, the resulting texts must be checked again manually to achieve acceptable level of consistency and accuracy.

### 3.4.2. Standardization of tokenized texts

There are several reasons that necessitate manual checking and standardizing of tokenized texts. First, the morphological analyzer follows the implemented segmentation rules, which may not serve the particular need of the current multi-morphemic study. For example, ChaSen recognizes some particle compounds, back-channeling expressions, and greeting expressions as single units despite their multi-morphemic statuses (e.g., *to-iu* [QT-say], *ni-shitagat-te* [DAT-follow-TE], *soo-desu-ne* [so-COP-PP], *arigatoo-gozai-masu* [thank-exist-POL]). It also cannot segment morphemes which are not included in the dictionary (and of course it does not take into account contextual information). There are also cases of segmentation errors (e.g., *anmari* ‘not much’ wrongly segmented as *an-mari*). Due to these limitations, the tokenized texts must be checked manually. Second, Japanese orthography with three types of characters, hiragana, katakana, and kanji, yields some problems. That is, a single word may be written in kana at one point and in kanji at another (e.g., *watashi* ‘I’ is written in kana わたし or in kanji 私). Another orthographical issue is the representation of vowel lengthening; it can be transcribed with a hiragana vowel character or with a hyphen (e.g., *naa* can be written as なあ or なー and *nee* as ねえ or ねー). The third reason is the colloquial pronunciations and conventions for transcribing them. One frequently occurring word with this issue is *iulyuu* ‘say.’ Very often this word is pronounced *yuu* (ゆー), as in *yutta* ‘said’ and

*yuwanai* ‘doesn’t say,’ but nonetheless transcribed as *iu* (hiragana only いゝ or hiragana-kanji mixed 言ゝ), as in *itta* ‘said’ and *iwana* ‘doesn’t say.’ A further complication is that this word appears in lexicalized or pragmatized units such as *koo-yuu* ‘like this’ and *tte-yuu-ka* ‘or rather than saying that,’ and again there is no consistency in how the words are transcribed (e.g., for *koo-yuu*, こういゝ, こうゆゝ, こう言ゝ are all used).

The unsegmented morpheme units and wrongly segmented units were manually (re-) segmented. For other types of issues, a decision was made for each case. For the issue of the kana-kanji mixture, all tokens of a single morpheme were standardized to either kanji or kana transcription (e.g., *watashi* ‘I’ was standardized to the kanji notation 私, while 来る ‘come’ was standardized to the hiragana notation くる). The choices were mostly made for practical reasons. Also, to avoid the issue of having too many homonyms, non-past forms of verbs and adjectives were not segmented into the root and the suffix. For example, if we segmented all the non-past forms, at least six morphemic homonyms would result for the form い (i), namely, verb roots *i* in *i-u* ‘say’ and *i-ru* ‘exist,’ an adjective root *i* in *i-i* ‘good,’ an adjective non-past suffix *i* as in *taka-i* ‘expensive,’ and aspectual auxiliaries *i* as in *te-i-ru* and *te-i-ku*. The most important thing is that all the texts are uniformly standardized so that the n-gram analysis can be performed on the corpus.

### 3.4.3. N-gram analysis

N-gram analysis identifies every uninterrupted sequence of  $n$  units in a given text. For example, in the previous sentence, [n-gram + analysis + identifies + every], [analysis + identifies + every + sequence], [identifies + every + sequence + of] (and so on) are all identified as four-grams. As explained already, the present study adopts a morpheme-based n-gram model. Using

the software program, AntConc (version 3.2.2; Anthony 2011), I have performed 2-, 3-, 4-, 5-, and 6-gram analyses at a preliminary stage, and decided to focus on 4-grams in this study. The decision is based on the findings that 2- and 3-gram analyses yield too many bundle types with many overlaps and 5- and 6-gram analyses yield too few types. In order to be considered a morphemic bundle, a 4-gram must appear more than 80 times per million morphemes and across at least five different texts to avoid including idiosyncratic use by individual speakers.<sup>25</sup> The frequency cut-off is very conservative compared to past studies on lexical bundles<sup>26</sup> and morphemic bundles.<sup>27</sup> To further illustrate, let us look at the following Japanese utterance example given in Chapter 1 (originally numbered (1-6)).

(3-1) 4-gram analysis (Conversation: Ono\_Ryokoo)

好きっていうんじゃないんだよな。

*suki tte iu n ja nai n da yo na.*

like QT say NML COP NEG NML COP PP PP

‘It’s not that (he) likes (his job).’

In this sentence, there are seven 4-grams as listed below:

- 好きっていうん *suki tte iu n* [like QT say NML]
- っていうんじゃ *tte iu n ja* [QT say NML COP]
- いうんじゃない *iu n ja nai* [say NML COP NEG]
- んじゃないん *n ja nai n* [NML COP NEG NML]
- じゃないんだ *ja nai n da* [COP NEG NML COP]
- ないんだよ *nai n da yo* [NEG NML COP PP]
- なんだよな *n da yo na* [NML COP PP PP]

<sup>25</sup> Although the number of texts was different in each register (138 texts in Conversation, 86 texts in Interview, and 108 texts in Speech), the same criterion of five different texts was used because the reason for imposing this criterion was to avoid idiosyncratic use by single speakers. In Interview, which only had a few interviewees, an additional step was taken to ensure that a bundle was used by at least five different speakers.

<sup>26</sup> Biber et al. (1999) study was based on the frequency cut-off of 10 times per million word. Biber et al. (2004) study was based on the frequency cut-off of 40 times per million word.

<sup>27</sup> Sohn and Nam’s (2011) 5-gram study was based on the frequency cut-off point of 30 times per million morphemes.

Among these, only one sequence じゃないんだ *ja nai n da* appears more than 80 times per million morphemes (and across five different texts) in the present conversation corpus. Hence, only this 4 gram would be identified as a morphemic bundle. In the rest of the paper, morphemic bundles are written with plus signs between the morphemes, as in *ja + nai + n + da* [COP + NEG + NML + COP]. In the next chapter, I report the result of the n-gram analysis described here.

## Chapter 4

### Morphemic Bundles in Conversation, Interview, and Speech

#### 4.1. Frequency of morphemic bundles among three registers

Figure 4.1 presents the total number of different four-morphemic bundle types (i.e., type frequency) found in three registers. As we can see in the figure, Interview contains the largest number of different kinds of morphemic bundles (189 bundle types), followed by Speech (175 bundle types), and Conversation has the fewest (92 bundle types).

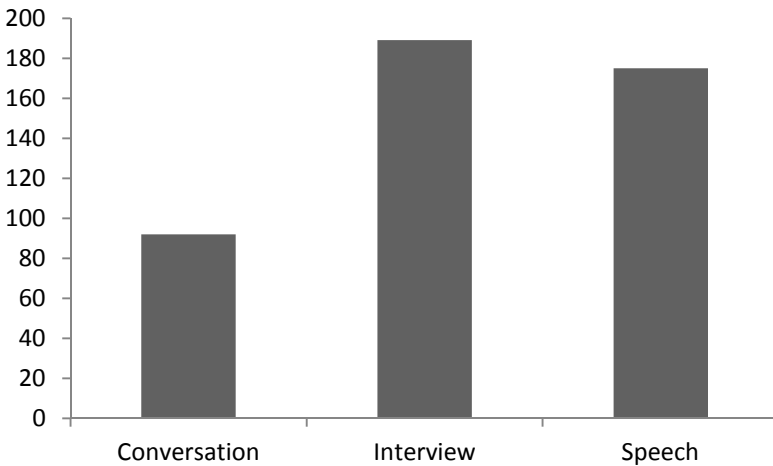


Figure 4.1. Number of different four-morphemic bundles across registers

Figure 4.2 presents the total number of four-morphemic bundle tokens (i.e., token frequency) in three registers. Speech has the largest number of bundle tokens (11,339 bundles), Interview has the second largest number (8,600 bundles), and Conversation has the least number (7016 bundles).

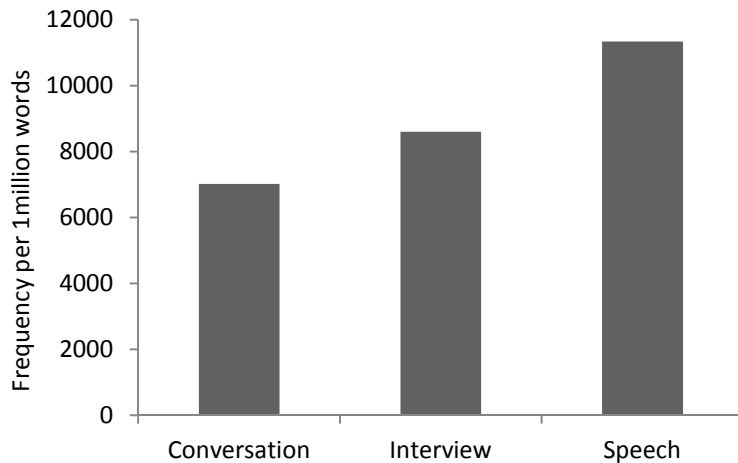


Figure 4.2. Overall frequency of four-morphemic bundles across registers

Figures 4.1 and 4.2 demonstrate that four-morphemic bundles appear more frequently in Interview and Speech than in Conversation. We can also see that Conversation relies on a limited set of bundles while Interview uses a large number of different morphemic bundles. To quantify how much repetition is involved in the use of morphemic bundles in each register, a type-token ratio can be measured. A low and high type-token ratio corresponds to the higher and the lower number of repetitions, respectively.

Conversation:  $92/7016 = .013$  (more repetition)

Interview:  $189/8600 = .022$  (fewer repetition)

Speech:  $175/11339 = .015$

As we can see above, Conversation has the lowest ratio and Interview has the highest. This tells us that Conversation relies on a limited set of bundles with a higher rate of repetition, whereas Interview uses a large number of different bundles.

## 4.2. Structural characteristics of the morphemic bundles

Just like English lexical bundles (Biber et al. 2004), most morphemic bundles do not represent complete structural units. In fact, none of the bundles in Speech are structurally complete, and only 5% and 7% of the bundles are structurally complete in Interview and Conversation, respectively. Below are the lists of structurally complete bundles found in Interview (4-1) and Conversation (4-2).

### (4-1) Structurally complete bundles in Interview (10/189; 5%)

<i>yoroshiku</i> favorably	<i>onagai shi</i> please do	<i>masu</i> POL		‘Please treat me kindly.’
<i>arigatoo</i> thank	<i>gozai</i> exist	<i>mashi</i> POL	<i>ta</i> PST	‘Thank you.’
<i>a</i> INJ	<i>soo</i> so	<i>desu</i> COP	<i>ka</i> Q	‘Oh, is that so?’
<i>aa</i> INJ	<i>soo</i> so	<i>desu</i> COP	<i>ka</i> Q	‘Oh, is that so?’
<i>soo</i> so	<i>na</i> COP	<i>n</i> NML	<i>da</i> COP	‘I see.’
<i>soo</i> so	<i>na</i> COP	<i>n</i> NML	<i>desu</i> COP	‘That’s right.’
<i>soo</i> so	<i>desu</i> COP	<i>yo</i> PP	<i>ne</i> PP	‘You are right. / I agree. / I understand.’
<i>soo</i> so	<i>desu</i> COP	<i>ne</i> PP	<i>hai</i> yes	‘That’s right. Yes.’
<i>hai</i> yes	<i>soo</i> so	<i>desu</i> COP	<i>ne</i> PP	‘Yes. That’s right.’
<i>a</i> INJ	<i>soo</i> so	<i>desu</i> COP	<i>ne</i> PP	‘Oh, that’s right.’

(4-2) Structurally complete bundles in Conversation (6/92; 7%)

<i>nan</i> what	<i>te</i> QT	<i>iu</i> say	<i>no</i> PP	‘How to say?’
<i>soo</i> so	<i>na</i> COP	<i>n</i> NML	<i>da</i> COP	‘I see.’
<i>soo</i> so	<i>da</i> COP	<i>yo</i> PP	<i>ne</i> PP	‘You are right. / I agree. / I understand.’
<i>soo</i> so	<i>na</i> COP	<i>no</i> NML	<i>yo</i> PP	‘That’s right.’
<i>a</i> INJ	<i>soo</i> so	<i>na</i> COP	<i>no</i> NML	‘Oh, I see.’
<i>aa</i> INJ	<i>soo</i> so	<i>na</i> COP	<i>no</i> NML	‘Oh, I see.’

In the lists above, we can see similarity as well as differences between the two registers. Most of the bundles in both Interview (8/10; 80%) and Conversation (5/6; 83%) involve the sequence, *soo + da/desu/na* [so + COP], which is used as various kinds of response tokens (e.g., ‘I see,’ ‘that’s right,’ ‘I agree,’ ‘Is that so?’). On the other hand, there are register-specific bundles; the two greeting expressions *yoroshiku + onegai + shi + masu* ‘Please treat me kindly’ and *arigatoo + gozai + mashi + ta* ‘Thank you’ are only found in Interview, while a word search expression *nan + te + iu + no* ‘How to say?’ is only found in Conversation. In addition, polite forms (i.e., the polite suffix *masu* and the polite form of the copula *desu*) only appear in Interview, which makes sense when we take into consideration the formality of Interview compared to the informality of Conversation. I will discuss functional aspects of these structurally complete bundles in Section 4.3.4.

The morphemic bundles found in the three registers are categorized into the following six structural types. Examples of morphemic bundles for the (sub-)categories are also presented.

Types of morphemes that must appear for the bundle to be categorized in the particular category are in bold.

1. Verb phrase related bundle (VP)

(i) Copula

<i>te</i>	<i>iru</i>	<i>n</i>	<i>desu</i>
TE	ASP	NML	<b>COP</b>

(ii) Quotative

<i>ta</i>	<i>to</i>	<i>omoi</i>	<i>masu</i>
PST	<b>QT</b>	<b>think</b> (or <b>say</b> )	POL

(iii) Verb

<i>kai</i>	<i>te</i>	<i>at</i>	<i>ta</i>
<b>Verb</b>	TE	ASP	PST

(iv) Clause-combining

<i>nai</i>	<i>n</i>	<i>desu</i>	<i>kedo</i>
NEG	NML	COP	<b>Conjunctive particle</b> (or <b>TE-form</b> )

2. Repetitive backchannel bundle<sup>28</sup> (RB)

<i>soo</i>	<i>soo</i>	<i>soo</i>	<i>soo</i>
so	so	so	so

<i>un</i>	<i>un</i>	<i>un</i>	<i>un</i>
yes	yes	yes	yes

3. Noun phrase related bundle (NP)

(i) Noun

<i>to/tte</i>	<i>iu</i>	<i>no</i>	<i>wa</i>
QT	say	<b>NML</b> (or <b>Noun</b> )	TOP

(ii) Phrasal particle<sup>29</sup>

<i>ni</i>	<i>taishi</i>	<i>te</i>	<i>wa</i>
DAT	face	TE	TOP

<sup>28</sup> Other repetitive backchannels found in the current corpus are: [*soo + ka + soo + ka*], [*so + kka + so + kka*], [*un + soo + soo + soo*] and [*u:n + un + un + un*].

<sup>29</sup> Phrasal particles (Kaiser, Ichikawa, Kobayashi & Yamamoto 2001) are also called “complex adjunct phrase” (Matsumoto 1998). Phrasal particle morphemic bundles found in the present corpus are as follows: *ni-taishi-te-wa* ‘in/with/toward’, *ni-tsui-te-wa* ‘concerning’, *ni-oi-te-wa* ‘as for’, *to-shi-mashi-te* ‘as’, *to-itashi-mashi-te* ‘as’, *to-shi-mashi-te* ‘as’, *itash-mashi-te-wa* ‘as’.

4. Adverb related bundle (AD)
 

<i>to</i>	<i>iu</i>	<i>yoo/fuu</i>	<i>ni</i>
QT	say	<b>like/way</b>	<b>COP:ADV</b>
  
5. Embedded clause bundle (EM)
 

<i>wa</i>	<i>nai</i>	<i>ka</i>	<i>to</i>
TOP	NEG	Q	<b>QT</b>
  
6. Others

As we can see in Table 4.1, a majority of bundles are VP-related in all three registers. Relatively speaking, VP-related bundles appear most frequently in Conversation (85.9%), followed by Interview (74.6%), and least frequently in Speech (52%). In contrast, NP-related bundles show a reverse tendency; they appear most frequently in Speech (35.4%), less frequently in Interview (18%), and least frequently in Conversation (3.3%). Other notable categories are Adverbial and Embedded clause bundles which are more frequent in Speech than in Interview or Conversation. Also, Repetitive backchannel bundles show a skewed distribution; they account for 5.4% and 1.1% of bundles in Conversation and Interview, respectively, but no instance is found in Speech.

Table 4.1. Distribution of morphemic bundle types across structural types

	Conversation		Interview		Speech	
	Type	%	Type	%	Type	%
VP	79	85.9	141	74.6	91	52.0
NP	3	3.3	34	18.0	62	35.4
AD	0	0.0	5	2.6	11	6.3
EM	3	3.3	4	2.1	10	5.7
RB	5	5.4	3	1.6	0	0.0
Other	2	2.2	2	1.1	1	0.6
Total	92	100.0	189	100.0	175	100.0

Let us look at two major structural categories, VP-related and NP-related bundles, in more detail in order to explain the overall distribution of the structural types.

#### 4.2.1. VP-related bundles

First, within VP-related bundles, Conversation and Interview show similar distributional patterns. In both registers, Copula is the most frequent VP-related morpheme type (41.8% in Conversation and 44% in Interview) followed by Clause-combining form and Quotative. In Speech, on the other hand, lexical Verb is the dominant VP-related bundle type (34.1%) and Copula is the least frequent type (11%). Interestingly, in all three registers, Clause-combining is the second frequent type (21.5% in Conversation, 22% in Interview, and 19.8% in Speech). Within Quotative, Interview and Speech share more similarity in that *omou* ‘think’ predominates the category than Conversation in which *iu* ‘say’ is the most frequent type.

Table 4.2. Distribution of VP-related morphemic bundle types

		Conversation		Interview		Speech	
		No.	%	No.	%	No.	%
Copula		33	41.8	62	44.0	10	11.0
Quotative	<i>omou</i>	5	6.3	15	10.6	9	9.9
	<i>iu</i>	6	7.6	1	0.7	1	1.1
	<i>kangaeru</i>	0	0.0	0	0.0	2	2.2
	bare	1	1.3	1	0.7	0	0.0
Verb		9	11.4	14	9.9	31	34.1
Clause-chaining		17	21.5	31	22.0	18	19.8
Other		8	10.1	17	12.1	20	22.0
Total		79	100.0	141	100.0	91	100.0

An overwhelming majority of Copula-related bundles in Conversation and Interview involve the sequence *n + da/desu* [NML + COP] (31/33 or 94% in Conversation and 56/62 or 90% in Interview). A majority of Copula-related bundles in Speech also include the sequence *n + desu* [NML + COP] (6/10 or 60%) although to a lesser degree than in Conversation and Interview. I will discuss the sequence in more detail in Section 4.3.2.2.

(4-3) VP-related bundles involving copula

*te*     *ta*     *n*     ***da***     (Conversation)  
 TE     PST     NML     COP

*na*     *n*     ***desu***     *yo*     (Interview)  
 COP     NML     COP     PP

*mase*     *n*     ***deshi***     *ta*     (Speech)  
 POL     NEG     COP     PST

While *omou* ‘think’ appears in all three registers, *iu* ‘say’ is the predominant verb in Conversation. Conversation includes *nan + te + iu + no* [what + QT + say + PP] which is a structurally complete unit meaning ‘how to say?’ (see Section 4.2 above). In addition, the utterance final, bare quotative *tte* appears in Conversation. Compared to the predominance of the verb *iu* ‘say’ in Conversation, there is only a single instance of *iu* ‘say’ in Interview; it is a fragment of the word search expression, that is, *nan + te + iu + n* [what + QT + say + NML] which is frequently followed by morphemic sequences such as *desu + ka* [COP + Q] and *desho + o + (ne)* [COP + AUX + (PP)]. Apart from *iu* ‘say’ and *omou* ‘think,’ *kangaeru* ‘think’ also appears in Speech. One distinctive feature of Speech is the use of passives, for instance, *omow + are* [think + PASS] ‘is thought’ and *kangae + rare* [think + Pass] ‘is thought’.

(4-4) VP-related bundles involving *omou* ‘think’

*da to omoi masu* (Interview)  
COP QT think POL

*to omow are masu* (Speech)  
QT think PASS POL

(4-5) VP-related bundles involving *iu* ‘say’

*nan te iu no* (Conversation)  
what QT say NML

*tte it te ta* (Conversation)  
QT say TE ASP:PST

(4-6) VP-related bundle involving *kangaeru* ‘consider’

*to kangae rare masu* (Speech)  
QT consider PASS POL

(4-7) VP-related bundles involving bare quotative *tte*

*ta n da tte* (Conversation)  
PST NML COP QT

*ta n desu tte* (Interview)  
PST NML COP QT

Speech has a variety of lexical verbs (i.e., *okonau* ‘conduct,’ *wakaru* ‘understand,’ *kangaeru*, ‘consider,’ *miru* ‘look,’ *deru* ‘go out,’ *kaku* ‘write,’ *omou* ‘think,’ *arawasu* ‘express,’ *motsu* ‘have’). Another point to note is that all three registers include the sequences *ni + naru* [COP + become] and *shi + te + ru/masu* [do + TE + ASP/POL].

(4-8) VP-related bundles involving lexical verb

*ni nat te ru* (Conversation)  
COP become TE ASP:NONPST

*ari masu yo ne* (Interview)  
 exist POL PP PP

*o okonai mashi ta* (Speech)  
 ACC conduct POL PST

Semantically, *kedo*, *keredo*, *keredomo*, and *ga* all mean ‘though’, but they appear in different registers. *Kedo* appears in Conversation and Interview while *kedomo*, *keredomo*, and *ga* appear in Interview and Speech. Thus, Conversation is dependent on a single form (*kedo*) while Interview utilizes all four forms, and Speech is in between the two (using three of the four forms). This is compatible with the overall characteristics of morphemic bundles in the three registers as revealed by the type-token ratio (see Section 4.1 above). The *te*-form appears in all three registers, but the conditional *to* only appears in Interview and Speech bundles.

(4-9) VP-related bundles involving clause-chaining forms

*ta n da kedo* (Conversation)  
 PST NML COP but

*na n desu keredomo* (Interview)  
 COP NML COP but

*te ori masu ga* (Speech)  
 TE ASP POL but

*kaet te ki te* (Conversation)  
 return TE ASP TE

*ga ari mashi te* (Interview)  
 NOM exist POL TE

*te mi masu to* (Speech)  
 TE see POL COND

#### 4.2.2. NP-related bundles

A majority of NP-related bundles across three registers involve formal nouns (nominalizers), namely, *koto*, *no*, *mono*, or *katachi* (3/3 or 100% in Conversation; 33/34 or 97% in Interview; 41/62 or 66% in Speech). Interview includes a single instance of a common noun, *imi* ‘meaning,’ in the bundle *soo + iu + imi + de* [so + say + meaning + COP] ‘in that sense.’ Speech also includes a single instance of a common noun, *kenkyuu* ‘research,’ in the bundle *hon + kenkyuu + de + wa* [this + research + LOC + TOP] ‘in this study.’

(4-10) NP-related bundles involving formal nouns

<i>tte</i>	<i>iu</i>	<i>no</i>	<i>wa</i>	(Conversation)
QT	say	NML	TOP	

<i>to</i>	<i>iu</i>	<i>koto</i>	<i>de</i>	(Interview)
QT	say	NML	COP	

<i>to</i>	<i>iu</i>	<i>mono</i>	<i>o</i>	(Speech)
QT	say	NML	ACC	

Speech also involves bundles with phrasal particles as in the following list.

(4-11) NP-related bundles involving phrasal particles

<i>ni</i>	<i>taishi</i>	<i>te</i>	<i>wa</i>	‘regarding’	(Speech)
DAT	face	TE	TOP		

<i>ni</i>	<i>tsui</i>	<i>te</i>	<i>wa</i>	‘concerning’	(Speech)
DAT	adhere	TE	TOP		

<i>to</i>	<i>shi</i>	<i>mashi</i>	<i>te</i>	‘as for’	(Speech)
QT	do	POL	TE		

### 4.3. Functional characteristics of the morphemic bundles

Given that each morpheme represents a smallest meaningful unit, all the multi-morphemic bundles identified in this study are potentially multi-functional. Nevertheless, it is possible to categorize the bundles into several functional types. The functional classification of the bundles enables us to get an overall picture of the interrelationships between the use of the multi-morphemic units and the particular registers. The functions are identified inductively based on a detailed examination of concordance lines and surrounding contexts, although the determination of the functional categories are influenced by previous literature. When a given bundle was found to be clearly multi-functional, I categorized the bundle according to its most frequent function.

There are five major functional categories: stance expressions, discourse organizers, referential expressions, socio-interactional expressions, and objective statements. The first three categories are based on the previous studies on lexical bundles (e.g., Biber et al. 2004; Biber & Barbieri 2007). Below are the definitions of these categories given in Biber et al. (2004: 348).

- (1) Stance expression — expresses attitudes or assessments of certainty that frame some other proposition
- (2) Discourse organizer — reflects relationships between prior and coming discourse
- (3) Referential expression — makes direct reference to physical or abstract entities, or to the textual context itself

I added two more categories to account for the bundles which do not belong to any of the three categories above. I define these categories as follows:

- (4) Socio-interactional expression — expresses various kinds of concerns for interactional activities and intersubjective understandings

(5) Objective statement — conveys action, event, or state of affairs objectively

Stance expressions include reporting (e.g., *na + to + omot + te* [PP + QT + think + TE], *tte + it + te + ta* [QT + say + TE + PST]), interpersonal (e.g., *na + n + da + yo* [COP:ATT + NML + COP + PP]), reactive (e.g., *un + un + un + un* [yes + yes + yes + yes]), word/expression search (e.g., *nan + te + iu + no* [what + QT + say + NML]), and inquiry (e.g., *ta + n + desu + ka* [PST + NML + COP + Q]). Examples of objective statements are *nat + te + i + masu* [become + TE + ASP + POL] and *o + okonai + mashi + ta* [ACC + conduct + POL + PST].

Table 4.3 presents the distribution of bundles across these five functional types. As we can see, stance expressions are not very common in all three registers whereas discourse organizers are more or less common in all three registers. Referential expressions are most frequent in Speech (38.9%), fairly frequent in Interview (22.8%), but they are infrequent in Conversation (4.3%). Socio-interactional expressions, on the other hand, are most frequently used in Conversation (54.3%) and Interview (40.7%), and least frequent in Speech (11.4%). Objective statements are approximately five times more frequent in Speech (24%) than in Conversation (4.3%) and Interview (4.8%).

Table 4.3. Distribution of morphemic bundle types across functional types

	Conversation		Interview		Speech	
	No.	%	No.	%	No.	%
Stance	7	7.6	8	4.2	13	7.4
Discourse organizer	27	29.3	52	27.5	32	18.3
Referential	4	4.3	43	22.8	68	38.9
Socio-interactional	50	54.3	77	40.7	20	11.4
Statement	4	4.3	9	4.8	42	24.0
Total	92	100.0	189	100.0	175	100.0

Table 4.4 lists representative bundles in the major and sub-functional categories (see Appendix for the complete list of the bundles). The subsequent subsections describe these categories and bundles in detail.

Table 4.4. Functional classification of common morphemic bundles across registers

*	80-129 per million morphemes
**	130-179 per million morphemes
***	180-229 per million morphemes
****	230-279 per million morphemes
*****	over 280 per million morphemes

		Conversation	Interview	Speech
(1) STANCE EXPRESSIONS				
(A) Epistemic stance				
<i>ta no ka na</i> PST NML Q PP	<uncertainty>	*		
<i>(doo) na n daro o</i> how COP:ATT NML COP AUX	<uncertainty>	*		
<i>(dochira) ka to iu to</i> which Q QT say COND	<relativeness>		*	
<i>kamo shire mase n</i> Q know POL NEG	<uncertainty>		*	*
<i>n ja nai ka</i> NML COP NEG Q	<uncertainty>	***	*****	**
<i>de wa nai ka</i> COP TOP NEG Q	<uncertainty>			*****
(B) Attitudinal stance				
<i>oo ka na to</i> VOL Q PP QT	<intention>	**		
<i>te iki tai to</i> TE go want QT	<desire>			**
<i>ta hoo ga ii</i> PST way NOM good	<recommendation>	***		

		Conversation	Interview	Speech
<i>nakere ba nara nai</i> NEG COND must NEG	<obligation>			*
<i>te mi mashi ta</i> TE look POL PST	<attempt>			****
<hr/>				
(2) DISCOURSE ORGANIZERS				
(A) Topic introduction/elaboration				
<i>hon kenkyuu de wa</i> this research LOC TOP	'in this research'			**
<i>na n desu ga</i> COP:ATT NML COP but	'as for'			*****
<i>na n desu keredomo</i> COP:ATT NML COP but	'as for'			*****
<i>to shi mashi te</i> QT do POL TE	'regarding'			****
<i>ni tsui te wa</i> DAT adhere TE TOP	'concerning'			*
<i>sore ni taishi te</i> that DAT oppose TE	'on the other hand'			**
<i>te mi masu to</i> TE ASP POL COND	'if (we) try'			**
(B) Cohesion				
<i>te ru n da/desu</i> TE ASP NML COP	<cohesion>	*****	*****	*
<i>te ta n da/desu</i> TE PST NML COP	<cohesion>	***	*****	
<i>dat ta n da/desu</i> COP PST NML COP	<cohesion>	**	*****	
(C) Background				
<i>ta n da/desu kedo</i> PST NML COP but	<background>	*****	*****	
<i>nai n da/desu kedo</i> NEG NML COP but	<background>	*****	****	

		Conversation	Interview	Speech
<i>omou n da/desu kedo</i> think NML COP but	<background>	**	*****	
<i>na n da/desu kedo</i> COP:ATT NML COP but	<background>	****	*****	
<i>ta n desu keredomo</i> PST NML COP but	<background>		*****	**
<i>ta n desu ga</i> PST NML COP but	<background>			***

### (3) REFERENTIAL EXPRESSIONS

#### (A) Conceptualization

<i>tte iu no wa</i> QT say NML TOP	<conceptualization>	*****	*****	
<i>to iu no wa</i> QT say NML TOP	<conceptualization>		*****	*****
<i>to iu koto de</i> QT say NML COP	<conceptualization>		****	*****
<i>to iu koto ni</i> QT say NML COP:ADV	<conceptualization>		*	*****
<i>tte/to iu koto desu</i> QT say NML COP	<conceptualization>		*	****

#### (B) Equivocation

<i>soo iu fuu ni</i> that say way COP:ADV	<equivocation>	**	*****	
<i>soo iu no wa</i> that way NML TOP	<equivocation>	*	**	
<i>tte/to iu fuu ni</i> QT say way COP:ADV	<equivocation>		*****	*****
<i>tte/to iu yoo na</i> QT say like COP:ATT	<equivocation>		*****	*****

		Conversation	Interview	Speech
<b>(4) SOCIO-INTERACTIONAL EXPRESSIONS</b>				
<b>(A) Reporting</b>				
<i>na to omot te</i> PP QT think TE	<stance reporting>	*****	*	
<i>oo to omot te</i> VOL QT think TE	<stance reporting>	***	*	
<i>tai to omoi masu</i> want QT think POL	<stance reporting>			*****
<i>ka to omoi masu</i> Q QT think POL	<stance reporting>			*****
<i>to omou n da/desu</i> QT think NML COP	<impersonal stance reporting>	**	*****	***
<i>da to omoi masu</i> COP QT think POL	<impersonal stance reporting>		****	**
<i>to kangae rare masu</i> QT think POT POL	<impersonal stance reporting>			**
<i>tte it te ta</i> QT say TE PST	<utterance reporting>	****		
<b>(B) Interpersonal</b>				
<i>n da/desu yo ne</i> NML COP PP PP	<interpersonal>	*****	*****	
<i>na n da/desu yo</i> COP:ATT NML COP PP	<interpersonal>	****	*****	
<i>ta n desu yo</i> PST NML COP PP	<interpersonal>	***	*****	
<i>n ja nai no</i> NML COP NEG PP	<interpersonal>	*****		
<i>ta n desu ne</i> PST NML COP PP	<interpersonal>		*****	
<b>(C) Reactive</b>				
<i>soo soo soo soo</i> so so so so	<reactive>	*****		
<i>un un un un</i> yes yes yes yes	<reactive>	*****	*****	

		Conversation	Interview	Speech
<i>soo na n da</i> so COP:ATT NML COP	<reactive>	*****	***	
<i>soo na n desu</i> so COP:ATT NML COP	<reactive>		*****	
<i>aa soo desu ka</i> INJ so COP Q	<reactive>		*****	
<i>soo da/desu yo ne</i> so COP PP PP	<reactive>	*****	*****	
(D) Word/expression search				
<i>nan te iu no</i> what QT say NML	<word/expression search>	*****		
<i>te iu no ka</i> QT say NML Q	<word/expression search>	*		
<i>iu no ka na</i> say NML Q PP	<word/expression search>	*		
<i>nan te iu n</i> what QT say NML	<word/expression search>		*****	
<i>te iu n desho</i> QT say NML COP	<word/expression search>		*	
<i>iu n desho o</i> say NML COP AUX	<word/expression search>		**	
(E) Inquiry				
<i>de irasshai masu ka</i> COP exist:HON POL Q	<status inquiry>		**	
<i>te rasshai masu ka</i> TE ASP:HON POL Q	<inquiry>		*	
<i>ta n desu ka</i> PST NML COP Q	<inquiry>		*****	
<i>yat te n no</i> do ASP NML PP	<inquiry>	*		
(F) Others				
<i>yoroshiku onegai shi masu</i> favorably please do POL	‘thank you in advance’		*****	
<i>arigatoo gozai mashi ta</i> thank exist:POL POL PST	‘thank you’		*****	

		Conversation	Interview	Speech
<i>doomo arigatoo gozai</i> <i>mashi (ta)</i> much thank exist:POL POL PST	'thank you very much'		*	
<i>s ase te itadaki</i> do CAUS TE receive:HUM	<announcement>			**
<i>ase te itadaki masu</i> CAUS TE ASP:HUM POL	<announcement>			**
<i>n de wa nai</i> NML COP TOP NEG	<proposal>			*
<i>nai desho o ka</i> NEG COP AUX Q	<proposal>			**
<i>to omoi masu ga</i> QT think POL but	<disclaimer>			*
<hr/>				
(5) OBJECTIVE STATEMENTS				
<i>kaet te ki ta</i> return TE ASP PST	<statement>	*		
<i>ni nat te ru</i> COP:ADV become TE ASP	<statement>	*	*	
<i>nat te i masu</i> become TE ASP POL	<statement>			*****
<i>shi te i masu</i> do TE ASP POL	<statement>			*****
<i>o okonai mashi ta</i> ACC conduct POL PST	<statement>			****
<i>are te i masu</i> PASS TE ASP POL	<statement>			*****
<i>s are te i</i> do PASS TE ASP	<statement>			*****
<i>mir are mashi ta</i> see PASS POL PST	<statement>			**

### 4.3.1. Stance expressions

Stance expressions express attitudes or assessments of certainty that frame some other proposition. Stance bundles can be further classified into two types: epistemic stance and attitudinal stance. Epistemic stance bundles comment on the knowledge status of the information in the proposition (Biber et al. 2004: 389), including possibility, impossibility, speculation, hearsay, inference, certainty, and uncertainty. Attitudinal stance bundles express speaker attitudes towards the actions or events described in the proposition (Biber et al. 2004: 389). The next two subsections examine the two types of stance bundles.

As seen in Table 4.3, compared to other functional types, the use stance expressions is limited in all three registers. This is contrary to our expectation that spoken discourse, especially interactive ones like Conversation and Interview, would include a lot of stance-related expressions.<sup>30</sup> There are two major explanations for the infrequent use of stance bundles in three spoken registers. One reason is that stance-laden linguistic items are often too short to be recognized as four-morphemic bundles. For example, the particle sequence *ka + na* [Q + PP] '(I) wonder' is frequently used in Conversation, but since it is a two-morpheme sequence, it is not recognized as a (four-morphemic) bundle in the present study. The second reason, which is more significant for the present study, is that stance-related expressions are regularly followed by the quoting particle *to* or *tte* and the quoting verb, *iu* 'say,' *omou* 'think,' or *kangaeru* 'think.' This tendency of stance expressions being quoted is lowest in Conversation and highest in Speech. As an illustration, I counted the number of bare and quoted *ka + na* [Q + PP] '(I) wonder' in each register. The following table shows the results.

---

<sup>30</sup> In fact, in lexical bundle studies on English, Spanish, and Korean, stance expressions were commonly used in interactive spoken registers (see Biber et al. 2010).

Table 4.5. Bare *ka + na* [Q + PP] vs. quoted *ka + na* [Q + PP] in three registers

	Conversation		Interview		Speech	
	No.	%	No.	%	No.	%
Bare [ <i>ka + na</i> ]	681	76.3	60	33.3	5	6.7
Quoted [ <i>ka + na</i> ]	212	23.7	120	66.7	70	93.3
Total	893	100.0	180	100.0	75	100.0

First of all, we can see that the uncertainty or speculation stance expression *ka + na* is much more commonly used in Conversation than in Interview and Speech.<sup>31</sup> In Conversation, it is more common for *ka + na* to appear utterance-finally, whereas in Interview and Speech, it tends to be quoted by *to + iu/omou/kangaeru* [QT + say/think/think]. The bare use of *ka + na* can be considered to be “self-directed,” and the quoted one with *to + omou* or *to + kangaeru* “addressee-directed” (see Hirose 1995). Self-directed expressions can be uttered to oneself without the presence of addressees. They simply represent what the speaker is thinking at the present moment. On the other hand, addressee-directed expressions “report” the speaker’s internal thoughts either at the present moment or in the past. It is important to note here that the bare and quoted *ka + na* belong to different functional categories: the bare ones (i.e., thinking out loud) are stance expressions while the quoted ones are socio-interactional expressions (see Section 4.3.4.1 for the analysis of the quoted *ka + na*). The fact that two different types of bundles, that is, stance and socio-interactional, overlap suggest that there is a larger pattern, which cannot be fully captured by the analysis of morphemic bundles alone (see Section 4.6 for further discussion).

<sup>31</sup> Five instances of non-quoted *ka + na* [Q + PP] in Speech all occur outside the main activity of giving an academic speech. The following is one example: *hait te n no ka na* [enter TE ASP NML Q PP] ‘(I) wonder if (my voice) is being recorded.’ This utterance is given at the beginning of a presentation. The speaker is not talking to the audience, but he is wondering out loud whether his voice is already being recorded or not.

### 4.3.1.1. Epistemic stance bundles

Epistemic stance bundles comment on the knowledge status of the information in the proposition (Biber et al. 2004: 389), including possibility, impossibility, speculation, hearsay, inference, certainty, and uncertainty. In Japanese, auxiliaries such as *rashii*, *yoo*, *mitai* (inference), *mai* (impossibility), *daroo* (speculation), and *soo* (hearsay) are often associated with epistemic stance (see Iwasaki 2002: 42). Most of the stance bundles in three registers belong to this type (5/7 or 71.4% in Conversation, 8/8 or 100% in Interview, and 10/13 or 76.9% in Speech). Below are some examples.

#### (4-12) Epistemic stance bundles

<i>ta</i>	<i>no</i>	<i>ka</i>	<i>na</i>		<uncertainty> (Conversation)	
PST	NML	Q	PP			
<i>(doo)</i>	<i>na</i>		<i>n</i>	<i>daroo</i>	<i>o</i>	<uncertainty> (Conversation)
how	COP:ATT		NML	COP:IMPF	AUX	
<i>(dochira)</i>	<i>ka</i>	<i>to</i>	<i>iu</i>	<i>to</i>		<relativeness> (Interview)
which	Q	QT	say	COND		
<i>kamo</i>	<i>shire</i>	<i>mase</i>	<i>n</i>			<uncertainty> (Interview and Speech)
Q	know	POL	NEG			
<i>n</i>	<i>ja</i>	<i>nai</i>	<i>ka</i>			<uncertainty> (Conversation, Interview and Speech)
NML	COP	NEG	Q			
<i>de</i>	<i>wa</i>	<i>nai</i>	<i>ka</i>			<uncertainty> (Speech)
COP	TOP	NEG	Q			

The majority of the second bundle in the list above, *na + n + daroo + o* [COP:ATT + NML + COP:IMPF + AUX], is preceded by *doo* ‘how’ (32/51; 62.7%). The third bundle *ka + to + iu + to* [Q + QT + say + COND] is immediately preceded by *dochira* ‘which’ in most cases (21/24;

87.5%). Thus, the five-morphemic sequence, *dochira + ka + to + iu + to* [which + Q + QT + say + COND], constitutes a formulaic expression meaning ‘if anything’ (lit. ‘if (I had to) say which’).

Regardless of the registers, the bundles in this category express uncertainty rather than certainty. Excerpts (4-13) and (4-14) include the examples of *ta + no + ka + na* [PST + NML + Q + PP]. The bundle in both examples express uncertainty on the part of the speaker.

(4-13) *ta + no + ka + na* [PST + NML + Q + PP] <uncertainty> (Conversation: BTS\_071)<sup>32</sup>

1. F161: ポポロンの一何味だったのかな。ブドウとあとあとなんだ  
つけ。あっあポポロンのアンズ味だったのかなー。
2. F062: うん。珍しい。

1. F161: *poporon no: nani aji dat ta no ka na. budoo to ato ato nan dak ke.*  
(name) GEN what flavor COP PST NML Q PP / grape and other other what COP Q  
*a a poporon no anzu aji dat ta no ka na:.*  
INJ INJ (name) GEN apricot flavor COP PST NML Q PP
2. F062: *un. mezurashii.*  
yes rare

1. F161: ‘(I) wonder what poporon’s flavor it was. It was grape and what was the other one? Oh, ah, (I) wonder if it was poporon’s apricot flavor.’
2. F062: ‘Yeah. (It)’s rare.’

(4-14) *ta + no + ka + na* [PST + NML Q + PP] <uncertainty> (Conversation: Work\_11F)

1. 11E: 見つかったのかな仕事。
2. 11A: あそこはお金持ちだから大丈夫でしょう。

1. 11E: *mitsukat ta no ka na shigoto.*  
found PST NML Q PP job
2. 11A: *asoko wa o kanemochi da kara daijoobu desho o.*  
there TOP PFX rich COP because fine COP AUX

1. 11E: ‘(I) wonder if (he) found a job.’
2. 11A: ‘He is fine because his family is rich.’

<sup>32</sup> The information about each excerpt is presented in parentheses as follows: (Register: Corpus\_excerpt ID)

Excerpts (4-15) and (4-16) include examples of (*doo*) + *na* + *n* + *daro* + *o* [(how) + COP:ATT + NML + COP:IMPF + AUX], expressing uncertainty.

(4-15) (*doo*) + *na* + *n* + *daro* + *o* [(how) + COP:ATT + NML + COP:IMPF + AUX]  
<uncertainty>(Conversation: BTS\_072)

1. F086: えーでも一応さーあの人英文学ができるとかそういうんじゃないの？
2. F046: 分かんない。 どう なんだろう。
1. F086: *e: demo ichioo sa: ano hito eibun gaku ga dekiru toka soo iu n ja nai no?*  
INJ but tentatively PP INJ person English literature NOM can such that say NML COP NEG PP
2. F046: *wakan nai. doo na n daro o.*  
understand NEG how COP NML COP AUX
1. F086: ‘Really? but isn’t it that the person can do English literature or something like that?’
2. F046: ‘I don’t know. (I) wonder.’

(4-16) (*doo*) + *na* + *n* + *daro* + *o* [(how) + COP:ATT + NML + COP:IMPF + AUX]  
<uncertainty> (Conversation: Callhome\_2004)

1. B: それはいくらいくらあのぐらいじゃなくてあのあれかしら年齢制限って  
いうのあるの？
2. A: どう なんだろうね。あんまりないんじゃないアメリカだったら。
1. B: *sore wa ikura ikura ano gurai ja naku te ano are kashira nenrei seigen tte iu no aru no?*  
that TOP how.much um about COP NEG TE um that PP age limit QT say NML exist PP
2. A: *doo na n daro o ne. anmari nai n ja nai amerika dat ta ra.*  
how COP NML COP AUX PP / much NEG NML COP NEG America COP PST COND
1. B: ‘How much how much is that, I mean um is it that, is there an age limit?’
2. A: ‘(I) wonder. If it’s America, maybe there isn’t much.’

The following excerpts (4-17) and (4-18) contain the five-morphemic sequence, *dochira* + *ka* + *to* + *iu* + *to* [which + Q + QT + say + QT]. It conveys the sense of relativeness in choosing one over the other, as in (4-17), or describing something in one way, as in (4-18).

(4-17) *dochira + ka + to + iu + to* [which + Q + QT + say + QT] <relativeness> (Interview: Hypermedia\_014)

1. N: そのまあ研究とは別にお暇な時間などの時のやはり本を読むことが多くてらっしゃるんですか？
2. O: そうですねあのー研究始める前はどちらかというとなんか訳された小説読むのが好きで
3. N: うん

1. N: *sono maa kenkyuu to wa betsuni o hima na jikan nado no toki no yahari hon o yomu koto ga ooku te rassharu n desu ka?*  
that um research COM TOP particularly PRF free COP:ATT time such GEN when GEN  
*yahari hon o yomu koto ga ooku te rassharu n desu ka?*  
after.all book ACC read NML many TE exist:HON NML COP Q
2. O: *soo desu ne ano: kenkyuu hajimeru mae wa dochira ka to iu to honyaku sa re ta shoosetsu yomu no ga suki de*  
that COP PP um research begin before TOP **which Q QT say QT** translation do PASS  
PST novel read NML NOM like COP
3. N: *un.*  
yes

1. N: ‘Besides your research, when you have free time, do you read books often?’
2. O: ‘Let me think, um, before I began my research, if I had to say which, I liked reading translated novels’
3. N: ‘Yes.’

(4-18) *dochira + ka + to + iu + to* [which + Q + QT + say + QT] <relativeness> (Interview: Hypermedia\_050)

1. C: 日本人はどちらかというとなんか以心伝心とか
2. M: うん うん
3. C: 分かってもらうのがき-なんていうのかしら期待していることもありま  
すし
4. M: うん

1. C: *nihonjin wa dochira ka to iu to ishindenshin toka*  
Japanese TOP **which Q QT say QT** telepathy such
2. M: *un un*  
yes yes
3. C: *wakat te morau no ga ki- nan te iu no kashira kitai shi te iru koto mo ari masu shi*  
understand TE receive NML NOM what TE say NML PP expect do TE ASP NML also exist POL and
4. M: *un*  
yes

1. C: ‘Japanese people are, if I had to say which, like direct communication from mind to mind’
2. M: ‘Yes, yes.’

3. C: ‘How to say, they expect to be understood’
4. M: ‘Yes.’

The following are examples of *kamo + shire + mase + n* [Q + know + POL + NEG] from Interview, (4-19), and Speech, (4-20). It is easy to see uncertainty expressed by the use of this bundle, meaning ‘maybe.’

(4-19) *kamo + shire + mase + n* [Q + know + POL + NEG] <uncertainty> (Interview: CSJ\_D01F0003)

1. R: まー多分元々教えることも好きだったの**かもしれません**
  2. L: うーんうーん
  3. R: うーん
  4. L: 教えんの向いてそうですね
  5. R: あそうですか
- 
1. R: *ma: tabun motomoto oshieru koto mo suki dat ta no kamo shire mase n.*  
INJ probably originally teach NML also like COP PST NML Q know POL NEG
  2. L: *u:n u:n*  
INJ INJ
  3. R: *u:n*  
INJ
  4. L: *oshien no mui te soo desu ne*  
teach NML suit TE seem COP PP
  5. R: *a soo desu ka*  
INJ that COP Q
- 
1. R: ‘I probably may have liked teaching as well from the beginning.’
  2. L: ‘Uh, I see.’
  3. R: ‘Yes.’
  4. L: ‘You seem to be suited for teaching.’
  5. R: ‘Ah do you think so?’

(4-20) *kamo + shire + mase + n* [Q + know + POL + NEG] <uncertainty> (Speech: CSJ\_A02M0076)

えーす- 全ての資料えー例えば今回扱った全ての資料がこの同じえー説明で  
いいかどうかということは考える必要がある**かもしれません**

*e: su- subete no shiryoo e: tatoeba konkai atsukat ta subete no shiryoo ga kono onaji e:*  
INJ all GEN data INJ for.example this.time use PST all GEN data NOM this same INJ

*setsumei de ii ka doo ka to iu koto wa kangaeru hitsuyoo ga aru kamo shire mase n.*  
 explanation COP good Q how Q QT say NML TOP think necessary NOM exist Q know POL NEG

‘Um, all the data, we may need to think whether the same explanation works for all the data that we used this time, for example.’

About 40% of the bundle *n + ja + nai + ka* [NML + COP + NEG + Q] in Interview is followed by the particle *na* [PP], and more than 70% of the sequences, *n + ja + nai + ka* [NML + COP + NEG + Q] and *n + ja + nai + ka + na* [NML + COP + NEG + Q + PP], are followed by the quotative particle *to* and the quoting verb *omou* ‘think.’ The tendency for the bundle to be followed by the quoting particle and the verb is also found in Speech. In addition, in Speech, the quotative particle *to* and the quoting verb *omou* ‘think’ or *kangaeru* ‘think’ tend to be interrupted by an equivocation sequence, *iu + fuu + ni* [say + way + COP:ADV] (see Section 4.3.3.2). This tendency also holds true for the bundle, *de + wa + nai + ka* [COP + TOP + NEG + Q], in Speech. Observe the following examples from Interview, (4-21) and (4-22), and Speech (4-23), (4-24), and (4-25).

(4-21) *n + ja + nai + ka* [NML + COP + NEG + Q] <uncertainty> (Interview: Hypermedia\_041)

1. M: ええまこの前のO J シンプソンの裁判のその一弁護士一側の弁護人の黒人の英語も非常に一つのリズムを持っています
  2. Y: ええ
  3. M: それにはその一かなり陪審員達はあ一影響されたんじゃないかと思うんですけれども
  4. M: ええ
- 
1. M: *ee ma kono mae no OJ shinpuson no saiban no sono: bengoshi gawa no bengonin no*  
 yes INJ this before GEN (name) GEN trial GEN INJ lawyer side GEN defense.lawyer GEN  
*kokujin eigo mo hijooni hitotsu no rizumu o mot te i te*  
 black.person GEN English also very one GEN rhythm ACC have TE ASP TE
  2. Y: *ee*  
 yes
  3. M: *sore ni wa sono: kanari baishin in tachi wa a: eikyoo sa re ta n ja nai ka to omou*  
 that DAT TOP INJ jury member PL TOP INJ influence do PASS PST NML COP NEG Q QT think

*n desu keredomo*

NML COP but

4. Y: *ee*  
yes

1. M: 'Yes, um, the defense lawyer at the O J Simpson's trial the other day, his black English also had a distinct rhythm,'
2. Y: 'Yes.'
3. M: 'I think the jury members may have been quite influenced by that,'
4. Y: 'Yes.'

(4-22) *n + ja + nai + ka* [NML + COP + NEG + Q] <uncertainty> (Interview:Hypermedia\_023)

1. S: ええとーすごい選手が気付かないうちに
2. M: うん
3. S: あの監督とかに
4. M: うん うん
5. S: 与えられたりするのがとても選手にとって悔しいことなんじゃないかなーと思います。

1. S: *eeto: sugoi senshu ga kizuka nai uchi ni*  
INJ really athlete NOM realize NEG while COP:ADV
2. M: *un*  
yes
3. S: *ano kantoku toka ni*  
INJ coach such DAT
4. M: *un un*  
yes yes
5. S: *atae rare tari suru no ga totemo senshu ni tot te kuyashii koto na n ja nai ka*  
give PASS such do NML NOM very athlete DAT concern TE regrettable NML COP NML COP NEG Q  
*na: to omoi masu.*  
PP QT think POL

1. S: 'Um, really, while the athletes don't realize'
2. M: 'Yes.'
3. S: 'Um, like by their coaches'
4. M: 'Yes, yes.'
5. S: 'I think it may be disappointing for them (the athletes) that they are given (the drugs).'

(4-23) *n + ja + nai + ka* [NML + COP + NEG + Q] <uncertainty> (Speech: CSJ\_A07M0201)

えじぞん-事前分布を使えるというね意味でもえーそれを有効にえーと分析に織り込むことができるという意味でもベイズを使うというのがいいんじゃないかという風に考えました

*e jizon- jizen bunpu o tuka eru to iu ne imi de mo e: sore o yuukoo ni e:to bunseki*  
 INJ prior distribution ACC use POT QT say meaning COP also INJ that ACC effective COP:ADV INJ analysis  
*ni orikomu koto ga dekiru to iu imi de mo beizu o tsukau no ga ii n ja nai*  
 DAT incorporate NML NOM can QT say meaning COP also (name) ACC use NML NOM good NML COP NEG  
*ka to iu fuu ni kangae mashi ta.*  
 Q QT say way COP:ADV think POL PST

‘Um, in the sense that we can use the prior distribution, and also in the sense that we can effectively incorporate that into the analysis, we thought it may be good to use Bayes.’

(4-24) *de + wa + nai + ka* [COP + TOP + NEG + Q] <uncertainty> (Speech: CSJ\_A01M0137)

えー までできれば三キロヘルツ以下の周波数で え 設定した方が いいんでは  
ないか という風に考えております

*e: ma dekire ba san kiro herutsu ika no shuuhasuu de e settei shi ta hoo ga ii n*  
 INJ INJ can COND three kilo hertz or.below GEN frequency COP INJ set do PST direction NOM good NML  
*de wa nai ka to iu fuu ni kangae te ori masu*  
 COP TOP NEG Q QT say way COP:ADV think TE ASP:HUM POL

‘Um, ah, if possible, we think it may be better to set up the frequency at or below three kilohertz.’

(4-25) *de + wa + nai + ka* [COP + TOP + NEG + Q] <uncertainty> (Speech: CSJ\_A05F0502)

このグループに関してはま 厳しくい うとまだ 発展 途上 とい うか 習得 途上 にあ  
 る 状況 ではないか という風に 思われます

*kono guruupu ni kanshi te wa na kibishiku iu to mada hatten tojoo to iu ka shuutoku tojoo ni*  
 this group DAT concern TE TOP harshly say COND yet development halfway QT say Q acquisition halfway DAT  
*aru jookyoo de wa nai ka to iu fuu ni omowa re masu*  
 exist situation COP TOP NEG Q QT say way COP:ADV think PASS POL

‘Regarding this group, to say harshly, we think they may be in the situation where they are still developing or acquiring.’

The tendency for the epistemic stance bundles to be followed by the quotative particle and verb suggest that there is a larger pattern than morphemic bundles that needs to be accounted for.

This issue is further discussed in Section 4.6.

### 4.3.1.2. Attitudinal stance bundles

The second type, attitudinal stance bundles express speaker attitudes towards the actions or events described in the proposition (Biber et al. 2004: 389). While the epistemic stance is concerned with the knowledge status of the information (e.g., certainty/uncertainty, possibility/impossibility), the attitudinal stance is related to the speaker attitudes other than the knowledge status. Biber et al. (2004) identify four sub-categories in the use of English lexical bundles: desire, obligation/directive, intention/prediction, and ability. In the present Japanese corpora, the following five sub-categories emerged: intention, desire, recommendation, obligation, and attempt as can be seen in (4-26).

#### (4-26) Attitudinal stance bundles

<i>oo</i>	<i>ka</i>	<i>na</i>	<i>to</i>	<intention>	(Conversation)
VOL	Q	PP	QT		
<i>te</i>	<i>iki</i>	<i>tai</i>	<i>to</i>	<desire>	(Speech)
TE	go	want	QT		
<i>ta</i>	<i>hoo</i>	<i>ga</i>	<i>ii</i>	<recommendation>	(Conversation)
PST	way	NOM	good		
<i>nakere ba</i>	<i>nara</i>	<i>nai</i>		<obligation>	(Speech)
NEG	COND	must	NEG		
<i>te</i>	<i>mi</i>	<i>mashi</i>	<i>ta</i>	<attempt>	(Speech)
TE	look	POL	PST		

We can observe that the first two attitudinal bundles, *oo + ka + na + to* [VOL + Q + PP + QT] and *te + iki + tai + to* [TE + go + want + QT], end with the quotative particle *to*. Examination of the concordance lines reveals that the quotative *to* is followed by the verb *omou* ‘think’ in Conversation and *omou* ‘think’ or *kangaeru* ‘think’ in Speech. If we take these larger segments into account, they do not simply express the speaker’s attitudinal stance but report the speaker’s

attitudinal stance in the past or at the present moment.<sup>33</sup> However, since these bundles do not include the quoting verbs, I have categorized them as the attitudinal stance bundles. There are other bundles which include stance-related morpheme(s) and the quoting sequence *to + omou/kangaeru* [QT + think]. These are categorized as socio-interactional bundles, which I discuss in Section 4.3.4.1.

The bundle *oo + ka + na + to* [VOL + Q + PP + QT] only appears in Conversation. In the following excerpts, (4-27) and (4-28) include the attitudinal stance at the present moment, while (4-29) includes the speaker's stance in the past.

(4-27) *oo + ka + na + to* [VOL + Q + PP + QT] <intention> (Conversation: Callhome\_2215)

1. A: あの日本のレストランがあんのよね。そこで働こうかなと思っているのよ二日ぐらい。
2. B: うん。

1. A: *ano nihon no resutoran ga an no yo ne. soko de hatarak oo ka na to omot te iru no*  
INJ Japan GEN restaurant NOM exist NML PP PP / there LOC work VOL Q PP QT think TE ASP NML  
*yo futsu ka gurai.*  
PP two day about

2. B: *un.*  
yes

1. A: 'Um, there is Japanese restaurant. I'm thinking of maybe working there, for about two days a week.'
2. B: 'Yeah.'

(4-28) *oo + ka + na + to* [VOL + Q + PP + QT] <intention> (Conversation: Calhome\_1263)

1. B: 今日だから今からその団地を見に行こうと思ってるんだけど。
2. A: うん。
3. B: 何着て行こうかなと思って。短パンで歩いたら変かなとかさ。
4. A: なるほどね。

---

<sup>33</sup> In Conversation, 70% of the bundle *oo + ka + na + to* [VOL + Q + PP + QT] report the speaker's stance at the present moment and 30% report the speaker's past stance. In Speech, all of the instances of *te + iki + tai + to* [TE + go + want + QT] report the speaker's want at the present moment.

1. B: *kyoo dakara ima kara sono danchi o mi ni ik oo to omot te ru n da kedo.*  
today because now from that apartment ACC look to go VOL QT think TE ASP NML COP but
2. A: *un.*  
yes
3. B: *nani ki te ik oo ka na to omot te. tanpan de arui ta ra hen ka na toka sa.*  
what wear TE ASP VOL Q PP QT think TE / shorts COP walk PST COND strange Q PP such PP
4. A: *naruhodo ne.*  
indeed PP

1. B: ‘So today, from now I’m thinking of going to look at the apartments.’
2. A: ‘Yeah.’
3. B: ‘I’m thinking what I shall wear, like, would it be strange if I walked in my shorts?’
4. A: ‘I see.’

(4-29) *oo + ka + na + to* [VOL + Q + PP + QT] <intention> (Conversation: Callhome\_2188)

1. A: 誰にしようかなと思ったの。始めね実家にかける時に使おうかなと思っただけ
2. B: うんうんうん。
3. A: 実家はなんか向こうからもかけてくるからそれじゃ友達にしようとか思っ

1. A: *dare ni shi yoo ka na to omot ta no. hajime ne jikka ni kakeru toki ni tsuka*  
who DAT decide VOL Q PP QT think PST PP / beginning PP parents’home DAT call when DAT use  
*oo ka na to omot ta n da kedo*  
VOL Q PP QT think PST NML COP but
2. B: *un un un.*  
yes yes yes
3. A: *jikka wa nanka mukoo kara mo kake te kuru kara sore ja tomodachi ni shiy oo toka omot*  
parents’home TOP DM they from also call TE ASP because that COP friend DAT decide VOL such think  
*te.*  
TE

1. A: ‘I wondered who I should call. At first, I thought of using (the calling card) when I call my parents’ home.’
2. B: ‘Yeah, yeah, yeah.’
3. A: ‘My parents’ home, like, they call me too, so I thought I should call my friend.’

The bundle *te + iki + tai + to* [TE + go + want + QT] only appears in Speech. Although the form *tai* ‘want’ conveys the speaker’s want or wish, the function of the sequence *te + iki + tai + to + omoi + masu* [TE + go + want + QT + think + POL] as a whole is not to report the speaker’s wish but is to make an announcement.

(4-30) *te + iki + tai + to* [TE + go + want + QT] <desire> (Speech: CSJ\_A05M0031)

続いてえ結果の方見ていきたいと思います

*tsuzui te e kekka no hoo mi te iki tai to omoi masu*

continue TE INJ result GEN direction look **TE ASP want QT** think POL

‘Next, we’d like to look at the results.’

(4-31) *te + iki + tai + to* [TE + go + want + QT] <desire> (Speech: CSJ\_A09F0600)

四つの観点からそれぞれ見ていきたいと思います

*yottsu no kanten kara soresore mi te iki tai to omoi masu*

four GEN perspective from each look TE ASP want QT think POL

‘We’d like to look at it from four perspectives one by one.’

(4-32) *te + iki + tai + to* [TE + ASP + want + QT] <desire> (Speech: CSJ\_A01M0115)

この効果についても引き続き検討していきたいと考えております

*kono kooka ni tsui te mo hikitsuzuki kentoo shi te iki tai to kangae te ori masu*

this effect DAT adhere TE also continuously consider do **TE ASP want QT** think TE ASP:HUM POL

‘We’d also like to keep considering this effect.’

The previous two types of attitudinal stance bundles, which are being followed by the quotative *to + omou/kangaeru* [QT + think], suggest, just like some of the epistemic stance bundles, that a larger sequence or pattern exists. This larger pattern involves a longer sequence than what the four-gram analysis can identify. I address what this means for the status of morphemic bundles in Section 4.6.

The bundle *ta + hoo + ga + ii* [PST + direction + NOM + good], which literally means ‘it is better to do ...,’ and it only appears in Conversation. The bundle is used to present a recommended action; the action may be desirable for the speaker him/herself, as in (4-33), for

the addressee, as in (4-34), for both the speaker and the addressee, as in (4-35), or more in general.

(4-33) *ta + hoo + ga + ii* [PST + direction + NOM + good] <recommendation> (Conversation: BTS\_72)

1. F086: やっぱり 資格 とか 取れる やつは 全部 とつといた方が いいかな  
と思って。

2. F046: そうだよー。

1. F086: *yappari shikaku toka tor eru yatsu wa zenbu tot to i ta hoo ga ii ka na to*  
after.all qualification such take POT thing TOP take TE ASP PST direction NOM good Q PP  
*omot te.*  
think TE

2. F046: *soo da yo ne:.*  
that COP PP PP

1. F086: ‘I think after all it may be better to get all the qualifications that I can.’

2. F046: ‘I agree.’

(4-34) *ta + hoo + ga + ii* [PST + direction + NOM + good] <recommendation> (Conversation: Callhome\_2196)

1. A: よく考えた方が いいよ そういうのは。

2. B: うん。

1. A: *yoku kangae ta hoo ga ii yo soo iu no wa.*  
well think PST direction NOM good PP that say NML TOP

2. B: *un.*  
yes

1. A: ‘It’s better to think about such thing well.’

2. B: ‘Yeah.’

(4-35) *ta + hoo + ga + ii* [PST + direction + NOM + good] <recommendation> (Conversation: BTS\_34)

1. F144: でも なんか これって 一緒には 食べづらいね。

2. F148: うーん これだけだとね。 おのおの 別に 食べた方が いいかもしれ  
ない。

3. F144: そうかもしれない。

1. F144: *demo nanka kore tte issho ni wa tabe zurai ne.*  
but DM this QT together COP:ADV TOP eat hard PP
  2. F148: *u:n kore dake da to ne. onoono betsu ni tabe ta hoo ga ii kamo shire*  
INJ this only COP COND PP / each separate COP:ADV eat **PST direction NOM good Q** know  
*nai.*  
NEG
  3. F144: *soo kamo shire nai.*  
that Q know NEG
1. F144: ‘But, like, this is hard to eat together, don’t you think?’
  2. F148: ‘Um, if it’s just this one. It may be better to eat them separately.’
  3. F144: ‘It maybe so.’

The bundle *nakere + ba + nara + nai* [NEG + COND + must + NEG] only appears in Speech and presents some action as obligatory. In both (4-36) and (4-37), what the bundle indicates as obligatory is the action of considering some issue.

(4-36) *nakere + ba + nara + nai* [NEG + COND + must + NEG] <obligation> (Speech: CSJ\_A07F0159)

検討して え-いかなければならない課題ではないだろうかとう-そういう風に思われます

*kentoo shi te e: ika nakere ba nara nai kadai de wa nai daro o ka to u- soo iu fuu ni*  
consider do TE INJ ASP NEG COND must NEG issue COP TOP NEG COP AUX Q QT that say way COP:ADV  
*omow are masu*  
think PASS POL

‘We think that (it) is an issue that we must continue to consider.’

(4-37) *nakere + ba + nara + nai* [NEG + COND + must + NEG] <obligation> (Speech: CSJ\_A08M0206)

... という所から検討を始めなければならぬのではないかと思います

*... to iu tokoro kara kentoo o hajime nakere ba nara nai no de wa nai ka to omoi masu*  
QT say place from consideration ACC begin NEG COND must NEG NML COP TOP NEG Q QT think POL

‘We think that we must begin our consideration from ...’

The last bundle *te + mi + mashi + ta* [TE + look + POL + PST] also appears in Speech only. It follows verbs such as *shiraberu* ‘investigate,’ *miru* ‘look at,’ *matomeru* ‘summarize,’ and *kangaeru* ‘think,’ and indicates the speaker’s stance that the activity specified by the verb is merely an attempt or somehow imperfect. It makes sense that the bundle co-appears with the hesitation marker *chotto* ‘a little’ as in the two of the excerpts, (4-39) and (4-40), below.

(4-38) *te + mi + mashi + ta* [TE + look + POL + PST] <attempt> (Speech: CSJ\_A08F0323)

日本でえ出版されてるものをいくつか えー 拾いましてその定義をまとめて  
みました

*nihon de e shuppan sa re te ru mono o ikutu ka e: hiroi mashi te sono teigi o matome te mi*  
Japan LOC INJ publish do PASS TE ASP NML ACC some Q pick POL TE definition ACC summarize TE ASP  
***mashi ta***  
**POL PST**

‘We picked some that are published in Japan and summarized their definitions.’

(4-39) *te + mi + mashi + ta* [TE + look + POL + PST] <attempt> (Speech: CSJ\_A07F0366)

リスク型一般型後将来型のものという形でえー商品をちょっと分けてみま  
した

*risuku gata ippan gata ato shoorai gata no mono to iu katachi de e: shoohin o chotto wake te mi*  
risk model regular model and future model GEN NML QT say shape COP INJ product ACC little classify TE ASP  
**POL PST**  
***mashi ta***

‘We tried classifying the products into risk model, regular model, and future model.’

(4-40) *te + mi + mashi + ta* [TE + look + POL + PST] <attempt> (Speech: CSJ\_A05F0154)

あの無理やりま英語との接点ていうことでちょっと今お話ししてみました

*ano muriyari ma eigo to no setten te iu koto de chotto ima o hanashi shi te mi mashi ta*  
INJ forcibly INJ English with GEN commonality QT say NML COP little now PRF talk do TE ASP POL PST

‘We just tried talking about the commonality with English sort of forcibly.’

### 4.3.1.3. Summary of stance expressions

In this section, I presented two sub-types of stance expression bundles: epistemic and attitudinal. In both cases, stance bundles tended to be followed by the quoting particle, *to*, and the verb, *omou* ‘think’ or *kangaeru* ‘think.’ I also attributed the rarity of the overall number of stance bundles in Japanese spoken discourse to the propensity of stance expressions to be quoted, especially in more formal registers, namely, Interview and Speech. This suggests that there is a larger pattern of fixedness or semi-fixedness than morphemic bundles, involving stance expressions and quoting constructions. I will come back to this point in Section 4.6.

### 4.3.2. Discourse organizers

Discourse organizers “reflect relationships between prior and coming discourse” (Biber et al. 2004: 384). The following three major types of discourse organizers are inductively identified based on close examination of the concordance lines and larger context: (1) Topic introduction/elaboration, (2) Cohesion, and (3) Background. The three sub-types of discourse organizers have structural correlates as well as register associations.<sup>34</sup> Topic introduction/elaboration bundles are only found in Speech and about half of these bundles are associated with the use of phrasal particles. Cohesion bundles are found in all three registers and they are associated with the morphemic sequence *n + da/desu* [NML + COP]. Background bundles are also found in all three registers, and they include the sequence *n + da/desu* [NML + COP], followed by the concessive conjunctive particle *kedo* ‘but’ or its variants (i.e., *keredomo*, *ga*) at the end of the sequence.

---

<sup>34</sup> (1) Topic introduction/elaboration: 23/32 (71.9%) in Speech, 0% in Conversation and Interview; (2) Cohesion: 16/27 (59.3%) in Conversation, 31/51 (45.1%) in Interview, 6/32 (18.8%) in Speech; (3) Background: 11/27 (40.7%) in Conversation, 20/51 (39.2%) in Interview, 3/32 (9.4%) in Speech.

### 4.3.2.1. Topic introduction/elaboration bundles

In Speech, topic introduction/elaboration is the most common sub-type of discourse organizing bundle. This contrasts significantly with Interview and Conversation which have no instances of topic introduction bundles.<sup>35</sup> Topic introduction bundles introduce new topics into discourse. Topic elaboration bundles are used to present contrastive idea. Below are some representative bundles in this sub-category of discourse organizing bundles. There is only one topic elaboration bundle in the present data, *sore + ni + taishi + te* [that + DAT + oppose + TE] ‘on the other hand.’

#### (4-41) Topic introduction/elaboration bundles

<i>hon</i> this	<i>kenkyuu</i> research	<i>de</i> LOC	<i>wa</i> TOP	‘in this research’	(Speech)
<i>na</i> COP:ATT	<i>n</i> NML	<i>desu</i> COP	<i>ga</i> but	‘as for’	(Speech)
<i>na</i> COP:ATT	<i>n</i> NML	<i>desu</i> COP	<i>keredomo</i> but	‘as for’	(Speech)
<i>to</i> QT	<i>shi</i> do	<i>mashi</i> POL	<i>te</i> TE	‘regarding’	(Speech)
<i>ni</i> DAT	<i>tsui</i> adhere	<i>te</i> TE	<i>wa</i> TOP	‘concerning’	(Speech)
<i>sore</i> that	<i>ni</i> DAT	<i>taishi</i> oppose	<i>te</i> TE	‘on the other hand’	(Speech)
<i>te</i> TE	<i>mi</i> ASP	<i>masu</i> POL	<i>to</i> COND	‘if (we) try ...’	(Speech)

<sup>35</sup> This means that there is no bundle type in Conversation and Interview that has topic introduction function in a majority of cases (in other words, it has a more predominant function). Therefore, no bundle is classified as topic introduction in these two registers.

The first bundle, *hon + kenkyuu + de + wa* [this + study + LOC + TOP] ‘in this study,’ is used quite regularly in Speech. More specifically, among 108 texts, a quarter (25/108) of them included this bundle. This bundle is used after the presentation of some background information against which the given study is being introduced. The conjunction *sokode* ‘because of that’ or ‘therefore’ frequently precedes the bundle as can be seen in the following excerpts, (4-42) and (4-43).

(4-42) *hon + kenkyuu + de + wa* [this + study + LOC + TOP] ‘in this research’ (Speech: CSJ\_A03M0010)

この文法は可能な構文構造を列挙するというだけで曖昧性解消のモジュールは持っていません/え そこで えー **本研究では** この文法を係り先の候補を絞る為に利用します

*kono bunpoo wa kanoo na koobun koozoo o rekkyo suru to iu dake de aimai sei*  
 this grammar TOP possible COP:ATT construction structure ACC list do QT say only COP ambiguous nature  
*kaishoo no mojuuru wa mot te i mase n / e sokode e: hon kenkyuu de wa kono bunpoo o*  
 resolution GEN module TOP have TE ASP POL NEG / INJ thus INJ **this study LOC TOP** this grammar ACC  
*kakari saki no kooho o shiboru tame ni riyoo shi masu*  
 relation point GEN candidate ACC narrow.down for DAT use do POL

‘This grammar only lists possible construction structures and does not have the module to resolve the ambiguity. Um, therefore, um, **in this study**, we use this grammar to narrow down a candidate for a related element.’

(4-43) *hon + kenkyuu + de + wa* [this + study + LOC + TOP] ‘in this research’ (Speech: CSJ\_A03M0045)

複数の種類のタームの中からえー必要なものを選んで検索する必要があると 考えます/ま そこで **本研究では** え このようなタームの選択手法というのを提案することを目的といたします

*fukusuu no shurui no ta:mu no naka kara e: hitsuyoo na mono o eran de kensaku suru hitsuyoo*  
 several GEN kind GEN term GEN among from INJ necessary COP NML ACC choose TE search do necessity  
*ga aru to kangae masu / ma sokode hon kenkyuu de wa e kono yoo na ta:mu no sentaku shuhoo*  
 NOM exist QT think POL / INJ thus **this study LOC TOP** INJ this like COP term GEN selection method  
*to iu no o teian suru koto o mokuteki to itashi masu*  
 QT say NML ACC propose do NML ACC purpose QT do:HUM POL

‘We think that we need to choose and search the terms that we need from several terms. Um, therefore, **in this study**, our purpose is to propose a method to select such terms.’

The next sub-class of topic introduction bundles in Speech has the form *na + n + desu + keredo/keredomo/ga* [COP:ATT + NML + COP + but], which also appears in Interview and Conversation. However, in the latter two registers, this bundle has the function of presenting background information instead of topic introduction (see Section 4.3.2.3). Below are examples of topic introducing *na + n + desu + ga* [COP:ATT + NML + COP + but] in Speech.

(4-44) *na + n + desu + ga* [COP:ATT + NML + COP + but] ‘as for’ (Speech: CSJ\_A01M0056)

でまず訓練のお話**なんです**が えー生後四歳の雄の日本猿一頭です / んで正刺激に対してボタンを押した場合に報酬として水を与えるということで反応を強化しました

*de mazu kunren no o hanashi na n desu ga e: seigo yon sai no osu no nihon zaru it too*  
and first training GEN PRF talk COP NML COP but INJ after.birth four year GEN male GEN Japan monkey one  
CLS *desu / nde sei shigeki ni taishi te botan o oshi ta baai ni hooshuu to shi te mizu o ataeru to iu*  
COP / and stimulus DAT oppose TE button ACC push PST case DAT reward QT do TE water ACC give QT say  
*koto de hannoo o kyooka shi mashi ta*  
NML COP reaction ACC strengthen do POL PST

‘And, first, **as for** the talk about the training, um, it’s a male four year old Japanese monkey. And we strengthened his reaction by giving him water as a reward when he pushed the button in response to a stimulus.’

(4-45) *na + n + desu + ga* [COP:ATT + NML + COP + but] ‘as for’ (Speech: CSJ\_A07F0399)

で事例の背景と目的**なんです**が えーと 乳脂肪分と無脂乳固形分と殺菌法の効果は食品学的にはある程度もう解明されています / でアメリカでは...

*de jirei no haikai to mokuteki na n desu ga e: to nyuu shiboo bun to mu shi nyuu kokei bun to*  
and case GEN background and purpose COP NML COP but INJ milk fat content and no fat milk solid content and  
*sakkin hoo no kooka wa shokuhin gaku teki ni wa aru teido moo kaimei s are te i*  
sterilize method GEN effect TOP food study perspective DAT TOP exist degree already explicate do PASS TE ASP  
*masu / de amerika de wa ...*  
POL and America LOC TOP

‘And **as for** the background of the case and the purpose, um, the effect of the milk fat, non-fat milk substance and the sterilization method has already been explicated from a sitological perspective. And in America ...’

(4-46) *na + n + desu + ga* [COP:ATT + NML + COP + but] ‘as for’ (Speech: CSJ\_A03F0153)

そして結果 **なんです**が えーと 機械 翻訳 のように大体あの一 言語 処理 学会 で  
は大抵出てくるであろうという ような テーマ については ...

*soshite kekka na n desu ga e:to kikai honyaku no yoo ni daitai ano: gengo shori*  
and result COP NML COP but INJ machine translation GEN like COP generally INJ language management  
*gakkai de wa taitei de te kuru de ar oo to iu yoo na te:ma ni tui te wa ...*  
conference LOC TOP usually appear TE ASP COP exist AUX QT say like COP theme DAT adhere TE TOP

‘And **as for** the result, um, concerning themes which are likely to appear in language management conferences such as machine translation ...’

In the excerpt (4-46), *na + n + desu + ga* [COP:ATT + NML + COP + but] first introduces a major topic, ‘the result of the study,’ and subsequently the topic is narrowed down with *ni + tsui + te + wa* [DAT + adhere + TE +TOP], which is another topic introducing bundle.

The following are examples of *na + n + desu + keredomo* [COP:ATT + NML + COP + but], which also introduces a major topic into the discourse.

(4-47) *na + n + desu + keredomo* [COP:ATT + NML + COP + but] ‘as for’ (Speech: CSJ\_A07F0221)

でこの えーっと 研究 の 目的 **なんです**けれども ちょっと 若干 鉄道 の こと について  
御説明 させていただく こと になります が あの 運転 関係 従事員 と 申しまして  
運転士 や 車掌 さん と いった 方々は

*de kono e:tto kenkyuu no mokuteki na n desu keredomo chotto jyakkan tetsudoo no koto ni tui te*  
and this INJ study GEN purpose COP NML COP but little some railroad GEN NML DAT adhere TE  
*go setsumei s ase te itadaku koto ni nari masu ga ano unten kankei juuji in to mooshi*  
PRF explain do CAUS TE ASP:HUM NML COP become POL but INJ drive relation engage staff QT say:HUM  
*mashi te unten shi ya shashoo san to it ta katagata wa ...*  
POL TE drive person and conductor SFX QT say PST people TOP

‘And **as for** the purpose of this, um, study, we have to explain something about railroad for a moment, um, drivers and conductors, who are called operation engaged staff, ...’

(4-48) *na + n + desu + keredomo* [COP:ATT + NML + COP + but] ‘as for’ (Speech: CSJ\_A07F0445)

で 報告の目的 **なんですけれども** その一牛乳パックのリサイクル行動に影響を与える要因にどんなものがあるのかというのをま探索していくというのが今日の報告の目的となります

*de hookoku no mokuteki na n desu keredomo sono: gyuunyuu pakku no risaikuru koodoo ni*  
and report GEN purpose COP NML COP but INJ milk carton GEN recycle action DAT  
*eikyoo o ataeru yooiin ni donna mono ga aru no ka to iu no o ma kensaku shi te iku*  
influence ACC give factor DAT what.kind NML NOM exist NML Q QT say NML ACC INJ investigate do TE ASP  
*to iu no ga kyoo no hookoku no mokuteki to nari masu*  
QT say NML NOM today GEN report GEN purpose QT become POL

‘And **as for** the purpose of the report, investigating the factors that influence the action of recycling um milk cartons is the purpose of today’s report.’

(4-49) *na + n + desu + keredomo* [COP:ATT + NML + COP + but] ‘as for’ (Speech: CSJ\_A01M0131)

えーまず今回 あのー実験で用いた学習データー **なんですけれども** えー1991年の四月からま98年の九月までのニュースを用いました

*e: mazu konkai ano: jikken de mochii ta gakushuu de:ta: na n desu keredomo e: 1999 nen no shi*  
INJ first this.time INJ experiement LOC use PST learning deta COP NML COP but INJ 1999 year GEN four  
*gatsu kara ma 98 nen no ku gatsu made no nyu:su o mochii mashi ta*  
month from INJ 98 year GEN nine month until GEN news ACC use POL PST

‘Um, first, **as for** the learning data used in the experiment, we used news from April 1999 to September 1998.’

The last sub-class of topic introduction bundles involves so-called phrasal particles. The first set of excerpts includes the bundle *to + shi + mashi + te* [QT + do + POL + TE].

(4-50) *to + shi + mashi + te* [QT + do + POL + TE] ‘regarding’ (Speech: CSJ\_A01M0070)

でえっとー比較手法としまして えーま話者適応手法の一つであります...

*de etto: hikaku shushoo to shi mashi te e: ma washa tekioo shuhoo no hitotsu de ari masu ...*  
and INJ comparison method QT do POL TE INJ INJ speaker adaptation method GEN one COP exist POL

‘And um, **regarding** a comparison method, um, one of the methods for speaker adaptation ...’

(4-51) *to + shi + mashi + te* [QT + do + POL + TE] ‘regarding’ (Speech: CSJ\_A01M0147)

えーとよく知られてる方法としましては えーと 深林先生によって提案されております

*e:to yoku shir are te ru hoohoo to shi mashi te wa e:to fukabayashi sensei ni yot te teian s are*  
INJ often know PASS TE ASP method **QT do POL TE TOP INJ** (name) teacher DAT owe TE propose do PASS  
*te ori masu ...*  
TE ASP:HUM POL

‘Um, **regarding** a well known method, um, the one that is proposed by Professor Fukagawa ...’

(4-52) *to + shi + mashi + te* [QT + do + POL + TE] ‘regarding’ (Speech: CSJ\_A03F0109)

で本発表の構成としまして えー二つう一話があります / ます ...

*de hon happyoo no koosei to shi mashi te e: futatsu u: hanashi ga ari masu / mazu ...*  
and this presentation GEN structure **QT do POL TE** um two INJ talk NOM exist POL first

‘And, **regarding** the structure of this presentation, there are two talks. First ...’

The second set of excerpts includes the bundle *ni + tsui + te + wa* [DAT + adhere + TE + TOP]. As we saw above, *ni + tsui + te + wa* is often used to introduce a narrower or more focused topic than *na + n + desu + keredo/keredomo/ga* [COP:ATT + NML + COP + but] type.

(4-53) *ni + tsui + te + wa* [DAT + adhere + TE + TOP] ‘concerning’ (Speech: CSJ\_A03F0153)

そして結果 なんですが えーと 機械翻訳のように大体あの一言語処理学会では大抵出てくるであろうというようなテーマについては ...

*soshite kekka na n desu ga e:to kikai honyaku no yoo ni daitai ano: gengo shori*  
and result COP NML COP but INJ machine translation GEN like COP generally INJ language management  
*gakkai de wa taitei de te kuru de ar oo to iu yoo na te:ma ni tsui te wa ...*  
conference LOC TOP usually appear TE ASP COP exist AUX QT say like COP theme **DAT adhere TE TOP**

‘And **concerning** the result, um, concerning themes which are likely to appear in language management conferences such as machine translation ...’

(4-54) *ni + tsui + te + wa* [DAT + adhere + TE + TOP] ‘concerning’ (Speech: CSJ\_A03M0005)

で あのー 今日のお話の目的 なんですけれども ... というようなことを探りたい  
と思ひまして えーま特に今回 については あー ... というようなことについて  
ちょっと実験結果等を交えてえお話ししたいと思ひます

*de ano: kyoo no o hanashi no mokuteki na n desu keredomo ... to iu yoo na koto o saguri tai to*  
and INJ today GEN PRF talk GEN purpoe COP NML COP but ... QT say like COP NML ACC explore QT  
*omoi mashi te e: ma tokuni konkai ni tsui te wa a: ... to iu yoo na koto ni tsui te chotto*  
think POL TE INJ especially this.time **DAT adhere TE TOP** INJ ... QT say like COP NML DAT adhere TE little  
*jikken kekka too o majie te e ohanashi shi tai to omoi masu*  
experiment result such ACC mix TE INJ talk do want QT think POL

‘And, um, **concerning** the purpose of today’s talk, we’d like to explore ..., um, concerning this time in particular, we’d like to talk about ... including the result of experiments.’

The third set of excerpts includes the bundle *sore + ni + taishi + te* [that + DAT + oppose + TE]. While all the topic-related bundles presented above are topic-introduction bundles, this bundle is a topic-elaboration bundle. The bundle is used to present a contrastive idea. In (4-55), two novels by the same author, Natsume Sooseki, are being contrasted. In (4-56), the speaker contrasts his own study with a previous study by other researchers.

(4-55) *sore + ni + taishi + te* [that + DAT + oppose + TE] ‘on the other hand’ (Speech: CSJ\_A08F0686)

戦前の中等学校国語教科書においては「草枕」に次いで「我が輩は猫である」  
の採録が多かったのです/それに対して「坊っちゃん」はわずかに昭和十年  
代に一種の教科書でしか採録されていませんでした

*sen zen no chuutoo gakoo kokugo kyookasho ni oi te wa “kusamakura” ni tsui de “wa-ga-hai-*  
war before GEN middle school national.language textbook DAT put TE TOP (name) DAT follow TE (name)  
*wa-neko-de-aru” no sairoku ga ookat ta no desu / sore ni taishi te “botchan” wa wazuka ni*  
GEN record NOM many PST NML COP / **that DAT oppose TE** (name) TOP little COP:ADV  
*shoowa juu nen dai ni isshu no kyookasho de shika sairoku s are te i mase n deshi ta.*  
Showa ten year time DAT one kind GEN textbook LOC only record do PASS TE ASP POL NEG COP PST

‘In middle school national language textbooks before the war, “I am a cat” was the most frequently recorded novel following “Kusamakura.” **On the other hand**, “Botchan” was only recorded in one textbook during the first decade of the Showa period.’

(4-56) *sore + ni + taishi + te* [that + DAT + oppose + TE] ‘on the other hand’ (Speech: CSJ\_A03M0061)

特に え 系統的な 与え方 は えー この ど-む- 研究では え 考えられて いません / え **それに対して** 本研究では ここで え 示している ような 素性構造 指標 という えー ものを 用いて え データーの 特徴を 系統的に 捉える ことを 考えます

*tokuni e tookei teki na atae kata wa e: kono do- mu- kenkyuu de wa e kangae rare te i mase n /*  
especially INJ systematic like COP give way TOP INJ this study LOC TOP think PASS TE ASP POL NEG  
*e sore ni taishi te hon kenkyuu de wa koko de e shimeshi te iru yoo na sujoo koozoo shihyoo to*  
INJ that DAT oppose TE this study LOC TOP here LOC INJ indicate TE ASP like COP origin structure index QT  
*iu e: mono o mochii te e de:ta: no tokuchoo o teitooteki teki ni toraeru koto o kangae masu*  
say INJ NML ACC use TE INJ data GEN characteristic ACC systemaic like COP grasp NML ACC think POL

‘In particular, this study does not consider a systematic way of giving. Um, **on the other hand**, in this (our) study, we use what is called an original structure index as shown here, and consider grasping the characteristics of the data systematically.’

The last topic introduction bundle is *te + mi + masu + to* [TE + ASP + POL + COND] ‘if (we) try and ...’ The majority of the bundles are preceded by the verb *miru* ‘look,’ as shown in the following excerpts. This morphemic sequence is similar to the English lexical bundle “if you look at.” However, the morphemic sequence does not include a second person pronoun and hence may not invite the addressee’s participation like the English equivalent bundle does (see Biber et al. 2004: 392-393 for the analysis of the lexical bundle “if you look at”).

(4-57) *te + mi + masu + to* [TE + ASP + POL + COND] ‘if (we) try’ (Speech: CSJ\_A07F0445)

コストと牛乳パックのリサイクル行動の関係を 見てみますと やはり コスト感  
が高い 人程 リサイクルを 実行しない とか 合理的な 考え方を する 人程 リサイ  
クルを 実行しない という ような えー 結果が 得られました

*kosuto to gyuunyuu paku no risaikuru koodoo no kankei o mi te mi masu to yahari kosuto kan*  
cost and milk carton GEN recycle behavior GEN relation ACC look **TE ASP POL COND** as.expected cost feeling  
*ga takai hito hodo risaikuru o jikkoo shi nai toka goori teki na kangae kata o suru hito*  
NOM high person degree recycle ACC practice do NEG such rationality like COP thought way ACC do person  
*hodo risaikuru o jikkoo shi nai to iu yoo na e: kekka ga e rare mashi ta*  
degree recycle ACC pracitce do NEG QT say like COP INJ result NOM get PASS POL PST

‘If (we) try looking at the relationship between the cost and the milk carton recycling behavior, as we expected, we got the result that higher they feel the recycling cost is and more rational their way of thinking is, they do not practice recycling.’

(4-58) *te + mi + masu + to* [TE + ASP + POL + COND] ‘if (we) try’ (Speech: CSJ\_A08F0552)

五年生の十月の詩の単元を見てみますと 学習活動としてまず 秋の山遊びをする となっており 経験主義的であることが特徴となっています

*go nen sei no juu gatsu no shi no tange n o mi te mi masu to gakushuu katsudoo to shi te*  
five year student GEN ten month GEN poem GEN unit ACC look TE ASP POL COND learning activity QT do TE  
*mazu aki no yama asobi o suru to nat te ori keiken shugi teki de aru koto ga*  
first autumn GEN mountain play ACC do QT become TE ASP experience principle like COP exist NML NOM  
*tokuchoo to nat te i masu*  
characteristic QT become TE ASP POL

‘If (we) try looking at the fifth grade’s October poetry unit, it has playing in an autumn mountain as a learning activity, thus empiricism is its characteristics.’

In the excerpts above, the sequence, *o + mi + te + mi + masu + to* [ACC + look + TE + ASP + POL + COND] ‘if (we) try looking at,’ introduces a topic, that is, ‘the relationship between the cost and the milk carton recycling behavior’ in (4-57), and ‘the fifth grade’s October poetry unit’ in (4-58).

#### 4.3.2.2. Cohesion bundles

Cohesion bundles have the function of relating a clause to other parts of a discourse. They are associated with the bundle final sequence *n + da/desu* [NML + COP]. Iwasaki (1985) attributes the cohesive power of *n + da/desu* clause to the ability of *n + da/desu* “to mark a proposition or information as unchallengeable due to the structural characteristics it shares with a relative clause construction” (p.134). The cohesion bundles are mostly found in Conversation and Interview, but they also appear in Speech. In a majority of cases, however, *n + da/desu* [NML + COP] is further followed by the conjunctive particle *keredo* ‘but’ or its variants, and in a

fewer cases, by the pragmatic particles *yo* or *ne*, or the quotative particle *to/tte*. To explicate the function of the cohesion bundles, only the bundles that are not followed by other particles (though they are extremely rare) are considered in this section. The following are the most frequent cohesion bundles.

(4-59) Cohesion bundles

*te ru n da/desu* <cohesion> (Conversation, Interview and Speech)  
TE ASP NML COP

*te ta n da/desu* <cohesion> (Conversation and Interview)  
TE PST NML COP

*dat ta n da/desu* <cohesion> (Conversation and Interview)  
COP PST NML COP

The next excerpt from Interview, (4-60), includes the bundles *te + ta + n + desu* [TE + PST + NML + COP] and *te + ru + n + desu* [TE + ASP + NML + COP]. It also includes other instances of the sequence *n + desu* [NML + COP] though they are not part of the bundles.

(4-60) *te + ta + n + desu* [TE + PST + NML + COP] and *te + ru + n + desu* [TE + ASP + NML + COP] <cohesion> (Interview: Tetsuko\_4)

1. I: あのー歌を歌うという番組がえーお正月に放送されるのが
2. K: うん。
3. I: 年末に作られてたんです。でその録画が終わった後に忘年会も兼ねてみんなでパーティーをやるんですね。でその時に全員立食で立ってるんです。
4. K: ええ。
5. I: それで一つだけテーブルがあって美空ひばりさんとえー会社の社長とそれからお母様とひばりさんの3人だけが着席なんですねテーブルに。
6. K: へー。
7. I: でその前にあのーその年の新人さんが並んでご挨拶をするんです。
8. K: ええ。
9. I: でまああのー今年何々でデビューをしたまあ「かくれんぼ」でデビューした石川さゆりですって...

1. I: *ano: uta o utau to iu bangumi ga e: o shoogatsu ni hoosoo s are ru no ga*  
INJ song ACC sing QT say program NOM INJ PRF new.year DAT broadcast do PASS NML NOM
  2. K: *un.*  
yes
  3. I: *nenmatsu ni tsukur are te ta n desu. de sono rokuga ga owat ta ato ni boonen kai mo*  
year.end DAT make PASS **TE PST NML COP** / and that record NOM end after DAT year.end party also  
*kane te minna de pa:ti: o yaru n desu ne. de sono toki ni zenin risshoku de*  
combine TE everyone by party ACC do NML COP PP / and that time DAT everyone stand-up.meal COP  
*tat te ru n desu.*  
stand.up **TE ASP NML COP**
  4. K: *ee.*  
yes
  5. I: *sorede hitotu dake te:buru ga at te misora hibari san to e: kaisha no shachoo to sorekara*  
and one only table NOM exist TE (name) and INJ company GEN president and and  
*okaasama to hibari san no san nin dake ga chakuseki na n desu ne te:buru ni*  
mother and (name) HON GEN three CLS only NOM seat COP NML COP PP table
  6. K: *he:.*  
INJ
  7. I: *de sono mae ni ano: sono toshi no shinjin san ga naran de go aisatsu o suru n desu.*  
and that before DAT INJ that year GEN new.face HON NOM line TE HON greeting ACC do NML COP
  8. K: *ee.*  
yes
  9. I: *de ma ano: kotoshi naninani de debyu: o shi ta maa "kakurenbo" de debyu shi ta*  
and INJ INJ this.year what.what TE debut ACC do PST INJ (name) TE debut do PST  
*ishikawa sayuri desu tte ...*  
(name) COP QT
- 
1. I: 'Um, a TV program in which singers sing their songs was being produced at the end of the year, and it was to be broadcasted in the New Year'
  2. K: Yes.
  3. I: 'And after that recording was done, we had a party with everyone which combined the the year end party. And at that time, since it was a stand-up meal, people are standing.'
  4. K: Yes.
  5. I: 'And there was only one table, and only Hibari Misora, the company president and his mother, three of them were being seated, at the table.'
  6. K: Really.
  7. I: 'And in front of them, that year's new faces line up and greet them.'
  8. K: Yes.
  9. I: 'And um, um, I said "my name is Sayuri Ishikawa, and I made my debut with blank, um, "Kakurenbo" ...'

In the above excerpt, (4-60), the interviewee, who is a well-known singer, recalls an end-of-the-year party in which she participated when she was a newly debuted singer. We see the use of *te + ta + n + desu* [TE + PST + NML + COP] and *te + ru + n + desu* [TE + ASP + NML + COP]

in line 3. We can also notice that she ends other utterances with *n + desu* [NML + COP] until just before line 9 in which she gives the main story of what happened at the party. By using the cohesion bundles, the speaker projects further talk to which the current utterance with the cohesion bundle is related.

Prior to the following excerpt, (4-61), the interviewee O talked about his volunteer experience in the United States where he studied. Then the interviewer N asked him if he also had a volunteer experience in Japan.

(4-61) *te + ru + n + desu* [TE + ASP + NML + COP] <cohesion> (Interview: Hypermedia\_009)

1. O: こっちでは
2. N: うん
3. O: あの公民館ですか地域の
4. N: はいはい
5. O: 家の近くに公民館があるんですけど
6. N: はい
7. O: そこにあ日本在住の外国人の方々が
8. N: うん
9. O: 大勢いるんで
10. N: はい
11. O: その方々があのお友達を作りたいって言って
12. N: うん
13. O: 公民館に集まってくるんですよね。
14. N: ええ
15. O: でそういう人と一緒に日本人もその外国人同士も交流しようっていうことで
16. N: はあー
17. O: そういうサロンを開いてるんです。
18. N: はい
19. O: そこにまボランティアとしてって行って
20. N: うん

1. O: *kotchi de wa*  
here LOC TOP
2. N: *un*  
yes
3. O: *ano koominkan desu ka chiiki no*

- INJ community.center COP Q district GEN
4. N: **hai hai**  
yes yes
5. O: *uchi no chikaku ni koominkan ga aru n desu kedo*  
home GEN near LOC community.center NOM exist NML COP but
6. N: *hai*  
yes
7. O: *soko ni a nihon zaijuu no gaikokujin no katagata ga*  
there LOC Japan living GEN foreigner GEN people NOM
8. N: *un*  
yes
9. O: *oozei iru n de*  
many exist NML COP
10. N: *hai*  
yes
11. O: *sono katagata ga ano o tomodashi o tsukuri tai tte it te*  
that people NOM INJ HON PRF friend ACC make want QT say TE
12. N: *un*  
yes
13. O: *koominkan ni atsumat te kuru n desu yo ne.*  
community.center DAT gather TE ASP NML COP PP PP
14. N: *ee*  
yes
15. O: *de soo iu hito tachi to issho ni nihon jin mo sono gaikokujin dooshi mo kooryuu shiy oo*  
and that say person PL with together DAT Japan person also INJ foreigner fellow also exchange do VOL  
*tte iu koto de*  
QT say NML COP
16. N: *ha:*  
INJ
17. O: *soo iu saron o hirai te ru n desu.*  
that say salon ACC open **TE ASP NML COP**
18. N: *hai*  
yes
19. O: *soko ni ma Oorantia to shi te tte- it te*  
there DAT INJ volunteer QT do TE go TE
20. N: *un*  
yes

1. O: 'Here'
2. N: 'Yes.'
3. O: 'Um, is it a community center, in the local area'
4. N: 'Yes yes.'
5. O: 'There is a community center close to my house'
6. N: 'Yes.'
7. O: 'There, um, foreigners living in Japan'
8. N: 'Yes.'
9. O: 'There are many of them'
10. N: 'Yes.'
11. O: 'Those people said they wanted to make friends'
12. N: 'Yes.'

13. O: 'They gather at the community center.'  
 14. N: 'Yes.'  
 15. O: 'And to exchange with those people, both with Japanese and among foreigners themselves'  
 16. N: 'Uh huh.'  
 17. O: 'They are opening such salon.'  
 18. N: 'Yes.'  
 19. O: 'Um, so I go there as a volunteer'  
 20. N: 'Yes.'

In line 17, O uses the cohesion bundle to indicate that the utterance is related to what follows.

In the subsequent utterance, O talks about an activity that directly addresses N's inquiry about O's volunteer experience in Japan. From the addressee's perspective, upon receiving an utterance with the cohesion bundle, he/she searches the surrounding discourse to which the current utterance is being related. If the relevance of the information has not been justified at the point of the utterance (e.g., answering a question, providing an account), the addressee expects the speaker to continue his/her talk. In the excerpt above, information given in line 19 not only answers N's question but it also provides a justification for the relevance of the information given in the preceding context with the cohesion bundle. Notice that two other utterances by O in lines 5 and 13 also include the sequence *n + desu* [NML + COP]. It suggests that the cohesion bundles themselves are not formulaic per se, but they include a fixed chunk, *n + desu* [NML + COP].

In Speech, the cohesion bundles are used mostly to signal that the information is related to the previously given information. In the next excerpt (4-62), for example, the speaker first says that children these days have a triangular face shape compared to older generation's pentagonal face shape like the baseball home base. He then sums up the information by saying *tsumari ago ga hosot te ki te ru n desu* 'in short, their jaw is getting smaller (thinner)' with the cohesion bundle *te + ru + n + desu* [TE + ASP + NML + COP].

(4-62) *te + ru + n + desu* [TE + ASP + NML + COP] <cohesion> (Speech: CSJ\_A05M0040)

昔の顔というのは例えばホームベースというか五角形で書けたんですね/今  
そうじゃないです/三角形です/よく見てください/最近のあのー小顔といわ  
れてるようなか人達含めてですねえ三角形つまり顎が細ってきてるんで  
す

*mukashi no kao to iu no wa tatoeba ho:mu be:su to iu ka gokakukei de kak e ta n desu ne. ima*  
old.time GEN face QT say NML TOP for.example home base QT say Q pentagon by draw POT PST COP PP / now  
*soo ja nai desu. sankakukei desu. yoku mi te kudasai. saikin ano: ko gao to iw are te ru yoo na*  
that COP NEG COP / triangle COP / well look TE please / recent INJ small face Q say PASS TE ASP like COP  
*ka- hitotachi fukume te desu ne e: sankakukei tsumari ago ga hosot te ki te ru n desu.*  
people include TE COP PP INJ triangle in.short jaw NOM narrow TE ASP TE ASP NML COP

‘We could draw the older faces with pentagonal shape like a home base. Now, it is not. It is triangle. Please take a look carefully. Recently, including those people who are called small faces, their faces are triangle; in short their jaw is getting smaller.’

In the next excerpt, (4-63), the speaker gives a long utterance that begins with *chuumoku su beki koto wa* ‘what is notable is’ and ends with the cohesion bundle *koto + na + n + desu* [NML + COP:ATT + NML COP]. The use of the cohesion bundle, together with the use of the nominalizer *koto*, which also appears at the beginning of the utterance, function to indicate that the information identifies what *chuumoku su beki koto* ‘noteworthy fact’ is.

(4-63) *koto + na + n + desu* [NML + COP:ATT + NML + COP] <cohesion> (Speech: CSJ\_A08F0709)

で注目すべきことはそのす-う-次なんですけれども今日は粘土中止ま今日は  
ペンキ中止などというラベルでその注意事項を知るように子供達の習慣をつ  
-作るということなんです

*de chuumoku su beki koto wa sono su- u- tsugi na n desu keresomo kyoo wa nendo chuushi ma*  
and notable do should NML TOP INJ next COP NML COP but today TOP clay TOP cancel INJ  
*kyoo wa penki chuushi nado to iu raberu de sono chuui jikoo o shiru yoo ni kodomo tachi no*  
today TOP paint cancel such QT say label INS that notice attention matter ACC know like COP child PL GEN  
*shuukan o tsukuru to iu koto na n desu.*  
habit ACC make QT say NML COP NML COP

‘And what is notable is, it is the next one, they make children develop the habit of finding out matters that require their attention from the labels such as “clay is canceled today” and “paint is canceled today.”

The next set of excerpts is from Conversation. Prior to the segment shown in (4-64), two friends were talking about the difference between people living in Tokyo and in Osaka. F114 told F147 if some stranger seemed lost at the station, people in Osaka would talk to the person. F147 asked whether age was not a factor, and F147 initially said younger people might not talk to a stranger like that. Then, as seen in the next segment, F114 recalls her encounter with a relatively young woman at a station in Osaka.

(4-64) *te + ta + n + da* [TE + PST + NML + COP] <cohesion> (Conversation: BTS\_65)

1. F114: 若い人でもね一回あたしがねこの急行はねなんか枚方駅止まる  
 んですかみたいなことを聞いてんであ止まりますよって行って  
 乗ってでなんか結構若い女の人に聞いてで乗ったの。で枚方に  
 着く直前にここ枚方ですよってまた話しかけてくれたんだ。あた  
 し分かってたんだちゃんと枚方だなって。でもわざわざ行って  
 くれたからすごい親切だなと思った。
2. F147: いいね。気持ちいいね。
1. F114: *wakai hito de mo ne ik kai atashi ga ne kono kyuukoo wa ne nanka maikata eki*  
 young person COP also PP one time I NOM PP this express TOP PP DM (name) station  
*tomaru n desu ka mitai na koto o kii te nde a tomari masu yo tte it te not te e*  
 stop NML COP Q like COP NML ACC ask TE and INJ stop POL PP QT say TE ride TE INJ  
*nanka kekkoo wakai onna no hito ni kii te de not ta no. de maikata ni tsuku*  
 DM relatively young woman GEN person DAT ask TE and ride PST PP / and (name) DAT arrive  
*chokuzen ni koko maikata desu yo tte mata hanashi kake te kure ta n da. atashi*  
 right.before DAT here (name) COP PP QT again talk ASP TE ASP PST NML COP / I  
*wakat te ta n da chanto maikata da na tte. demo wazawaza it te kure ta kara*  
 know TE PST NML COP surely (name) COP PP QT / but specially say TE ASP PST because  
*sugoi shinsetsu da na to omot ta.*  
 very kind COP PP QT think PST
2. F147: *ii ne. kimochi ii ne.*  
 good PP / feeling good PP
1. F114: ‘Even young people, one time, I asked a relatively young woman, “does this  
 express train stop at Maikata station,” and she told me “It stops there,” and I got  
 on. And right before the train arrived at Maikata, she talked to me and said “this

is Maikata.” I knew surely that that it was Maikata. But I thought she was really kind to take the trouble of telling me.’

2. F147: ‘That’s good. That makes one feel good.’

At the beginning, the speaker uses *te*-form, as in *kii te* ‘asked,’ *it te* ‘said,’ *not te* ‘got on.’ She then switches to the *n + da* forms, as in *hanashi kake te kure ta n da* ‘(she) kindly spoke to me’ and *wakat te ta n da* ‘(I) knew.’ By using *n + desu* endings, the speaker explicitly indicates that the information is related to some other part of discourse, in this case, what follows. From the addressee’s perspective, it creates a sense of suspension in that the addressee must wait to hear additional utterance(s) to understand the relevance of the *n + desu* attached information. In the immediately subsequent utterance, the speaker indicates what she thought of the woman, *sugoi shinsetsu da na to omot ta* ‘(I) thought (the woman) was very kind.’ Based on the preceding context, the core message the speaker wanted to express was that young woman was kind. Upon hearing this main message, which also justifies the relevance of the *n + desu* attached information, the addressee F147 responds with an aligned assessment.

The next excerpt, (4-65), portrays the conversational use of a cohesion bundle. A and B are talking about their mutual acquaintance, Shannon. In line 5, B reports that Shannon appeared in the show called Beauty and the Beast at the Disneyland. In the next turn, A says *a mada yat te ta n da are* ‘oh, that was still playing,’ displaying that she received a new piece of information and that she was surprised to find out that the show was still playing. The cohesion bundle or the sequence *n + da* in this utterance does not simply indicate that the information is linked to the surrounding discourse. Instead, it signals the speaker’s attitude that she expects further explanation from B about the show still being played. This interpretation is only appropriate because the *n + da* attached information is something A is unfamiliar with and B is familiar with.

After A's *n + da* utterance, B confirms the fact and provides further explanation that the show was still playing when she went there.

(4-65) *te + ta + n + da* [TE + PST + NML + COP] <cohesion> (Conversation: Callhome\_2236)

1. B: ディズニーでポールと
2. A: うん。
3. B: なんだっけ？ シャノンだっけ？
4. A: うん うん うん うん。
5. B: シャノンが Beauty and Beast に出た。
6. A: あっまだやっ **て た ん だ** あれ！
7. B: うん。あたしが行った時にはまだやっ **て**。でもシャノンと特別親しくしてる訳じゃないから ...

1. B: *dizuni: de po:ru to*  
disney LOC paul with
2. A: *un.*  
yes
3. B: *nan dak ke? shanon dak ke?*  
what COP Q / (name) COP Q
4. A: *un un un un.*  
yes yes yes yes
5. B: *shanon ga "beauty and beast" ni de te ta.*  
(name) NOM (name) DAT appear TE ASP
6. A: *a mada yat te ta n da are!*  
INJ still do **TE PST NML COP** that
7. B: *un. mada atashi ga it ta toki ni wa mada yat te te. demo shanon to tokubetsu shitashiku*  
yes still I NOM go PST when DAT TOP still do TE ASP / but (name) with particularly friendly  
*shi te ru wake ja nai kara ...*  
do TE ASP reason COP NEG because

1. B: 'At Disneyland, with Paul'
2. A: 'Yeah.'
3. B: 'What is it? Is it Shannon?'
4. A: 'Yeah yeah yeah yeah.'
5. B: 'Shannon was in "Beauty and the Beast."'
6. A: 'Oh, that was still playing!'
7. B: 'Yeah. When I went there, it was still playing. But I'm not that close to Shannon, so ...'

### 4.3.2.3. Background bundles

Background bundles are also found in all three registers, and they include the sequence *n* + *da/desu* [NML + COP] followed by the concessive conjunctive particle *kedo* ‘but’ or its variants (i.e., *keredo*, *keredomo*, *ga*) at the end of the sequence. As I mentioned in the previous section, a majority of the cohesion bundles ending in *n+da/desu* are actually followed by the particle *kedo* ‘but’ or its variants (see Section 5.3 for statistical measures of collocational strength between *n +da/desu* and *kedo*-type conjunctive particles). There are a few sub-types of the background bundles, but the main function is to present background information and indicate that the information is related to some other part of discourse (whether explicit or implicit) which conveys the main or overriding message of the speaker. In other words, the background bundles present pragmatically or interactionally subordinate information. Below are some of the most frequent background bundles.

#### (4-66) Background bundles

<i>ta</i>	<i>n</i>	<i>da/desu</i>	<i>kedo</i>	<background> (Conversation and Interview)
PST	NML	COP	but	
<i>nai</i>	<i>n</i>	<i>da/desu</i>	<i>kedo</i>	<background> (Conversation and Interview)
NEG	NML	COP	but	
<i>omou</i>	<i>n</i>	<i>da/desu</i>	<i>kedo</i>	<background> (Conversation and Interview)
think	NML	COP	but	
<i>na</i>	<i>n</i>	<i>da/desu</i>	<i>kedo</i>	<background> (Conversation and Interview)
COP:ATT	NML	COP	but	
<i>ta</i>	<i>n</i>	<i>desu</i>	<i>keredomo</i>	<background> (Interview and Speech)
PST	NML	COP	but	
<i>ta</i>	<i>n</i>	<i>desu</i>	<i>ga</i>	<background> (Speech)
PST	NML	COP	but	

In Speech, some of the background bundles convey the contrastive meaning which derives from the meaning of the conjunctive particle *keredomo* ‘but’ or its variants. In the next excerpts (4-67) and (4-68), the bundle clearly conveys the meaning of ‘although.’

(4-67) *ta + n + desu + keredomo* [PST + NML COP + but] <background/contrastive> (Speech: CSJ\_A04M0026)

今回でも色々なことが明らかにはなったな**ったんですけれども** えこの点に  
ち- に関してはですねまだあー私共も経験が浅いので えー引き続き えー 検討課  
題として えー いきたいと考えております

*konkai de mo iroiro na koto ga akiraka ni wa nat ta nat ta n desu keredomo e kono ten*  
this.time COP also various COP NML NOM clear COP become PST become PST NML COP but INJ this point  
*ni chi- kanshi te wa desu ne mada a: watashi domo mo keiken ga asai no de e: hikitsuzuki e:*  
DAT relate TE TOP COP PP still INJ I PL also experience NOM shallow NML COP INJ continuously INJ  
*kentoo kadai to shi te e: iki tai to kangae te ori masu.*  
consideration issue QT do TE INJ go want QT think TE ASP POL

‘Although various things became clear this time, regarding this point, because our experience is still limited, we’d like to take it as an issue to consider continuously.’

(4-68) *ta + n + desu + ga* [PST + NML COP + but] <background/contrastive> (Speech: CSJ\_A07F0844)

二千百もあのサンプルがあ**ったんですが** その中で四百四十九しか回収され  
てないという恐ろしい状況になりました

*ni sen hyaku mo ano sanpuru ga at ta n desu ga sono naka de yon hyaku yonjyuu kyuu*  
two thousand hundred also INJ sample NOM exist **PST NML COP but** that among LOC four hundred forty nine  
*shika kaishuu s are te nai to iu osoroshii jookyoo ni nari mashi ta*  
only collect do PASS TE NEG QT say terrible situation COP become POL PST

‘Although there were as many as two thousand one hundred samples, we have a terrible situation that, only four hundred forty nine of those were collected.’

Other instances of the background bundles only have the function of presenting background information in Speech. In some cases, the background information is related to the context of the speech environment, as in (4-69), and in others, the background information is related to the context of the study being reported, as in (4-70).

(4-69) *ta + n + desu + keredomo* [PST + NML COP + but] <background/background> (Speech: CSJ\_A05F0154)

最初にお話をいただいた時にまそういう風にお断わりしたんですけれどもあの一ま 日本語の えー ネイティブスピーカーとしての発言も欲しいということで えーそういう立場から少し えー発言させていただきます

*saisho ni o hanashi o itadai ta toki ni ma soo iu fuu ni o kotowari shi ta n desu keredomo ano*  
first DAT PRF talk ACC receive PST when DAT INJ that say way DAT decline do **PST NML COP but** INJ  
*ma nihongo no e: neitibu supi:ka: to shi te no hatsugen mo hoshii to iu koto de e: soo iu tachiba*  
INJ Japanese GEN INJ native speaker QT do TE GEN utterance also want QT say NML COP INJ that say position  
*kara sukoshi e: hatsugen s ase te itadaki masu.*  
from little INJ utterance do CAUS TE ASP POL

‘When I was first given the opportunity, I declined in such a way. Since they also want talks from native Japanese speakers’ perspective, I will talk a little from such position.’

(4-70) *ta + n + desu + keredomo* [PST + NML COP + but] <background/background> (Speech: CSJ\_A07F0366)

この理由一からえーと八のものがあつたんですけれどもこの理由にもタイプ別に分かれるものがあるのではないかという形で ちょっと 見てみました

*kono riyuu ichi kara e: to hachi no mono ga at ta n desu keredomo kono riyuu ni mo taipu betsu*  
this reason one from INJ eight GEN NML NOM exist **PST NML COP but** this reason DAT also type according.to  
*ni wakareru mono ga aru no de wa nai ka to iu katachi de chotto mi te mi mashi ta.*  
COP divide NML NOM exist NML COP TOP NEG Q QT say NML COP little look TE ASP POL PST

‘There were reasons from one to, um, eight, and we took a little look thinking that there may be ones that can be divided according to types.’

In Interview, some of the background bundles carry the contrastive meaning (excerpt (4-71)) and some only have the function of presenting background information (excerpt (4-72)). In addition, there are some bundles that have an interactional function (excerpts (4-73) and (4-74)). The interactional function is related to offering interactional options to the addressee (Haugh 2008). That is, the speaker gives the addressee a primary right to interpret the particular relevance of the information in the current discourse context. This interactional function with the use of the background bundle is not found in Speech.

(4-71) *ta + n + desu + kedo* [PST + NML COP + but] <background/contrastive> (Interview: Tetsuko\_2)

1. M: 最初はナンパだったんですけどもちろん最初は声かけた時はそういう気持ちだったんですけどやっぱ歌を聞いてそのガラッと変わっちゃいましたねその彼女に対する気持ちとどうか
2. T: なるほどね。

1. M: *saisho wa nanpa dat ta n desu kedo mochiron saisho wa koe kake ta toki wa soo iu*  
first TOP girl.hunting COP PST COP but of.course first TOP voice call PST when TOP that say  
*kimochi dat ta n desu kedo yappa uta o kii te sono garat to kawat chai mashi ta ne sono*  
feeling COP PST NML COP but surely song ACC listen TE INJ ONM QT change ASP POL PST PP INJ  
*kanojo ni taisuru kimochi to iu ka*  
she DAT face feeling QT say Q
2. T: *naruhodo ne.*  
indeed PP

1. M: ‘**Although** it was a girl hunting at first, of course I had that kind of mind set when I spoke to her first, but my feeling toward her completely changed after listening to her singing.’
2. T: ‘I see.’

(4-72) *ta + n + desu + kedo* [PST + NML COP + but] <background/background> (Interview: CSJ\_D01F0023)

1. L: えーっと一海外旅行に行ったのが中国が初めてっていうことだったんですけど
2. R: はい
3. L: わりところ最近若いうちにいっぱい色んなところ行くじゃないですか
4. R: うんうん はいはい
5. L: なぜな-どこにも行かなかったんですか

1. L: *e:tto: kaigai ryokoo ni it ta no ga chuugoku ga hajimete tte iu koto dat ta n desu kedo*  
INJ overseas travel DAT go PST NML NOM China NOM first QT say NML COP PST NML COP but
2. R: *hai*  
yes
3. L: *warito koo saikin wakai uchi ni ippai ironna toko iku ja nai desu ka*  
comparatively this recent young while DAT many various place go COP NEG COP Q
4. R: *un un hai hai*  
yes yes yes yes
5. L: *naze na- doko ni mo ika nakat ta n desu ka.*  
why where DAT also go PST PST NML COP Q

1. L: ‘U:m, you said China was your first travel to foreign countries’
2. R: ‘Yes.’

3. L: ‘Comparatively people go to many different places while they are young recently, you know.’
4. R: ‘Yeah yeah, yes yes.’
5. L: ‘Why didn’t you go anywhere?’

In the following two excerpts, the bundle *ta + n + desu + kedo* [PST + NML COP + but] has the function of giving the addressee an interactional option of how to treat and react to the information provided with the bundle. In the first excerpt (4-73) below, the interviewer K assumes that the interviewee S was born in Ooita, but the interviewee S corrects that information, saying *chuugoku de umare ta n desu kedo* ‘I was born in China.’ Here, the speaker is using the bundle to give the interviewer an option of how to interpret the relevance of the information and how to respond to the utterance.

(4-73) *ta + n + desu + kedo* [PST + NML COP + but] <background/interactional option>  
(Interview: Tetsuko\_6)

1. K: でその九州の大分のお生まれ
  2. S: 中国で生まれたんですけど
  3. K: あお生まれは中国なんですか！
  4. S: そうなんです。
  5. K: あそうなんだ。
  6. S: で父の里が大分だったので大分に行って小学校に上がる前に滋賀県に寄って母方の所へ行って中学まで。
- 
1. K: *de sono kyuushuu no ooita no o umare*  
and that (name) GEN (name) GEN HON birth
  2. S: *chuugoku de umare **ta n desu kedo***  
China LOC born **PST NML COP but**
  3. K: *a o umare wa chuugoku na n desu ka !*  
INJ HON birth TOP China COP NML COP Q
  4. S: *soo na n desu.*  
that COP NML COP
  5. K: *a soo na n da.*  
INJ that COP NML COP
  6. S: *de chichi no sato ga ooita dat ta node ooita ni it te shoogakkoo ni*  
and father GEN hometown NOM (name) COP PST because (name) DAT go TE elementary.school DAT  
*agaru mae ni shiga ken ni yot te haha gata no tokoro e it te chuugaku made.*  
go before DAT (name) prefecture DAT go TE mother side GEN place ALL go TE middle.school until

1. K: 'And, um, you were born in Ooita in Kyushu'
2. S: 'I was born in China'
3. K: 'Oh, you were born in China!'
4. S: 'That's right.'
5. K: 'Oh, I see.'
6. S: 'And, my father's hometown was Ooita, so I went there, and before I went to elementary school, I went to Shiga prefecture, which was where my mother was, until the middle school.'

Prior to the next excerpt (4-74), the interviewee N has told the interviewer H that his home is in Suginami (a city in Tokyo). Then in the segment shown below, the interviewer H says she likes Suginami and she used to live there when she lived in Tokyo in lines 1 and 3. She then asks N about a more specific area in line 5. N answers H's question in line 6 with the bundle *na + n + desu + kedo* [COP:ATT + NML COP + but]. Here, N does not exactly know the relevance this information has in the current context because it is H who introduced this topic and is asking for this specific information. Thus, the bundle is used to indicate that the speaker is giving the addressee an interactional option for how to respond to the information. H also uses the bundle in line 9 when she tells N that it is a "local" talk. Here, H is using the bundle to give N an interactional option as to whether or not to give a go ahead sign for H to pursue the direction of the discourse she is suggesting.

(4-74) *na + n + desu + kedo* [COP:ATT + NML COP + but] <background/interactional option>  
(Interview: Hypermedia\_046)

1. H: 私もね杉並に結構好きで
2. N: はい
3. H: 東京にいた時住んでたんですよ。
4. N: あそうなんですか
5. H: どの辺ですか。
6. N: えっと松庵っていう所なんですけど。
7. H: あ私もそうなんです。
8. N: そうなんですか
9. H: はい。あの一なんかローカルな話なんですけど

10. N: はい  
 11. H: ね  
 12. N: はい  
 13. H: あの駅はどっち。

1. H: *watashi mo ne suginami ni kekkoo suki de*  
 I also PP (name) DAT quite like COP  
 2. N: *hai*  
 yes  
 3. H: *tookyoo ni i ta toki ni sun de ta n desu yo.*  
 Tokyo DAT exist PST when DAT live TE PST NML COP PP  
 4. N: *a soo na n desu ka*  
 INJ that COP NML COP Q  
 5. H: *dono hen desu ka*  
 which area COP Q  
 6. N: *etto shooan tte iu tokoro na n desu kedo.*  
 INJ (name) QT say place COP NML COP but  
 7. H: *a watashi mo soo na n desu.*  
 INJ I also that COP NML COP  
 8. N: *soo na n desu ka*  
 that COP NML COP Q  
 9. H: *hai. ano: nanka rookaru na hanashi na n desu kedo.*  
 yes INJ DM local COP talk COP NML COP but  
 10. N: *hai*  
 yes  
 11. H: *ne*  
 PP  
 12. N: *hai*  
 yes  
 13. H: *ano: eki wa dotchi.*  
 INJ station TOP which

1. H: 'I also like Suginami quite a bit'  
 2. N: 'Yes.'  
 3. H: 'I lived there when I was in Tokyo.'  
 4. N: 'Oh, really.'  
 5. H: 'Where in Suginami?'  
 6. N: 'Um, the place is called Shooan.'  
 7. H: 'Oh, I also lived in Shooan.'  
 8. N: 'I see.'  
 9. H: 'Yes. Um, it's a local talk'  
 10. N: 'Yes.'  
 11. H: 'You know.'  
 12. N: 'Yes.'  
 13. H: 'Um, which station?'

The background bundles in Conversation are similar to the ones in Interview. All three sub-types that are found in Interview, that is, concessive, background, interactional, are also found in Conversation. On the other hand, there are some instances in Conversation which do not belong to any of the three sub-types. These bundles express a certain attitude or stance of the speaker.

Just as in Interview, some background bundles in Conversation convey the contrastive meaning in addition to the sense of backgroundness. Excerpts (4-75) and (4-76) are examples of this sub-type.

(4-75) *ta + n + da + kedo* [PST + NML + COP + but] <background/contrastive> (Conversation: Callhome\_1263)

1. B: で私はすごい腹が立っただけど
2. A: ああ。
3. B: でも怒っちゃいけないとか思っさ。

1. B: *de watashi wa sugoi hara ga tat ta n da kedo*  
and I TOP very stomach NOM rise **PST NML COP but**
2. A: *aa.*  
INJ
3. B: *demo okot cha ike nai toka omot te sa.*  
but get.angry ASP go NEG such think TE PP

1. B: ‘And **although** I was so upset’
2. A: ‘Yeah.’
3. B: ‘But I thought I shouldn’t show it.’

(4-76) *ta + n + da + kedo* [PST + NML + COP + but] <background/contrastive> (Conversation: Callhome\_2208)

1. B: バンビとか何とかウォルトディズニーのがたくさんあるけど。
2. A: あっディズニーのはねーこっちで安いよね。まあ英語なんだけど。
3. B: うん。

1. B: *banbi toka nan toka woruto dizuni: no ga takusan aru kedo.*  
(name) such what such (name) GEN NOM many exist but
2. A: *a dizuni: no wa ne: kotchi de yasui no ne. maa eigo na n da kedo.*  
INJ (name) GEN TOP PP here LOC cheap NML PP / INJ English **COP NML COP but**

3. B: *un.*  
yes

1. B: ‘There are many Walt Disney videos like Bambi and others.’
2. A: ‘Oh, Disney videos are cheap here **though** they are in English.’
3. B: ‘Yeah.’

Some background bundles in Conversation simply present background or supplementary information that is related to another part of discourse. In the next excerpt, (4-77), E1F presents the background information that she was riding a bicycle, with the bundle.

(4-77) *ta + n + da + kedo* [PST + NML + COP + but] <background/background>  
(Conversation: Sakura\_09)

E1F: なんかチャリで通って**ただ**けどあネコだっと思ってふーって見てたらすごい見てくるのねじーって。だからなんか目が離せなくなってじーって見ててお互いなんか通りすぎるまでじーって。そしたらガーって追いかけてきて。

E1F: *nanka chari de toot te ta n da kedo a neko da tte omot te fu: tte mi te ta ra*  
DM bicycle INS pass TE **PST NML COP but** INJ cat COP QT think TE ONM QT look TE PST COND  
*sugoi mi te kuru no ne ji: tte. dakara nanka me ga hanase naku nat te ji: tte mi te*  
very look TE ASP NML PP ONM QT / so DM eye NOM turn.away NEG become TE ONM TE look TE  
*te otagai nanka toori sugiru made ji: tte. soshitara ga: tte oikeke te ki te.*  
ASP each.other DM pass pass until ONM QT / then ONM QT chase TE ASP TE

E1F: ‘Like, I was riding a bicycle, and I spotted a cat and looked at her, and she stared at me. So, like, I couldn’t turn my eyes away, and we were staring at each other until we passed by. Then the cat came running after me.’

In the next excerpt, (4-78), B tells A in line 1 that cherry blossoms are now in full bloom where she lives. Then, she mentions a gathering she participated in, again to talk about the cherry blossoms. In lines 3 and 7, she uses the bundle *ta + n + da + kedo* [PST + NML + COP + but] to indicate the backgroundness of the information she is presenting.

(4-78) *ta + n + da + kedo* [PST + NML + COP + but] <background/background>  
(Conversation: Callhome\_3008)

1. B: あの桜が今満開昨日今日。

2. A: えっそうなんだ。

(Several lines omitted.)

3. B: あのお食事して桜の- 駅前の白十寺の上のビラっていう所でやっ  
た  
ただけどね

4. A: あっうんうんうんうん。

5. B: だからずっと国校まで歩いて国校の校舎を外から眺めて

6. A: うん。

7. B: それで皆してお茶飲んで別れたんだけど

8. A: うん

9. B: もう桜すごかったよ。

1. B: *ano sakura ga ima mankai kinoo kyoo*  
INJ cherry NOM now full.bloom yesterday today

2. A: *e soo na n da.*  
INJ that COP NML COP

(Several lines omitted.)

3. B: *ano o shokuji shi te sakura no- eki mae no hakujuuji no ue no bira tte iu tokoro de yat*  
INJ HON meal do TE cherry station front GEN (name) GEN above GEN (name) QT say place LOC do  
*ta n da kedo ne*  
PST NML COP but PP

4. A: *a un un un un.*  
INJ yes yes yes yes

5. B: *dakara zutto kok koo made arui te kok koo no koosha o soto kara nagame te*  
do all.the.way (name) school until walk TE (name) school GEN campus ACC outside from look TE

6. A: *un.*  
yes

7. B: *sorede mina shi te o cha non de wakare ta n da kedo*  
and everyone do TE HON tea drink TE separate PST NML COP but

8. A: *un.*  
yes

9. B: *moo sakura sugokat ta yo.*  
already cherry terrific PST PP

1. B: 'That cherry blossoms are now in full bloom, yesterday, today.'

2. A: 'Oh, really.'

(Several lines omitted.)

3. B: 'Um, we had a meal, at the place called Villa above Hakujuuji in front of the station.'

4. A: 'Oh, yeah yeah yeah yeah.'

5. B: 'So we walked all the way to the school, and looked at the campus from outside.'

6. A: 'Yeah.'

7. B: 'And we drank tea and parted ways'

8. A: 'Yeah.'

9. B: 'The cherry blossoms were already in full bloom.'

In the next excerpt, (4-79), B answers a question by A about some students in line 3, and gives further supplemental information in line 5.

(4-79) *na + n + da + kedo* [COP:ATT + NML + COP + but] <background/background>  
(Conversation: Callhome\_1725)

1. A: 日本人みたいなのだ。
2. B: それりさが教えている人なの？
3. A: それはあたしの生徒じゃないのね。
4. B: うん。
5. A: 私の他の友達の-が持っている人達**なんだけど**。
6. B: ああなるほどなるほど。

1. A: *nihonjin mitai na no.*  
Japanese like COP PP
2. B: *sore risa ga oshie te iru hito na no?*  
that (name) NOM teach TE ASP person COP PP
3. A: *sore wa atashi no seito ja nai no ne.*  
that TOP I GEN student COP NEG NML PP
4. B: *un.*  
yes
5. A: *watashi no hoka no tomodachi no- ga mot te ru hito tachi **na n da kedo.***  
I GEN other GEN friend NOM have TE ASP person PL **COP NML COP but**
6. B: *aa naruhodo naruhodo.*  
INJ indeed indeed

1. A: 'They are like Japanese.'
2. B: 'Are those the students you are teaching?'
3. A: 'Those are not my students.'
4. B: 'Yeah.'
5. A: 'Those are the ones my other friend has.'
6. B: 'Oh, I see, I see.'

Some background bundles in Conversation have an interactional function in addition to presenting the background information and linking it to some other part of discourse. In the next excerpt, F114 tells her friend F147 in line 5 that she began working at the hotel next to the one where F147 works. In doing so, she uses the background bundle. The use of the bundle gives the addressee an interactional option, namely, how to deal with the information given to her and

how to respond to the utterance. F147 displays her surprise, *uso* ‘no kidding.’ Only after hearing F147’s response, does F114 continue to provide more information in the subsequent discourse.

(4-80) *ta + n + da + kedo* [PST + NML COP + but] <background/interactional option>  
(Conversation: BTS\_65)

1. F114: ね F147 ちゃんまだあのバイトやってる？
  2. F147: ん？
  3. F114: オークラ。
  4. F147: うん。
  5. F114: ね偶然にも私隣のホテルで働きだしたんだけど。
  6. F147: うそ。
  7. F114: 赤坂プリンス。
  8. F147: まじで。
- 
1. F114: *ne F147 chan mada ano baito yat te ru?*  
PP (name) dear still that part.time.job do TE ASP
  2. F147: *n?*  
huh
  3. F114: *o:kura.*  
(name)
  4. F147: *un.*  
yes
  5. F114: *ne guuzen ni mo watashi tonari no hoteru de hataraki dashi ta n da kedo.*  
PP coincidence DAT also I next GEN hotel LOC work begin **PST NML COP but**
  6. F147: *uso.*  
lie
  7. F114: *akasaka purinsu.*  
(name)
  8. F147: *maji de.*  
real COP
- 
1. F114: ‘F147, are you still doing that part time job?’
  2. F147: ‘Huh?’
  3. F114: ‘Okura.’
  4. F147: ‘Yeah.’
  5. F114: ‘You know, I began working at the hotel next to it by coincidence.’
  6. F147: ‘No kidding.’
  7. F114: ‘Akasaka Prince.’
  8. F147: ‘For real?’

In the next excerpt, (4-81), B gives A her opinion or advice regarding A's personal problem. To do that, B first says *atashi omou n da kedo* 'I think,' with the background bundle. This is a preliminary activity of securing the floor and getting the addressee's "go ahead" sign before actually stating her opinion. The use of the background bundle here not only signals that the utterance is linked to what to follow, but it also gives the addressee a chance to respond with a continuer to show that she is okay with the speaker proceeding to the main action of stating her opinion.

(4-81) *omou + n + da + kedo* [think + NML + COP + but] <background/background>  
(Conversation: Callhome\_1593)

1. A: そういう人の力が必要だなんていうかちょっと病気だなんて思うもん自分で。こりゃちょっと病気だなんて。
2. B: うん。あたし思うんだけど
3. A: うん。
4. B: あのこの前のその五月とかの状況ね。
5. A: うん。
6. B: なんていうのすごく客観的に見れる人がいればね
7. A: うん。

1. A: *soo iu hito no chikara ga hitsuyoo da na tte iu ka chotto byooki da na tte omou mon*  
that say person GEN power NOM necessary COP PP QT say Q little sick COP PP QT think PP  
*jibun de. korya chotto byooki da na tte.*  
oneself by this.TOP little sick COP PP QT
2. B: *un. atashi omou n da kedo.*  
yes I think NML COP but
3. A: *un.*  
yes
4. B: *ano kono mae no sono go gatsu toka no jookyoo ne.*  
INJ this before GEN that five month such GEN situation PP
5. A: *un.*  
yes
6. B: *nan te iu no suggoku kyakkan teki ni mir eru hito ga ire ba ne*  
what QT say PP very objective like DAT look POT person NOM exist COND PP
7. A: *un.*  
yes

1. A: 'It's like I need the help of those people, I feel like I'm sick myself. This is an illness.'
2. B: 'Yeah. I think'
3. A: 'Yeah.'

4. B: 'Um, like the recent situation in May'
5. A: 'Yeah.'
6. B: 'How to say, if there was someone who can see it very objectively'
7. A: 'Yeah.'

In the next excerpt, (4-82), F002 tells F080 about an event in which she and other students went to lunch with their haiku teacher. In line 1, F002 explains that she and other students had to treat the teacher to lunch. She then says the amount of money she had to pay to treat the teacher was no big deal, with the background bundle. The background bundle functions to give an interactional option to the addressee. From the speaker's perspective, it allows the speaker to wait to see how the addressee responds before proceeding to express her own evaluation. Following the background bundle, the addressee F080 gives her assessment of the situation as *fuyukai* 'displeasing' in line 2. Only after hearing F080's assessment, does F002 give her own evaluation of the situation, namely that the teacher should have paid for her lunch herself because that was supposed to be included in her pay.

(4-82) *nai + n + da + kedo* [NEG + NML + COP + but] <background/interactional option>  
(Conversation: BTS\_31)

1. F002: でね地下で900円のスパゲッティを食べたの。そしたら先生の方を持つ訳。900円が1000円になってま額としてはねどうってことは**ないんだけど**。
2. F080: ないけどねー。不愉快よねそんなの。
3. F002: うん。だってそれはだって報酬の中に入ってるはずだし
4. F008: そうよそうよ
5. F080: 自分が払うべきものでしょう。

1. F002: *de ne chika de 900 en no supagetti o tabe ta no. soshitara sensei no bun o motu*  
and PP basement LOC 900 yen GEN spaghetti ACC eat PST PP then teacher GEN part ACC have  
*wake. 900 en ga 1000 en ni nat te ma gaku to shi te wa ne doo tte koto **nai***  
PP 900 yen NOM 1000 yen COP become TE INJ amount QT do TE TOP how QT NML **NEG**  
***n da kedo.***  
**NML COP but**
2. F080: *nai kedo ne.: fuyukai yo ne sonna no.*  
NEG but PP / unpleasant PP PP such NML

3. F002: *un. datte sore wa datte hooshuu no naka ni hait te ru hazu da shi*  
yes since that TOP since pay GEN among DAT include TE ASP must COP because
  4. F008: *soo yo soo yo*  
that PP that PP
  5. F080: *jibun ga harau beki mono desho o.*  
oneself NOM pay should NML COP AUX
1. F002: ‘And, we ate 900 yen spaghetti in the basement. And we treated our teacher to the lunch. 900 yen became 1000 yen, and um, the amount was not much at all.’
  2. F080: ‘That’s not much. That’s displeasing.’
  3. F002: ‘Yeah. Because that should be included in her pay.’
  4. F008: ‘That’s right. That’s right.’
  5. F080: ‘She should pay herself.’

Some of the background bundles in Conversation express the speakers’ attitude in addition to indicating the supplementary nature of the information. These bundles often appear with emphatic or extreme expressions such as *honto ni* ‘really,’ *choo* ‘extremely,’ and *ari e nai* ‘impossible (lit. cannot exist).’ The stance expressed has to do with the feeling of disbelief or of distance between what the speaker thinks/feels and some other information expressed in the prior context.

Prior to the next fragment, (4-83), F114 told F147 that she could not find a microphone she borrowed from her teacher no matter how hard she looked for it. Then, she said that last night she found a microphone, which was not quite the one she borrowed but was very similar to it, on the chest of drawers in her room. In line 1, she expresses her puzzlement that the similar-looking microphone was found where she had looked for the missing microphone earlier. In line 7, she uses the background bundle and says *imi fumei na n da kedo honto ni* ‘(it)’s really incomprehensible.’ The use of the bundle indicates the speaker’s feeling of disbelief that something that is incomprehensible actually happened to her.

(4-83) *na + n + da + kedo* [COP:ATT + NML + COP + but] <background/stance>  
 (Conversation: BTSJ\_65)

1. F114:  でさあたし さ あんなに
2. F147:  何でー？
3. F114:  一生懸命探した場所にさ 一何でこれがぽつんと置いてあるのとか  
 思って。
4. F147:  うん。えーっ。
5. F114:  そう。
6. F147:  何だろう。
7. F114:  意味不明 **なん** **だけ** **けど** **ほん** **と** **に**。
8. F147:  いやお父さんがこうそつと置いてってくれたんじゃないの。
9. F114:  いやお父さん 単身赴任だから大阪行っちゃってるもん。

1. F114:  *de sa atashi sa anna ni*  
 and PP I PP that.much COP
2. F147:  *nande:?*  
 why
3. F114:  *isshookenmei sagashi ta basho ni sa: nande kore ga potsun to oi te aru no toka*  
 as.hard.as look.for PST place LOC PP why this NOM ONM QT put TE ASP PP such  
*omot te.*  
 think TE
4. F147:  *un. e:.*  
 yes INJ
5. F114:  *soo.*  
 so
6. F147:  *nan daro o.*  
 what COP AUX
7. F114:  *imi fumei **na n da kedo** honto ni.*  
 meaning unclear **COP NML COP but** really COP
8. F147:  *iya ootoosan ga koo sotto oi te tte kure ta n ja nai no.*  
 no father NOM this quietly put TE ASP ASP PST NML COP NEG PP
9. F114:  *iya ootoosan tanshin funin da kara oosaka it chat te ru mon.*  
 no father alone set.off COP because (name) go ASP TE ASP PP

1. F114:  ‘And, I, that much’
2. F147:  ‘How come?’
3. F114:  ‘I wondered why this was at the place where I looked for (the microphone) so  
 hard.’
4. F147:  ‘Yeah, What?’
5. F114:  ‘Yeah.’
6. F147:  ‘Why was that?’
7. F114:  ‘It’s really incomprehensible.’
8. F147:  ‘No, maybe your dad had put it there without telling you.’
9. F114:  ‘No, my dad has gone to Osaka for work.’

In the next excerpt, (4-84), two friends at a college, are talking about a long winter break at their school. In lines 1-4, they agree that the break is long, and they continue to talk about exactly how long the break is. In line 7, F086 says *ari e nai n da kedo* '(it)'s impossible,' with the background bundle. Again, the bundle indicates the speaker's disbelief in how long their winter break is compared to a normal length of a winter break based on her experience.

(4-84) *ta + n + da + kedo* [PST + NML COP + but] <background/stance> (Conversation: BTS\_72)

1. F086: うんうん。だって超長いでしょ春休みー。
2. F046: なーがい
3. F086: 一番長いよねー多分。
4. F046: なーがい長い。だって三か月ぐらいない？
5. F086: そんなにある。だって普通にさ1月始まったってさ1月の最後らへんに終わるんでしょ？
6. F046: 終わる終わるー。
7. F086: ありえないんだけど。
8. F046: ないないない。

1. F086: *un un. datte choo nagai desho haru yasumi:*  
yes yes since very long COP spring break
2. F046: *na:gai*  
long
3. F086: *ichiban nagai yo ne: tabun.*  
best long PP PP probably
4. F046: *na:gai nagai. datte san kagetsu gurai nai?*  
long long since three months about NEG
5. F086: *sonna ni aru. datte futsuu ni sa ichi gatsu hajimat ta tte sa ichi gatsu no saigo*  
that.much COP exist / since usually COP PP one month begin PST QT PP one month GEN end  
*ra hen ni owaru n desho?*  
about area DAT end NML COP
6. F046: *owaru owaru:.*  
end end
7. F086: *ari e nai n da kedo.*  
exist POT NEG NML COP but
8. F046: *nai nai nai.*  
NEG NEG NEG

1. F086: 'Yeah yeah. The spring break is extremely long, you know.'
2. F046: 'It's long.'
3. F086: 'It's the longest, probably.'
4. F046: 'It's so long. Isn't it about three months?'

5. F086: 'It's that long. You know, it (the quarter) begins in January but ends at the end of January.'
6. F046: 'Yeah it does.'
7. F086: 'It's impossible.'
8. F046: 'You are right.'

In the next excerpt, (4-85), in line 3, F086 says that she probably cannot ride a vertical drop ride. In the next line, F046 agrees that the vertical drop ride is scary, but then continues to say *watashi demo zekkyoo kei choo suki na n da kedo* 'but I really like thrill rides,' with the background bundle. This background bundle is interactional in that it gives the addressee an option of how to respond. As we can see in the next line 5, F086 gives a qualification that she can ride the ones that go *gaa* (onomatopoeia for the sound of 'high speed rollercoaster'). In line 6, F046 brings back the topic of the vertical drop ride (a specific one called 'Tower Hacker'). F086 again expresses that she cannot ride the vertical drop ride. Then in line 8, F046 says *are choo tanoshii n da kedo* 'that (Tower Hacker) is so much fun,' with the sequence *n da kedo*. In this case, the background bundle is used to indicate the speaker's feeling of distance between her own assessment of the ride and the addressee's assessment of the ride which had already been expressed in the previous turn.

(4-85) *na + n + da + kedo* [COP:ATT + NML COP + but] <background/interactional option & stance> (Conversation: BTS\_72)

1. F086: 90度に落ちるやつ。サーッと頂上まで行って。
2. F046: あああるあるあるある。うんうん。
3. F086: 私ああいうのは多分無理だなー。
4. F046: 怖ーい。私でも絶叫系超好きなんだけど。
5. F086: 私こういうガーッとか行くやつだったら平気だけどー。
6. F046: でそこでガッと落ちるタワーハッカー？
7. F086: そうそう。あっだめあれ。
8. F046: あれ超楽しいんだけど。
9. F086: うそー？

1. F086: *90 do ni ochiru yatsu. sa: tto choojoo made it te.*  
90 degree DAT drop thing / ONM QT top until go TE
2. F046: *aa aru aru aru aru. un un.*  
INJ exist exist exist exist / yes yes
3. F086: *watashi aa iu no wa tabun muri da na:.*  
I INJ say NML TOP probably impossible COP PP
4. F046: *kowa:i. watashi demo zekkyoo kei choo suki **na n da kedo.***  
scary / I but scream kind very like **COP NML COP but**
5. F086: *watashi koo iu ga: toka iku yatsu dat ta ra heiki da kedo:.*  
I this say ONM such go thing COP PST COND fine COP but
6. F046: *de soko de ochiru tawa: hakka:?*  
and there LOC drop toward hacker
7. F086: *soo soo. a dame are.*  
so so / INJ not.good that
8. F046: *are choo tanoshii **n da kedo.***  
that very fun **NML COP but**
9. F086: *uso:?*  
lie

1. F086: 'The one that drops vertically. It goes to the top.'
2. F046: 'Yeah, I know.'
3. F086: 'I probably cannot ride that kind of ride.'
4. F046: 'It's scary. But I really like thrill rides.'
5. F086: 'I can ride the ones that goes like gaa.'
6. F046: 'And the tower hacker that drops there.'
7. F086: 'Right right. I cannot ride that one.'
8. F046: 'That one is so much fun.'
9. F086: 'Really?'

#### 4.3.2.4. Summary of discourse organizers

In this section, we have identified three major sub-types of discourse organizing bundles:

(1) topic introduction/elaboration bundles, (2) cohesion bundles, and (3) background bundles.

We have seen that the sub-functions (as well as finer sub-types of the sub-functions) are closely related to the registers in which these bundles are used. Register variations in the use of discourse organizing bundles are further discussed in Section 4.4. Another finding is that the functions identified for the cohesion bundles and background bundles are not attributed to the bundles themselves, but to the shorter sequences, *n + da/desu* [NML + COP] and *n + da/desu +*

*kedo/keredo/keredomo/ga* [NML + COP + but]. The issue of the formulaic status of morphemic bundles is addressed in more detail in Section 4.6.

### 4.3.3. Referential expressions

Referential expressions make direct reference to physical or abstract entities, or to the textual context itself.<sup>36</sup> Referential bundles appear most frequently in Speech (68/175; 38.9%), followed by Interview (43/189; 22.8%). They are least frequent in Conversation (4/92; 4.3%). Two major functional sub-types are further identified: (1) Conceptualization and (2) Equivocation.

#### 4.3.3.1. Conceptualization bundles

Conceptualization bundles account for a half of the referential bundles in Conversation (2/4), 60% in Interview (26/43), and 72% in Speech (49/68). Conceptualization bundles can be further divided into two types: *no*-type and *koto*-type. The *no*-type involves the nominalizer *no*, while the *koto*-type involves the nominalizer *koto*. While many previous studies compared the uses of *no* and *koto* from various perspectives (e.g., speaker's certainty in the truth of the information, S. Suzuki 1996b; event vs. proposition, Horie 1993; abstractness of the information, Kuno 1973), none of these studies provide positional characteristics of these nominalizers based on empirical data. As we see below, the examination of the bundles explicates the positional divergence between *no* and *koto*. Below are the examples of the most frequent *no*-type conceptualization bundles.

---

<sup>36</sup> The major function of referential expressions has to do with the identity and status of the entity/text itself, while the major function of cohesion bundles is to indicate the link between preceding and following discourse.

(4-86) *No*-type conceptualization bundles

*tte*    *iu*    *no*    *wa*    <conceptualization>    (Conversation and Interview)  
QT    say    NML    TOP

*to*    *iu*    *no*    *wa*    <conceptualization>    (Interview and Speech)  
QT    say    NML    TOP

Although the bundles above involve the nominalizer, *no*, their function is not to nominalize the preceding clause or phrase. In fact, in many cases, the preceding elements are nouns or noun phrases, which need not be nominalized. As the name suggests, the conceptualization bundles conceptualize the preceding entity, state, or action as an abstract notion. By using the form of quotation, *tte/to iu* [QT say], the speaker presents the preceding information as “one step further from the immediate discourse” (Maynard 1996: 213). In some cases, the conceptualization bundles seem to signal that “either speaker himself or the addressee, according to the speaker’s assessment, has lower certainty” about the preceding information (Iwasaki 2002: 200). However, this does not hold true for all the conceptualization bundles, as can be seen in the examples below. What is common among these examples is that the speaker is objectifying the preceding information and distancing himself/herself from it (see S. Suzuki 1996a).

In the following examples, (4-87) and (4-88), the conceptualization bundles frame the preceding noun/nominal phrase as a general concept rather than a particular referential entity. By conceptualizing the nominal entity as a general concept, the speaker excludes himself/herself from being personally involved in what is expressed about the entity. In (4-87), the interviewee E, who is an actor, is talking about another actor who was being hospitalized. By using the bundle, E frames the concept of actor as an abstract notion and says ‘an actor’s job is to do something in front of people.’ In this way, E excludes himself from being personally included in the characterization of actors in general.

(4-87) *tte + iu + no + wa* [QT + say + NML + TOP] <conceptualization> (Interview: Tetsuko\_10)

1. E: だから一俳優っていうのは本当に自分が人前で何かをするっていうことが仕事でしょ？

2. K: うん。

1. E: *dakara: haiyuu tte iu no wa hontoo ni jibun ga hito mae de nanika o suru tte iu koto*  
so actor QT say NML TOP real COP oneself NOM person front LOC something ACC do QT say NML  
*ga shigoto desho?*  
NOM job COP

2. K: *un.*  
yes

1. E: ‘So an actor’s job is to do something in front of people, isn’t it?’

2. K: ‘Yes.’

In (4-88), the interviewer M asks the interviewee S about a typical day for a housewife. In doing so, M uses the conceptualization construction to generalize the concept of housewife. Although she knows the interviewee is a housewife, the question is formulated in a way that separates the interviewee S from the general concept of a housewife.

(4-88) *tte + iu + no + wa* [QT + say + NML + TOP] <conceptualization> (Interview: Hypermedia\_019)

1. M: あのーご家庭の主婦っていうのはだいたい一日はどういう風にお過ごしになってなるんですか？

2. S: そうですねまあわたくしはごくごく一般的ですけども

1. M: *ano: go katei no shufu tte iu no wa daitai ichi nichi wa doo iu fuu ni o sugoshi*  
INJ HON family GEN housewife QT say NML TOP usually one day TOP how say way DAT HON spend  
*ni nat te ru n desu ka?*  
COP become TE ASP NML COP Q

2. S: *soo desu ne maa watakushi wa gokugoku ippan teki desu keredomo*  
so COP PP INJ I TOP ONM general like COP but

1. M: ‘Um, how do housewives usually spend one day?’

2. S: ‘Let me see, um, my case is ordinary’

In (4-89), the speaker at an academic conference describes the characteristics of a Japanese accent. The speaker uses the conceptualization bundle to make explicit that she is talking about Japanese accents in general, not a particular type of accent.

(4-89) *to + iu + no + wa* [QT + say + NML + TOP] <conceptualization> (Speech: CSJ\_A05F0502)

先程の発表にありましたように日本語のアクセント **というのは** ピッチの下落目があるかないかそれからあったとすれば...

*sakihodo no happyoo ni ari mashi ta yoo ni nihongo no akusento to iu no wa pitchi no*  
 while.ago GEN presentatioin DAT exist POL PST like COP Japanese GEN accent **QT say NML TOP** pitch GEN  
*sagarime ga aru ka nai ka sorekara at ta to sure ba ...*  
 falling.point NOM exist Q NEG Q and exist PST QT do COND

‘As was in the previous presentation, Japanese accent is whether or not there is a pitch fall and if there is, ...’

The *koto*-type conceptualization bundle tends to constitute a predicate or comment as can be seen in the use of a copula at the end of the bundle. This contrasts with the *no*-type which constitutes a topic or a subject (and in some cases an object). The *koto*-type conceptualization bundle encapsulates the preceding information as a conceptual unit. The conceptualized information tends to be a proposition rather than a nominal phrase. The following are the most frequent *koto*-type bundles.

(4-90) *Koto*-type conceptualization bundles

*to iu koto de* <conceptualization> (Interview and Speech)  
 QT say NML COP:TE

*to iu koto ni* <conceptualization> (Interview and Speech)  
 QT say NML COP:ADV

*tte/to iu koto desu* <conceptualization> (Interview and Speech)  
 QT say NML COP:POL

In (4-91), the interviewee F tells the interviewer M what her family did during the summer break. The clause preceding the bundle, *kaigai ryokoo wa shibaraku nai daro o* ‘(we) probably won’t travel abroad for a while’ represents what M’s family members discussed or the conclusion reached through the discussion. The conceptualization bundle encapsulates this information into a conceptual unit, and presents the whole idea as given, unchallengeable information, which is then used as a reason for their travel to Honk Kong during the summer.

(4-91) *to + iu + koto + de* [QT + say + NML + COP:TE] <conceptualization> (Interview: Hypermedia\_034)

1. F: 海外旅行はしばらくないだろうということであの香港に行って来たり
2. M: ええええ
3. F: 夏休みを利用してね
4. M: ええ
5. F: 参りましたですけれども。

1. F: *kaigai ryokoo wa shibaraku nai daro o to iu koto de ano honkon ni it te ki tari*  
overseas travel TOP for.a.while. NEG COP AUX QT say NML COP INJ (name) DAT go TE ASP such.as
2. M: *ee ee*  
yes yes
3. F: *natsu yasumi o riyoo shi mashi te ne*  
summer break ACC use do POL TE PP
4. M: *ee*  
yes
5. F: *mairi mashi ta desu keredomo.*  
go:HUM POL PST COP but

1. F: ‘We decided that we probably wouldn’t travel abroad for a while, and went to such place as Hong Kong’
2. M: ‘Yes yes.’
3. F: ‘We used the summer break’
4. M: ‘Yes.’
5. F: ‘We went.’

The *koto*-type conceptualization bundle is also used by interviewers to recapitulate what the interviewees said in the prior context and link it to the next topic. In (4-92), for example, the interviewer M uses the conceptualization bundle to recapitulate the information given by S in the

prior context that S was studying abroad until March. By encapsulating the information as a conceptual unit, the interviewer use it as a preface to the next topic, namely, what S found surprising when he return to Japan from overseas.

(4-92) *to + iu + koto + de* [QT + say + NML + COP:TE] <conceptualization> (Interview: Hypermedia\_002)

1. M: 三月まで留学してらしたということ

2. S: はい

3. M: 四月に日本に帰ってきて

4. S: はい

5. M: なんか

6. S: はは

7. M: なんかびっくりしたことってありますか？

1. M: *san gatsu made ryuugaku shi te rashi ta to iu koto de*  
three month until study.abroad do TE ASP PST QT say NML COP

2. S: *hai*  
yes

3. M: *shi gatsu ni nihon ni kaet te ki te*  
four month DAT Japan DAT return TE ASP TE

4. S: *hai*  
yes

5. M: *nanka*  
DM

6. S: *haha*  
INJ

7. M: *nanka bikkuri shi ta koto tte ari masu ka?*  
DM surprise do PST NML QT exist POL Q

1. M: ‘You were studying abroad until March’

2. S: ‘Yes.’

3. M: ‘You came back to Japan in April’

4. S: ‘Yes.’

5. M: ‘Um’

6. S: ‘Yea’

7. M: ‘Um, was there anything surprising to you?’

In the next excerpt from Speech, (4-93), the speaker uses the bundle to conceptualize the information as one of the three points concerning ‘the educational situation at that time.’

(4-93) *to + iu + koto + de* [QT + say + NML + COP:TE] <conceptualization> (Speech: CSJ\_A08F0498)

その当時の時代的 教育状況についてまとめてみますと 三つの点が挙げられると思います 一つには 学校制度の整備が進んで 義務教育の制度が 完成段階であったということでありす

*sono tooji no jidai teki kyooiku jookyoo ni tsui te matome te mi masu to mittsu no ten*  
 that then GEN time like education situation DAT adhere TE summarize TE ASP POL COND three GEN point  
*ga age rareru to omoi masu / hitotsu ni wa gakkoo seido no seibi ga susun de gimu*  
 NOM give POT QT think POL / first DAT TOP school system GEN maintenance NOM advance TE compulsory  
*kyooiku no seido ga kansei dankai de at ta to iu koto de ari masu*  
 education GEN system NOM completion step COP exist PST QT say NML COP exist POL

‘If we summarize the educational situation at that time, there are three points. The first is that the development of the school system was advanced and the compulsory education system was at the completion stage.’

#### 4.3.3.2. Equivocation bundles

The second referential bundle sub-type is equivocation. Equivocation bundles account for a half of the referential bundles in Conversation (2/4), 40% in Interview (17/43), and 28% in Speech (19/63). The equivocation bundles indicate that given information is not exact. The following represent the most frequent morphemic bundles of this sub-type.

(4-94) Equivocation bundles

<i>tte/to</i>	<i>iu</i>	<i>fuu</i>	<i>ni</i>	<equivocation>	(Interview and Speech)
QT	say	way	COP:ADV		
<i>soo</i>	<i>iu</i>	<i>fuu</i>	<i>ni</i>	<equivocation>	(Conversation and Interview)
that	say	way	COP:ADV		
<i>soo</i>	<i>iu</i>	<i>no</i>	<i>wa</i>	<equivocation>	(Conversation and Interview)
that	say	NML	TOP		
<i>tte/to</i>	<i>iu</i>	<i>yoo</i>	<i>na</i>	<equivocation>	(Interview and Speech)
QT	say	like	COP:ATT		

The morphemic bundle *tte/to + iu + fuu + ni* [QT + say + way + COP:ADV] appears with quoting verbs such as *iu* ‘say,’ *omou* ‘think,’ and *kangaeru* ‘think’ as in the following examples. In all of the examples below, *tte/to iu fuu ni* can be replaced by just the quotative particle, *tte/to*. When the following verb is *omou* ‘think,’ as in (4-95), or *kangaeru* ‘think,’ as in (4-99), the resulting sequence parallels the combination of stance expression and the quotative particle and verb, as described in Section 4.3.1. The equivocation bundle makes the quoted information equivocal and noncommittal. The following three excerpts are from Interview.

In (4-95), the interviewee A, who is a famous ice skater, talks about her transition from competitive figure skating to ice shows and exhibitions. In line 5 and 7, she uses the equivocation bundle with stance expressions to indicate that the quoted expressions are not exact.

(4-95) *tte + iu + fuu + ni* [QT + say + way + COP:ADV] <equivocation> (Interview: Tetsuko\_5)

1. A: そうですねちょうどその少し前にアイスショーを初めて生で見ることはありませんして
  2. K: ええ。
  3. A: であんな世界もあるんだっていうすごく衝撃を受けたんですね。
  4. K: ええ。
  5. A: で私もこういう世界で滑れたらいいなっていう風に思いましてでアイスショーにはルールもありませんし
  6. K: うーん。
  7. A: やはりあのエンターテイメントが-性が強いなっていう風を感じてここで滑れたら気持ちいいだろうなっていうのがあって
  8. K: うん。
  9. A: そこにすごく憧れをいただいたんですね。
- 
1. A: *soo desu ne choodo sono sukoshi mae ni aisu sho: o hajimete nama de miru koto ga ari*  
so COP PP just that little before DAT ice show ACC first real COP watch NML NOM exist  
*mashi te*  
POL TE
  2. K: *ee.*  
yes
  3. A: *de a konna sekai mo aru n da tte iu sugoku shoogeki o uke ta n desu ne.*  
and INJ this.kind world also exist NML COP QT say very shock ACC receive PST NML COP PP

4. K: *ee.*  
yes
5. A: *de watashi mo koo iu sekai de sube re tara ii na tte iu fuu ni omoi mashi te de aisu sho:*  
and I also this say world LOC skate POT COND good PP QT say way COP think POL TE and ice show  
*ni wa ru:ru mo ari mase n shi*  
DAT TOP rule also exist POL NEG because
6. K: *u:n.*  
INJ
7. A: *yahari ano enta:teimento ga- sei ga tsuyoi na tte iu fuu ni kanji te koko de sube re*  
as.expected INJ entertainment nature NOM strong PP QT say way COP feel TE here LOC skate POT  
*tara kimochi ii daro o na tte iu no ga at te*  
COND feeling good COP AUX PP QT say GEN NOM exist TE
8. K: *un.*  
yes
9. A: *soko ni sugoku akogare o idai ta n desu ne.*  
there DAT very admiration ACC have PST NML COP PP
1. A: ‘Let me see, I had a chance to see an ice show live for the first time a little bit before that’
2. K: ‘Yes.’
3. A: ‘And, I was really shocked and thought oh this kind of world exists.’
4. K: ‘Yes.’
5. A: ‘And I thought it would be good if I could skate in this kind of world, and the ice show doesn’t have rules’
6. K: ‘Mhm.’
7. A: ‘I felt that the entertainment factor is strong, and felt that it would feel good if I could skate here’
8. K: ‘Yes.’
9. A: ‘I had strong admiration for that.’

In the next excerpt, the interviewee R talks about a professional musician who came to teach her high school brass band. The bundle is used in line 1 with the quotation of what the musician said. The bundle shows that the exact wording of the musician is not the point; what is important is the general idea conveyed by the musician.

(4-96) *tte + iu + fuu + ni* [QT + say + way + COP:ADV] <equivocation> (Interview: CSJ\_D01F0030)

1. R: 確かボランティアでいっていう風におっしゃって来てもらったように思うんですけども
2. L: へー
3. R: えーとープロの演奏家の方なんですけどそういう
4. L: ええ

5. R: ちい-す-小さいっていかま 青少年の育成に非常に
6. L: うーん
7. R: 熱意のある方で

1. R: *tashika borantia de ii tte iu fuu ni osshat te ki te morat te ta yoo ni omou*  
probably volunteer COP good QT say way COP say:HON TE come TE ASP TE ASP like COP think  
*n desu kedomo*  
NML COP but
2. L: *he:*  
INJ
3. R: *e:to: puro no ensooka no kata na n desu ga soo iu*  
INJ pro GEN player GEN person COP NML COP but that say
4. L: *ee*  
yes
5. R: *chii- su- chiisai tte iu ka ma seishoonen no ikusei ni hijoo ni*  
small QT say Q young.people GEN upbringing DAT extreme COP
6. L: *u:n*  
INJ
7. R: *netsui no aru kata de*  
enthusiasm GEN exist person COP

1. R: 'If I remember correctly, I think he came to us saying it would be fine to volunteer'
2. L: 'Wow'
3. R: 'Um, he was a professional musician'
4. L: 'Yes'
5. R: 'For the bringing up of small, or, young people, he was very'
6. L: 'Mhm'
7. R: 'He was very enthusiastic'

In (4-97), the equivocation bundle is used by the interviewer M twice. In line 1, M uses the bundle in quoting what the interviewee A said earlier. In line 5, M again uses the bundle in quoting A's students' wish to enter school in Japan. The both cases are similar to the previous example in that the bundle is used to indicate the exact wording is not important.

(4-97) *tte + iu + fuu + ni* [QT + say + way + COP:ADV] <equivocation> (Interview: Hypermedia\_001)

1. M: あの韓国の学生さんが多いっていう風にさっきおっしゃいましたよねー。
  2. A: はい
- (Four turns omitted.)

3. M: そういう学生さん達ってというのはどういう目的でその日本 就学生って  
おっしゃったから日本で大学に行きたいとか
4. A: ええ
5. M: まあ学校に入りたいということですよね。どういう目的であのーその  
日本の学校についていう風に思ってるんじゃないかな。

1. M: *ano kankoku no gakusei san ga ooi tte iu fuu ni sakki osshai mashi ta yo ne:*  
INJ korea GEN student SFX NOM many QT say way COP before say:HON POL PST PP PP

2. A: *hai*  
yes

(Four turns omitted.)

3. M: *soo iu gakusei san tachi tte iu no wa doo iu mokuteki de sono nihon shuugakusei tte*  
that say student SFX PL QT say GEN NML TOP how say purpose COP INJ Japan pre-college.student QT  
*osshat ta kara nihon de daigaku ni iki tai toka*  
say:HON PST because Japan LOC college DAT go want such

4. A: *ee*  
yes

5. M: *maa gakkoo ni hairi tai to iu koto desu yo ne. doo iu mokuteki de ano: sono nihon*  
INJ school DAT enter want QT say NML COP PP PP / how say purpose COP INJ INJ Japan  
*no gakkoo ni tte iu fuu ni omot te rassharu n desho o ne:*  
GEN school DAT QT say way COP think TE ASP NML COP AUX PP

1. M: 'Um, you said before that there are many students from Korea'

2. A: 'Yes'

(Four turns omitted.)

3. M: 'That kind of students, you said they are pre-college students, so for what purpose do  
they want to go to Japanese colleges'

4. A: 'Yes'

5. M: 'Um, they want to enter the schools. For what purpose, um, do you think they want to go  
to Japanese schools?'

The next two excerpts, (4-98) and (4-99), are from Speech. The excerpts all contain the  
bundle *to + iu + fuu + ni* [QT + say + way + COP:ADV]. The function of the bundle is the  
same as in Interview, that is, indicating the equivocalness of the quoted information.

(4-98) *to + iu + fuu + ni* [QT + say + way + COP:ADV] <equivocation> (Speech: A09F0600)

え第一にタイでは末子継承慣行の下一般的には年齢順に他出しえ一嫡系傍  
系の区別であるとか男女の区別による流出形態の違いは生まれにくい  
風にえーいわれています

*e dai ichi ni tai de wa masshi keishoo kankoo no moto ippan teki ni wa nenrei*

INJ the first DAT Thailand LOC TOP youngest.child succession custom GEN under general like DAT TOP age

*jun ni tashutsushi e: chaku kei bookei no kubetsu de aru toka dan jo no*  
 order DAT go.out INJ legitimate line collateral line GEN distinction COP exist such male female GEN  
*kubetsu ni yoru ryuushutsu keitai no chigai wa umare nai to iu fuu ni e: iw are te i masu*  
 distinction DAT depend outflow form GEN difference TOP appear NEG **QT say way COP** say PASS TE ASP POL

‘Um, first, in Thailand, it is said that under the custom of the youngest child succession, in general, people leave their homes in the order of their ages, and the distinction between the legitimate line and the collateral line or between male and female does not give arise to a difference in the outflow.’

(4-99) *to + iu + fuu + ni* [QT + say + way + COP:ADV] <equivocation> (Speech: CSJ\_A01M0074)

今後はまーあの何といつても複数の被験者を検討したいという風に考えて  
 おります / 以上です

*kongo wa ma: ano nan to it te mo fukusuu no hikensha o kentoo shi tai to iu fuu ni kangae te*  
 after.this TOP INJ INJ what QT say TE also several subject ACC consider want **QT say way COP** think TE  
*ori masu. ijoo desu.*  
 ASP:HUM POL / foregoing COP

‘In the future, more than anything, we’d like to consider several subjects. That’s it.’

The equivocation bundle *soo + iu + fuu + ni* [that + say + way + COP:ADV] is similar to *tte/to + iu + fuu + ni* [QT + say + way + COP:ADV] in that it makes the quoted information equivocal and noncommittal. One difference between the two types of bundles is that what is being referred to by *soo + iu + fuu + ni* [that + say + way + COP:ADV] is often an utterance by someone else or information whose source is someone other than the speaker. This contrasts with *tte/to + iu + fuu + ni* [QT + say + way + COP:ADV] whose quoted information often belongs to the speaker. This accounts for the fact that the bundle *soo + iu + fuu + ni* [that + say + way + COP:ADV] does not appear in Speech where no actively participating co-participant is present. Below are three examples from Conversation, (4-100), (4-101), (4-102), and two from Interview (4-103) and (4-104).

(4-100) *soo + iu + fuu + ni* [that + say + way + COP:ADV] <equivocation> (Conversation: Callhome\_2004)

1. A: 最後まで努力したらね
2. B: うん。
3. A: 絶対その効果はあるからって励ましてあげて。
4. B: そうそう。そういう風にあってんだけど。やっただけのことはあるからってさ。

1. A: *saigo made doryoku shi ta ra ne*  
end until make.effort do PST COND PP
2. B: *un.*  
yes
3. A: *zettai sono kooka wa aru kara tte hagemashi te age te.*  
surely that effect TOP exist because QT encourage TE ASP TE
4. B: *soo soo. soo iu fuu ni it te n da kedo. yat ta dake no koto wa aru kara tte sa.*  
so so / that say way COP say TE NML COP but / do PST only GEN NML TOP exist because QT PP

1. A: 'If he keeps working hard to the end'
2. B: 'Yeah.'
3. A: 'It will definitely bring some good results, please tell him that to encourage him.'
4. B: 'Right right. I'm telling him that way, like "you are not studying for nothing."'

(4-101) *soo + iu + fuu + ni* [that + say + way + COP:ADV] <equivocation> (Conversation: Callhome\_1425)

1. A: でも日本も結構不景気なんでしょう？
2. B: そうよ。うん。
3. A: ねー。そうそういう風に聞いているけどさ。
4. B: うん。

1. A: *demo nihon mo kekkoo fukeiki na n desho o?*  
but Japan also fairly bad.economy COP NML COP AUX
2. B: *soo yo. un.*  
so PP / yes
3. A: *ne.: soo iu fuu ni kii te ru kedo sa.*  
PP / that say way COP hear TE ASP but PP
4. B: *un.*  
yes

1. A: 'But Japan's economy is in a pretty bad shape, too, right?'
2. B: 'Right. Yeah.'
3. A: 'Right? I've heard it that way.'
4. B: 'Yeah.'

(4-102) *soo + iu + fuu + ni* [that + say + way + COP:ADV] <equivocation> (Conversation: Callhome\_0743)

1. A: だからもう積極的にエクササイズをしようと思ってんだけど。
2. B: うん うん うん うん。
3. A: うん。でもなんか最近今日なんかすごく下がってきたような気がして。
4. B: ああそう。まああの臨月に近くなればそういう風になってくるよ。

1. A: *dakara moo sekkyoku teki ni ekusasaizu o shiyo o to omot te n da kedo.*  
so already active like COP exercise ACC do AUX QT think TE NML COP but
2. B: *un un un un.*  
yes yes yes yes
3. A: *un. demo nanka saikin kyoo nanka sugoku sagat te ki ta yoo na ki ga shi te.*  
yes but DM recently today DM very come.down TE ASP PST like COP feeling NOM do TE
4. B: *aa soo. maa ano ringetsu ni chikaku nare ba soo iu fuu ni nat te kuru yo.*  
INJ so / INJ INJ due.month DAT near become COND **that say way COP** become TE ASP PP

1. A: 'So I'm thinking of doing exercises actively.'
2. B: 'Yeah yeah yeah yeah.'
3. A: 'Yeah. But, like, recently, today, I'm feel like the baby is coming down a lot.'
4. B: 'Oh, yeah? Well, when the due month is nearing, it gets to be that way.'

(4-103) *soo + iu + fuu + ni* [that + say + way + COP:ADV] <equivocation> (Interview: Tetsuko\_3)

1. H: 女の子にしては筋肉はつきやすいタイプみたいで
2. T: うん。
3. H: なのであまりこう筋トレとかジムとかは行かないでー
4. T: うん。
5. H: 走ることだったり泳ぐことであの一ちょう-調整したりとか気分転換程度ですねー。
6. T: 全然そういう風にはお見受けしませんけどねー。

1. H: *onna no ko ni shi te wa kinniku wa tsuki yasui taipu mitai de*  
female GEN child DAT do TE TOP muscle TOP get easy type seem COP
2. T: *un.*  
yes
3. H: *nanode amari koo kin tore toka jimmu toka wa ika nai de:*  
so not.much this muscle training such gym such TOP go NEG COP
4. T: *un.*  
yes
5. H: *hashiru koto dat tari oyogu koto de ano: choosei shi tari toka kibun tenkan teido desu ne:.*  
run NML COP such.as swim NML COP INJ adjustment such.as such mood change degree COP PP
6. T: *zenzen soo iu fuu ni wa o miuke shi mase n kedo ne:.*  
at.all **that say way COP** TOP PFX look do POL NEG but PP

1. H: 'For a girl, it's easy for me to develop muscles'
2. T: 'Yeah.'
3. H: 'So I don't go to a muscle training or gym much'
4. T: 'Yeah.'
5. H: 'I just do such things as running and swimming for a change.'
6. T: 'You don't look that way at all.'

(4-104) *soo + iu + fuu + ni* [that + say + way + COP:ADV] <equivocation> (Interview: Hypermedia\_18)

1. K: あのひととひととの関係
2. N: うんうん。
3. K: みたいのをあのーそのー研究できたらなと思って
4. N: うん。
5. K: その割と世界的な
6. N: うん。
7. K: 視点ーに興味があるんです。あの国際機構とか
8. N: はい。
9. K: 国際法とかですね。
10. N: うん。うん。うん。
11. K: はい。
12. N: それはあの最近そういう風に思い始めたんですか。

1. K: *ano hito to hito to no kankei*  
INJ person and person and GEN relationship
2. N: *un un.*  
yes yes
3. K: *mitai no o ano: sono: kenkyuu deki tara na to omot te*  
like NML ACC INJ INJ research do:POT COND PP QT think TE
4. N: *un.*  
yes
5. K: *sono wari to sekai teki na*  
INJ comparative COP world like COP
6. N: *un.*  
yes
7. K: *shiten: ni kyoomi ga aru n desu. ano kokusai kikoo toka*  
perspective DAT interest NOM exist NML COP / INJ interactional organization such
8. N: *hai.*  
yes
9. K: *kokusai hoo toka desu ne.*  
international law such COP PP
10. N: *un un un.*  
yes yes yes
11. K: *hai.*  
yes

12. N: *sore wa ano saikin soo iu fuu ni omoi hajime ta n desu ka.*  
that TOP INJ recently **that say way COP** think begin PST NML COP Q

1. K: 'Um the relationship between people and people'
2. N: 'Yeah yeah.'
3. K: 'I hope I can study, um, that kind of thing'
4. N: 'Yeah.'
5. K: 'Um, relatively, global'
6. N: 'Yeah.'
7. K: 'I'm interested in that perspective. Um, like international organization'
8. N: 'Yes.'
9. K: 'Or like international law.'
10. N: 'Yeah yeah yeah.'
11. K: 'Yes.'
12. N: 'Did you begin to think that way recently?'

The bundle *soo + iu + no + wa* [that + say + NML + TOP] or *soo + iu + no + ga* [that + say + NML + NOM] involves the deictic expression, *soo*. However, the function of the whole bundle is not merely to identify a previously introduced referent/proposition, but to make the preceding referent or proposition ambiguous (i.e., making it less specific). As we can see in the following examples, the bundles often co-occur with other expressions that denote lack of specification, such as *tari* 'doing such thing as ...' and *toka* 'such as' (see Taylor 2010). These bundles signal that the references given are not exact. In this way, this type of referential bundle allows the speaker to keep the discourse moving forward without giving specific references. The excerpts (4-105) and (4-106) below are from Conversation, and (4-107) and (4-108) are from Interview.

(4-105) *soo + iu + no + wa* [that + say + NML + TOP] <equivocation> (Conversation: Callhome\_2219)

その打ち合せで昨日一昨日かな全日空ホテル下見に行ったりそういうのはしてるけども学校はないじゃんみんな受験だから。

*sore no uchiawase de kinoo ototoi ka na zennikuu hoteru shitami ni it tari soo iu*  
that GEN arrangements LOC yesterday day.before.yesterday (name) hotel take.a.look DAT go such.as **that say**

*no wa shi te ru kedomo gakkoo wa nai jan minna juken da kara.*

NML TOP do TE ASP but school TOP NEG AUX everyone exam.preparation COP because

‘For the arrangements, she went to the ANA hotel to take a look yesterday or a day before that, so she is doing things like that, but there is no school because everyone is preparing for the entrance examination.’

(4-106) *soo + iu + no + wa* [that + say + NML + TOP] <equivocation> (Conversation: Callhome\_1370)

肺に肺にあの菌が入ってる とか そういうのはないの？

*hai ni haini ano kin ga hait te ru toka soo iu no wa nai no?*

lung DAT lung DAT INJ bacteria NOM enter TE ASP such that say NML TOP NEG PP

‘Bacteria are in the lung; that kind of thing is not happening?’

(4-107) *soo + iu + no + wa* [that + say + NML + TOP] <equivocation> (Interview: Tetsuko\_10)

じゃあ仕事はどのくらい休めばいい とか そういうのはあったんですか？

*jaa shigoto wa dono kurai yasume ba ii toka soo iu no wa at ta n desu ka?*

then work TOP how about stay.away COND good such that say NML TOP exist PST NML COP Q

‘Then, was there something like how long you have to take time off from work?’

(4-108) *soo + iu + no + ga* [that + say + NML + NOM] <equivocation> (Interview: CSJ\_D04F0011)

あの一脳の部位の名前を覚えたり とか そういうのがやっぱり大変ですし

*ano: noo no bui no namae o oboe tari toka soo iu no ga yappari taihen desu shi*

INJ brain GEN part GEN name ACC memorize such.as such that say NML NOM as.expected hard COP because

‘Um, like memorizing the names of the brain parts, that kind of thing is hard as expected.’

The bundle *tte + iu + yoo + na* [QT + say + like + COP:ATT] is similar to the bundle *tte/to + iu + fuu + ni* [QT + say + way + COP:ADV] in that it makes the quoted information equivocal and noncommittal. The difference is that while *tte/to + iu + fuu + ni* [QT + say + way + COP:ADV] is followed by quoting verbs such as *iu* ‘say,’ *omou* ‘think,’ and *kangaeru* ‘think,’

*tte + iu + yoo + na* [QT + say + like + COP:ATT] is followed by a noun. The most frequent noun is the nominalizer *koto*. In Interview, *tte + iu + yoo + na* is often used to leave interpretive options for the addressee. On the other hand, in Speech, this bundle is used to disclaim the responsibility on the part of the speaker for the preciseness of the information being presented.

This bundle also overlaps in form with the conceptualization bundle, introduced earlier in Section 4.3.3.1. In the following Interview excerpts, (4-109), (4-110), and (4-111), the equivocation bundle adds a sense of vagueness to the status of the quote. The sense of vagueness, together with the sense of abstractness, which is conveyed by the use of the conceptualization bundle, works to leave interpretive options to the addressee.

(4-109) *tte + iu + yoo + na* [QT + say + like + COP:ATT] <equivocation/interpretive option>  
(Interview: Hypermedia\_026)

1. M: 何か特にこれだけはっていうようなことがおありでしたか？
2. T: やはり一つはあののびのびと育てて欲しいということと
3. M: ああふんふん
4. T: もう一つはやっぱりあの一信用の置ける人柄になって欲しい
5. M: うん うん
6. T: といいいますか
7. M: うん
8. T: それぐらいのことしか気を使っておりませんでした。

1. M: *nanika toku ni kore dake wa tte iu yoo na koto ga o ari deshi ta ka?*  
something special COP this only TOP QT say like COP NML NOM SFX exist COP PST Q
2. T: *yahari hitotsu wa ano nobinobi to sodat te hoshii to iu koto to*  
after.all one TOP INJ carefree QT grow TE want QT say NML QT
3. M: *aa fun fun*  
INJ yes yes
4. T: *moo hitotsu wa yappari ano: shinyoo no okeru hitogara ni nat te hoshii*  
another one TOP after.all INJ trust GEN keep personality COP become TE want
5. M: *un un*  
yes yes
6. T: *to ii masu ka*  
QT say POL Q
7. M: *un*  
yes
8. T: *sore gurai no koto shika ki o tsukat te ori mase n deshi ta.*  
that about GEN NML only mind ACC use TE ASP POL NEG COP PST

1. M: 'Was there anything in particular you hope for your children?'
2. T: 'One thing is that I want my children to grow carefree, and'
3. M: 'Uh, hmm hmm.'
4. T: 'Another thing is that I want them to be trustworthy'
5. M: 'Yeah yeah'
6. T: 'Something like that'
7. M: 'Yeah.'
8. T: 'I only paid attention to such things.'

(4-110) *tte + iu + yoo + na* [QT + say + like + COP:ATT] <equivocation/interpretive option>  
(Interview: CSJ\_D01M0042)

1. L: あのー先程 ちょっと 子供の方が 熱心だ っていう ような ことを おっしゃっ  
ていた んです けれど も
2. R: うん うん
3. L: そういう 子供 っていう のは やはり こ-自ら 子供が やりたくて それとも 親  
が こう やり なさい よって いう 子 が 多い ン です か ね
1. L: *ano: sakihodo chotto kodomo no hoo ga nesshin da tte iu yoo na koto o osshat*  
INJ while.ago little child GEN direction NOM enthusiastic COP QT say like COP NML ACC say:HON  
*te i ta n desu keredomo*  
TE ASP PST NML COP but
2. R: *un un*  
yes yes
3. L: *soo iu kodomo tte iu no wa yahari ko- mizukara kodomo ga yari taku te soreto mo oya ga*  
that say child QT say NML TOP after.all oneself child NOM dowant TE or parent NOM  
*koo yari nasai yo tte iu ko ga ooi n desu ka ne.*  
this do AUX PP QT say child NOM many NML COP Q PP
1. L: 'Um, you said a while ago that children are more enthusiastic or something like that'
2. R: 'Yeah yeah'
3. L: 'Those children are, are most of them want to do it themselves or are told by their parents to do it.'

(4-111) *tte + iu + yoo + na* [QT + say + like + COP:ATT] <equivocation/interpretive option>  
(Interview: Hypermedia\_028)

1. M: あの 外国 人 に 日本語 を 教えて らして
2. N: はい。
3. M: あのー まあ 面白い とか
4. N: はい。

5. M: あの こう いう 点 と とも あ の ー えー 興 味 が あ っ た っ て いう よう な エ ピ ソ ー ド と し て は い く つ か あ り ま す か 。 何 か 心 に 残 る も の 。

1. M: *ano gaikoku jin ni nihongo o oshie te rashi te*  
INJ foreign person DAT Japanese ACC teach TE ASP TE

2. N: *hai.*  
yes

3. M: *ano: maa omoshiroi toka*  
INJ INJ interesting such

4. N: *hai.*  
yes

5. M: *ano koo iu ten totemo ano: e: kyoomi ga at ta tte iu yoo na episo:do to shi te wa ikutsuka*  
INJ this say point very INJ INJ interest NOM exist PST QT say like COP episode QT do TE TOP several  
*ari masu ka. nanika kokoro ni nokoru mono.*  
exist POL Q / something mind DAT remain NML

1. M: ‘Um, as you teach Japanese to foreigners’

2. N: ‘Yes.’

3. M: ‘Um, well like it’s interesting’

4. N: ‘Yes.’

5. M: ‘Um, are there some episodes that you were interested in this point? Something you remember.’

In the next excerpts from Speech, (4-112), (4-113), and (4-114), the bundle is used to disclaim the responsibility on the part of the speaker for the preciseness of the information being presented.

(4-112) *to + iu + yoo + na* [QT + say + like + COP:ATT] <equivocation/disclaiming responsibility> (Speech: CSJ\_A08M0186)

ま 何 か あ の 例 え ば えー う - 先 週 の 日 曜 日 えー 経 験 し た こ と を 話 し な さ い と い う よう な えー まー あ の ー 課 題 を 与 え ら れ て えー 二 人 の 子 供 が 対 話 し て る ん だ と 思 い ま す が

*ma nanika ano tatoeba e: u- senshuu no nichiyoo bi e: keiken shi ta koto o hanashi nasai to iu*  
INJ something INJ for.example INJ last.week GEN Sunday INJ experience do PST NML ACC talk AUX QT say  
*yoo na e: ma: ano: kadai o atae rare te e: futari no kodomo ga taiwa shi te ru n da to*  
like COP INJ INJ INJ assignment ACC give PSS TE INJ two GEN child NOM dialogue do TE ASP NML COP QT  
*omoi masu ga*  
think POL but

‘Well, like, um, for example, um, I think the two children are conversing after being given an assignment to talk about their experiences last Sunday.’

(4-113) *to + iu + yoo + na* [QT + say + like + COP:ATT] <equivocation/disclaiming responsibility> (Speech: CSJ\_A01M0115)

え 例 例 ば あー 高 齢 者 の よ う な 場 合 に は 発 声 速 度 が 著 し く 遅 い 高 齢 者 が い る  
と い う よ う な こ と が こ れ ま で の 研 究 で 知 ら れ て お り ま す

*e tatoeba a: koorei sha no yoo na baai ni wa hassei sokudo ga ichijirushiku osoi koorei*  
INJ for.example INJ old person GEN like COP case DAT TOP vocalization speed NOM remarkably slow old  
*sha ga iru to iu yoo na koto ga kore made no kenkyuu de shir are te ori masu*  
person NOM exist QT say like COP NML NOM this until GEN research COP know PASS TE ASP POL

‘Um, for example, um, it is known that, in the case of older people, from past studies, that there are older people whose vocalization speed is remarkably slow.’

(4-114) *to + iu + yoo + na* [QT + say + like + COP:ATT] <equivocation/disclaiming responsibility> (Speech: CSJ\_A07F0445)

口 頭 や 文 書 に よ っ て 私 が リ サ イ ク ル し ま す と い う よ う に リ サ イ ク ル の 実 行 宣  
言 を し て も ら う 更 に シ ス テ ム と し て 回 収 回 数 を 増 加 す る と か 回 収 場 所 を 近  
く す る と い う そ う い う 変 化 が えー 行 動 に ど う 影 響 を 与 え る か と い う よ う な  
タ イ プ の 研 究 が 行 な わ れ て い ま す

*kootoo ya bunsho ni yot te watashi ga risaikuru shi masu to iu yoo ni risaikuru no jikkoo sengen*  
verbal and document DAT depend TE I NOM recycle do POL QT say like COP recycle GEN practice declaration  
*o shi te morau sarani shisutemu to shi te kaishuu kaisuu o zooka suru toka kaishuu basho o*  
ACC do TE ASP moreover system QT do TE collect number ACC increase do such collect place ACC  
*chikaku suru to iu soo iu henka ga e: koodoo ni doo eikyoo o ataeru ka to iu yoo na taipu no*  
close do QT say change NOM INJ behavior DAT how influence ACC give Q QT say like COP type GEN  
*kenkyuu ga okonaw are te i masu*  
research NOM conduct PASS TE ASP POL

‘The kind of research that studies how having people declare verbally or in writing that they will recycle and increasing the number of collections or making the collection places closer (to their homes) influence their recycling behavior is being conducted.’

#### 4.3.3.3. Summary of referential expressions

The analysis in this section indicates that the referential bundles in Japanese spoken discourse are closely related to the quoting constructions and the nominalizers. Unlike English referential lexical bundles in spoken English (see Biber et al. 2004), most of which are used to

identify/focus on a referent, specify an attribute of an entity, or to refer to a particular time, place or location in the text, the Japanese referential morphemic bundles are used to generalize and/or obscure the status of the information. While most bundles in the referential category represent formulaic units on their own, two types of the equivocation bundles, namely the ones involving the string *yoo + na* [like + COP] and the ones involving the string *fuu + ni* [QT + say + way + COP] are found to be part of larger sequences. The string *yoo + na* [like + COP] is part of the longer sequence, *tte/to + iu + yoo + na + koto* [QT + say + like + COP + NML], and the string *fuu + ni* is part of the larger sequence, *tte/to + iu + fuu + ni + omoi/kangaeru* [QT + say + way + COP + think]. As we can see, both of these longer sequences are beyond what can be identified through a four-morphemic bundle analysis. These larger sequences are further discussed in Section 4.6.

#### 4.3.4. Socio-interactional expressions

Socio-interactional bundles express various kinds of concerns for interactional activities and intersubjective understandings. As stated earlier, socio-interactional expressions are predominantly found in Conversation (50/92; 55.3%) and in Interview (77/189; 40.7%), but they are much less common in Speech (20/175; 11.4%). Socio-interactional expressions can be categorized into several sub-types according to their specific functions identified through examination of concordance lines and discourse contexts. Sub-types are as follows: (1) reporting, (2) interpersonal, (3) reactive, (4) word/expression search,<sup>37</sup> (5) inquiry, and (6) other functions.

---

<sup>37</sup> Word/expression search bundles may seem, at first glance, to belong to the category of discourse organizers or referential expressions. However, as the analysis in Section 4.3.4.4 shows, the primary function of the word/expression search bundles is socio-interactional in nature. Rather than indicating the relationship between prior and coming discourse (discourse organizers) or making direct reference to an entity (referential expressions), the word/expression search bundles interactionally display the speaker's carefulness and deliberateness in choosing an appropriate word/expression or that the speaker is trying to access a concept or knowledge.

This section describes each of these sub-categories with examples. Table 4.6 gives the breakdown of these sub-types across three registers.

Table 4.6. Socio-interactional expression sub-types across three registers

	Conversation		Interview		Speech	
	Type	%	Type	%	Type	%
Reporting	15	30	21	27.3	11	73.3
Interpersonal	18	36	20	26.0	0	0
Reactive	13	26	14	18.2	0	0
Word/expression search	3	6	3	3.9	0	0
Inquiry	1	2	16	20.8	0	0
Other	0	0	3	3.9	4	26.7
Total	50	100	77	100.0	15	100.0

#### 4.3.4.1. Reporting bundles

Reporting bundles quote and report the speaker’s speech, thoughts, or stance, or someone else’s speech.<sup>38</sup> They are the most frequent sub-type in Interview (21/77; 27.3%) and in Speech (11/15; 73.3%), and they are the second most frequent sub-type in Conversation (15/50; 30%).

There are three major types of reporting bundles. The first type includes “self-directed” speech elements, such as *na* [PP], *oo* [VOL], and *tai* [want], before the quoting particle *to* and the verb *omou* ‘think’ (see Section 5.2 for the analysis of this type as a semi-fixed thought-reporting construction). The second type includes non-self-directed stance elements (e.g., *hitsuyoo* + *da* [necessary + COP], *taisetsu* + *da* [important + COP]) or other non-stance related elements before the quoting particle *to* and the verb *omou* ‘think’ or *kangaeru* ‘think.’ The third type, which only appears in Conversation and Interview, includes the quoting particle *to* and the verb *iu* ‘say.’

Below are the most frequently used bundles of the first type.

<sup>38</sup> For a bundle to be categorized as a reporting bundle, it has to include both a quotative particle and a quoting verb such as *omou* ‘think,’ *kangaeru* ‘think,’ or *iu* ‘say,’ with the exception of some utterance reporting bundles (e.g., *ta* + *n* + *da/desu* + *tte* [PST + NML + COP + QT]).

(4-115) Stance-related reporting bundles

<i>na</i> PP	<i>to</i> QT	<i>omot</i> think	<i>te</i> TE	<stance reporting>	(Conversation and Interview)
<i>oo</i> VOL	<i>to</i> QT	<i>omot</i> think	<i>te</i> TE	<stance reporting>	(Conversation and Interview)
<i>tai</i> want	<i>to</i> QT	<i>omoi</i> think	<i>masu</i> POL	<stance reporting>	(Speech)
<i>ka</i> Q	<i>to</i> QT	<i>omoi</i> think	<i>masu</i> POL	<stance reporting>	(Speech)

As explained in Section 4.3.1., stance-related expressions are regularly followed by the quoting particle *to* or *tte* and the quoting verb, *iu* ‘say,’ *omou* ‘think,’ or *kangaeru* ‘think.’ This means that many of the bundles that are classified under stance bundles are also, in part, reporting bundles. For example, the first bundle listed above, *na + to + omot + te* [PP + QT + think + TE] shares some morphemes with some of the stance bundles, *oo + ka + na + to* [VOL + Q + PP + QT] (Conversation), *ta + no + ka + na* [PST + NML + Q + PP] (Conversation), *ja + nai + ka + na* [COP + NEG + Q + PP] (Conversation and Interview), *nai + ka + na + to* [NEG + Q + PP + QT] (Interview), and *no + ka + na + to* [NML + Q + PP + QT] (Interview). Observe the following instance of the overlapping of the stance bundle *oo + ka + na + to* [VOL + Q + PP + QT] and the reporting bundle *na + to + omot + te* [PP + QT + think + TE].

(4-116) *oo + ka + na + to* [VOL + Q + PP + QT] and *na + to + omot + te* [PP + QT + think + TE] <stance reporting> (Conversation: Callhome\_1263)

1. B: 今日だから今からその団地を見に行こうと思ってるんだけど。
2. A: うん。
3. B: 何着て行こうかなと思って。短パンで歩いたら変かなとかさ。
4. A: なるほどね。まだ暑いんでしょう？
5. B: 暑いね。でもまだ大分ましなんだって。

1. B: *kyoo dakara ima kara sono danchi o mi ni ik oo to omot te ru n da kedo.*  
today so now from INJ apartment ACC look to go VOL QT think TE ASP NML COP but
  2. A: *un.*  
yes
  3. B: *nani ki te ik oo ka na to omot te. tanpan de arui tara hen ka na toka sa.*  
what wear TE ASP VOL Q PP QT think TE / shorts INS walk COND strange Q PP such PP
  4. A: *naruhodo ne. mada atsui n desho o ?*  
indeed PP / yet hot NML COP AUX
  5. B: *atsui ne. demo mada daibu mashi na n da tte.*  
hot PP / but still quite better COP NML COP QT
- 
1. B: ‘Today, so, I’m thinking of going to the apartment now to take a look.’
  2. A: ‘Yeah.’
  3. B: ‘I’m wondering what to wear. I’m thinking it may be strange if I walk in shorts.’
  4. A: ‘I see. It’s still hot, isn’t it?’
  5. B: ‘It is. But I heard it’s much better.’

As we can see in the modified example (4-117’) below, it is possible to say this expression without the accompaniment of *to + omou* [QT + think]. In fact, in Conversation, it is more common to find non-quoted self-directed stance expressions. The conversational participants have much more intimate relationship with each other than interviewers/interviewees and speech givers/audience, and they tend to display their private-self more freely than interviewees and speech givers in other two registers when expressing immediate thoughts or feelings. Nevertheless, the fact that the co-occurrence of private thought and the sequence *to omou* [QT think] is observed as part of morphemic bundles tells us that speakers in Conversation, as well as in Interview and Speech, frequently display “private-self” and “interactional-self” simultaneously.<sup>39</sup>

---

<sup>39</sup> For example, for the suffix *ka na* ‘wonder,’ in Conversation 201/902 instances (22.28%) are quoted, in Interview 89/190 (49.44%) are quoted, and in Speech 70/75 (93.33%) are quoted. Thus, in Conversation, about a quarter of the private-thoughts (represented by the suffix *ka na*) are quoted. In stark contrast, in Speech, more than 90% of the private-thoughts are quoted. Interview is somewhere between the two registers; half is quoted and the other half is non-quoted.

(4-117) Addressee-directed stance reporting (Conversation: Callhome\_1263)

何着て行こうかなと思って  
*nani ki te ik oo ka na to omot te*  
what wear TE go VOL Q PP QT think TE  
'(I'm) thinking (I) wonder what to wear.'

(4-117') Self-directed stance expression (Conversation: Calhome\_1263; modified)

何着て行こうかな  
*nani ki te ik oo ka na*  
what wear TE go VOL Q PP  
'I wonder what to wear.'

It is somewhat surprising and even puzzling that speakers in Speech also use this pattern frequently, as in (4-118) or (4-119) below. In Speech, it seems possible and appropriate to just make a statement without *tai* 'want' or *to + omou* [QT + think], as shown in (4-119') below. The simple announcement of what the speaker is going to do is information-oriented and sounds impersonal. As far as morphemic bundles are concerned, speakers in Speech seem to prefer making an announcement in the form of quoting and reporting their private thoughts. Unlike Conversation, however, it is pragmatically inappropriate, that is, it comes across as blunt or child-like, to omit the quotation with *to + omou* in Speech, as shown in (4-119''). The utterance *chigai o setsumei shite iki tai desu* '(I) want to explain the difference now' in (4-119'') merely expresses the speaker's wish that he wants to explain the difference. In the original utterance (4-119), on the other hand, the speaker makes an announcement that he is going to explain the difference. The difference between the self-quotation of the speaker's private thought, as in (4-119), and the simple statement, as in (4-119'), can be said to be socio-pragmatic in nature. That is, while both accomplish the same interactional action of making an announcement, they project different social images of the speakers. The former portrays the speaker as polite and personal whereas the latter depicts the speaker as impersonal and information-oriented.

(4-118) *tai + to + omoi + masu* <stance reporting/announcement> [want + QT + think + POL]  
(Speech: CSJ\_A03M0004)

えー基準を用いた場合のえ三つの要約例を示しその違いを説明していき  
たいと思います / では判断基準を日本語らしさのみとした場合の要約例を示  
します

*e: kijun o mochii ta baai no e mittsu no yooyaku rei o shimeshi sono chigai o setsumei shi*  
INJ standard ACC use PST case GEN INJ three GEN summary example ACC show that difference ACC explain do  
*te iki tai to omoi masu / dewa handan kijun o nihongo rashi sa nomi to shi ta baai*  
TE ASP want QT think POL / then judge standard ACC Japanese like SFX only QT do PST case GEN summary  
*no yooyaku rei o shimeshi masu*  
GEN example ACC show POL

‘Um, we’d like to show three summary examples which used the standard and explain their differences. Then, we will show the summary example which only used Japanese-ness as the judging criteria.’

(4-119) Announcement through stance reporting (Speech: CSJ\_A03M0004)

違いを説明していきたいと思います  
*chigai o setsumei shi te iki tai to omoi masu.*  
difference ACC explain do TE ASP want QT think POL  
‘(I) would like to explain the difference now.’

(4-119’) Simple announcement (Speech: CSJ\_A03M0004; modified\_1)

違いを説明していきます  
*chigai o setsumei shite iki masu.*  
difference ACC explain do TE ASP POL  
‘(I) will explain the difference now.’

(4-119’’) Simple stance expression (Speech: CSJ\_A03M0004; modified\_2)

違いを説明していきたいです  
*chigai o setsumei shi te iki tai desu.*  
difference ACC explain do TE go want COP:POL  
‘(I) want to explain the difference now.’

The sense of politeness derives from the effect of quoting one’s private thought with *to omou* (or *to kangaeru*). The direct quotation of one’s private thought indicates that the speaker

feels comfortable to show his or her private-self, and at the same time, it shows that the speaker is able to distance himself or herself from it to highlight his or her interactional-self. Here, *to + omou* is not a mere expression of the speaker's mental state, but it is part of the construction which serves as a politeness strategy. I will discuss this type of construction and its socio-interactional import in Chapter 5.

The second type of the reporting bundle includes impersonal stance elements (e.g., *hitsuyoo + da* [necessary + COP], *taisetsu + da* [important + COP]) or other non-stance related elements before the quoting particle *to* and the verb *omou* 'think' or *kangaeru* 'think.' Most frequently occurring bundles of the second type are listed below.

(4-120) Impersonal stance and non-stance reporting bundles

<i>to</i> QT	<i>omou</i> think	<i>n</i> NML	<i>da/desu</i> COP	<impersonal stance reporting>	(Conversation, Interview and Speech)
<i>da</i> COP	<i>to</i> QT	<i>omoi</i> think	<i>masu</i> POL	<impersonal stance reporting>	(Interview and Speech)
<i>to</i> QT	<i>kangaerare</i> think	<i>masu</i> POT	<i>masu</i> POL	<impersonal stance reporting>	(Speech)

These bundles report some information or evaluation as one's opinion. Unlike the first type, the omission of the quoting sequence *to + omou* [QT + think] in the second type does not turn the expression into self-directed speech, which displays one's private-self. Instead, the omission of the quoting sequence (*to + omou*) makes the utterance an assertive statement, as in (4-122') and (4-124') below. Therefore, the second type of the reporting bundle functions to qualify information or evaluation as one's opinion. This may also lead to downgrading or mitigating effect. Observe the following two examples, (4-121)/(4-122) and (4-123)/(4-124), with non-quoted counterparts, (4-122') and (4-124').

(4-121) *da + to + omoi + masu* [COP + QT + think +POL] <impersonal stance reporting>  
(Interview: Hypermedia\_024)

1. S: あの犯罪とかえー麻薬みたいなものがあの一教育の現場に入ってきてるっていうことで
2. N: ああああ
3. S: それほんとに深刻な問題だと思いますねー。
4. N: なるほど。

1. S: *ano hanzai toka e: mayaku mitai na mono ga ano: kyooiku no genba ni hait te ki te ru tte*  
INJ crime such INJ drug like COP NML NOM INJ education GEN scene DAT enter TE ASP TE ASP QT  
*iu koto de*  
say NML COP
2. N: *aa aa*  
INJ INJ
3. S: *sore honto ni shinkoku na mondai da to omoi masu ne:.*  
that real COP serious COP issue COP QT think POL PP
4. N: *haruhodo.*  
indeed

1. S: ‘Um, things such as crimes and, um, drugs are coming into the education scenes’
2. N: ‘Uh uh’
3. S: ‘I think that is a very serious problem.’
4. N: ‘I see.’

(4-122) Impersonal stance reporting (Interview: Hypermedia\_024)

ほんとに深刻な問題だと思いますねー

*honto ni shinkoku na mondai da to omoi masu nee.*  
really COP serious COP:ATT issue COP QT think POL PP  
‘(I) think (it) is a very serious issue.’

(4-122’) Simple expression of impersonal stance (Interview: Hypermedia\_024; modified)

ほんとに深刻な問題ですねー  
*honto ni shinkoku na mondai desu nee.*  
really COP serious COP:ATT issue COP:POL PP  
‘(It) is a very serious issue.’

(4-123) *da + to + omoi +masu* [COP + QT + think +POL] <impersonal stance reporting>  
(Speech: CSJ\_A09M0204)

最後 なんですが最もあのこの仮説をですね 今後 経験的に検証していくと え  
ーいうことが必要 **だ** と思います

*saigo na n desu ga mottomo ano kono kasetsu o desu ne kongo keiken teki ni kenshoo shi te*  
last COP NML COP but most INJ this hypothesis ACC COP PP after.this experience perspective DAT verify do TE  
*iku to e: iu koto ga hitsuyoo da to omoi masu*  
ASP QT INJ say NML NOM necessary COP QT think POL

‘Lastly, we think what is most necessary is that we verify the hypothesis empirically in the future.’

(4-124) Impersonal stance reporting (Speech: CSJ\_A09M0204)

検証していくと いうことが必要 **だ** と思います

*kenshoo shi te iku to iu koto ga hitsuyoo da to omoi masu*  
verify do TE ASP QT say NML NOM necessary COP QT think POL  
‘(We) think (it) is necessary that we verify (the hypothesis).’

(4-124’) Simple expression of impersonal stance (Speech: CSJ\_A09M0204; modified)

検証していくと いうことが必要 です  
*kenshoo shi te iku to iu koto ga hitsuyoo desu*  
verify do TE ASP QT say NML NOM necessary COP:POL  
‘(It) is necessary that we verify (the hypothesis).’

The third sub-type of the reporting bundle, which only appears in Conversation and Interview, includes the quoting particle *to* and the verb *iu* ‘say.’ Below is the most frequently used bundle from Conversation.

(4-125) Utterance reporting bundle

<i>tte</i>	<i>it</i>	<i>te</i>	<i>ta</i>	<utterance reporting>	(Conversation)
QT	say	TE	PST		

The utterance reporting bundle reports what someone else has said, and in some cases, works to convey one's uncertainty, as in (4-126) and (4-127), or to put blame on someone, as in (4-128).

In (4-126), the speaker B reports to A in line 1 that their mutual acquaintance Mika has quit her job. A expresses her surprise in lines 2, and asks for a confirmation in line 4. In line 5, B simply repeats the quotative particle and verb to attribute the source of information to Mika, avoiding confirming the information as true. By using the bundle, B conveys her uncertainty whether Mika has actually quit her job.

(4-126) *tte + it + te + ta* [QT + say + TE + PST] <utterance reporting/uncertainty>  
(Conversation: Callhome\_1538)

1. B: なんかねみかちゃんは仕事辞めたとかいったよ。
2. A: うそ！
3. B: うん。
4. A: 辞めちゃったの。
5. B: っていってたよ。
6. A: 看病しているのお婆さんの？
7. B: 分かんないよく

1. B: *nanka ne mika chan wa shigoto yame ta toka it te ta yo.*  
DM PP (name) SFX TOP work quit PST such say TE PST PP
2. A: *uso!*  
lie
3. B: *un.*  
yes
4. A: *yame chat ta no.*  
quit ASP PST PP
5. B: *tte it te ta yo.*  
QT say TE PST PP
6. A: *kanbyoo shi te iru no obasan no?*  
nurse do TE ASP PP aunt GEN
7. B: *wakan nai yoku*  
know NEG well

1. B: 'You know, Mika said she quit her job.'
2. A: 'No way!'
3. B: 'Yeah.'
4. A: 'She quit?'
5. B: 'She said so.'
6. A: 'Is she nursing her aunt?'

7. B: 'I don't know well.'

In (4-127), the speaker C reports the information that a parcel should arrive in Japan within one week or so as a hearsay, which conveys a lack of confidence in C in regards to the validity of the information.

(4-127) *tte + it + te + ta* [QT + say + TE + PST] <utterance reporting/uncertainty>

(Conversation: Callhome\_1731)

1. C: なんかいつもねファックスで送ると
2. A: うん。
3. C: 日本にも一週間ぐらいで届いちゃうんですって。
4. A: ああそう！うん。
5. C: うん。だからね多分届くんじゃないかっていってたんだけど。
6. A: そうよね。

1. C: *nanka itsumo ne fakkusu de okuru to*  
DM always PP fax INS send COND
2. A: *un.*  
yes
3. C: *nihon ni mo is shuukan gurai de todoi chau n desu tte.*  
Japan DAT also one week about COP arrive ASP NML COP QT
4. A: *aa soo! un.*  
INJ so / yes
5. C: *un. dakara ne tabun todoku n ja nai ka tte it te ta n da keredo.*  
yes / so PP probably arrive NML COP NEG Q QT say TE PST NML COP but
6. A: *soo yo ne.*  
so PP PP

1. C: 'Like, when she sends her order by fax'
2. A: 'Yeah.'
3. C: 'She said it always arrives in Japan within one week or so.'
4. A: 'Oh really! Yeah.'
5. C: 'Yeah. So she said it probably arrives.'
6. A: 'I agree.'

In (4-128), F062 blames her teacher E for telling her that cakes at a certain shop are delicious.

(4-128) *tte + it + te + ta* [QT + say + TE + PST] <utterance reporting/blaming> (Conversation: BTS\_71)

1. F062:      なんかさー E 先生はさーおいしいっていったからさーそうやっぱ。
2. F161:      私も超あの人のこと信用して食べたらーうんうーんて感じで。
3. F062:      そううんまあおいしいーにはおいしいけどまー他にもおいしいのはあるよねみたいなそんな感じだった。

1. F062:      *nanka sa: E sensei wa sa: oishii tte it te ta kara sa: soo yappa.*  
DM PP (name) teacher TOP PP delicious QT say TE PST because PP so after.all
2. F161:      *watashi mo choo ano hito no koto shinyoo shi te tabe tara: un u:n te kanji de.*  
I also very that person GEN NML trust do TE eat COND yes INJ QT feel COP
3. F062:      *soo un maa oishii: hi wa oishii kedo ma: hoka ni mo oishii no wa aru yo ne*  
so yes INJ delicious DAT TOP delicious but INJ other DA also delicious NML TOP exist PP PP  
*mitai na sonna kanji dat ta.*  
like COP that.kind feel COP PST

1. F062:      ‘Like, Ms. E said it is delicious, you know, so after all.’
2. F161:      ‘I also trusted that person and ate it, but it was so-so.’
3. F062:      ‘Right, yeah. Well, it was tasty alright but there are other tasty things like that.’

#### 4.3.4.2. Interpersonal bundles

Interpersonal bundles are the most frequent interactional bundle sub-type in Conversation (18/50; 43.6%) and the second most frequent sub-type in Interview (20/77; 26%). It is not surprising that the interpersonal bundles do not appear at all in Speech. Interpersonal bundles involve so-called interactional or pragmatic particles or sequences, *yo*, *ne*, *yo + ne*, *ja + nai*, and *ja + nai + no*.

(4-129) Interpersonal bundles

<i>n</i>	<i>da/desu</i>	<i>yo</i>	<i>ne</i>	<interpersonal>	(Conversation and Interview)
NML	COP	PP	PP		

<i>na</i>	<i>n</i>	<i>da/desu</i>	<i>yo</i>	<interpersonal>	(Conversation and Interview)
COP:ATT	NML	COP	PP		

*ta n desu yo* <interpersonal> (Conversation and Interview)  
 PST NML COP PP

*n ja nai no* <interpersonal> (Conversation)  
 NML COP NEG PP

*ta n desu ne* <interpersonal> (Interview)  
 PST NML COP PP

Below are three examples of the use of interpersonal bundles. Two excerpts are from Conversation and one is from Interview.

(4-130) *n + da + yo + ne* [NML + COP + PP + PP] <interpersonal> (Conversation: Callhome\_2237)

1. A: あれあたし朝早く起きれない質だからいいわ。
2. B: うん。あっそうだ。そうだよね時差があるんだよね考えてみたら。
3. A: うん。そう。時差ボケでねうんよく寝てます朝は。
4. B: うんうんうん。

1. A: *are atashi asa hayaku okir e nai shitsu da kara ii wa.*  
 that I morning early get.up POT NEG nature COP because good PP
2. B: *un. a soo da. soo da yo ne jisa ga aru n da yo ne kangae te mi tara.*  
 yes INJ so COP so COP PP PP time.difference NOM exist NML COP PP PP think TE ASP COND
3. A: *un. soo. jisa boke de ne un yoku ne te masu asa wa.*  
 yes / so / time.difference lag COP PP yes well sleep TE POL morning TOP
4. B: *un un un.*  
 yes yes yes

1. A: 'I cannot get up early in the morning, so that is fine.'
2. B: 'Yeah. Oh, that's right. That's right, there is time difference when I come to think of it.'
3. A: 'Yeah, right. I will have a jet lag, so I will be sleeping well in the morning.'
4. B: 'Yeah yeah yeah.'

(4-131) *n + desu + yo + ne* [NML + COP + PP + PP] <interpersonal> (Interview: Tetsuko\_10)

1. K: スギムラ先生亡くなった時どんな気持ちでした？
2. E: 僕はちょうどその芝居をしてたんですよね
3. K: うん。
4. E: 奈良で訃報を聞いたんですけどねでその前の入院してらっしゃる時に見舞いにも行ったんですよ何回かね。
5. K: うんうん。

1. K: *sugimura sensei nakunat ta toki donna kimochi deshi ta?*  
(name) teacher pass.away PST when what.kind feeling COP PST
2. E: *boku wa choodo sono shibai o shi te ta n desu yo ne*  
I TOP just INJ theater ACC do TE PST NML COP PP PP
3. K: *un.*  
yes
4. E: *nara de fuhoo o kii ta n desu kedo ne de sono mae no nyuuin shi te*  
(name) LOC death.news ACC hear PST NML COP but PP and that before GEN hospitalize do TE  
*rassharu toki ni mimai ni mo it ta n desu yo nan kai ka ne.*  
ASP:HON time DAT visit DAT also go PST NML COP PP how time Q PP
5. K: *un un.*  
yes yes

1. K: 'How did you feel when Mr. Sugimura passed away?'
2. E: 'I was just performing in a play at that time.'
3. K: 'Yeah.'
4. E: 'I heard the news of his death in Nara, and before that, while he was being hospitalized, I went to visit him a few times.'
5. K: 'Yeah yeah.'

(4-132) *n + ja + nai + no* [NML + COP + NEG + PP] <interpersonal> (Conversation: Japan\_Ojoosama)

1. M: そんな政略結婚させるような親の心配なんかしなくてもいいんじゃないの？
  2. K: でもやっぱり親は親っていうかー
  3. M: おととととと うん。
1. M: *sonna seiryaku kekkon s aseru yoo na oya no shinpai nanka shi naku te mo ii n*  
that.kind political.tactics marriage do CAUS like COP parent GEN worry DM so NEG TE also good NML  
**ja nai no?**  
COP NEG PP
  2. K: *demo yappari oya wa oya tte iu ka:*  
but after.all parent TOP parent QT say Q
  3. M: *ottottotto un.*  
oops yes
1. M: 'She doesn't have to worry about such parents who marry her off for political reasons, don't you think?'
  2. K: 'But after all, they are her parents.'
  3. M: 'Oops, yeah.'

Even though the bundles that end with one of the pragmatic particles are categorized as

interpersonal bundles, the interpersonal functions are attributed to the particle themselves and not

to the whole bundle. It implies a limitation of the morphemic bundle analysis. This issue is addressed in Section 4.6.

#### 4.3.4.3. Reactive bundles

Reactive bundles are the third most frequent sub-type of the socio-interactional bundle in Conversation (13/50; 26%) and the fourth most frequent sub-type in Interview (14/77; 18.2%). They do not occur in Speech. Reactive bundles are also referred to as *aizuchi* or back-channeling, which is employed by the addressees to show their engaged listenership. The following are the most frequent bundles in this category.

##### (4-133) Reactive bundles

<i>soo</i> so	<i>soo</i> so	<i>soo</i> so	<i>soo</i> so	<reactive>	(Conversation)
<i>un</i> yes	<i>un</i> yes	<i>un</i> yes	<i>un</i> yes	<reactive>	(Conversation and Interview)
<i>soo</i> so	<i>na</i> COP:ATT	<i>n</i> NML	<i>da/desu</i> COP	<reactive>	(Conversation and Interview)
<i>aa</i> INJ	<i>soo</i> so	<i>desu</i> COP	<i>ka</i> Q	<reactive>	(Interview)
<i>soo</i> so	<i>da/desu</i> COP	<i>yo</i> PP	<i>ne</i> PP	<reactive>	(Conversation and Interview)

The bundle *soo + soo + soo + soo* [so + so + so + so] only appears in Conversation. It is the second most frequent bundle type in Conversation. In many cases, it has the function of confirming the previously stated information. In (4-134), for example, 11G asks for a confirmation about some information in line 1. In line 2, 11H confirms the information given by 11G through the use of the bundle.

(4-134) *soo + soo + soo + soo* [so + so + so + so] <reactive/confirmation> (Conversation: Work\_11F)

1. 11 G: あそうか 20 日出発で 25 日だけ。
2. 11 H: そう そう そう そう。
3. 11 A: あそんな早かったんだ。
4. 11 H: そうだねー。

1. 11 G: *a soo ka hatsu ka shuppatsu de nijuugo nichi dak ke.*  
INJ so Q twenty date departure COP twenty.five date COP Q
2. 11 H: *soo soo soo soo.*  
*so so so so*
3. 11 A: *a sonna ni hayakat ta n da.*  
INJ that.much DAT early PST NML COP
4. 11 H: *soo da ne:.*  
so COP PP

1. 11 G: ‘Oh, right, the departure is on the 20<sup>th</sup> and the 25<sup>th</sup>, isn’t it?’
2. 11 H: ‘Right right right right.’
3. 11 A: ‘Oh it was that early.’
4. 11 H: ‘Yeah, right.’

The bundle *un + un + un + un* [yes + yes + yes + yes] appears in both Conversation and Interview. Unlike the confirmation function of *soo + soo + soo + soo*, the bundle *un + un + un + un* functions as a continuer, signaling to the current main speaker that he/she may go on. In (4-135), F086 provides information about a type of bread in line 1. In line 2, F046 responds with the bundle, which is functioning as a continuer for F086 to continue her talk about the bread.

(4-135) *un + un + un + un* [yes + yes + yes + yes] <reactive/continuer> (Conversation: BTS\_072)

1. F086: そう。しかも夕張メロンのパンでねー。
2. F046: うん うん うん うん
3. F086: 中にクリームが入ってるの。
4. F046: マジで？超おいしそう。
5. F086: 超うまいよ

1. F086: *soo. shikamo yuubari meron no pan de ne:.*  
so moreover (name) melon GEN bread COP PP

2. F046: *un un un un.*  
yes yes yes yes
3. F086: *naka ni kuri:mu ga hait te ru no.*  
inside LOC cream NOM enter TE ASP PP
4. F046: *maji de? choo oishi soo.*  
real COP / very delicious seem
5. F086: *choo umai yo.*  
very delicious PP
1. F086: ‘Right. Moreover, it’s Yuubari melon bread.’
2. F046: ‘Yeah yeah yeah yeah.’
3. F086: ‘There is cream inside.’
4. F046: ‘Really? It sounds so delicious.’
5. F086: ‘It’s so delicious.’

The bundle *soo + na + n + da/desu* [that + COP:ATT + NML + COP] appears in Conversation and Interview, but further inspection reveals that the bundles with the short and long forms of the copula have different collocational patterns and functions. Interview contains the bundle *soo + na + n + da* as well as *soo + na + n + desu*. The following two Interview excerpts, (4-136) and (4-137), each involve both types of the bundles appearing in sequence. In both excerpts, *soo + na + n + desu*, with the long form of the copula *desu*, is used to confirm the information which belongs to the speaker’s territory of information (Kamio 1997). On the other hand, *soo + na + n + da*, with the short form of the copula *da*, is used to indicate the speaker’s change of state in his/her knowledge status about the information belonging to the addressee’s territory of information. The bundle *soo + na + n + da* frequently co-appears with the change-of-state token *a* ‘oh’ (Tanaka 1999; cf. Ikeda 2007) and indicates the speaker’s new awareness based on the information provided by the addressee in the previous context.

(4-136) *soo + na + n + desu* [that + COP:ATT + NML + COP] <reactive/confirmation>; *soo + na + n + da* [that + COP:ATT + NML + COP] <reactive/change of state> (Interview: Tetsuko\_6)

1. K: *でその九州の大分のお生まれ*
2. S: *中国で生まれたんですけど*

3. K: あお生まれは中国 なんですか！
4. S: そうなんです。
5. K: あ そうなんだ。

1. K: *de sono kyuushuu no ooita no o umare*  
and INJ (name) GEN (name) GEN PFX birth
2. S: *chuugoku de umare ta n desu kedo*  
China LOC born PST NML COP but
3. K: *a o umare wa chuugoku na n desu ka!*  
INJ PFX birth TOP China COP NML COP Q
4. S: ***soo na n desu.***  
**so COP NML COP**
5. K: ***a soo na n da.***  
**INJ so COP NML COP**

1. K: ‘And you were born in Oita in Kyushu’
2. S: ‘I was born in China’
3. K: ‘Oh, you were born in China!’
4. S: ‘That’s right.’
5. K: ‘Oh, I see.’

(4-137) *soo + na + n + da* [that + COP:ATT + NML + COP] <reactive/change of state>; *soo + na + n + desu* [that + COP:ATT + NML + COP] <reactive/confirmation> (Interview: CSJ\_D01F0023)

1. R: うち妹はすごくねあの一海外好きで
2. L: へーええええ
3. R: うー行ってるんですけど私は結局後にも先にもあの中国にしか行ってなくて
4. L: あ そうなんだ
5. R: そうなんです

1. R: *uchi imooto wa sugoku ne ano: kaigai suki de*  
home younger.sister TOP very PP INJ overseas like COP
2. L: *he: ee ee*  
INJ yes yes
3. R: *u: it te ru n desu kedo watashi wa kekkyoku ato ni mo saki ni mo ano chuugoku ni shika*  
INJ go TE ASP NML COP but I TOP after.all later DAT also early DAT also INJ China DAT only  
*it te naku te*  
go TE NEG TE
4. L: ***a soo na n da***  
**INJ so COP NML COP**
5. R: ***soo na n desu***  
**so COP NML COP**

1. R: ‘My younger sister really likes overseas’

2. L: 'Oh, yes yes.'
3. R: 'Um, she is going to foreign countries, but I have only been to China'
4. L: 'Oh I see.'
5. R: 'That's right.'

The bundle *aa + soo + desu + ka* [ah + so + COP + Q] is the most frequently appearing bundle in Interview. Interestingly, the bundle co-appears with another so-called change-of-state token *aa* 'ah' much more frequently than with *a* 'oh.'<sup>40</sup> Examination of actual contexts reveals that *aa + soo + desu + ka* is frequently used by interviewers to display their attentiveness and provide acknowledging receipt of the information provided by the addressee in the prior turn. Moreover, after the production of this bundle, the interviewer tends to continue with a related question. It suggests that the collocational patterns between specific interjectory tokens (e.g., *a* and *aa*) and reactive expressions (e.g., *soo + na + n + da* and *soo + desu + ka*) are more or less fixed and that the patterns are attributed to the particular interactional functions they play as a whole sequence.

(4-138) *aa + soo + desu + ka* [INJ + so + COP + Q] <reactive/acknowledging receipt>  
(Interview: Hypermedia\_006)

1. M: えっと何を専攻してらっしゃいますか。
  2. Y: はい今の専攻はあの一言語学とか語-語学ですね。英語を専攻しようと思ってますけど。
  3. M: ああそうですか。
  4. Y: はい。
  5. M: えっとどうして英語を勉-ま専攻なさりたいっていう風にお思いになったんですか。
- 
1. M: *etto nani o senkoo shi te rasshai masu ka.*  
INJ what ACC major do TE ASP POL Q
  2. Y: *hai ima no senkoo wa ano: gengo gaku te ka go- gogaku desu ne. eigo o senkoo shiyo o*  
yes now GEN major TOP INJ language study QT Q language.study COP PP / English ACC major do VOL

<sup>40</sup> The token *aa* 'ah' is more than three times frequent than *a* 'oh.' The two change-of-state tokens, *aa* and *a*, are not usually distinguished (e.g., Tanaka 1999: 38-39; Saft 2011). I recognize the difficulty in distinguishing the two tokens phonetically (vowel length). The distinction made in the present paper is based on the provided transcription. It requires a detailed acoustic analysis of the tokens as well as a more comprehensive qualitative analysis of their uses to warrant the claim made in the paper.

*to omot te masu kedo.*

QT think TE POL but

3. M: *aa soo desu ka.*  
INJ so COP Q
4. Y: *hai.*  
yes
5. M: *e dooshite eigo o ben- ma senkoo nasari tai tte iu fuu ni o omoi ni nat ta n desu*  
INJ why English ACC INJ major do:HON want QT say way COP PFX think COP become PST NML COP  
*ka?*  
Q

1. M: 'Um, what are you majoring in?'
2. Y: 'Yes, my current major is linguistics, or language. I'm thinking of majoring in English.'
3. M: 'Ah I see.'
4. Y: 'Yes.'
5. M: 'Uh, why did you think you want to major in English?'

(4-139) *aa + soo + desu + ka* [INJ + so + COP + Q] <reactive/acknowledging receipt>  
(Interview: Tetsuko\_9)

1. O: ま そう いった プログラム が あった ので
2. K: うん。
3. O: 参加 さ せて いただ いた ん です けれど も。
4. K: ああ そ う です か。
5. O: ええ。
6. K: いかが でした ？
7. O: ああ 楽し かった です ね。

1. O: *ma soo it ta puroguramu ga at ta node*  
INJ that say PST program NOM exist PST because
2. K: *un.*  
yes
3. O: *sanka s ase te itadai ta n desu keredomo.*  
participate do CAUS TE ASP PST NML COP but
4. K: *aa soo desu ka.*  
INJ so COP Q
5. O: *ee.*  
yes
6. K: *ikaga deshi ta?*  
how COP PST
7. O: *aa tanoshikat ta desu ne.*  
INJ fun PST COP PP

1. O: 'Well, there was such program'
2. K: 'Yeah.'
3. O: 'I participated in it.'
4. K: 'Ah I see.'

5. O: 'Yes.'
6. K: 'How was it?'
7. O: 'Ah, it was fun.'

The last bundle type to be examined in this sub-section of reactive bundles is *soo + da/desu + yo + ne* [so + COP + PP + PP]. In terms of the territory of information, this bundle is used to display an agreement with the addressee regarding the information which the speaker thinks belongs to the addressee's territory of information. The first two excerpts below, (4-140) and (4-141), are from Interview, and the next two, (4-142) and (1-143) are from Conversation.

(4-140) *soo + desu + yo + ne* [so + COP + PP + PP] <reactive/agreement> (Interview: Tetsuko\_4)

1. K: でもほんとお母様お元気そうですね。
2. I: そうですね。やはりもう母が元気なのが一番うれしいです。
3. K: ほんとに**そう**ですよ。

1. K: *demo honto okaasama o genki soo desu ne.*  
but really mother PFX fine seem COP PP
2. I: *soo desu ne. yahari moo haha ga genki na no ga ichiban ureshii desu.*  
so COP PP / after.all DM mother NOM fine COP NML NOM most happy COP
3. K: *honto ni soo desu yo ne.*  
real DAT so COP PP PP

1. K: 'But your mother seems fine.'
2. I: 'Right. After all, what makes me the happiest is that my mother is fine.'
3. K: 'You are really right.'

(4-141) *soo + desu + yo + ne* [so + COP + PP + PP] <reactive/agreement> (Interview: CSJ\_D01F0049)

1. R: 最近ほらあの津軽三味線とかはやってるじゃないですか
2. L: うんうんうんはやってますよね
3. R: あんなに弾けたら本当楽しいだろ うなと うーん
4. L: **そう**ですよ
5. R: 思います

1. R: *saikin hora ano tsugaru shamisen toka hayat te ru ja nai desu ka*  
recently INJ INJ (name) samisen such be.popular TE ASP COP NEG COP Q

2. L: *un un un hayat te masu yo ne*  
yes yes yes be.popular TE POL PP PP
3. R: *anna ni hike tara honto tanoshii daro o na to u:n*  
that.kind DAT play COND really fun COP AUX PP QT INJ
4. L: *soo desu yo ne*  
**so COP PP PP**
5. R: *omoi masu ne*  
think POL PP

1. R: ‘You know, recently, um, Tsugaru shamisen is popular, isn’t it?’
2. L: ‘Yeah yeah yeah, it’s popular.’
3. R: ‘If I could play that much, it would be so much fun, um’
4. L: ‘That’s right.’
5. R: ‘I think.’

(4-142) *soo + da + yo + ne* [so + COP + PP + PP] <reactive/agreement> (Conversation: Callhome\_1263)

1. B: うんでも困ったもんだよ。
2. A: 本当？
3. B: 大きな子供が二人いるみたいな感じ。
4. A: 本当。
5. B: うん。
6. A: そうだよね。
7. B: そうだよ。

1. B: *un demo komat ta mon da yo.*  
yes but have.hard.time PST NML COP PP
2. A: *hontoo?*  
really
3. B: *ooki na kodomo ga futari iru mitai na kanji.*  
big COP child NOM two exist like COP impression
4. A: *hontoo.*  
really
5. B: *un.*  
yes
6. A: *soo da yo ne.*  
**so COP PP PP**
7. B: *soo da yo.*  
so COP PP

1. B: ‘Yeah, but it’s troubling.’
2. A: ‘Really?’
3. B: ‘It’s like there are two big kids.’
4. A: ‘Really.’
5. B: ‘Yeah.’

6. A: 'That's right.'  
 7. B: 'Right.'

(4-143) *soo + da + yo + ne* [so + COP + PP + PP] <reactive/agreement> (Conversation: Japan\_Broccoli)

1. A: そんな調べ-でも多分-このまま植えるんじゃないかと思うよ。  
 2. B: そうだよね。  
 3. A: うん。だから時期あるじゃん。それにもう咲いたばかりだから-今すぐ植えても多分もうすぐには出ないんじゃないか。  
 4. B: そうだよね。

1. A: *sonna shirabe- demo tabun: kono mama ueru n ja nakat ta to omou yo.*  
 that.much but probably this as plant NML COP NEG PST QT think PP  
 2. B: *soo da yo ne.*  
 so COP PP PP  
 3. A: *un. dakara jiki aru jan. soreni moo sai ta bakkari da kara: ima sugu ue te mo tabun*  
 yes so time exist AUX / moreover already bloom PST just COP because now soon plant TE also probably  
*moo sugu ni wa de nai n ja nai ka.*  
 already soon DAT TOP come.out NEG NML COP NEG Q  
 4. B: *soo da yo ne.*  
 so COP PP PP

1. A: 'That much, but I don't think we should probably plant it as is.'  
 2. B: 'That's right.'  
 3. A: 'Yeah. So there is an appropriate time to plant. Moreover, it has just bloomed, so even if we plant it now, it probably won't come out soon.'  
 4. B: 'That's right.'

#### 4.3.4.4. Word/expression search bundles

Word/expression search bundles also appear in Conversation and Interview. In

Conversation, the following three bundles are observed.

(4-144) Word/expression search bundles in Conversation

*nan te iu no* <word/expression search>  
 what QT say NML

(*nan*) *te iu no ka* <word/expression search>  
 QT say NML Q

<i>iu</i>	<i>no</i>	<i>ka</i>	<i>na</i>	<word/expression search>
say	NML	Q	PP	

More than 60% of the first bundle *nan + te + iu + no* [what + QT + say + NML] is free standing (i.e., used as a chunk on its own, meaning ‘how to say’), and close to 27% is part of the longer sequence, *nan + te + iu + no + ka + na* [what + QT + say + NML + Q + PP] ‘(I) wonder how to say.’ The other two bundles, *te + iu + no + ka* [QT + say + NML + Q] and *iu + no + ka + na* [say + NML + Q + PP], constitute part of the longer sequence, *nan + te + iu + no + ka + na* [what + QT + say + NML + Q + PP] ‘(I) wonder how to say.’ The next excerpts, (4-145) and (4-146) involve the shorter *nan + te + iu + no*, while (4-147) and (4-148) involve the longer sequence *nan + te + iu + no + ka + na*. Both types of expression tend to co-appear with emphatic expressions, such as *sugoi* ‘really,’ *shikamo* ‘in addition,’ and *honto ni* ‘really.’

(4-145) *nan + te + iu + no* [what + QT + say + NML] <expression search> (Conversation: Callhome\_1593)

1. A: で自分の英語がすごいなんていうの自信がなくて
2. B: うん。
3. A: はたしてああいう所で働けるかどうか。
4. B: ああそれは大丈夫よ。

1. A: *de jibun no eigo ga sugoi nan te iu no jishin ga nakut te*  
and oneself GEN English NOM very what QT say NML confidence NOM NEG TE
2. B: *un.*  
yes
3. A: *hatashi te aa iu tokoro de hatarak eru ka doo ka.*  
fulfill TE that say place LOC work POT Q how Q
4. B: *aa sore wa daijoobu yo.*  
INJ that TOP fine PP

1. A: ‘And, I really don’t have, how to say, confidence in my English’
2. B: ‘Yeah.’
3. A: ‘Whether I can really work at such place.’
4. B: ‘Ah, that’s fine.’

(4-146) *nan + te + iu + no* [what + QT + say + NML] <expression search> (Conversation: BTS\_72)

1. F086: そしたらなんかすごい広くてね—しかもなんか **なんていうの**二人つきりになれる部屋みたいのがあんの。

2. F046: あっすごいうん。

1. F086: *soshitara nanka suggoi hiroku te ne: shikamo nanka nan te iu no futari kkiri ni*  
then DM very large TE PP moreover DM **what QT say NML** two only COP  
*nar eru heya mitai no ga an no.*  
become POT room like NML NOM exist PP

2. F046: *a sugoi un.*  
INJ great yes

1. F086: ‘Then, you know, there is like a room that is so large, and moreover, like, how to say, where you two can be alone.’

2. F046: ‘Wow, that’s great yeah.’

(4-147) *nan + te + iu + no + ka + na* [what + QT + say + NML + Q +PP] <expression search> (Conversation: Callhome\_2199)

1. A: それで一番上の息子だから

2. B: うん。うん。

3. A: 親がすごいなんて **なんていうの**かな期待が大きい訳。

4. B: うん。

5. A: うん。だからその親の期待にね

1. A: *sorede ichiban ue no musuko da kara*  
and most top GEN son COP because

2. B: *un un.*  
yes yes

3. A: *oya ga sugoi nan te nan te iu no ka na kitai ga ookii wake.*  
parent NOM really **what QT what QT say NML Q PP** expectation NOM big PP

4. B: *un.*  
yes

5. A: *un. dakara sono oya no kitai ni ne*  
yes so that parent GEN expectation

1. A: ‘And because he is the oldest son’

2. B: ‘Yeah yeah.’

3. A: ‘His parents, really, how to say, they have high expectations of him.’

4. B: ‘Yeah.’

5. A: ‘Yeah, so for the parents’ expectations’

(4-148) *nan + te + iu + no + ka + na* [what + QT + say + NML + Q + PP] <expression search>  
(Conversation: BTS\_33)

1. F002: よかったわねー。
2. F082: ええ。でもこのごろねー私はほんとになんていうのかなあの一  
とてもいい出会いが多いんですよね。

1. F002: *yokat ta wa ne:*.  
good PST PP PP
2. F082: *ee. demo kono goro ne: watashi wa honto ni nan te iu no ka na ano: totemo ii*  
yes but this about PP I TOP real DAT what QT say NML Q PP INJ very good  
*deai ga ooi n desu yo ne.*  
encounter NOM many NML COP PP PP

1. F002: 'I'm glad for you.'
2. F082: 'Yeah. But lately, I, really, how to say, um, I have many good encounters.'

It is interesting that the word/expression search bundles often co-appear with emphatic expressions, such as *sugoi* 'really' and *honto ni* 'really' (see above excerpts). When the speaker is emotionally involved in or feels strongly about what is to be expressed, he/she seems to become more careful and deliberate in choosing words. The bundle *nan + te + iu + no* [what + QT + say + NML] and the longer sequence *nan + te + iu + no + ka + na* [what + QT + say + NML + Q +PP] function to indicate that the speaker is currently in the process of searching for an appropriate expression. This differs from searching for a concept or knowledge. In fact, there seems to be a functional division between the sequence *nan + te + iu + no* [what + QT + say + NML] (or the longer sequence *nan + te + iu + no + ka + na* [what + QT + say + NML + Q +PP]) and a similar expression *nan + dak + ke* [what + COP + AUX]. As shown in the following examples, *nan + dak + ke* [what + COP + AUX] is used to signal that the speaker is trying to access a concept or knowledge. While in the use of *nan + te + iu + no* [what + QT + say + NML], the addressee waits for the speaker to resolve the issue on his or her own, in the case of *nan + dak + ke* [what + COP + AUX], the search is sometimes done collaboratively (see excerpts (4-151) and (4-152)). This is because the addressee understands that the speaker is

simply trying to access knowledge, not trying to find/formulate an appropriate wording or expression.

(4-149) *nan + dak + ke* [what + COP + AUX] <concept search> (Conversation: Callhome\_2085)

1. A: なんの癌？
2. B: ええとねーえー
3. A: 膵臓癌かなんか？
4. B: ああ肺癌かな？
5. A: あっそう。
6. B: あっ違う違うあれあの**なんだ**っけ？えー食腸癌。
7. A: あっそう。

1. A: *nan no gan?*  
what GEN cancer
2. B: *eeto ne: e:*  
INJ PP INJ
3. A: *suizoo gan ka nanka?*  
pancreas cancer Q something
4. B: *aa hai gan ka na?*  
INJ lung cancer Q PP
5. A: *a soo.*  
INJ so
6. B: *a chigau chigau are ano nan dak ke? e: shokuchoo gan.*  
INJ different different that INJ **what COP AUX** / INJ intestine cancer
7. A: *a soo.*  
INJ so

1. A: 'What cancer?'
2. B: 'Um, uh'
3. A: 'Is it pancreatic cancer or something?'
4. B: 'Uh, is it lung cancer?'
5. A: 'Oh, I see.'
6. B: 'Oh, no no, that, um, what is it, um, intestine cancer.'
7. A: 'Oh I see.'

(4-150) *nan + dak + ke* [what + COP + AUX] <concept search> (Conversation: Callhome\_2004)

1. B: アメリカへ行ってお兄さんがアメリカに留学してたんだよ。
2. A: うん。
3. B: あのうニューヨークじゃなくて**なんだ**っけ？ロスロス。
4. A: へえ！

1. B: *amerika e it te oniisan ga amerika ni ryuugaku shi te ta n da yo.*  
America ALL go TE older.brother NOM America DAT study.abroad do TE PST NML COP PP
2. A: *un.*  
yes
3. B: *anoo nyu:yo:ku ja naku te nan dak ke? rosu rosu.*  
INJ (name) COP NEG TE what COP AUX / (name) (name)
4. A: *hee!*  
INJ

1. B: 'His brother was studying abroad in America.'
2. A: 'Yeah.'
3. B: 'Um, it was not New York, what was it, Los Angeles Los Angeles.'
4. A: 'Wow!'

(4-151) *nan + dak + ke* [what + COP + AUX] <concept search> (Conversation: Japan\_Geshuku)

1. C: あれ **なんだ** っけ？米が入ってるの **なんだ** っけ？
2. T: ん 玄米 茶。
3. C: そう そう それ 飲んだんだよ。

1. C: *are nan dak ke? kome ga hait te ru no nan dak ke?*  
that what COP AUX / rice NOM enter TE ASP NML what COP AUX
2. T: *n genmai cha.*  
what brown.rice tea
3. C: *soo soo sore non da n da yo.*  
so so that drink PST NML COP PP

1. C: 'That, what is it, what is the one which has rice in it?'
2. T: 'What? brown rice tea.'
3. C: 'Right right, I drank that.'

(4-152) *nan + dak + ke* [what + COP + AUX] <concept search> (Conversation: Callhome\_1069)

1. B: えーあの **なんだ** っけ。
2. A: アットのマークね。はい。
3. B: そう そう そう そう。でインシアド。

1. B: *e: ano nan dak ke.*  
INJ INJ what COP AUX
2. A: *atto no ma:ku ne. hai.*  
at GEN mark PP / yes
3. B: *soo soo soo soo. de inshiado.*  
so so so so / and (name)

1. B: 'Um, uh, what is it.'



feels strongly about the subject. The sequence displays the speaker's carefulness or deliberateness in verbalizing the sensitive matter.

(4-154) *nan + te + iu + n + desu + ka* [what + QT + say + NML + COP + Q] <expression search> (Interview: Hypermedia\_015)

1. M: あの生活面でね韓国の学生さんが日本で生活する時にあのアパートを借りたりする時に問題が起こるってことは全然ありませんか？
2. W: あの前一
3. M: うん
4. W: そうですねご-あのー56年前ですかあの勤めていた学校ではあのーとても大変であの不動産屋に断られるケースが
5. M: うん
6. W: 度々あってあのそれですごくなんていうんですかショックを受けた学生ーに対する対応ーうまあ大変でしたけれども最近もうあまりそういう話は聞きません。

1. M: *ano sekatsu men de ne kankoku no gakusei san ga nihon de seikatsu suru toki ni ano*  
INJliving side LOC PP Korea GEN student SFX NOM Japan LOC living do time DAT INJ  
*apa:to o kari tari suru toki ni mondai ga okoru tte koto wa zenzen ari mase n ka?*  
apartment ACC rent such time DAT problem NOM happen QT NML TOP at.all exist POL NEG Q
2. W: *ano mae:*  
INJ before
3. M: *un*  
yes
4. W: *soo desu ne go- ano: go roku nen mae desu ka ano tsutome te i ta gakkoo de wa ano:*  
so COP PP INJ five six year before COP Q INJ work TE ASP PST school LOC TOP INJ  
*totemo taihen de ano fudoosan ya ni kotowar areru ke:su ga*  
very hard COP INJ real.estate shop DAT decline PASS case NOM
5. M: *un*  
yes
6. W: *tabitabi at te ano sorede sugoku nan te iu n desu ka shokku o uke ta gakusei: ni taisuru*  
often exist TE INJ and very what QT say NML COP Q shock ACC get PST student DAT oppose  
*taioo: maa taihen deshi ta keredomo saikin moo amari soo iu hanashi wa kiki mase n.*  
support INJ hard COP PST but lately already much that say story TOP hear POL NEG

1. M: 'Um, from living perspective, are there any problems when Korean students rent an apartment to live in Japan?'
2. W: 'Um, before'
3. M: 'Yeah'
4. W: 'Let me see, five or six years ago, at the school where I was working, um it was very difficult, and there were cases in which students were being declined by the real estate agent'
5. M: 'Yeah'

6. W: ‘There were often cases like that, and, um, really, how to say, it was difficult to support those students who were shocked, but lately I don’t hear such stories.’

(4-155) *nan + te + iu + n + desu + ka* [what + QT + say + NML + COP + Q] <expression search> (Interview: Hypermedia\_024)

1. S: 日本一で仕事をしていた時も大体帰るのがそうすね 9 時とか 10 時一でしたので
2. N: うん
3. S: で一まあ友人の中には 11 時とかいましたし一 あの一 ちよっと あの一 なんていうんですかね一
4. N: ええ
5. S: 非人間的というか
6. N: ああ一
7. S: あの一ちよっと普通の生活じゃないとその当時から思っていましたので

1. S: *nihon: de shigoto o shi te i ta toki mo daitai kaeru no ga soo su ne ku ji toka*  
Japan LOC work ACC do TE ASP PST time also usually return NML NOM so COP PP nine o'clock such  
*juu ji: deshi ta node*  
ten o'clock COP PST because
2. N: *un*  
yes
3. S: *de: maa yuujin no naka ni wa juuichi ji toka i mashi ta shi: ano: chotto ano: nan te*  
and INJ friend GEN among DAT TOP eleven o'clock such exist POL PST because INJ little INJ what QT  
*iu n desu ka ne:*  
say NML COP Q PP
4. N: *ee*  
yes
5. S: *hi ningen teki to iu ka*  
non human like QT say Q
6. N: *aa:*  
INJ
7. S: *ano: chotto futsuu no seikatsu ja nai to sono tooji kara omot te mashi ta node*  
INJ little ordinal GEN living COP NEG QT that time from think TE POL PST because

1. S: ‘When I was working in Japan, it was usually 9 or 10 o’clock when we went home’
2. N: ‘Yeah’
3. S: ‘And, well, among my friends, some went home at 11 o’clock, um, little, um, how to say’
4. N: ‘Yes’
5. S: ‘It’s like inhuman’
6. N: ‘Oh’
7. S: ‘Um, little, I was thinking it’s not a normal life from that time’

(4-156) *nan + te + iu + n + desu + ka* [what + QT + say + NML + COP + Q] <expression search> (Interview: Tetsuko\_10)

1. E: 僕が行った時にお嬢さんに「もう一度聞いてみてくれる？」とって  
そしたら会ってくださったんですよ。
2. K: うん。
3. E: それだからその時に会えたのは僕ぐらいでしたでしょうからやっぱり  
とても なんていうんですか ベットの枕元にね次に出る芝居の台本を  
置いてらしたからね
4. K: ねー
5. E: うーん。それはかわいそうでしたねとってもね

1. E: *boku ga it ta toki ni ojoosan ni "moo ichi do kii te mi te kureru?" to it te soshitara at te*  
I NOM go PST time DAT daughter DAT more one time ask TE ASP TE ASP QT say TE then meet TE  
*kudasut ta n desu yo.*  
ASP PST NML COP PP
2. K: *un.*  
yes
3. E: *sore dakara sono toki ni a e ta no wa boku gurai deshi ta desho o kara yappari totemo*  
that so that time DAT meet POT PST NML TOP I about COP PST COP AUX because after.all very  
*nan te iu n desu ka betto no makura moto ni ne tsugi ni deru shibai no daihon o oi te*  
**what QT say NML COP Q** bed GEN pillow by DAT PP next COP appear play GEN script ACC put TE  
*rashi ta kara ne*  
ASP PST because PP
4. K: *ne:*  
PP
5. E: *u:n. sore wa kawai soo deshi ta ne tottemo ne*  
INJ / that TOP pitiful COP PST PP very PP

1. E: 'When I went to visit him, I asked his daughter "Can you ask him again?" and he let me see him.'
2. K: 'Yeah.'
3. E: 'So, I think it was just about myself who could see him at that time, so after all, very, how to say, he had the script for the next play in which he was supposed to appear by his pillow'
4. K: 'Yeah.'
5. E: 'Um, I felt really sorry for him'

The second sequence *nan + te + iu + n + desho + o* [what + QT + say + NML + COP + AUX] in Interview is functionally similar to *nan + dak + ke* [what + COP + AUX] in Conversation. It co-appears with fillers such as *sono* 'um' and *eeto* 'um.' This collocational pattern contrasts with that of *nan + te + iu + n + desu + ka* which co-appears with *ano* 'um' and

*chotto* ‘um.’ At first glance, these fillers or interjections all seem to have the same function, but there is a functional divergence between *sono/eeto* type and *ano/chotto* type. For instance, Takubo and Kinsui (1997) state that *eeto* signals that the speaker is searching for a concept or knowledge while *ano* indicates that the speaker is trying to formulate an appropriate verbal expression for the concept or knowledge.<sup>41</sup> Appearing with *sono/eeto*, the sequence *nan + te + iu + n + desho + o* indicates that the speaker is trying to access a concept or knowledge, as can be seen in the following excerpts, (4-157) and (4-158).

(4-157) *nan + te + iu + n + desho + o* [what + QT + say + NML + COP + AUX] <concept search> (Interview: Hypermdia\_005)

1. N: 大学でもスポーツがさかんでそのいわゆる全日本で
2. I: ああ
3. N: 大学レベルでの一そのなんていうんでしょう 大会とか
4. I: ええ
5. N: 行われますよねー
6. I: うん
7. N: それはあの一興味持ってらっしゃいますか？

1. N: *daigaku de mo supo:tsu ga sakan de sono iwayuru zen nihon de*  
college LOC also sports NOM strong COP INJ so-called all Japan COP
2. I: *aa*  
INJ
3. N: *daigaku reberu de no: sono: nan te iu n desho o taikai toka*  
college level COP GEN INJ what QT say NML COP AUX meet such
4. I: *ee*  
yes
5. N: *okonaw are masu yo ne:*  
hold PASS POL PP PP
6. I: *un*  
yes
7. N: *sore wa ano: kyoomi mot te rasshai masu ka?*  
that TOP INJ interest have TE ASP POL Q

1. N: ‘At college, sports are popular, and um, all Japan’

<sup>41</sup> The function of *ano* seems to be register-dependent, however. In Conversation, *ano* co-appears with *nan + dak + ke* [what + COP + AUX] which has the function of signaling that the speaker is searching for a concept or knowledge, not a suitable verbal form for a concept/knowledge. Thus, the function of *ano* to indicate that the speaker is trying to find an appropriate verbal expression for a given concept/knowledge does not seem to be applicable to Conversation.

2. I: 'Uh'
3. N: 'At college level, um, what do you call, such as meets'
4. I: 'Yes'
5. N: 'They are being held'
6. I: 'Yeah'
7. N: 'Are you interested in such things?'

(4-158) *nan + te + iu + n + desho + o* [what + QT + say + NML + COP + AUX] <concept search> (Interview: CSJ\_D04M0056)

1. R: えーとピッチ変化っていうのはえーピッチっていうのは音の高さなんですけど
2. L: はい
3. R: えーとー なんていうんでしょう 主観的な音の高さ
4. L: 主観的な音の高さっていうのは
5. R: えー物理的な音の高さではなくて
6. L: はい

1. R: *e:to pitchi henka tte iu no wa e: pitchi tte iu no wa oto no taka sa na n desu*  
INJ pitch change QT say NML TOP INJ pitch QT say NML TOP sound Gen high SFX COP NML COP  
*kedo*  
but
2. L: *hai*  
yes
3. R: *e:to: nan te iu n desho o shukan teki na oto no taka sa*  
INJ what QT say NML COP AUX subjective like COP sound GEN high SFX
4. L: *shukan teki na oto no taka sa tte iu no wa*  
subjective like COP sound GEN high SFX QT say NML TOP
5. R: *e: butsuri teki na oto no taka sa de wa naku te*  
INJ physics like COP sound GEN high SFX COP TOP NEG TE
6. L: *hai*  
yes

1. R: 'Um the pitch change is, um, pitch is the highness or lowness of the sound'
2. L: 'Yes'
3. R: 'Um, how do I say, the subjective highness or lowness of the sound'
4. L: 'What is the subjective highness or lowness of the sound'
5. R: 'Um, it is not the physical highness or lowness'
6. L: 'Yes'

#### 4.3.4.5. Inquiry bundles

Inquiry bundles are the third most frequent sub-type of socio-interactional bundles in Interview (16/77; 20.8%). There is one bundle of this type in Conversation (see below). Some of the bundles include honorific expressions as in the first two examples below. The inclusion of honorific elements is due to the decision to focus on four-morphemic bundles. While questions are usually identified through two-gram expressions at the end of an utterance (e.g., *desu ka*, *masu ka*), the four-gram analysis recognizes larger fixed chunks.

##### (4-159) Inquiry bundles

<i>de</i>	<i>irasshai</i>	<i>masu</i>	<i>ka</i>	<status inquiry>	(Interview)
COP	exist:HON	POL	Q		
<i>te</i>	<i>rasshai</i>	<i>masu</i>	<i>ka</i>	<inquiry>	(Interview)
te	ASP:HON	POL	Q		
<i>ta</i>	<i>n</i>	<i>desu</i>	<i>ka</i>	<inquiry>	(Interview)
PST	NML	COP	Q		
<i>yat</i>	<i>te</i>	<i>n</i>	<i>no</i>	<inquiry>	(Conversation)
do	ASP	NML	PP		

The bundle *de + irasshai + masu + ka* [COP:TE + exist:HON + POL + Q] is used by the interviewer mostly to inquire about the interviewee's current status, as in (4-160), and/or year in school, as in (4-161).

(4-160) *de + irasshai + masu + ka* [COP:TE + exist:HON + POL + Q] <status inquiry>  
(Interview: Hypermedia\_027)

1. N: あ田中さんでいらっしゃいますか。田中さんは学生さんでいらっしゃいますか？
  2. T: はいそうです。
1. N: *a tanaka san de irasshai masu ka. tanaka san wa gakusei san de irasshai masu ka?*  
INJ (name) SFX COP exist POL Q / (name) SFX TOP student SFX COP ASP POL Q

2. T: *hai soo desu.*  
yes so COP

1. N: ‘Ah, are you Mr. Tanaka? Are you a student?’

2. T: ‘Yes I am.’

(4-161) *de + irasshai + masu + ka* [COP:TE + exist:HON + POL + Q] <status inquiry>  
(Interview: Hypermedia\_032)

1. M: あのー えっと 今何年生でいらっしゃいますか？

2. D: 今四年です。

1. M: *ano: etto ima nan nensei de irasshai masu ka?*  
INJ INJ now what year COP exist POL Q

2. D: *ima yo nen desu.*  
now four year COP

1. M: ‘Um, well, what year are you?’

2. D: ‘I’m the fourth year now.’

The bundle *te + rasshai + masu + ka* [TE + ASP:HON + POL + Q] is also used by the interviewer to inquire what the interviewee does for living, as in (4-162), or the interviewee’s opinion about some topic, as in (4-163).

(4-162) *te + rasshai + masu + ka* [TE + ASP:HON + POL + Q] <inquiry/work> (Interview: Hypermedia\_019)

1. M: あのえーと 佐々木さん失礼ですけども あのーお仕事してらっしゃいますか？

2. S: いえ。家庭の主婦です。

1. M: *ano e:to sasaki san shitsurei desu kedomo ano: o shigoto shi te rasshai masu ka?*  
INJ INJ (name) SFX excuse COP but INJ PFX work do TE ASP POL Q

2. S: *ie. katei no shufu desu.*  
no / house GEN house.wife COP

1. M: ‘Um, well, Ms. Sasaki, excuse me but are you working?’

2. S: ‘No, I’m a house wife.’

(4-163) *te + rasshai + masu + ka* [TE + ASP:HON + POL + Q] <inquiry/opinion> (Interview: Hypermedia\_030)

1. N: 受験まあ今の受験制っていうんですか
2. G: はい
3. N: システムについて
4. G: ええ
5. N: どういう風に考えてらっしゃいますか？
6. G: あーやっぱりなんかまあ塾で教えたことがあるんですけども
7. N: あっそうですかうん

1. N: *juken maa ima no juken sei tte iu n desu ka*  
entrance.exam INJ now GEN entrance.exam system QT say NML COP Q
2. G: *hai*  
yes
3. N: *shisutemu ni tsui te*  
system DAT adhere TE
4. G: *ee*  
yes
5. N: *doo iu fuu ni kangae te rasshai masu ka?*  
how say way DAT think **TE ASP POL Q**
6. G: *a: yappari nanka maa juku de oshie ta koto ga aru n desu kedomo*  
INJ after.all DM INJ cram.school LOC teach PST NML NOM exist NML COP but
7. N: *a soo desu ka un*  
INJ so COP Q yes

1. N: ‘Entrance exam, well, the entrance exam system now’
2. G: ‘Yes’
3. N: ‘Regarding the system’
4. G: ‘Yes’
5. N: ‘How do you think about it?’
6. G: ‘Um, after all, I have taught in cram schools’
7. N: ‘Oh, I see. Yeah.’

The bundle *ta + n + desu + ka* [PST + NML + COP +Q] is used by the interviewer to ask for further information or explanation about what has been stated previously.

(4-164) *ta + n + desu + ka* [PST + NML + COP +Q] <inquiry/explanation> (Interview: Hypermedia\_001)

1. M: どうしてそのその作品に感銘を受けたんですか。
2. A: 彼女の使う言葉が
3. M: うん

4. A: あの南部の英語一のしゃべってる言葉を
5. M: うん
6. A: そのままー
7. M: ああああ
8. A: 使ってるんですけどなんていうかこう

1. M: *dooshite sore no sono sakuhin ni kanmei o uke ta n desu ka.*  
why that GEN that work DAT impression ACC get **PST NML COP Q**
2. A: *kanojo no tsukau kotoba ga*  
she GEN use word NOM
3. M: *un*  
yes
4. A: *ano nanbu no eigo: no shabet te ru kotoba o*  
INJ south GEN English GEN talk TE ASP language ACC
5. M: *un*  
yes
6. A: *sono mama*  
that as
7. M: *aa aa*  
INJ INJ
8. A: *tsukat te ru n desu kedo nan te iu ka koo*  
use TE ASP NML COP but what QT say Q this.way

1. M: 'Why were you impressed by that work?'
2. A: 'The words that she used'
3. M: 'Yeah'
4. A: 'The words used in the South'
5. M: 'Yeah'
6. A: 'As they were'
7. M: 'Oh oh'
8. A: 'She used them, how to say, um'

(4-165) *ta + n + desu + ka* [PST + NML + COP + Q] <inquiry/further information> (Interview: CSJ\_D01F0055)

1. L: どんな風においしかったんですか
2. R: かなり衝撃的だったんですけど
3. L: うーん
4. R: その衝撃が大き過ぎて味はどんなだったか覚えてないんですがとにかくおいしかったんですね

1. L: *donna fuu ni oishikat ta n desu ka*  
what.kind way COP delicious **PST NML COP Q**
2. R: *kanari shoogeki teki dat ta n desu kedo*  
quite shocking like COP PST COP but

3. L: *u:n*  
INJ
4. R: *sono shoogeki ga ooki sugi te aji wa donna dat ta ka oboe te nai n desu ga tonikaku*  
that shock NOM big exceed TE flavor TOP what.kind COP PST Q remember TE NEG NML COP but  
*oishikat ta n desu ne*  
anyway delicious PST NML COP PP
1. L: ‘In what way was it delicious?’
2. R: ‘It was quite shocking’
3. L: ‘Um’
4. R: ‘The shock was too big that I don’t remember what kind of flavor it was, but anyway it was delicious.’

One inquiry bundle in Conversation is *yat + te + n + no* [do + ASP + NML + PP]. The following is one example in context.

(4-166) *yat + te + n + no* [do + ASP + NML + PP] <inquiry> (Conversation: BTS\_107)

1. F152: Cさんって今何やってんの？
2. F111: 非常勤。
3. F152: あっそうなんだ。
1. F152: *C san tte ima nani yat te n no?*  
(name) SFX QT now what **do ASP NML PP**
2. F111: *hijookin.*  
part-time work
3. F152: a soo na n da.  
INJ so COP NML COP
1. F152: ‘What is C doing now?’
2. F111: ‘A part-time work.’
3. F152: ‘Oh, I see.’

#### 4.3.4.6. Other socio-interactional bundles

There are other types of socio-interactional bundles appearing in Interview and in Speech. In Interview, there are three “greeting” bundles.

(4-167) Greeting bundles in Interview

*yoroshiku*    *onegai*    *shi*    *masu*    ‘Thank you in advance.’  
favorably    please    do    POL

*arigatoo*    *gozai*    *mashi ta*    ‘Thank you.’  
thank    exist:POL    POL    PST

*doomo*    *arigatoo*    *gozai*    *mashi (ta)*    ‘Thank you very much.’  
much    thank    exist:POL    POL    PST

The first greeting bundle *yoroshiku + onegai + shi + masu* [favorably + please + do + POL] appears at the beginning of the interview. The expression is exchanged between the interviewer and the interviewee in sequence. Sometimes the polite form *itashi* [do:POL] is used in place of *shi* [do], as in (4-170).

(4-168) *yoroshiku + onegai + shi + masu* [favorably + please + do + POL] ‘thank you in advance’ (Interview: Hypermedia\_047)

1. H: じゃあ 始めます。
  2. T: はい
  3. H: よろしく お願い します。
  4. T: あよろしく お願い します。
  5. H: 私は日野と申しますけれどもお名前は。
- 
1. H: *jaa hajime masu.*  
then begin POL
  2. T: *hai*  
yes
  3. H: *yoroshiku onegai shi masu.*  
favorably please do POL
  4. T: *a yoroshiku onegai shi masu.*  
INJ favorably please do POL
  5. H: *watashi wa hino to moo shi masu keredomo o namae wa.*  
I TOP (name) QT say do POL but PFX name TOP
- 
1. H: ‘Then we’ll begin.’
  2. T: ‘Yes.’
  3. H: ‘**Thank you in advance.**’
  4. T: ‘Ah, **thank you in advance.**’
  5. H: ‘My name is Hino, your name is?’

(4-169) *yoroshiku + onegai + shi + masu* [favorably + please + do + POL] ‘thank you in advance’ (Interview: CSJ\_D01F0023)

1. L: じゃよろしくお願ひします
2. R: よろしくお願ひします
3. L: あのまず
4. R: はい

1. L: *ja yoroshiku onegai shi masu*  
then favorably treat do POL
2. R: *yoroshiku onegai shi masu*  
favorably treat do POL
3. L: *ano mazu*  
INJ first
4. R: *hai*  
yes

1. L: ‘Then, **thank you in advance.**’
2. R: ‘**Thank you in advance.**’
3. L: ‘Um, first’
4. R: ‘Yes.’

(4-170) *yoroshiku + onegai + shi + masu* [favorably + please + do + POL] ‘thank you in advance’ (Interview: Hypermedia\_014)

1. N: はじめまして。わたくし根津と申します。
2. O: あっあのわたくし小沢と申します。よろしくお願ひします。
3. N: よろしくお願ひいたします。小沢さんは学生さんでいらっしゃるか？

1. N: *hajime mashi te. watakushi nezu to mooshi masu.*  
begin POL TE / I (name) QT say POL
2. O: *a ano watakushi ozawa to mooshi masu. yoroshiku onegai shi masu.*  
INJ INJ I (name) QT say POL / favorably treat do POL
3. N: *yoroshiku onegai itashi masu. ozawa san wa gakusei san de irasshai masu ka?*  
favorably treat do POL / (name) SFX TOP student SFX COP exist POL Q

1. N: ‘How do you do? My name is Nezu.’
2. O: ‘Ah, um, my name is Ozawa. **Thank you in advance.**’
3. N: ‘Thank you in advance. Are you a student?’

The second greeting bundle in Interview, *arigatoo + gozai + mashi + ta* [thank + exist:POL + POL + PST], characteristically appears at the end of the interview. This expression

is also exchanged between the interviewer and the interviewee, as shown in the following three excerpts.

(4-171) *arigatoo + gozai + mashi + ta* [thank + exist:POL + POL + PST] ‘thank you’  
(Interview: Tetsuko\_03)

1. T: *そうですか。ご成功祈ってますね。*
2. K: *ありがとうございます。*
3. T: *ありがとうございました。*
4. H: *ありがとうございました。*

1. T: *soo desu ka. go seikoo inot te masu ne.*  
so COP Q / SFX success pray TE POL PP
2. K: *arigato gozai masu.*  
thank exist POL
3. T: ***arigatoo gozai mashi ta.***  
**thank exist POL PST**
4. H: ***arigatoo gozai mashi ta.***  
**thank exist POL PST**

1. T: ‘I see. I’m praying for your success.’
2. K: ‘Thank you.’
3. T: ‘**Thank you.**’
4. H: ‘**Thank you.**’

(4-172) *arigatoo + gozai + mashi + ta* [thank + exist:POL + POL + PST] ‘thank you’  
(Interview: Hypermedia\_045)

1. M: *えー今日はお忙しい所ほんとにどうもありがとうございました。*
2. T: *いえいえ どうもこちらこそ。*
3. M: *どうも。*
4. T: *どうもありがとうございました。*

1. M: *e: kyoo wa o ishogashii tokoro honto ni doomo arigatoo gozai mashi ta.*  
INJ today TOP SFX busy place really DAT much **thank exist POL PST**
2. T: *ie ie doomo kochira koso.*  
no no thank here just
3. M: *doomo.*  
much
4. T: *doomo arigatoo gozai mashi ta.*  
much **thank exist POL PST**

1. M: ‘Um, today, while you are busy, **thank you so much.**’
2. T: ‘No no, it is I who should say so.’

3. M: ‘Thanks.’
4. T: ‘**Thank you** very much.’

(4-173) *arigatoo + gozai + mashi + ta* [thank + exist:POL + POL + PST] ‘thank you’  
(Interview: CSJ\_D04M0056)

1. L: ああ そうですか
2. R: はい
3. L: 分かりました どうも ありがとう ございました
4. R: ありがとう ございました

1. L: *aa soo desu ka*  
INJ so COP Q
2. R: *hai*  
yes
3. L: *wakari mashi ta / doomo arigatoo gozai mashi ta*  
understand POL PST / much thank exist POL PST
4. R: *arigatoo gozai mashi ta*  
thank exist POL PST

1. L: ‘Oh I see.’
2. R: ‘Yes.’
3. L: ‘I understand. **Thank you** very much.’
4. R: ‘**Thank you**.’

The third bundle in Interview is a more polite version of the second bundle. It has an additional morpheme *doomo* [much] before *arigatoo + gozai + mashi + ta* [thank + exist:POL + POL + PST]. Since this expression includes five morphemes, the bundle itself, which is four-morphemes long, is not complete. Two of the excerpts above, (4-172) and (4-173), include the longer expression *doomo + arigatoo + gozai + mashi + ta* [much + thank + exist:POL + POL + PST].

In Speech, there are five other interactional bundles. The first two bundles are used to make an “announcement.” The third and fourth bundles are used to make a “proposal,” and the fifth bundle is used to insert a disclaimer.

(4-174) Other socio-interactional bundles in Speech

<i>s</i>	<i>ase</i>	<i>te</i>	<i>itadaki</i>		<announcement>
do	CAUS	TE	receive:HUM		
	<i>ase</i>	<i>te</i>	<i>itadaki</i>	<i>masu</i>	<announcement>
	CAUS	TE	ASP:HUM	POL	
<i>n</i>	<i>de</i>	<i>wa</i>	<i>nai</i>		<proposal>
NML	COP	TOP	NEG		
		<i>nai</i>	<i>desho</i>	<i>o</i>	<i>ka</i>
		NEG	COP	AUX	Q
<i>to</i>	<i>omoi</i>	<i>masu</i>	<i>ga</i>		<disclaimer>
QT	think	POL	but		

A majority of the two bundles used for making an announcement are parts of the same sequence *s + ase + te + itadaki + masu* [do + CAUS + TE + ASP:HUM + POL] ‘Please let (me) do...’ In a majority of instances, this expression is used towards the very beginning of a speech to make a formal announcement that the speaker is going to present his/her research. The bundle is preceded by *happyoo* [present], *happyoo + o* [presentation ACC], or *ohanashi o* [talk ACC], and the whole chunk is further preceded by a quoting clause ending in *te*-form, specifying the title of the presentation (e.g., *to + daishi + mashi + te* [QT + title (V) + POL + TE], *to + iu + dai + de* [QT + say + title (N) + COP], *to + iu + koto + de* [QT + say + NML + COP]).

We have seen before another sequence, namely, *te + iki + tai + to + omoi + masu* [TE + go + want + QT + think + POL], which has the function of making an announcement in Speech. The difference between them is that the sequence *s + ase + te + itadaki + masu* [do + CAUS + TE + ASP:HUM + POL] is predominantly used to begin the presentation, whereas the sequence *te + iki + tai + to + omoi + masu* [TE + go + want + QT + think + POL] is used either in the

middle of the presentation to announce what comes next or towards the end of the presentation to announce the remaining issues or future goals.

The following three excerpts present the use of the sequence, *s + ase + te + itadaki + masu* [do + CAUS + TE + ASP:HUM + POL], to formally begin one's presentation.

(4-175) *s + ase + te + itadaki + masu* [do + CAUS + TE + ASP:HUM + POL] <announcement>  
(Speech: CSJ\_A01M0099)

えー × × × と題しまして発表させていただきます / えーとまず研究の背景ですが ...

*e: XXX to daishi mashi te happyoo s ase te itadaki masu / e: to mazu kenkyuu no haikai desu ga ...*  
INJ (name) QT title POL TE present **do CAUS TE ASP POL** / INJ first study GEN background COP but

'Um, let us present our study titled XXX. Um, first, the background of the study ...'

(4-176) *s + ase + te + itadaki + masu* [do + CAUS + TE + receive:HUM + POL]  
<announcement> (Speech: CSJ\_A03M0005)

× × × という一題でえー発表させていただきます / であのー今日のお話の目的なんですけれども ...

*XXX to iu u: dai de e: happyoo s ase te itadaki masu / de ano: kyoo no o hanashi no mokuteki na*  
(name) QT say INJ title COP INJ present **do CAUS TE ASP POL** / and INJ today GEN PFX GEN purpose COP  
*n desu keredomo ...*  
NML COP but

'Let us present our study titled XXX. And, um, the purpose of today's talk ...'

(4-177) *s + ase + te + itadaki + masu* [do + CAUS + TE + receive:HUM + POL]  
<announcement> (Speech: CSJ\_A02F0082)

えー本日は × × × というのでえー発表させていただきます / まず一初めにと  
して簡単に研究の動機を述べさせていただきます

*e: honjitsu wa XXX to iu koto de e happyoo s ase te itadaki masu / mazu ichi hajime ni to shi te*  
INJ today TOP (name) QT say NML COP INJ present **do CAUS TE ASP POL** / first one first DAT QT do TE  
*kantan ni kenkyuu no dooki o nobe s ase te itadaki masu*  
brief DAT study GEN motivation ACC talk **do CAUS TE ASP POL**

‘Today, let us present our study titled XXX. First, as number one introduction, let us talk briefly about the motivation for the study.’

The third bundle *nai + desho + o + ka* [NEG + COP + AUX + Q] is preceded by *de + wa* [COP + TOP], and it is used to state the speaker’s opinion in the form of a proposal.

Presenting one’s opinion in the form of a proposal, rather than simply stating it, makes a stronger appeal to the audience.

(4-178) *de + wa + nai + desho + o + ka* [COP + TOP + NEG + COP + AUX + Q] <proposal>  
(Speech: CSJ\_A08F0313)

コミュニケーション能力として聞く話す力を育てる為には えーじょ-その状況の中で具体的な方略として教育内容その場のから帰納的に自覚化させることが大事なのではないでしょうか

*komyunike:shon nooryoku to shi te kiku hanasu chikara o sodateru tame ni wa e: jo- sono*  
communication ability QT do TE listen speak ability ACC develop for DAT TOP INJ that  
*jookyoo no naka de gutaiteki na hooryaku to shi te kyooiku naiyoo sono ba no kara kinou teki ni*  
situation GEN inside LOC concrete COP scheme QT so TE education content that place from induction like DAT  
*jikaku ka s aseru koto ga daiji na no de wa nai desho o ka.*  
awareness -ize do CAUS NML NOM important COP NML COP TOP **NEG COP AUX Q**

‘In order to develop their listening and speaking as communication ability, um, as a concrete scheme in the environment, it is important to make them self aware inductively from the educational content itself.’

(4-179) *de + wa + nai + desho + o + ka* [COP + TOP + NEG + COP + AUX + Q] <proposal>  
(Speech: CSJ\_A09F0859)

新しい協同組合活動による試みが1990年代の自治体制度改革の誘因の一つであるとすれば九十年代の分権社会はライフポリティクス場の一つとして存在し得るのではないでしょうか

*atarashii kyoodoo kumiai katsudoo ni yoru kokoromi ga 1990 nen dai no jichitai seido kaikaku*  
new cooperative association activity DAT depend trial NOM 1900 year time GEN government system reform

*no yuwin no hitotsu de aru to sure ba kyuujuu nen dai no bunken shakai wa raifu poritikkusu no*  
GEN cause GEN one COP exist QT do COND 90 year time GEN decentralization society TOP life politics GEN  
*ba no hitotsu to shi te sonzai shi eru no de wa nai desho o ka*  
place GEN one QT do TE exist do POT NML COP TOP **NEG COP AUX Q**

‘If the new attempts by the cooperative activities are one cause for the local government reform of the 1990s, the decentralized society of the 90s can exist as one place for “life politics.”’

The fourth bundle *n + de + wa + nai* [NML + COP + TOP + NEG] is part of the longer sequence *n + de + wa + nai + ka* [NML + COP + TOP + NEG + Q]. The sequence also is used to make a proposal, but unlike the previous proposal bundle, it is frequently being quoted, that is, followed by *to + omou/kangaeru* [QT + think], and hence it sounds more indirect. The whole sequence is a combination of the proposal bundle and the reporting bundle: {*n + de + wa + nai + ka*} [NML + COP + TOP + NEG + Q] and {*to + (iu + fuu + ni) + omoi + masu*} [QT + (say + way + COP:ADV) + think + POL].

(4-180) *n + de + wa + nai + ka* [NML + COP + TOP + NEG + Q] <proposal> (Speech: CSJ\_A01M0137)

えー までできれば三キロヘルツ以下の周波数で え設定した方が いいんではないか という風に考えております

*e: ma dekire ba san kiro herutsu ika no shuuhasuu de e settei shi ta hoo ga ii n de*  
INJ INJ can COND three kilo hertz below GEN frequency COP INJ set so PST direction NOM **good NML COP**  
***wa nai ka to iu fuu ni kangae te ori masu***  
TOP NEG Q QT say way DAT think TE ASP POL

‘Um, well, we are thinking that it would be better to set the frequency at or below three kilo hertz if possible.’

(4-181) *n + de + wa + nai + ka* [NML + COP + TOP + NEG + Q] <proposal> (Speech: CSJ\_A07M0223)

あのーテスト理論のお方達があのーん-実にこう貢献できる ところがあるんではないか という風に私はあのおもっ-思う訳です

*ano: tesuto riron no o kata tachi ga ano: n- jitsu ni koo kookan dekiru koo ba ga aru n de wa*  
INJ test theory GEN PFX person PL NOM INJ real DAT this contribute can this place NOM NML COP TOP  
***nai ka to iu fuu ni watashi wa ano omo- omou wake desu***  
NEG Q QT say way DAT I TOP INJ think NML COP

‘Um, I think that there is a place where test theorists can really contribute.’

(4-182) *n + de + wa + nai + ka* [NML + COP + TOP + NEG + Q] <proposal> (Speech: CSJ\_A08M0186)

あのパブリックな音調 というものを子供達に身に付けさせることが必要な  
んではないかという風にこう考えております

*ano paburikku na onchoo to iu mono o kodomo tachi ni mi ni tsuke s aseru koto ga hitsuyoo*  
INJ public COP intonation QT say NML ACC child PL DAT body DAT adhere do CAUS NML NOM necessary  
*na n de wa nai ka to iu fuu ni koo kangae te ori masu*  
COP NML COP TOP NEG Q QT say way DAT this think TE ASP POL

‘We are thinking that, um, it is necessary to have the children master the tone in public speaking.’

The fifth bundle *to + omoi + masu + ga* [QT + think + POL + but] is used to present a disclaimer as can be seen in the following excerpts. In (4-183), before the speaker begins explaining the device called dynamic paratograph, he gives a disclaimer saying that everybody (in the audience) knows the device well.

(4-183) *to + omoi + masu + ga* [QT + think + POL + but] <disclaimer> (Speech: CSJ\_A05M0031)

であえーっとこれもあの皆さんよく御存じのことと 思いますがあのダイナミックパラトグラフについてえ説明いたします

*de a e:tto kore mo ano minasan yoku gozonji no koto to omoi masu ga ano: dainamikku*  
and INJ INJ this also INJ everyone well know GEN NML QT think POL but INJ dynamic  
*paratografu ni tsui te e setsumei itashi masu.*  
(name) DAT adhere TE INJ explain do POL

‘And, uh, well, we think everyone knows this very well too, but we will explain, um, regarding the dynamic paratograph.’

In a similar manner, in (4-184), the speaker gives the disclaimer that he probably does not need to explain the device called EMMA in particular, right before he begins explaining it.

(4-184) *to + omoi + masu + ga* [QT + think + POL + but] <disclaimer> (Speech: CSJ\_A01M0074)

でまーあのEMMAというのは特に説明する必要もないと**思います**がこういう装置でございましてえー大事なのはこのコイルの配置でございまして

*de ma: ano EMMA to iu no wa tokuni setsumei suru hitsuyoo mo nai to omoi masu ga koo iu*  
and INJ INJ (name) QT say NML TOP especially explain do necessity also NEG **QT think POL but** this say  
*soochi de gozai mashi te e: daiji na no wa kono koiru no haichi de gozai masu*  
equipment COP exist POL TE INJ important COP NML TOP this coil GEN arrange COP ASP POL

‘And, well, um, we think we don’t need to explain what EMMA is in particular, but it is this kind of equipment, and, um, what is important is the arrangement of this coil.’

In (4-185), the disclaimer concerns the whole presentation. The speaker says that she began preparing for the presentation from the morning, so there may be shortcoming in her presentation.

(4-185) *to + omoi + masu + ga* [QT + think + POL + but] <disclaimer> (Speech: CSJ\_A01F0122)

えと何分原稿今朝受け取りましてあの一準備したという段階ですので至らない点がある**と****思います**がどうぞよろしくお願ひいたします

*eto nanibun genkoo kesa uketori mashi te ano: junbi shi ta to iu dankai desu node itara nai ten*  
INJ anyhow script this.morning receive POL TE INJ prepare do PST QT say stage COP because be.competent NEG  
*ga aru to omoi masu ga doozo yoroshiku onegai itashi masu*  
point NOM exist **QT think POL but** much favorably treat do POL

‘Um, anyhow, I received the script this morning and, um, prepared, so I think there are points where I’m incompetent, but thank you in advance.’

#### 4.3.4.7. Summary of socio-interactional expressions

In this section, we have seen that socio-interactional expressions are mostly used in interactive discourse, Conversation and Interview, to achieve various kinds of interactional and interpersonal actions and understandings. With respect to the formulaic status of morphemic bundles, I have noted that some of the functions are linked to units smaller than the four-morphemic bundles. In particular, the interpersonal functions are mostly attributed to the

utterance final pragmatic particles or other units, for example, *yo*, *yo + ne*, and *ja + nai*. At the same time, we have also seen larger patterns of formulaicity. The larger sequences were identified from the overlapping of two or more bundles, or the combination of two or more bundles. These patterns are discussed in more detail in Section 4.6.

#### 4.3.5. Objective statements

Objective statement bundles involve verbal elements and convey an action, event, or state of affairs objectively. The statement bundles are fairly common in Speech (42/175; 24%) but not in Conversation (4/92; 4.3%) or Interview (9/189; 4.8%). Among verbs that are used in the statement bundles, *naru* ‘become’ is the most common type in Interview and Speech, while in Conversation, *kaeru* ‘return’ is the most common verb. In many cases, the bundles involve aspectual elements such as *-te-(i)ru*, *-te-aru*, *-te-kuru*.<sup>42</sup> In Speech, 24% (10/42) of the statement bundles involve passive morphemes. Furthermore, 29% (12/42) of the statement bundles in Speech only contain grammatical elements, for instance, *are + te + i + masu* [PASS + TE + ASP + POL] and *te + ori + mashi + ta* [TE + ASP + POL + PST].<sup>43</sup> The following are some examples of the objective statement bundles.

#### (4-186) Statement bundles

<i>kaet</i>	<i>te</i>	<i>ki</i>	<i>ta</i>	<statement>	(Conversation)
return	TE	ASP	PST		
<i>ni</i>	<i>nat</i>	<i>te</i>	<i>ru</i>	<statement>	(Conversation and Interview)
COP:ADV	become	TE	ASP:NONPST		

<sup>42</sup> The bundles involving aspectual elements are 100% (4/4) in Conversation, 44% (4/9) in Interview, and 50% (21/42) in Speech.

<sup>43</sup> This contrasts with the statement bundles in Conversation, all of which contain lexical verbs, and also with the statement bundles in Interview, all but one of which include lexical verbs.

<i>nat</i> become	<i>te</i> TE	<i>i</i> ASP	<i>masu</i> POL	<statement>	(Speech)
<i>shi</i> do	<i>te</i> TE	<i>i</i> ASP	<i>masu</i> POL	<statement>	(Speech)
<i>o</i> ACC	<i>okonai</i> conduct	<i>mashi</i> POL	<i>ta</i> PST	<statement>	(Speech)
<i>are</i> PASS	<i>te</i> TE	<i>i</i> ASP	<i>masu</i> POL	<statement>	(Speech)
<i>s</i> do	<i>are</i> PASS	<i>te</i> TE	<i>i</i> ASP	<statement>	(Speech)
<i>mir</i> see	<i>are</i> PASS	<i>mashi</i> POL	<i>ta</i> PST	<statement>	(Speech)

The first bundle, *kaet + te + ki + ta* [return + TE + ASP + PST], only appears in Conversation. In Conversation, friends and family members tend to talk about each other's or their mutual acquaintances' whereabouts, which includes talks about 'returning' from some place. In (4-187), a daughter A asks her mother C whether her father has returned home from the hospital (i.e., discharged from the hospital).

(4-187) *kaet + te + ki + ta* [return + TE + ASP + PST] <statement> (Conversation: Callhome\_2206)

1. A: うん。パパ帰ってきたの？
2. C: うん？
3. A: パパ病院から帰ってきたの？
4. C: うん。

1. A: *un. papa kaet te ki ta no?*  
yes dad return TE ASP PST PP
2. C: *un?*  
yes
3. A: *papa byooin kara kaet te ki ta no?*  
dad hospital from return TE ASP PST PP
4. C: *un.*  
yes

1. A: ‘Yeah. Did dad come home?’
2. C: ‘Huh?’
3. A: ‘Did dad come back from the hospital?’
4. C: ‘Yeah.’

The second bundle *ni + nat + te + ru* [COP:ADV + become + TE + ASP:NONPST] appears in both Conversation and Interview. In Speech, the related bundle *nat + te + i + masu* [become + TE + ASP + POL] appears. The prevalence of this type of bundle is probably due to the fact that the bundles themselves do not have a fixed meaning, but they gain meaning from co-occurring elements as a whole.<sup>44</sup> For example, in (4-188), the sequence, *dekiru + yoo + ni + nat + te + ru* [can + like + COP + become + TE + ASP], means ‘it is set up so that (you) can ...’ In (4-189), the sequence, *taishoo + ni + nat + te + ru* [target + COP + become + ASP], means ‘... is the target of ...,’ and in (4-190), the sequence, *ooku + nat + te + i + masu* [many + become + TE + ASP + POL], means ‘... is many/high.’

(4-188) *ni + nat + te + ru* [COP:ADV + become + TE + ASP:NONPST] <statement>  
(Conversation: Work\_02F)

1. A: でなんかチェックインみたいのが自動で機械でできるようになってる  
でしょう。
2. C: ーんーんーんーんー。
  
1. A: *de nanka chekku in mitai no ga jidoo de kikai de dekiru yoo ni nat te ru desho*  
and DM check in like NML NOM automatic COP machine COP can like **COP become TE ASP COP**  
*o.*  
AUX
2. C: *n: n: n: n:.*  
INJ INJ INJ INJ
  
1. A: ‘And, like, it is set up so that you can check in automatically by machine, you know.’
2. C: ‘Mhm, mhm, mhm, mhm.’

---

<sup>44</sup> See Stubbs (1995) for the tendency of commonly used words to be delexicalized and gain meaning “only via their repeated co-occurrence with other words” (p. 387).

(4-189) *ni + nat + te + ru* [COP:ADV + become + TE + ASP:NONPST] <statement>  
(Interview: CSJ\_D04F0022)

1. L: 時間が経過することによってあの自動的にこうあの一カテゴリー分けして  
いっていったようなことも
2. R: はーあー
3. L: 研究の対象になってるんですか
4. R: そうですね

1. L: *jikan ga keika suru koto ni yot te ano jidooteiki ni koo ano: kategori: wake shi te iku tte*  
time NOM pass do NML DAT depend TE INJ automatically DAT this INJ category divide do TE ASP QT  
*iu yoo na koto mo*  
say like COP NML also
2. R: *ha: a:*  
INJ INJ
3. L: *kenkyuu no taishoo ni nat te ru n desu ka*  
study GEN target **DAT become TE ASP NML COP Q**
4. R: *soo desu ne*  
so COP PP

1. L: ‘Things like automatically categorizing them as the time passes’
2. R: ‘Uh huh’
3. L: ‘Are they also the target of your study?’
4. R: ‘That’s right’

(4-190) *nat + te + i + masu* [become + TE + ASP + POL] <statement> (Speech:  
CSJ\_A02F0038)

縮小減少型の自他の比率は他動詞がえー十六点じゅ-六十一.五パーセント  
で他の型に比べて多くなっています

*shukushoo genshoo gata no jita no hiritsu wa tadooshi ga e: juuroku ten ju- rokujuuichi ten go*  
reduction reduction pattern GEN intransitive.transitive GEN ratio TOP transitive.verb NOM INJ sixty one point five  
*pa:sento de hoka no kata ni kurabe te ooku nat te i masu*  
percent COP other GEN pattern DAT compare TE many **become TE ASP POL**

‘The ratio of reduced intransitives and transitives, the transitive verb is 16 points, 61.5%, which is higher compared to other patterns.’

The rest of the bundles in this category only appear in Speech. Below, I list two examples for each bundle type. Excerpts (4-191) and (4-192) present examples of *shi + te + i + masu* [do + TE + ASP + POL].

(4-191) *shi + te + i + masu* [do + TE + ASP + POL] <statement> (Speech: CSJ\_A07F0356)

本研究では衣類乾燥過程の快適性を定量的に評価するということを目的と  
しています

*hon kenkyuu de wa irui kansoo katei no kaiteki sei o teiryoo teki ni hyooka suru to iu*  
this study LOC TOP clothes dry process GEN comfort charactersitics ACC quantity like DAT evaluate do QT say  
*koto o mokuteki to shi te i masu*  
NML ACC purpose QT do TE ASP POL

‘In this study, we aim to quantitatively evaluate the comfort of the clothes drying process.’

(4-192) *shi + te + i + masu* [do + TE + ASP + POL] <statement> (Speech: CSJ\_A03F0153)

アルゴリズムはあの六番の参考文献と七番の参考文献を参考にして作った  
プログラムを利用しています

*arugorizumu wa ano roku ban no sankoo bunken to nana ban no sankoo bunken o*  
algorithm TOP INJ six number GEN reference document and seven number GEN reference document ACC *te*  
*sankoo ni shi tsukut ta puroguramu o riyoo shi te i masu*  
reference DAT do TE make PST program ACC use **do TE ASP POL**

‘As for algorithm, we are using the program which we made by referring to the references six  
and seven.’

The next two excerpts include the bundle, *o + okonai + mashi + ta* [ACC + conduct + POL +  
PST].

(4-193) *o + okonai + mashi + ta* [ACC + conduct + POL + PST] <statement> (Speech:  
CSJ\_A05F0043)

具体的には実験一から実験四までの四つの実験を行ないました

*gutai teki ni wa jikken ichi kara jikken yon made no yottsuo no jikken o okonai mashi*  
concrete like DAT TOP experiment one from experiment four until GEN four GEN experiment ACC conduct POL  
*ta*  
PST

‘Concretely, we conducted experiments one through four.’

(4-194) *o + okonai + mashi + ta* [ACC + conduct + POL + PST] <statement> (Speech: CSJ\_A07F0226)

得られた距離行列は多次元尺度法によって分析を行ないました

*e rare ta kyori gyooetsu wa ta jigen shakudo hoo ni yot te bunseki o okonai mashi*  
get POT PST distance line TOP many dimension measure method DAT depend TE analysis ACC conduct POL  
*ta*  
PST

‘We conducted the analysis on the gained distance line based on the multiple dimension method.’

As stated above, about a quarter of the statement bundles in Speech involve a passive morpheme. All the bundles types presented below include a passive morpheme, *are* [PASS] or *rare* [PASS]. The bundle *are + te + i + masu* [PASS + TE + ASP + POL] is most frequently preceded by the verb *iw* [say], as in (4-195), followed by *shirar* [know], as in (4-196).

(4-195) *are + te + i + masu* [PASS + TE + ASP + POL] <statement> (Speech: CSJ\_A06F0073)

その訂正方法を用いるとあの学習者が仮説検証を自由に行なうそのプロセスを阻害してしまうその為に問題があるといわれています

*sono teisei hoo hoo o mochiiru to ano gakushuusha ga kasetsu kenshoo o jiyuu ni okonau sono*  
that correct method ACC use COND INJ lean person NOM hypothesis verification ACC free DAT conduct that  
*purosesu o sogai shi te shimau sono tame ni mondai ga aru to iw are te i masu*  
process ACC obstruct do TE ASP that reason DAT problem NOM exist QT say PASS TE ASP POL

‘It is said that if we use that correction method, it prevents the learners from freely making hypotheses and verifying them, and because of that it has problems.’

(4-196) *are + te + i + masu* [PASS + TE + ASP + POL] <statement> (Speech: CSJ\_A01F0143)

え生理学的にはえー蝸牛神経核から大脳皮質に至る各段階にえ一周波数の上昇方向あるいは下降方向の周波数変化に選択的にん-応答するニューロンが存在することが知られています

*e seibutsugaku teki ni wa e: kagyuu shinkei kaku kara dainoo hishitsu ni itaru kaku dankai ni e:*  
INJ biology perspective DAT TOP INJ cochlear nerve nucleus from cerebral cortex DAT lead each step DAT INJ  
*shuuhasuu no jooshoo hookoo aruiwa kakoo hookoo no shuuha henka ni sentaku teki ni n- ootoo*  
frequency GEN increase direction or decrease direction GEN frequency change DAT selective like DAT INJ react

*suru nyuuron ga sonzai suru koto ga shir are te i masu*  
do neuron NOM exist do NML NOM know **PASS TE ASP POL**

‘Um, from a biological perspective, it is known that there are neurons which react selectively to the change in the increasing or decreasing frequency at each stage leading from the cochlear nucleus to the cerebral cortex.’

The bundle *s + are + te + i* [do + PASS + TE + ASP] is followed by *masu* [POL]. The most frequent preceding element is *shiteki* [point out], as in (4-197).

(4-197) *s + are + te + i* [do + PASS + TE + ASP] <statement> (Speech: CSJ\_A07F0445)

えー この 影響 外から与えられた要因の影響力というのが長く持続しないという  
ような問題点がえー指摘 されています

*e: kono eikyoo gai kara atae rare ta yoin no eikyoo ryoku to iu no ga nagaku jizoku shi nai to*  
INJ this influence outside from give PASS factor GEN influence power QT say NML NOM long keep do NEG QT  
*iu yoo na mondai ten ga e: shiteki s are te i masu*  
say like COP problem point NOM INJ point.out **do PASS TE ASP POL**

‘Um, the problematic point that the influential power of the factors given outside the influence does not continue long has been pointed out.’

(4-198) *s + are + te + i* [do + PASS + TE + ASP] <statement> (Speech: CSJA01F0067)

えピッチ知覚にはえ基底膜振動のピーク位置である場所情報それからえー  
聴神経の発火の周期性である時間情報の二つが関わっていると されています  
けれども

*e pitchi chikaku ni wa e kitei maku shindoo no pi:ku ichi de aru basho joohoo sorekara*  
INJ pitch perception DAT TOP INJ base membrane vibration GEN peak position COP exist place information and  
*e: choo shinkei no hakka no shuuki sei de aru jikan joohoo no futatsu ga kakawat te*  
INJ auditory nerve GEN ignition GEN period characteristics COP exist time information GEN two NOM relate TE  
*ru to s are te i masu keredomo*  
ASP QT **do PASS TE ASP POL** but

‘It is thought that the place information which is the peak position of the base membrane vibration and the time information which is the ignition periodicity for auditory nerve are related to the pitch perception.’

There are only two verbs that appear within the bundles with passive morphemes, *mi* [look], as in (4-199) and (4-200), and *s* [do], as in (4-198) above.

(4-199) *mi + rare + mashi + ta* [see + PASS + POL + PST] <statement> (Speech:  
CSJ\_A07F0247)

えー高得点者よりも非常にと-えー解決の方略が多様なばかりかそれがえ安定していないという傾向が見られました

*e: koo tokuten sha yori mo hijoo ni to- e: kaiketsu no hooryaku ga tayoo na bakari ka sore ga e*  
INJ high score person than also very DAT INJ solution GEN strategy NOM various COP only or that NOM INJ  
*antei shi te i nai to iu keikoo ga mi rare mashi ta*  
steady do TE ASP NEG QT say tendency NOM see PASS POL PST

‘The tendency observed was that they not only have more diverse solution strategies than high scorers but also the strategies are not stable.’

(4-200) *mi + rare + mashi + ta* [see + PASS + POL + PST] <statement> (Speech:  
CSJ\_A09F0600)

えー特に村内や移動先における身近な先例からの直接間接の情報といたしますのは重要な契機となっているケースが多く見られました

*e: tokuni son nai ya idoo saki ni okeru mijika na senrei kara no chokusetsu kansetsu no*  
INJ especially village inside and transfer place LOC at close COP precedent from GEN direct indirect GEN  
*joohoo to ii masu no wa juuyoo na keiki to nat te iru ke:su ga ooku mi rare*  
information QT say POL NML TOP important COP opportunity QT become TE ASP case NOM many see PASS  
*mashi ta*  
POL PST

‘Um, in particular, many cases were seen where direct and indirect information from the familiar precedents at villages or destinations are becoming an important opportunity.’

In this section, I presented examples of objective statement bundles. Most statement bundles only appear in Speech and many of them include the polite verb suffixes, *masu* [POL] or *mashi + ta* [POL + PST]. While both Interview and Speech are formal registers, and the speakers are expected to use polite verb suffixes, bundles with *masu* [POL] or *mashi + ta* [POL + PST] are much more common in Speech than in Interview (most of these have been identified as objective

statement bundles). In contrast to the popularity of *masu* and *mashi + ta* within the bundles in Speech, in Interview, the polite form of the copula, *desu* [COP] as part of the sequence *n + desu* [NML + COP] is much more popular. As we have seen, the sequence *n + desu* can also follow verbs. This suggests that the morphemic bundles are closely associated with the functional characteristics of the registers. In the next section, I focus on the register variations in the use of the morphemic bundles.

#### 4.4. Register variations in the use of the morphemic bundles

The distributional patterns as well as functional characteristics of the bundles across three registers suggest that the use of the morphemic bundles is closely related to the characteristics of the particular register. Figure 4.3 summarizes the overall distributional patterns of the morphemic bundles across functional categories.

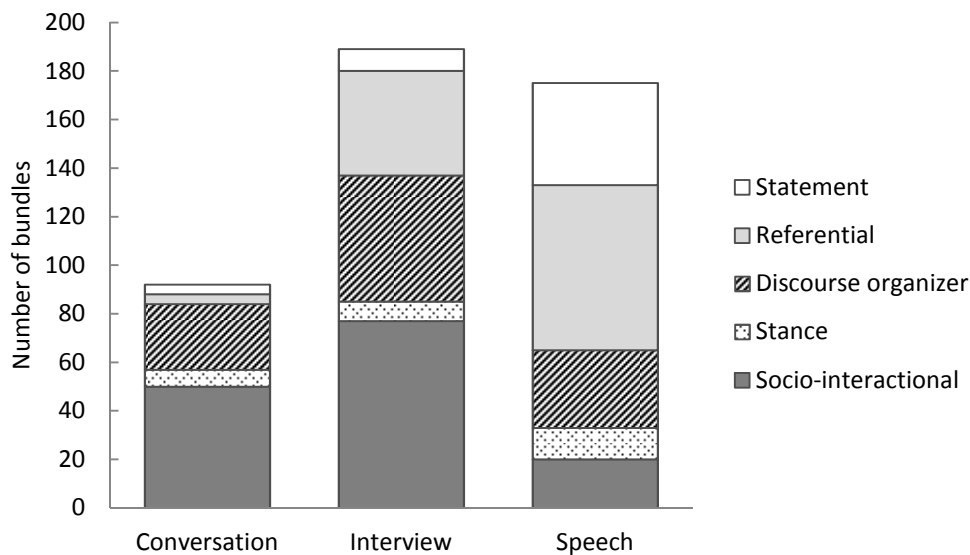


Figure 4.3. Distribution of morphemic bundles across functional categories

From the figure above, we can observe the following overall patterns.

1. Socio-interactive bundles are extremely common in Conversation and Interview.
2. Referential bundles are most common in Speech and fairly common in Interview.
3. Statement bundles are much more common in Speech than other two registers.
4. Stance bundles are not common in all three registers.
5. Discourse organizers are more or less common in all three registers.

Based on the closer examination of each functional type, the socio-interactive bundles are found to have functions (e.g., interpersonal, reactive, inquiry) that serve the needs of interactive and involved discourse, namely, Conversation and Interview. There are also register-specific socio-interactive bundles. Polite inquiry and greeting bundles are only found in Interview. These bundles are closely related to the formal and interactive nature of the interview discourse. Despite the overall rarity of the socio-interactive bundles in Speech, some functional types only appear in Speech: announcement, proposal, and disclaimer. These bundles are used to do some of the most basic actions in the formal academic presentation setting, such as making a formal announcement to begin one's presentation, stating one's opinion in a form of proposal, and inserting a disclaimer to avoid sounding too confident or presumptuous.

Referential bundles are common in both Speech and Interview, but their most frequently used sub-functional types differ. In Speech, conceptualization bundles are more common. By conceptualizing an entity, state, or action as an abstract notion, the speakers in Speech are able to objectify and distance themselves from the concept under discussion. In Interview, on the other hand, equivocation bundles, which indicate that the information being referred to is not exact, are

more common. The use of the equivocation bundles allows the speakers in Interview move forward with the flow of the ongoing dialogue without giving an exact reference.

Discourse organizers are more or less common in all three registers, but their functional sub-types differ. The majority of discourse organizing bundles are used to introduce a topic in Speech. Since Speech is more information oriented than other two registers, a clear identification of a topic is crucial in successfully conveying the often complex information and developing ideas. In Interview and Conversation, discourse organizing bundles are used to relate parts of discourse (cohesion) and to present background information (background). Because Interview and Conversation are interactive and under constraints of online production and comprehension without any visual aids (such as slides and handouts which are typical aids in academic Speech), indicating how parts of discourse are related is important in successfully carrying out a smooth interactive discourse. Some of the discourse organizing bundles also serve interactional functions. This suggests that the function of morphemic bundles is closely interrelated to the characteristics of the registers in which they are used. Both Interview and Conversation are much more interactive compared to Speech. Displaying concerns and attention to the co-participants is an important social activity to maintain the collaborative interaction.

Although stance bundles are infrequent in all three registers, in Conversation, some of the socio-interactional and discourse organizing bundles are found to serve stance-related functions, such as expressing uncertainty, disbelief, or emotional distance, as well. Again, this suggests that the particular functions of the bundles are being influenced by the functional needs of the speakers in the specific registers.

With respect to the overall frequency of the morphemic bundles, we have seen that the number of bundles is the largest in Interview. This can be accounted for by the fact that

Interview shares characteristics with both Conversation and Speech. Just like Conversation, Interview is interactive and more involved than Speech. At the same time, Interview also shares a formal discourse environment with Speech. These dual characteristics of Interview are reflected in the types of frequent morphemic bundles. As we can see in Figure 4.4 below, Interview shares socio-interactive and cohesion/background discourse organizing bundles with Conversation, and referential bundles with Speech.

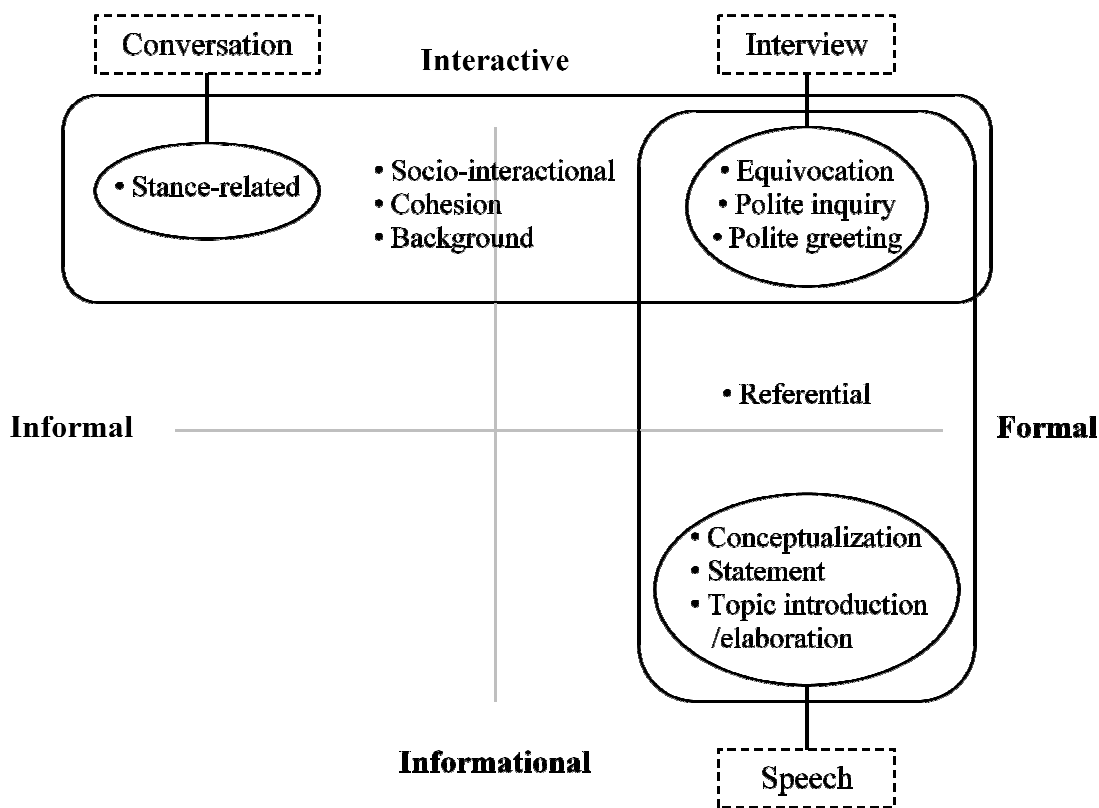


Figure 4.4. Register distribution of the bundle functional types

Another notable finding in terms of register variations is that form-function links are register-independent to certain extent. We have seen examples of some morphemic sequences being associated with different functions in different registers. First, the discourse organizing bundle *na + n + da/desu + kedo/keredo/keredomo/ga* [COP:ATT + NML + COP + but] is used

to introduce a topic in Speech in a majority of cases, while it is used to present background information in Interview and Conversation. A closer examination has also revealed that there are other interactional and stance-related functions in Interview and Conversation. In Interview, 19% of the bundle *na + n + desu + kedo* [COP:ATT + NML + COP + but] is used for interactional functions (see Section 4.3.2.3). In Conversation, 17% of the same bundle is used for interactional functions and 25% is used for stance-related functions (see Section 4.3.2.3). The apparent functional division, however, is not clear-cut. Though extremely rare, the topic introduction function is also found in Interview and Conversation, and the stance-related function is also found in Interview. Thus, the form-function links are more or less associated with the particular registers but these associations are not discrete.

The second example is the referential bundle *tte + iu + yoo + na* [QT + say + seem + COP:ATT]. In general, this bundle is used to signal that the given information is not exact (equivocal and noncommittal). Further analysis of each register has explicated that the speakers in each register employ the bundle for a particular effect. In Speech, the bundle is used to disclaim responsibility on the part of the speaker for the preciseness of the information being presented. In Interview, the bundle is sometimes used to leave interpretive options for the addressee. In Conversation, there are some cases in which the speaker expresses his/her affective stance through the use of this bundle. The register-specific functions may be considered pragmatic applications of the core equivocal function of the bundle. If the bundle is increasingly used to perform a specific pragmatic function in a register, there may be a stronger register-specific form-function connection.

The last example is the sequence *tai + to + omou* [want + QT + think]. The verb *omou* ‘think’ can take various forms such as *omoi + masu* [think + POL], *omot + te* [think + TE], and

*omou + n + desu* [think + NML + COP]. The sequence is used to make an announcement in Speech, whereas in Interview and Conversation, it is mostly used to report one's want or wish. In Interview, there are also some instances where interviewers use the sequence to make an announcement to begin or end the interviews.

In summary, the three bundle types exhibit different association patterns of functions and registers. The functions of the bundle *na + n + da/desu + kedo/keredo/keredomo/ga* [COP:ATT + NML + COP + but] are overlapped to a greater extent with Conversation and Interview (i.e., background information and interactional functions) than with Speech (topic introduction). In contrast, in the case of the sequence *tai + to + omou* [want + QT + think], Interview shares functions with both Speech (announcement) and Conversation (reporting one's wish). The bundle *tte + iu + yoo + na* [QT + say + seem + COP:ATT] shares the same basic function (equivocation) among all three registers, but each register also has a more specific function associated with the bundle. These divergent association patterns suggest that, while sharing the fundamental functional types (i.e., discourse organizing, interactional, and referential), the bundles are quite flexible and responsive to the needs of the particular registers.

One remaining issue concerns the formulaic status of the morphemic bundles. Throughout the examination of the functions of the morphemic bundles in this chapter, I have mentioned that at least some of the bundles do not seem to represent form-function unit. Section 4.6 deals with this topic.

#### **4.5. Cross-linguistic perspectives on the bundles**

This section presents cross-linguistic perspectives on the types and uses of the bundles found in English (see Biber 2004) and in Japanese. Although English and Japanese bundles

consist of different units (i.e., lexical units in English bundles and morphemic units in Japanese bundles), a comparison of the bundles brings to light some interesting characteristics of the two languages.

Structurally, bundles in both languages are incomplete, but the position in which they appear differ between the two languages. In Japanese, bundles tend to appear at clause or phrase finally. In English, on the other hand, bundles tend to appear clause or phrase initially (Biber et al. 2004). Many of the English bundles begin with personal pronouns such as *I* and *you*, while Japanese bundles incorporate copulas such as *da* and *desu*.

In terms of functions, there are three categories that are shared between English and Japanese bundles: stance expressions, discourse organizers, and referential expressions. While both English and Japanese epistemic stance bundles tend to express uncertainty rather than certainty, their associated forms differ. English epistemic stance bundles include the sequence, *I + don't + know*, or *I + think*, whereas Japanese epistemic stance bundles include the question particle, *ka*, the combination of the question and pragmatic particles, *ka + na*, or the combination of the copula and the auxiliary, *daro + o*. Although the direct translation of *think* in Japanese is *omou* 'think,' *omou* was found to have a socio-interactional reporting function rather than the function of marking the speaker's epistemic stance.

English discourse organizing bundles are concerned with topic-related functions (i.e., topic introduction, focus, elaboration, and clarification). This contrast sharply with Japanese discourse organizing bundles, most of which are used to link parts of discourse (cohesion) and present background information, especially in Conversation and Interview. This difference suggests that, so far as bundles are concerned, discourse organizing work is done through clear

identification of topics in English, while it is done through clear indication of how units of discourse are related in Japanese.

Particular functions of referential expressions also contrast between English and Japanese. In English, referential expressions are used to identify entities and specify particular attributes to the referents. In Japanese, in contrast, referential expressions are used to objectify and generalize the concepts as well as to equivocate the status of the information. Thus, referential bundles in the two languages work to opposite directions, that is, specification in English, and generalization in Japanese.

The pictures of the two languages that emerge from the observation above are as follows. Both English and Japanese tend to express uncertainty than certainty through linguistic strings, though with different frames/forms. The speakers in English discourse are concerned with identifying topics and referents clearly, whereas the speakers in Japanese discourse are more concerned with linking parts of discourse and maintaining the flow of discourse as well as smooth social interaction.

#### **4.6. Formulaic status of the morphemic bundles**

In the present study, we began with a form-based definition of formulaic language, and identified morphemic bundles, which are the most frequent multi-morphemic sequences in a given register. As we investigated their functions in this chapter, a question emerged regarding the formulaic status of the morphemic bundles. Here, we are talking about two different dimensions of the concept of formulaicity. In one aspect, formulaicity refers to the formal fixedness (see Section 1.3). In this sense of formulaicity, the morphemic bundles by definition are formulaic. However, in another sense of formulaicity, which is related to the fixedness in

terms of form-function association, only a quarter of the bundles represent formulaic units. The other three fourths of the bundles exhibit three different kinds of relationship with form-function formulaicity.

The first group of non-formulaic bundles (in the sense of form-function links) include shorter sequences that have particular functions. The following are three of examples of the shorter formulaicity found in the bundles.

- *ka + na* [Q + PP] (two morphemic sequence; part of epistemic stance bundle)
- *n + da/desu* [NML + COP] (two morphemic sequence; part of cohesion bundle)
- *yo + ne* [PP + PP] (two morphemic sequence; part of interpersonal bundle)

The second group of non-formulaic bundles are part of longer sequences that are linked to particular functions. This group is bigger than the first group. Below are some examples.

- *doo + na + n + daro + o* [how + COP:ATT + NML + COP:IMPF + AUX] ‘(I) wonder how’ (five morphemic sequence; extension of stance bundle)
- *dochira + ka + to + iu + to* [which + Q + QT + say + COND] ‘if (I) have to say which’ (five morphemic sequence; extension of stance bundle)
- *o + mi + te + mi + masu + to* [ACC + look + TE + ASP + POL + COND] ‘if (we) try and look at’ (six morphemic sequence; extension of topic-introducing bundle)
- *a + soo + na + n + da* [INJ + so + COP:ATT + NML + COP] ‘oh I see’ (five morphemic sequence; extension of reactive bundle)
- *nan + te + iu + no + ka + na* [what + QT + say + NML + Q + PP] ‘how to say’ (six morphemic sequence; extension of word/expression search bundle)
- *nan + te + iu + n + desu + ka* [what + QT + say + NML + COP + Q] ‘how to say’ (six morphemic sequence; extension of word/expression search bundle)
- *nan + te + iu + n + desho + o* [what + QT + say + NML + COP:IMPF + AUX] ‘how to say’ (six morphemic sequence; extension of word/expression search bundle)

- *de + wa + nai + desho + o + ka* [COP + TOP + NEG + COP + AUX + Q] (six morphemic sequence; extension of proposal bundle)

The last group of non-formulaic bundles are part of semi-fixed sequences. This group include the largest number of different morphemic bundles across three registers. There are three major sequences as represented below.

- (1) open slot + *to/tte/toka* + (*iu + fuu + ni*) + *omou/kangaeru*  
 private thought      QT      say way COP      think
- (2) open slot + *n* + *da/desu* + *kedo/keredo/keredomo/ga*  
 clause/phrase      NML      COP      but
- (3) open slot + *to/tte* + *iu* + (*yoo + na*) + *no/koto/mono/katachi* + *wa/ga/o/ni/de*  
 clause/phrase      QT      say      like COP      NML      PT/COP

These sequences are better characterized as semi-fixed constructions with open slots and variations in some of the morphemes as well as optional elements (see Section 5.1 for the definition of the semi-fixed construction). Since so many of the bundles from all three registers constitute these semi-fixed constructions, it is crucial that we further explore this type of formulaicity. The next chapter investigates the uses of these three semi-fixed constructions in more detail, with an aim of gaining a more realistic and comprehensive understanding of the native speakers' knowledge and use of language in relation to fixedness.

## Chapter 5

### Semi-Fixed Constructions in Spoken Japanese

#### 5.1. Introduction

Constructions are form-meaning/function pairs (e.g., Fillmore et al. 1988; Goldberg 1995) represented wholes in speakers' mind. Forms in constructions can be lexical or more complex patterns. Each construction can be fully lexically filled (most substantive), partially lexically filled, or fully abstract (most schematic). In the view of construction grammar, there is no strict division between lexicon and syntax as well as between semantics and pragmatics. Therefore, not only can a construction be a single word or a syntactic structure, information such as registers and contexts of use can also be represented in each construction.

In this chapter, we investigate the nature and functions of three partially filled, or semi-fixed, constructions in which many of the specific morphemic bundles take part. Forty-one percent of the bundles in Conversation (38/92), 49% of the bundles in Interview (93/189), and 53% of the bundles in Speech (92/175) are part of these constructions. Below are schematic representations of the three semi-fixed constructions. The label for each construction represents the construction's main function, which is further explicated in what follows.

#### (5-1) Thought-reporting construction

open slot + *to/tte/toka* + (*iu + fuu + ni*) + *omou/kangaeru*  
private thought      QT      say way COP      think

#### (5-2) Utterance-expansion construction

open slot + *n* + *da/desu* + *kedo/keredo/keredomo/ga*  
clause/phrase      NML      COP      but

(5-3) **Conceptualization construction**

open slot + *to/tte* + *iu* + (*yoo + na*) + *no/koto/mono/katachi* + *wa/ga/o/ni/de*  
clause/phrase      QT      say      like      COP                      NML                      PT/COP

For the purpose of investigating these constructions, examples with gaps within constructions are also considered. For instance, in the following example, two parts of the thought-reporting construction, *tai + to* [want + QT] and *omot + te* [think + TE] are separated by the supplemental information or qualification *nyuugaku tooji wa* ‘at the time of school entrance.’

(5-4) Gap within semi-fixed construction (Interview: Hypermedia\_009)

ずうっと 発展 途上 国 の 国際 教育 という もの に たず さ わ り たい と 入学  
当時は 思っ て いた ン だ す け ど

*zuutto hatten tojoo koku no kokusai kyooiku: to iu mono ni tazusawari tai*  
all.the.time development on.the.way country GEN international education QT say NML DAT involve **want**  
**to nyuugaku tooji wa omot te i ta n desu kedo ...**  
QT entrance then TOP **think TE** ASP PST NML COP but

‘I was thinking all the time, at the time of school entrance, that I wanted to get involved in the interactional education of developing countries ...’

Another thing to note while considering the above example is the relative positioning of the three constructions. As we can see in (5-4’), the conceptualization construction tends to be used towards the beginning of an utterance, the thought-reporting construction is used in the middle, and the utterance-expansion construction is used at the end of an utterance unit.

(5-4’) Three semi-fixed constructions (Interview: Hypermedia\_009)

ずうっと 発展 途上 国 の 国際 教育 という もの <Conceptualization>  
に たず さ わ り たい と 入学 当時は 思っ て <Thought-reporting>  
いた ン だ す け ど <Utterance-expansion>

zuutto hatten tojoo koku no kokusai kyooiku: to iu mono <Conceptualization>  
all.the.time development on.the.way country GEN international education QT say NML

ni tazusawari tai to nyuugaku tooji wa omot te <Thought-reporting>  
DAT involve want QT entrance then TOP think TE

i ta n desu kedo <Utterance-expansion>  
ASP PST NML COP but

In the previous chapter, the three constructions were described as having different functions in different registers. However, the fact that their use is widespread across all three registers begs a further investigation into the uses of these constructions. Intuitively, these constructions represent preferred ways of speaking in the Japanese socio-linguistic community. Rather than adding to the conceptual meaning of the utterances, the three constructions seem to contribute to smooth socio-pragmatic communication among discourse participants, whether they are friends having ordinary conversation, an interviewer and an interviewee in a formal interview, or a speech presenter and audience at an academic conference. The aim of this chapter is to provide a more precise account of the socio-pragmatic basis as well as significance of these semi-fixed constructions in the Japanese socio-linguistic community. The structure of this chapter is as follows. In Sections 5.2, 5.3, and 5.4, each construction is described in more detail. Then in Section 5.5, I discuss motivations for the use of the constructions and socio-pragmatic import of the constructions.

## 5.2. Thought-reporting construction

The first construction involves a self-directed, stance-related expression such as *tai* ‘want’ or *ka + na* [Q + PP] ‘(I) wonder’ and a quoting particle and verb *to/tte/toka + omou/kangaeru* [QT + think]. The construction optionally takes the sequence *iu + fuu + ni* [say + way +

COP:ADV]. The clause or phrase involving a self-directed, stance-related expression in the open slot is labeled “private thought” as in the schematic representation below.

(5-5) Thought-reporting construction

open slot + *to/tte/toka* + (*iu + fuu + ni*) + *omou/kangaeru*  
 private thought QT say way COP think

The following are the lists of stance-related expressions that appear at the end, within the private thought slot. They are listed in the order of frequency in each register.

(5-6) Stance-related expressions in the open slots in Conversation

Expression	No.	%	Gloss	Meaning
<i>ka + na(a)</i>	205	27.1%	[Q + PP]	<conjecture>
<i>na(a)</i>	172	22.7%	[PP]	<exclamation>
<i>(y)oo</i>	160	21.1%	[VOL]	<volition>
<i>ka</i>	120	15.9%	[Q]	<doubt>
<i>daro + o</i>	48	6.3%	[COP + AUX]	<speculation>
<i>ya</i>	12	1.6%	[PP]	<resignation>
<i>tai</i>	11	1.5%	[want]	<wish>
<i>e:</i>	11	1.5%	[INJ]	<surprise>
<i>nakya</i>	9	1.2%	[NEG+COND]	<obligation>
<i>kashira</i>	5	0.7%	[Q]	<doubt>
<i>cha</i>	2	0.3%	[TE+TOP]	<obligation>
<i>kya:</i>	1	0.1%	[INJ]	<obligation>
<i>uwa:</i>	1	0.1%	[INJ]	<surprise>

(5-7) Stance-related expressions in the open slots in Interview

Expression	No.	%	Gloss	Meaning
<i>na(a)</i>	71	24.9%	[PP]	<exclamation>
<i>(y)oo</i>	58	20.4%	[VOL]	<volition>
<i>ka</i>	54	18.9%	[Q]	<doubt>
<i>ka + na(a)</i>	46	16.1%	[Q + PP]	<conjecture>
<i>tai</i>	44	15.4%	[want]	<wish>
<i>daro + o</i>	12	4.2%	[COP + AUX]	<speculation>

(5-8) Stance-related expressions in the open slots in Speech

Expression	No.	%	Gloss	Meaning
<i>tai</i>	232	59.2%	[want]	<wish>
<i>ka</i>	110	28.1%	[Q]	<doubt>
( <i>y</i> ) <i>oo</i>	25	6.4%	[VOL]	<volition>
<i>ka</i> + <i>na</i> ( <i>a</i> )	15	3.8%	[Q + PP]	<conjecture>
<i>daro</i> + <i>o</i>	6	1.5%	[COP + AUX]	<speculation>
<i>na</i> ( <i>a</i> )	4	1.0%	[PP]	<exclamation>

Conversation has the most varied forms though some of these are rather rare. Interview and Speech both have six forms. While in Interview, the forms are more or less distributed equally, in Speech, one form, *tai* [want], predominates (close to 60%). We have seen in the previous chapter that the combination of *tai* [want] and the quotative *to* + *omou* [QT + think] is strongly associated with the function of making an announcement. It suggests that verb + *tai* + *to* + *omou* [verb + want + QT + think] has become a register-specific semi-fixed construction in Speech.

Interestingly, the most frequent private thought ending types differ among three registers. Conversation is characterized by conjecture and exclamation particles *ka* + *na* [Q + PP] and *na* [PP]. These particles are used most frequently at utterance final positions when one is talking to himself/herself (see Hasegawa 2010). About three fourths of the particle sequence *ka* + *na* and the particle *na* still appear utterance finally in Conversation, but the quarter of them are followed by the quotative sequence *to* + *omou* [QT + think]. The reported thought can be of the past, as in (5-9), or a current one, as in (5-10). The open slots are indicated by the curly brackets, { }.

(5-9) Thought-reporting construction (Conversation: Japan\_Bukatsu)

1. U: {他に俺何できるかな}って思ってバスケもバレーも無理じゃん。
2. M: んーうん。

1. U: {*hoka ni ore nani dekiru ka na*} *tte omot te basuke mo bare: mo muri jan.*  
other DAT I what can Q PP QT think TE basketball also volleyball also impossible AUX
2. M: *n: un.*  
INJ yes

1. U: 'I thought what else I could do. Basket ball and volleyball are impossible.'
2. M: 'Mhm, yeah.'

(5-10) Thought-reporting construction (Conversation: Callhome\_1263)

1. B: {何着て行こうかな} と思つて。短パンで歩いたら変かなとかさ。
2. A: なるほどね。

1. B: {*nani ki te ik oo ka na*} *to omot te. tanpan de arui ta ra hen ka na toka sa.*  
what wear TE ASP VOL Q PP QT think TE / shorts COP walk PST COND strange Q PP such PP
2. A: *naruhodo ne.*  
indeed PP

1. B: 'I'm thinking what I shall wear, like, would it be strange if I walked in my shorts?'
2. A: 'I see.'

In reporting one's inner speech, the speaker reveals his/her private thought, and at the same time, distances himself/herself from it. Whether the quoted inner speech represents past or present thought, the speaker is distancing himself/herself from his/her private self and foregrounding his/her interactional self (as one of the participants in discourse). The use of the thought-reporting construction enables the speaker to present his/her private thought as an interactional unit to the addressee. The interactional unit serves as the target of further interactional activities such as stance taking. One pragmatic effect of reporting one's inner thought as an interactional unit is a kind of mitigation or hedging. In the following excerpt, F002 tells her friend F080 that she and her classmates had to treat her teacher to lunch. She uses the thought-reporting construction in conveying her feeling of disgust at the teacher's attitude.

(5-11) *na + to + omou* [PP + QT + think] (Conversation: BTS\_31)

1. F002: だからいくら 100 円でもね {ちょっとそういうの嫌だなー} って  
思っ て。
2. F080: ああそうね。そういうのって私も嫌だと思っう。
  
1. F002: *dakara ikura 100 en de mo ne {chotto soo iu no iya da naa} tte omot te.*  
so how.much 100 yen COP also PP little that say NML disgusting COP PP QT think TE
2. F080: *aa soo ne. soo iu no tte watashi mo iya da to omou.*  
INJ so PP / that say NML QT I also disgusting COP QT think
  
1. F002: ‘So even it’s 100 yen, um, I think that kind of attitude is offensive.
2. F080: ‘Ah right. I think that kind of attitude is offensive, too.’

Compared to the direct conveyance of one’s inner thought in the modified version (5-11’) below, the thought-reporting construction in (5-11) modulates the strength of the disgust being conveyed.

(5-11’) Direct conveyance of inner thought (Conversation: BTS\_31; modified)

1. F002: だからいくら 100 円でもね ちよつとそういうの嫌だ なー。
2. F080: ああそうね。そういうのって私も嫌だ。
  
1. F002: *dakara ikura 100 en de mo ne chotto soo iu no iya da naa.*  
so how.much 100 yen COP also PP little that say NML disgusting COP **PP**
2. F080: *aa soo ne. soo iu no tte watashi mo iya da.*  
NJ so PP / that say NML QT I also disgusting COP
  
1. F002: ‘So even it’s 100 yen, um, that kind of attitude is offensive.’
2. F080: ‘Ah right. That kind of attitude is offensive to me, too.’

Another pragmatic effect achieved through the use of the construction is a sense of politeness.

The sense of politeness comes from the fact that the speaker is taking an extra step in quoting his/her own thought to present it as an interactional unit. The politeness effect is present in both Interview and Speech, as shown in the following excerpts.

In Interview, exclamation and volition are the two most frequently observed stance expressions. Compared to other two registers, however, five of the six types of expressions

found in Interview, namely, exclamation (*na*), volition ((*y*)*oo*), doubt (*ka*), conjecture (*ka + na*), and wish (*tai*), are more or less frequent. In the following excerpt, the interviewee O answers the interviewer N's inquiry about her major. In doing so, she uses the thought-reporting construction.

(5-12) Thought-reporting construction (Interview: Hypermedia\_009)

1. N: 専攻は？
  2. O: え 専攻は一
  3. N: うん
  4. O: {日本語 教育の方に進みたい}と思ってるんですけど
- 
1. N: *senkoo wa?*  
major TOP
  2. O: *e senkoo wa:*  
INJ major TOP
  3. N: *un*  
yes
  4. O: {*nihongo kyooiku no hoo ni susumi tai*} *to omot te ru n desu kedo*  
Japanese education GEN direction DAT advance want QT think TE ASP NML COP but
- 
1. N: 'Your major is?'
  2. O: 'Um, my major is'
  3. N: 'Yeah.'
  4. O: 'I'm thinking that I want to go into the field of Japanese language education'

It is possible for speaker O to convey the same message without the thought-quoting construction.

To do so, she can simply take out the quotative particle *to* [QT] and the verb phrase *omot + te + ru* [think + TE + ASP] as in the modified example (5-12') below.

(5-12') Direct conveyance of inner thought (Interview: Hypermedia\_009)

1. N: 専攻は？
  2. O: え 専攻は一
  3. N: うん
  4. O: 日本語 教育の方に進みたいんですけど
- 
1. N: *senkoo wa?*  
major TOP

2. O: *e senkoo wa:*  
INJ major TOP
  3. N: *un*  
yes
  4. O: *nihongo kyooiku no hoo ni susumi tai n desu kedo*  
Japanese education GEN direction DAT advance **want** NML COP but
1. N: ‘Your major is?’
  2. O: ‘Um, my major is’
  3. N: ‘Yeah.’
  4. O: ‘I want to go into the field of Japanese language education’

The original version with the thought-quoting construction, *susumi tai to omot te ru n desu kedo* [advance want QT think TE ASP NML COP but], conveys more personal, polite tone than the one without the construction, *susumi tai n desu kedo* [advance want NML COP but]. The latter sounds direct and impersonal, just expressing one’s wish. Again, the sense of politeness derives from distancing from the private-self and foregrounding of the interactional-self.

When the optional sequence *iu + fuu + ni* [say + way + COP] is included, the hedging effect is amplified. In (5-13) below, the interviewee A, who is a famous ice skater, talks about her transition from competitive figure skating to ice shows and exhibition. In line 5, she uses the thought-reporting construction with *iu + fuu + ni* to mitigate the strength of expressing her wish that she wants to skate in this kind of world (i.e., ice show). The use of mitigation here portrays the speaker as modest and socially mature because she is an accomplished figure ice skater who was new to professional ice shows.

(5-13) Thought-reporting construction with *iu + fuu + ni* (Interview: Tetsuko\_5)

1. A: そうですねちょうどその少し前にアイスショーを初めて生で見ることがありまして
2. K: ええ。
3. A: であんな世界もあるんだっていうすごく衝撃を受けたんですね。
4. K: ええ。
5. A: で{私もこういう世界で滑れたらいいな}っていう風に思いまして

6. A: そこにすごく 憧れを いただいた んですね。
1. A: *soo desu ne choodo sono sukoshi mae ni aisu sho: o hajimete nama de miru koto ga ari*  
so COP PP just that little before DAT ice show ACC first real COP watch NML NOM exist  
*mashi te*  
POL TE
2. K: *ee.*  
yes
3. A: *de a konna sekai mo aru n da tte iu sugoku shoogeki o uke ta n desu ne.*  
and INJ this.kind world also exist NML COP QT say very shock ACC receive PST NML COP PP
4. K: *ee.*  
yes
5. A: *de {watashi mo koo iu sekai de sube re tara ii na} tte iu fuu ni omoi mashi te*  
and I also this say world LOC skate POT COND good PP QT say way COP think POL TE
1. A: ‘Let me see, I had a chance to see an ice show live for the first time a little bit before that’
2. K: ‘Yes.’
3. A: ‘And, I was really shocked and thought oh this kind of world exists.’
4. K: ‘Yes.’
5. A: ‘And I thought it would be good if I could skate in this kind of world’

In Speech, the expression of wish with *tai* [want] predominates in the open slot within the thought-reporting construction. As mentioned above, the thought-reporting construction with *tai* [want] has the function of making an announcement.

(5-14) Thought-reporting construction (Speech: CSJ\_A07F0445)

えー {今回は特に牛乳パックのリサイクルについて えー 焦点を当てて  
検討していき たい} と 思い ます

*e: {konkai wa tokuni gyuunyuu pakku no risaikuru ni tsui te e: shuuten o ate te kentoo shi te*  
INJ this.time TOP especially milk carton GEN recycle DAT adhere TE INJ focus ACC target TE consider do TE  
*iki tai} to omoi masu*  
ASP want QT think POL

‘Um, this time, we’d like to consider recycling behavior, um, focusing on the milk carton recycling in particular.’

As discussed earlier in Section 4.3.4.1, the speakers in Speech prefer to use the thought-quoting construction over the simple statement (see (5-14’) below) in making an announcement. The use

of the thought-reporting construction conveys a sense of politeness, portraying the speaker as personal and socially aware at the same time. The simple statement accomplishes the same interactional function of making an announcement, but the social image it portrays differs from the one depicted by the thought-reporting construction. The simple statement portrays the speaker as impersonal and information-oriented, as shown in (5-14').

(5-14') Simple statement (Speech: CSJ\_A07F0445; modified)

えー 今回 は 特に 牛乳 パック の リサイクル について えー 焦点 を 当て て 検討 して いきます

*e: konkai wa tokuni gyuunyuu pakku no risaikuru ni tsui te e: shuuten o ate te kentoo shi te*  
INJ this.time TOP especially milk carton GEN recycle DAT adhere TE INJ focus ACC target TE consider do TE  
*iki masu*  
ASP POL

'Um, this time, we will consider recycling behavior, um, focusing on the milk carton recycling in particular.'

The second most frequent stance expression in Speech is doubt *ka* [Q]. A closer examination of the instances reveals that the thought-reporting construction with *ka* [Q] also conveys a sense of politeness and mitigates the strength of a claim. Many of the thought-reporting construction instances with *ka* actually include a longer sequence, *no de wa nai ka* [NML COP TOP NEG Q], as in (5-15) below.

(5-15) Thought-reporting construction (Speech: CSJ\_A03M0059)

えー {これにより 統合された表が えー 検索 の 際 えー 一種 の ディレクトリー と して の 役割 が 期待 できる のではないか} と 考えて います

*e: kore ni yori toogoo s are ta hyoo ga e: kensaku no sai e: is shu no direkutori: to*  
INJ this DAT depend unify do PASS PST chart NOM INJ search GEN case INJ one kind GEN directory QT  
*shi te no yakuwari ga kitai dekiru no de wa nai ka to kangae te i masu*  
do TE GEN role NOM expect can NML COP TOP NEG Q QT think TE ASP POL

‘Um, in this way, we are thinking that the unified chart may be able to serve as a kind of directory at the time of searching.’

The sequence *no de wa nai ka* [NML COP TOP NEG Q] can be omitted altogether and the remaining utterance still conveys the same message, as in (5-15’). The difference between (5-15) and the modified (5-15’) is the strength of the claim being made. Utterance (5-15’) conveys a higher degree of certainness in the claim that the chart can serve as a kind of directory.

(5-15’) Non-thought-reporting construction (Speech: CSJ\_A03M0059; modified\_1)

えーこれにより統合された表がえー検索の際 えー種のディレクトリーとしての役割が期待できると考えています

*e: kore ni yori toogoo s are ta hyoo ga e: kensaku no sai e: is shu no direkutori: to*  
INJ this DAT depend unify do PASS PST chart NOM INJ search GEN case INJ one kind GEN directory QT  
*shi te no yakuwari ga kitai dekiru to kangae te i masu*  
do TE GEN role NOM expect can QT think TE ASP POL

‘Um, in this way, we are thinking that the unified chart can serve as a kind of directory at the time of searching.’

It is also possible to convey the same message without the quotative particle *to* [QT] and the verb *kangaeru* [think]. The resulting utterance is a simple statement as shown in (5-15’). The simple statement (5-15’) expresses the highest degree of confidence in the claim that the chart can serve as a directory, compared to (5-15) and (5-15’).

(5-15’) Simple statement (Speech: CSJ\_A03M0059; modified\_2)

えーこれにより統合された表がえー検索の際 えー種のディレクトリーとしての役割が期待できます

*e: kore ni yori toogoo s are ta hyoo ga e: kensaku no sai e: is shu no direkutori: to*  
INJ this DAT depend unify do PASS PST chart NOM INJ search GEN case INJ one kind GEN directory QT  
*shi te no yakuwari ga kitai deki masu*  
do TE GEN role NOM expect can POL

‘Um, in this way, the unified chart can serve as a kind of directory at the time of searching.’

Now, it is important to note that the thought-reporting construction with *ka* [Q] does not express the speaker's doubt in his/her claim but simply mitigates the strength of his/her claim for social purposes. In academic conferences where scholars report the findings from their own studies in their field of expertise, we should expect them to be confident of what they talk about. The thought-reporting construction does not decrease the credibility of their findings. Rather, the construction contributes to depicting a particular social persona of the speaker as being modest and socially mature.

The next example of the thought-reporting construction with *ka* [Q] does not include the longer sequence *no de wa nai ka* [NML COP TOP NEG Q] as in (5-15) above, yet the mitigating and politeness effects are still observed.

(5-16) Thought-reporting construction (Speech: CSJ\_A03M0045)

{まだえこの点に関しては改善の余地があるか}と思われ  
ます

{*mada e kono ten ni kanshi te wa kaizen no yochi ga aru ka*} *to omow are masu*  
yet INJ this point DAT concern TE TOP improvement GEN space NOM exist Q QT think PASS POL

'We think that there may be still room for improvement regarding this point.'

(5-16') Non-thought-reporting construction (Speech: CSJ\_A03M0045; modified\_1)

まだえこの点に関しては改善の余地があると  
思われ  
ます

*mada e kono ten ni kanshi te wa kaizen no yochi ga aru to omow are masu*  
yet INJ this point DAT concern TE TOP improvement GEN space NOM exist QT think PASS POL

'We think that there is still room for improvement regarding this point.'

(5-16'') Simple statement (Speech: CSJ\_A03M0045; modified\_2)

まだえこの点に関しては改善の余地があります

*mada e kono ten ni kanshi te wa kaizen no yochi ga ari masu*  
 yet INJ this point DAT concern TE TOP improvement GEN space NOM exist POL

‘There is still room for improvement regarding this point.’

In this section, we have seen that the thought-reporting construction has the function of expressing politeness and mitigating the strength of (and softening the directness of) a claim, opinion, wish, intention, or affective expression. These socio-interactional functions derive from the distancing effect of reporting one’s inner speech. By distancing oneself from one’s private thought, the speaker foregrounds his/her role as a social, interactional participant. In a larger socio-linguistic perspective, the successful use of the construction indexes the socio-pragmatic competence of the speaker. The construction is not typical of children’s speech because they have not had any or enough experience in the socio-linguistic community outside of the individual families.

### 5.3. Utterance-expansion construction

The utterance-expansion construction has the following structure. The open slot is located prior to the nominalizer *n* [NML] where a clause or phrase followed by the attributive form of a copula, *na*, appears. The open slot is followed by the nominalizer *n* and the copula *da/desu*. The final element *kedo* [but] has four variant forms.

(5-17) Utterance-expansion construction

<u>open slot</u>	+	<i>n</i>	+	<i>da/desu</i>	+	<i>kedo/keredo/keredomo/ga</i>
clause/phrase		NML		COP		but

The sequence *n(o) + da/desu* [NML + COP] has been recognized as a unit by many linguists (e.g., Iwasaki 1985; Maynard 1992; Noda 1997; Najima 2007). While there are various

pragmatic uses of the *n + da/desu* sequence, as Iwasaki (1985) states, and as we have seen above in section 4.3.2.2, the fundamental characteristics of the sequence is that it has the cohesive power to relate a clause or phrase to other parts of a discourse. Formal fixedness of the functional unit is also affirmed through the analysis of the morphemic bundles above and additional statistical calculations which measure the collocational strength between *n* [NML] and *da/desu* [COP].<sup>45</sup>

In this section, I provide evidence and analysis that the utterance-expansion construction with the longer sequence *n + da + kedo* [NML + COP + but] represents a type of form-function unit, that is not found in the *n + da* sequence or the particle *kedo* (or its variants). First, statistical evidence is provided to establish the structural formulaicity of the sequence *n + da + kedo* across three registers. Then, uses of the utterance-expansion construction are scrutinized, and compared to those of the *n + da* sequence and the utterance-final *kedo*-type conjunctive particles, to illustrate the unique role of the construction in spoken discourse.

The co-occurrence of the nominalizer *n*, the copula *da/desu*, and the conjunctive particle *kedo* (or its variants) is extremely frequent. The collocational strength among these morphemes in each register is measured using two statistical formulae, MI (mutual information) and z-score.<sup>46</sup> We have already established the collocational status of the sequence *n + da/desu*. Therefore, for the purpose of measuring the collocational strength between the string *n+da/desu* and following items, the sequence *n+da/desu* is taken for granted as a single item in the measures. The following tables show the most frequent right collocates of the sequence *n+da/desu* in the three registers (for Conversation, *n+da* is used as the node, and for Interview

---

<sup>45</sup> Two statistical tests, MI (mutual information) and z-test, are conducted to test the collocational strength between the two items. The MI values and z-scores for *n* [NML] and *desu* [COP] are 5.14 / 203.22 (Speech), 4.14 / 216.42 (Interview), and 4.77 / 133.27 (Conversation). The MI values and z-scores for *n* [NML] and *da* [COP] are 3.64 / 59.05 (Interview) and 4.31 / 229.77 (Conversation).

<sup>46</sup> See McEnery et al. (2006) for other statistical formulae used to test collocational strength.

and Speech, *n+desu* is used as the node). The collocates are ranked by their z-score values. Following Hunston (2002), a MI score of 3 or higher is considered an indicator of a significant collocation. For the z-score, I avoid using standard thresholds but only use the scores to rank collocates. The tables demonstrate that the *kedo*-type conjunctive particles are strongly associated with the sequence *n+da/desu* in all three registers.

Table 5.1. Collocates of *n+da* ( $F(n)^{47} = 2941$ ) in Conversation (span +1)

Rank	Collocate	F(c) <sup>48</sup>	F(n,c) <sup>49</sup>	MI value	z-score value
1	<b><i>kedo</i></b>	<b>2899</b>	<b>1024</b>	<b>5.70</b>	<b>226.14</b>
2	<i>yo</i> [PP]	6045	596	3.86	86.83
3	<b><i>kedomo</i></b>	<b>144</b>	<b>60</b>	<b>5.94</b>	<b>59.01</b>
4	<i>mon</i> [PP]	649	73	4.05	32.56
5	<i>kke</i> [Q]	438	50	4.07	27.15
6	<b><i>keredo</i></b>	<b>69</b>	<b>16</b>	<b>5.09</b>	<b>22.64</b>
7	<b><i>keredomo</i></b>	<b>53</b>	<b>13</b>	<b>5.17</b>	<b>21.29</b>
8	<i>tte</i> [QT]	5534	161	2.10	20.13

Table 5.2. Collocates of *n+desu* ( $F(n) = 1773$ ) in Interview (span +1)

Rank	Collocate	F(c)	F(n,c)	MI value	z-score value
1	<b><i>kedo</i></b>	<b>1165</b>	<b>311</b>	<b>5.09</b>	<b>123.97</b>
2	<b><i>keredomo</i></b>	<b>626</b>	<b>255</b>	<b>5.70</b>	<b>112.59</b>
3	<i>ka</i> [Q]	4663	482	3.72	74.05
4	<b><i>kedomo</i></b>	<b>182</b>	<b>90</b>	<b>5.98</b>	<b>73.93</b>
5	<b><i>keredo</i></b>	<b>92</b>	<b>42</b>	<b>5.86</b>	<b>48.42</b>
6	<i>yo</i> [PP]	1130	150	4.11	47.31
7	<i>ne</i> [PP]	4720	202	2.45	27.23
8	<i>ga</i> *	4458	129	1.88	15.94

\**ga* includes both nominative case particles and conjunctive particles.

<sup>47</sup> F(n) = Frequency of the node

<sup>48</sup> F(c) = Frequency of the collocate

<sup>49</sup> F(n,c) = Co-occurrence of the node and the collocate

Table 5.3. Collocates of *n+desu* (F(n) = 1249) in Speech (span +1)

Rank	Collocate	F(c)	F(n,c)	MI value	z-score value
1	<i>keredomo</i>	<b>842</b>	<b>346</b>	<b>6.86</b>	<b>198.35</b>
2	<i>kedomo</i>	<b>249</b>	<b>144</b>	<b>7.35</b>	<b>152.03</b>
3	<i>kedo</i>	<b>118</b>	<b>60</b>	<b>7.17</b>	<b>91.44</b>
4	<i>ga</i> *	9644	493	3.85	79.44
5	<i>ne</i> [PP]	1077	94	4.62	46.12
6	<i>keredo</i>	<b>20</b>	<b>11</b>	<b>6.74</b>	<b>34.48</b>
7	<i>yo</i> [PP]	160	22	5.28	28.27
8	<i>ka</i> [Q]	2573	33	1.86	7.92

\**ga* includes both nominative case particles and conjunctive particles.

The construction, clause/phrase + *n + da + kedo* [clause/phrase + NML + COP + but], shares the function of linking parts of discourse with the sequence *n+da* [NML+COP], as well as the function of expressing contrastive meaning with the conjunctive particle *kedo* [but].

However, the function of the construction is not simply a sum of its components, the sequence *n+da* and the conjunctive particle *kedo*. To demonstrate that the construction constitutes a unique form-function unit, in this section, I compare the function of the construction with that of the *n+da* sequence and of the particle *kedo* in detail.

First, while both the construction, clause/phrase + *n + da + kedo*, and the sequence, *n+da*, have the function of relating a clause/phrase to other parts of a discourse, only the construction conveys the subordinate nature of the clause/phrase in the open slot. This difference results in different usages of the two. As we have seen in Section 4.3.2.2, the sequence *n+da* is used to project further talk to which the clause/phrase is related (Interview and Conversation), to give a summary of or explanation for the information in the prior context (Speech), or to ask for further explanation (Interview and Conversation). On the other hand, as we have seen in Section 4.3.2.3, the sequence *n + da + kedo* is used to present background information (three registers), to give

interactional options for how to interpret the relevance of the information and how to respond to the information to the addressee (Interview and Conversation), to express the feeling of disbelief or distance (Conversation), or to introduce a topic (Speech). All of these functions derive from the function of the construction to indicate the subordinate/background nature of the information in the open slot. The sense of backgroundness is obtained because the construction entails a main clause, which may or may not have been expressed at the point of the utterance. The introduction of a topic can also be regarded as the presentation of background information. The function of giving interactional options to the addressee is observed where the main thrust of an utterance has not been clearly explicated. Finally, the expression of the speaker's particular stance, such as disbelief or distance, also assumes the existence of the foreground information, from which the speaker emotionally distances himself/herself. Therefore, while the sequence *n+desu* merely indicates that the clause is related to some other parts of discourse, the construction, clause/phrase + *n + da + kedo*, further specifies the kind of relationship the information has to the other part of discourse (i.e., background-foreground or subordinate-main relationship).

It seems natural to think that the sense of backgroundness comes from the meaning of the conjunctive particle *kedo* whose original function is to connect clauses that express opposing or contrasting ideas. Nakayama and Nakayama (1997), for example, claim that the non-contrastive use of *kedo* in spoken narrative has the function of presenting background information. However, three of the four examples they present actually contain the sequence *n + da + kedo* (in other words, the utterance-expansion construction). These three examples indeed present background information. On the other hand, one example with the particle *kedo* has a slightly different function. *Kedo*, besides being used to indicate a contrastive meaning, is used to add information

interpolatively or incrementally. This is different from the background information presentation function of the utterance-expansion construction, which entails the existence of a main message or information. In fact, conversational examples of *kedo*-attached utterances presented in Nakayama and Nakayama (1997) express the speakers' main messages, which contrast with the supplemental nature of the information expressed through the utterance-expansion construction.

Below are three narrative examples from Nakayama and Nakayama (1997) with the sequence *n + da + kedo*. As they explain, all of these examples have the function of presenting background information, though, in my view, the function is attributed to the utterance-expansion construction with the sequence *n + da + kedo*, not the particle *kedo*. In the excerpts, the curly brackets, { }, indicate the open slots. In (5-18), the speaker first gives the background situation with the utterance-expansion construction, then moves on to talk about a specific event.

(5-18) Background information: *n+desu+kedo* ((2) in Nakayama and Nakayama 1997: 609)

{佐藤君 高橋君 なんかと、えー、よくつるんでた} **ん** **です** **けど**、例えば、新しい  
 出会いを求めて行ったーのにも関わらず、ただ日焼けだけして帰ってきたこ  
 うず島のツアーとかですね、...

{*sato kun takahashi kun nanka to, e:, yoku tsuru n de ta*} ***n desu kedo***, *tatoeba, atarashii deai o*  
 (name) SFX (name) SFX DM with INJ often hang.out NML TE PST NML COP but new encounter ACC  
*motome te it ta no ni mo kakawara zu, tada hiyake dake shi te kaet te ki ta koozu shima no tua:*  
 seek TE go PST NML DAT also concern NEG only sun.tan only do TE return TE ASP PST (name) island GEN tour  
*toka desu ne, ...*  
 such COP PP

‘We used to hang out quite a lot with Mr. Sato and Mr. Takahashi, and others — for example, we made a trip to the Koozu Island where we hoped to meet someone but only got a sun tan ...’

In (5-19), the construction is again used by the speaker to present background information before proceeding to tell the main story line.

(5-19) Background information: *n+desu+kedo* ((3) in Nakayama and Nakayama 1997: 610)

{「キャシー、アイラブユー」っていう台詞があった}んですけど、それがですね、何回やってもできないんですね、...

{“Kathy, I love you,” tte iu serifu ga at ta} **n desu kedo**, sore ga desu ne, nan kai yat te mo deki nai (quote) QT say line NOM exist PST **NML COP but** that NOM COP PP how time do TE also can NEG  
*n desu ne*, ...  
NML COP PP

‘(he) had a line that said “Kathy, I love you,” and (he) couldn’t (say) it no matter how many times (he) tried, ...’

The segment in (5-20) precedes that of (5-19). Again, the utterance-expansion construction is used to present background information about the girl under discussion (that she is not attending the reception today).

(5-20) Background information: *n+desu+kedo* ((5) in Nakayama and Nakayama 1997: 611)

あの、ある- ここに、{この場にはちょっとー今日は出席してない}んですけど、ある女性の、まあ我々の同期の女の子にですね、えー、「キャシーアイラブユー」っていう台詞があったんですけど、...

*ano, aru- koko ni*, {*kono ba ni wa chotto: kyoo wa shusseki shi te nai*} **n desu kedo**, *aru josei no*, INJ here LOC this place LOC TOP little today TOP attend do TE NEG NML COP but one lady GEN  
*maa wareware no dooki no onnanoko ni desu ne*, e: “Kathy, I love you” tte iu serifu ga at ta *n desu kedo*, ...  
COP but

‘Well, (he) had a line that said “Kathy, I love you” to one — one who is not attending this reception — girl who was in our class ...’

One narrative example in Nakayama and Nakayama (1997), which only contains the particle *kedo*, is presented below. This example demonstrates the information adding function of the particle *kedo* (*keredomo* in this particular case), which differs from the backgrounding function of the utterance-expansion construction with *n + da + kedo*. In this particular example,

the information is added interpolatively in the middle of producing an utterance. The interpolative clause is indicated by the square brackets, [ ].

(5-21) Parenthetical information: *keredomo* ((4) in Nakayama and Nakayama 1997: 610)

まあ今流行の遠距離恋愛というのを何年間か、[まあ遠距離といたしても、かなりの遠距離ですけれども]、なさってたわけですけれども、...

*maa ima hayari no en kyori renai to iu no o nan nen kan ka, [maa en kyori to ii*  
INJ now trend GEN long distance relationship QT say GEN ACC how year duration Q INJ long distance QT say  
*mashi te mo, kanari no en kyori desu keredomo], nasat te ta wake desu keredomo, ...*  
POL TE also extreme GEN long distance COP **but** do TE PST NML COP but

‘Well, they were in a trendy long-distance, for some years, — it was truly long-distance in their case — relationship.’

What the speaker is doing in (5-21) is not presenting background information for some other part of discourse, but interpolatively adding information, or doing a self-repair on what has been said just prior.

In conversational interaction, both the utterance-expansion construction with *n + da + kedo* and the particle *kedo* are used for stance-related functions. We have seen in Section 4.3.2.3 that the *n+da+kedo* sequence expresses the speaker’s disbelief or distance from what has been expressed in the prior context. *Kedo*, on the other hand, functions to soften the assertive tone of the utterance or mitigate the force of illocutionary force (Nakayama and Nakayama 1997). In the following example from Nakayama and Nakayama (ibid, p. 612), the speaker uses the utterance final *kedo* to attenuate the tone of his assertion. This utterance is made in the context where the discourse participants are gossiping about a student who turned down the offer from their department. The utterance below is a response to another speaker’s utterance ‘Wasn’t she from UCSD?’

(5-22) Softening assertion with *kedo* ((6) in Nakayama and Nakayama 1997: 612)

Y: サンディエゴステイトって噂を聞いたけど。

Y: *San Diego State tte uwasa o kii ta kedo.*  
(name) QT rumor ACC hear PST but

Y: ‘I heard (she) is from San Diego State (University).’

What is important for our discussion is that this utterance expresses the speaker’s main message whose assertive tone he softens with the particle *kedo*. The utterance-final use of the particle *kedo* does not express the backgroundness of the information.

In contrast, in the use of the utterance-expansion construction in (5-23) below, the speaker expresses her disbelief or emotional distance from the foreground information that the school’s spring break is three-months long. The utterance-expansion construction does not mitigate the strength of the stance expression, but it conveys the background status of the information in the open slot, and at the same time, suggests the existence of the foreground information to which the background information is linked. In this way, the use of the construction in this context signals the distance between the two pieces of information.

(5-23) Utterance-expansion construction with stance expression (Conversation: BTS\_72)

1. F086: うん うん。だって超長いでしょ春休みー。
  2. F046: なーがい
  3. F086: 一番長いよねー多分。
  4. F046: なーがい長い。だって三か月ぐらいない？
  5. F086: そんなにある。だって普通にさ1月始まったってさ1月の最後らへんに終わるんでしょ？
  6. F046: 終わる終わるー。
  7. F086: {ありえない}んだけど。
  8. F046: ないないない。
- 
1. F086: *un un. datte choo nagai desho haru yasumi:*  
yes yes since very long COP spring break

2. F046: *na:gai*  
long
3. F086: *ichiban nagai yo ne: tabun.*  
best long PP PP probably
4. F046: *na:gai nagai. datte san kagetsu gurai nai?*  
long long since three months about NEG
5. F086: *sonna ni aru. datte futsuu ni sa ichi gatsu hajimat ta tte sa ichi gatsu no saigo*  
that.much COP exist / since usually COP PP one month begin PST QT PP one month GEN end  
*ra hen ni owaru n desho?*  
about area DAT end NML COP
6. F046: *owaru owaru:.*  
end end
7. F086: *{ari e nai} n da kedo.*  
exist POT NEG NML COP but
8. F046: *nai nai nai.*  
NEG NEG NEG
9. F086: ‘Yeah yeah. The spring break is extremely long, you know.’
10. F046: ‘It’s long.’
11. F086: ‘It’s the longest, probably.’
12. F046: ‘It’s so long. Isn’t it about three months?’
13. F086: ‘It’s that long. You know, it begins in January but ends at the end of January.’
14. F046: ‘Yeah it does.’
15. F086: ‘It’s impossible.’
16. F046: ‘You are right.’

In this section, I have explicated the unique function of the utterance-expansion construction by comparing and contrasting it with the functions of the sequence *n+da/desu* and the utterance-final use of the particle *kedo* (or its variants). The utterance-expansion construction objectifies the information in the open slot and indicates the subordinate status of the information. The construction serves as a kind of discourse frame for bridging and indicating relationship between units of discourse. The use of the construction is crucial for smooth on-line production and comprehension of extended spoken discourse in Japanese.

#### 5.4. Conceptualization construction

A conceptualization construction conceptualizes the preceding entity, state, or action as an abstract notion. It distances the speaker from the entity/state/event, and encourages the

addressee to grasp the entity/state/action as a whole, not from a point of view of a concrete subject/object. The construction has an open slot for a clause or phrase, followed by the quotative *to/tte* + *iu* [QT + say], an optional inclusion of the sequence *yoo+na* [like + COP:ATT], a nominalizer *no/koto/mono*, and the particle *wa/ga/o/ni* or the copula *de*.

(5-24) Conceptualization construction

open slot + *to/tte* + *iu* + (*yoo + na*) + *no/koto/mono/katachi* + *wa/ga/o/ni/de*  
 clause/phrase      QT      say      like   COP                      NML                      PT/COP

Since the conceptualized idea is a generalized concept, it can be accessed by both the speaker and the addressee in the same manner. This brings various interactional effects, as can be seen in the following examples. In the first excerpt below, the interviewer M has the information that the interviewee S is a housewife. When the interviewer asks the interviewee about a typical day for a housewife, she uses the conceptualization construction to generalize the concept of housewife. Although she knows the interviewee is a housewife, the question is formulated in a way that separates the interviewee from the general concept of a housewife. In this way, the conceptualization construction allows them to talk about the topic abstractly. It also avoids directly labeling S a housewife (the use of the topic particle only, i.e., *go katei no shufu wa* [HON family GEN housewife TOP], would be the direct labeling of M as a house wife), which can sound presumptuous or even offensive.

(5-25) Conceptualization construction (Interview: Hypermedia\_019)

1. M: あのー {ご家庭の主婦} っていうのはだいたい一日はどういう風にお過ごしになってるんですか？
2. S: そうですねまあわたくしはごくごく一般的ですけれども

1. M: *ano: {go katei no shufu} tte iu no wa daitai ichi nichi wa doo iu fuu ni o sugoshi*  
 INJ HON family GEN housewife QT say NML TOP usually one day TOP how say way DAT HON spend

*ni nat te ru n desu ka?*

COP become TE ASP NML COP Q

2. S: *soo desu ne maa watakushi wa gokugoku ippan teki desu keredomo*  
so COP PP INJ I TOP ONM general like COP but
1. M: 'Um, how do housewives typically spend a day?'
2. S: 'Let me see, um, my case is ordinary'

The conceptualization construction is also used in Conversation where the participants have rather close relationships with one another, such as friends and family. In the following excerpt, the speaker U tells his friend M about his after school sports club experience in middle school. He uses the conceptualization construction in telling M that he could not see any meaning in continuing his club activity when he was not even a regular team member. Here, the act of continuing a club activity when one is not a regular team member is being conceptualized and generalized. By doing so, the speaker characterizes such deed as something any member of the Japanese socio-linguistic community can identify with (whether positively or negatively) rather than a specific event in his particular case. Thus, the use of the construction makes the socially and culturally shared knowledge relevant for understanding the information in the open slot.

(5-26) Conceptualization construction (Conversation: Japan\_Bukatsu)

1. U: {レギュラーでもないのに続ける}っていうのは
2. M: んん。
3. U: 何の意味も見いだせなかったのね。
1. U: {*regyura: de mo nai no ni tuzukeru*} ***tte iu no wa***  
regular COP also NEG though continue QT say NML TOP
2. M: *nn.*  
INJ
3. U: *nan no imi mo miidase nakat ta no ne.*  
what GEN meaning also find NEG PST NML PP
1. U: 'Continuing (the club activity) even though I was not a regular member'
2. M: 'Mhm.'

3. U: ‘I couldn’t find any meaning in it.’

In the next two examples, the clauses in the open slots contain stance-related expressions, *darō + o* [COP + AUX] ‘probably’ and *hoshii* [want]. Both examples also include the copula *de*, which is preceded by the nominalizer *koto*. In (5-27), the interviewee F tells the interviewer M what her family did during the summer break. The clause in the open slot, *kaigai ryokoo wa shibaraku nai darō o* ‘(we) probably wouldn’t travel abroad for a while’ represents what F and F’s family members discussed or the conclusion reached through the discussion. While the use of *darō + o* ‘probably’ implies uncertainty on the part of F and F’s family whether or not they would travel abroad in near future, the conceptualization construction abstracts the core idea and presents the whole idea as given. The use of the construction makes specific information irrelevant, such as who actually said that they probably would not travel abroad in the near future or whether or not they all agreed on this point. By using the conceptualization construction, the speaker is able to provide a reason for going to Hong Kong in a succinct manner without giving specific details or having the risk of being asked further questions about the reason.

(5-27) Conceptualization (Interview: Hypermedia\_034)

1. F: {海外旅行はしばらくない だろう} という こと であの 香港 に 行っ て 来 た り

2. M: ええ ええ

3. F: 夏 休 み を 利 用 し ま し て ね

4. M: ええ

5. F: 参 り ま し た で す け れ ど も 。

1. F: {*kaigai ryokoo wa shibaraku nai darō o*} **to iu koto de ano honkon ni it te ki tari**  
overseas travel TOP for.a.while. NEG COP AUX QT say NML COP INJ (name) DAT go TE ASP such.as

2. M: *ee ee*  
yes yes

3. F: *natsu yasumi o riyoo shi mashi te ne*  
summer break ACC use do POL TE PP

4. M: *ee*  
yes
5. F: *mairi mashi ta desu keredomo.*  
go:HUM POL PST COP but
1. F: ‘We decided that we probably wouldn’t travel abroad for a while, and went such place as Hong Kong’
2. M: ‘Yes yes.’
3. F: ‘We used the summer break and’
4. M: ‘Yes.’
5. F: ‘We went.’

The next excerpt is from Speech. In (5-28), the speaker provides a disclaimer before presenting her research. Prior to the segment shown below, the speaker says her specialization is neither Japanese dialects nor Japanese phonology, which are the major topic of the conference, but it is English phonology. Then, in the segment shown below, she expresses what the organizer(s) of the conference or the panel expressed to her, using the conceptualization construction. If she just wanted to express what she was told, she could have used a simple quotation. What the speaker wanted to express instead is the reason she decided to present her research even though she declined first. The conceptualization construction works to this aim as it objectifies the speech event and abstracts the core point. Again, the use of the construction makes specific details of who actually said this to the speaker or the particular words or expression used totally irrelevant.

(5-28) Conceptualization construction (Speech: CSJ\_A05F0154)

最初にお話をいただいた時にまそういう風にお断わりしたんですけれども  
あの一ま {日本語の えー ネイティブスピーカーとしての発言も 欲しい} という  
ことで えーそういう立場から少し えー発言させていただきます

*saisho ni o hanashi o itadai ta toki ni ma soo iu fuu ni o kotowari shi ta n desu keredomo ano ma*  
first DAT PRF talk ACC receive PST when DAT INJ that say way DAT decline do PST NML COP but INJ INJ  
{*nihongo no e: neitibu supi:ka: to shi te no hatsugen mo hoshii*} **to iu koto de e: soo iu tachiba**  
Japanese GEN INJ native speaker QT do TE GEN utterance also want QT say NML COP INJ that say position

*kara sukoshi e: hatsugen s ase te itadaki masu.*  
from little INJ utterance do CAUS TE ASP POL

‘When I was first given the opportunity, I declined in such way. Since they also want talks from a Japanese native speaker’s perspective, I will talk a little from such position.’

When the optional sequence *yoo+na* [like + COP:ATT] is included, the degree of abstractness is increased by the added sense of equivocality and vagueness. In Interview, the conceptualization construction with *yoo+na* is used by the interviewer to leave interpretive options for the addressee. In the next excerpt, the interviewer M asks the interviewee T if there was anything in particular she hoped for her children. In doing so, the interviewer uses the conceptualization construction with *yoo+na*. It gives the addressee T an option to talk about other things besides what she especially hoped for her children.

(5-29) *tte + iu + yoo + na* [QT + say + like + COP:ATT] (Interview: Hypermedia\_026)

1. M: {何か特にこれだけは} っていうようなことがおありでしたか？
2. T: やはり一つはあののびのびと育てて欲しいということと
3. M: ああふんふん

1. M: {*nanika toku ni kore dake wa*} *tte iu yoo na koto ga o ari deshi ta ka?*  
something special COP this only TOP QT say like COP NML NOM SFX exist COP PST Q
2. T: *yahari hitotsu wa ano nobinobi to sodat te hoshii to iu koto to*  
after.all one TOP INJ carefree QT grow TE want QT say NML QT
3. M: *aa fun fun*  
INJ yes yes

1. M: ‘Was there anything in particular you hope for your children?’
2. T: ‘One thing is that I want my children to grow carefree, and’
3. M: ‘Uh, hmm hmm.’

The conceptualization construction conceptualizes the information in the open slot into general or socio-culturally loaded abstract notion. The construction allows the speaker to avoid giving specific details about the information in the open slot for whatever the reason in the particular context (e.g., avoid labeling the addressee, talk about the information as a socio-

culturally loaded abstract notion, extract the core message to be used as a reason, account, background, etc. for what to follow). In this way, the use of the conceptualization construction contributes to the smooth social interaction.

## **5.5. Motivations for the use of the constructions**

In the previous three sections, we have observed that the three semi-fixed constructions have the following major functions.

### Thought-reporting construction

- Distances the speaker from the quoted inner thought or speech
- Mitigates the strength of a claim, opinion, wish, intention, etc., and conveys a sense of politeness
- Depicts the speaker as modest and socially mature

### Utterance-expansion construction

- Objectifies the information and indicates the subordinate status of the information
- Presents background or supplemental information and simultaneously links it to a main part of the utterance (which may be expressed in the prior context, to be expressed subsequently, or just implied)
- Contributes to the smooth flow of extended spoken discourse

### Conceptualization construction

- Distances the speaker from the quoted information and objectifies it
- Conceptualizes the information into general or socio-culturally loaded abstract notion
- Allows the speaker to avoid giving details about the information for various reasons
- Contributes to the smooth social interaction

From the summaries of the functions above, we can see that the thought-reporting construction is associated with a socially-oriented function, whereas the utterance-expansion construction is associated with a discourse-oriented function. The function of utterance-expansion construction is both socially-oriented and discourse-oriented. Despite the differences in the core aspects of the functions, the three constructions share underlying motivations which account for the frequent use of these constructions in spoken discourse.

First, all three constructions, but especially the thought-reporting and conceptualization constructions, allow for various forms to appear in the open slots. The quotative particle *to/tte* can follow forms representing self-directed speech as well as other abstract ideas represented by clauses, phrases, or simple verbs, nouns, adjectives, and so on. The syntactic flexibility is useful in spoken language where the speaker must produce and process language on-line. The next two examples are a case in point. In (5-30), the self-directed speech, ending in the form *tai* [want] appears in the open slot, while in (5-31), only a topic phrase, *nanika toku ni kore dake wa* [something special COP:ADV this only TOP] ‘anything, just this in particular,’ appears in the open slot.

(5-30) Thought-reporting construction (Interview: Hypermedia\_009)

{日本語 教育 の 方 に 進みたい} と 思っ て る ん です けど

{*nihongo kyooiku no hoo ni susumi tai*} *to omot te ru n desu kedo*

Japanese education GEN direction DAT advance want **QT think** TE ASP NML COP but

‘I’m thinking that I want to go into the field of Japanese language education’

(5-31) Conceptualization construction (Interview: Hypermedia\_026)

{何か特にこれだけは} っ て い う よ う な こ と が お あ り で し た か ？

*nanika toku ni {kore dake wa} tte iu yoo na koto ga o ari deshi ta ka?*  
something special COP this only TOP QT say like COP NML NOM SFX exist COP PST Q

‘Was there anything in particular you hope for your children?’

The three constructions also work as a fluency device (Nattinger & DeCarrico 1992) which gives the speaker more time to plan for what to say next. This tendency is stronger for the utterance-expansion and conceptualization constructions which include longer sequences of morphemes than the thought-reporting constructions. As we can see in the following examples, the constructions tend to be preceded by and/or followed by so-called fillers. This tendency shows that the constructions are treated as single chunks, and that the constructions are used at discourse boundaries where the speaker plans for the next utterance. In this way, the constructions, along with other functions described above, work to generate more time for the speaker to think what to say next. The constructions also give more time for the addressee to interpret what has been expressed by the speaker and prepare for what comes next. Below are examples from Speech, (5-32), and Interview, (5-33). In both excerpts, the fillers are underlined.

(5-32) Utterance-expansion construction (Speech: CSJ\_A03M0005)

で あのー {今日のお話の目的な} んですけれども あのー ま 概念体系って  
いうものが...

*de ano: {kyoo no o hanashi no mokuteki na} n desu keredomo ano: ma gainen taikei tte*  
and INJ today GEN PRF talk GEN purpose COP NML COP but INJ INJ concept system QT  
*iu mono ga ...*  
say NML NOM

‘And, um, as for the purpose of today’s talk, um, uh, what is called ontology ...’

(5-33) Conceptualization construction (Interview: Hypermedia\_034)

{海外旅行はしばらくはないだろう} ということであの 香港に行って来たり ...

{*kaigai ryokoo wa shibaraku nai daro o*} *to iu koto de ano honkon ni it te ki tari ...*  
overseas travel TOP for.a.while. NEG COP AUX QT say NML COP INJ (name) DAT go TE ASP such.as

‘We decided that we probably wouldn’t travel abroad for a while, and went to a place such as Hong Kong ...’

Lastly, the frequent use and the functions of the three constructions are closely interrelated with the core socio-cultural values and characteristics of communication styles within the Japanese socio-linguistic community. As mentioned above, the successful use of the semi-fixed constructions in speech portrays the speaker as socio-pragmatically mature and competent. Although patterns of speaking and interaction in Japanese cannot be explained by cultural values and concepts alone (see Saft 2011), if we are to gain a more realistic and comprehensive account of the use of formulaic language by (adult) native speakers in real social contexts, it is important to touch upon the socio-cultural basis for, as well as import of, the frequent use of the constructions.

First, the socio-cultural virtue of entering into and understanding another’s feelings, thoughts, or intentions is linked to the practice of reporting one’s inner feelings and thoughts through the use of the thought-reporting construction as well as the practice of conceptualizing one’s direct speech through the use of the conceptualization construction. This socio-cultural value is captured in the words/expressions such as *omoiyari* ‘empathy’ and *kimochi/kokoro o kumu* ‘feel for; sympathize with.’ The extreme case of such social practice is *ishindenshin* ‘telepathic communication; reading another’s mind without saying anything.’ However, within the context of the spoken discourse we examined, speakers often display their feelings, thoughts, and direct speech for others to empathize and sympathize with. Even in formal interview or speech where one is expected to use polite and formal language, the speakers still display their private thoughts, feelings, and direct speech, through the uses of the thought-reporting

construction (see excerpt (5-30)) and the conceptualization constructions (see excerpt (5-31)). The extreme frequency of the quotatives in Japanese spoken discourse is therefore linked to the preferred way of speaking in Japanese, which has socio-cultural foundations.

The second characteristic of the Japanese socio-linguistic community is the tendency to grasp and express an event as a whole rather than from the point of view of a concrete subject/object (Ikegami 1981). This tendency is reflected in the use and function of the conceptualization construction to abstract general, socio-culturally loaded ideas from a concrete event. By presenting information as a whole, a socio-culturally loaded abstract idea, the conceptualization construction makes details irrelevant and depicts the event as if it is naturally realized. The tendency to grasp an event as a whole is also reflected in the use of the utterance-expansion construction, which objectifies the whole event as given background information. Focusing on the abstractness and wholeness of an event not only contributes to the smooth flow of discourse and smooth social interaction, but it also contributes to maintaining a harmonious social relationship (or *wa* 'harmony') among discourse participants by avoiding unnecessary conflicts and arguments about details of the information.

The function of the utterance-expansion construction is also interrelated with the tendency of Japanese discourse to leave the most important point towards the end.<sup>50</sup> In some cases, a main thrust of an utterance is not expressed at all (i.e., it is left unsaid), which is also related to the socio-cultural concept of *omoiyari* 'empathy.' The socio-interactive activity of leaving interactional options to the addressee, through the use of the utterance-expansion construction, illuminates the interrelationship between socio-cultural values/practices and the use of formulaic language. More generally, the function of the utterance-expansion construction to

---

<sup>50</sup> This is in part due to the predicate-final characteristics of the Japanese language.

indicate how the units of discourse are related contributes to the successful and smooth progression of extended discourse.

Rather than adding to the conceptual meaning of utterances, the three constructions play important interactional and social roles which shape and are shaped by some of the core socio-cultural values and practices in the Japanese linguistic community. The multiple functions served by the constructions, some of which are register-dependent, explain their frequent use in Japanese spoken discourse across different registers.

## Chapter 6

### Conclusions

This dissertation explored the patterns and nature of formulaicity in Japanese spoken discourse. We began with the form-based definition of formulaic language and adopted a frequency-driven approach to formulaicity. The identification and analysis of morphemic bundles revealed some characteristics of fixedness in Japanese spoken language. Structurally, most bundles are incomplete and they appear at clause or phrase final positions in Japanese. This contrasts with both English lexical bundles, which tend to appear clause/phrase initially (Biber et al. 2004), and Korean lexical bundles, which tend to appear sentence/utterance finally (Y-J. Kim 2009). We have also seen functional divergences between English lexical bundles and Japanese morphemic bundles (see Section 4.5), which are closely linked to different discourse and interactional concerns the speakers of the two languages have.

In terms of register variations, overall, the largest number of different bundles are found in Interview, followed by Speech, and a fewest bundles are found in Conversation. In Conversation, copula-based bundles are most common, while in Speech, NP-based bundles are most common. In Interview, both copula-based and NP-based bundles are common. Functionally, in interactive registers, Conversation and Interview, socio-interactional bundles are commonly used. In non-interactive, information-oriented register, Speech, statement bundles are commonly found. Referential bundles are commonly found in Speech followed by Interview. Within discourse organizers, topic introduction bundles are exclusively used in Speech, whereas background bundles are commonly used in Conversation and Interview. The structural and

functional characteristics are found to be closely associated with the particular needs of the different registers.

Although morphemic bundles represent formal fixed units (based on frequency criteria), the findings in this study indicate that they do not necessarily represent form-function units. We have seen both shorter and longer patterns of formulaicity involving the morphemic bundles. In particular, three types of partially-filled constructions are found to encompass many of the morphemic bundles. These semi-fixed constructions serve as discourse frames for fluent and idiomatically natural production of utterances in spoken discourse. Given the pervasiveness of these constructions in all three registers and the cognitive and socio-interactional foundations for their use as discussed in Chapter 5, the semi-fixed constructions must be recognized as an important type of formulaicity in Japanese spoken discourse. Rather than being linked to concrete meanings/functions, the constructions exhibit flexibility in adjusting their functions and pragmatic meanings to fit the particular needs of the contexts in which they are used (e.g., the utterance-expansion constructions are used to introduce a topic in Speech, whereas they are used to present supplemental/background information in Conversation and Interview). The constructions also exhibit formal flexibility as they accommodate various types of structural units in the open slots. The versatility of the constructions partially explains their ubiquitous occurrences in a wide range of context.

The discussions so far can be summarized by considering the three semi-fixed constructions with respect to four different dimensions of formulaic language. As described in Chapter 1, four dimensions are: formal, functional, cognitive, and native speaker preference.

Form: semi-fixed constructions with open slots, variations in some forms, and optional elements; not totally compositional but not completely fixed

- Function: rooted in discourse and socio-interactional concerns; specific functions are associated with register needs and determined by context
- Cognitive: stored as an prepackaged discourse frame, and accessed with an economy of effort
- Preference: nativelike, idiomatic choice by native speakers, in preference to other ways of communicating/speaking

Even though the constructions are perceptually not as salient as other types formulaic language such as idioms and greeting expressions, they are significant building blocks of native speakers' natural speech.

The identification and analysis of morphemic bundles and semi-fixed constructions in this dissertation also confirm the pervasiveness of nominalization (e.g., utterance-expansion construction and conceptualization construction) as well as quotatives (e.g., thought-quoting construction) in Japanese discourse, both of which have been important topics in Japanese linguistic literature. I hope that the present study contributes to a further understanding of these notions by providing statistical evidence based on large empirical data sets, register analysis, and degrees of fixedness associated with these phenomena.

Final discussion concerns applications and implications of the findings for Japanese as a foreign language education, focusing on the semi-fixed constructions. The discussion addresses the following three points: (1) authenticity of language taught, (2) discourse units to be focused on, and (3) top-down teaching vs. discovery-based learning.

The first point concerns the authenticity of language taught in Japanese textbooks and classrooms. Although the three semi-fixed constructions are very commonly used in spoken discourse, they are rarely introduced in Japanese textbooks. For example, instead of the utterance-expansion construction, the shorter sequence, *n + desu* [NML + COP], which was

found to be extremely rare in the present corpora, is introduced in commonly used Japanese textbooks (e.g., Chapter 12 in Genki I, Banno, Ikeda, Ohno & Shinagawa 2011; Chapter 4 in Yookoso, Tohsaku 2006; Chapter 8 in Nakama 1, Hatasa, Hatasa & Makino 2010). The sequence, *n + desu + ga*, is merely introduced as part of an idiomatic expression, *kore tsumaranai mono na n desu ga* ‘this is a trifling thing’ used when giving a gift to someone (e.g., Aoki & Okamoto 1988). A large corpus, like the one used in the present study, can provide samples of authentic language use, which may not be intuitively salient to native speakers.

The second point is related to the discourse units that are often used in language teaching. Even in intermediate and advanced classes, student-student or student-teacher exchanges tend to be limited to a few short turns. In the ACTFL (American Council on the Teaching of Foreign Languages) proficiency guidelines for speaking,<sup>51</sup> learners at the advanced level must be able to give narrations and descriptions through connected sentences/utterances in paragraph-level discourse. In order to develop competence and fluency in producing paragraph-level utterances, a mastery of the type of formulaic language represented by the three semi-fixed constructions seems essential. Since these constructions have strong socio-cultural foundations and their use is linked to the preferred way of speaking and communicating in Japanese socio-linguistic community, it is also crucial to take a dynamic approach in which language use is viewed as a socio-interactive activity.

The third and final point speaks directly to the effective and practical use of corpora in language learning and teaching. Gavioli and Aston (2001) state that corpora can serve as resources not only for teachers and material designers to decide what and how to teach, but also

---

<sup>51</sup> In the new 2012 ACTFL proficiency guidelines, there are five major proficiency levels for speaking: Novice, Intermediate, Advanced, Superior, and Distinguished. In the Japanese oral proficiency interview (OPI), however, only the first four levels are currently recognized. The three of the major levels (novice, intermediate, and advanced) are further divided into Low, Mid and High sub-levels.

for learners to make discoveries themselves. With the use of corpus tools, learners can discover the meanings and patterns of language use through exposure to multiple examples in real contexts, instead of being handed prescriptive rules by the teachers. The discovery-based learning with corpus tools is especially important at the intermediate and advanced levels when learning the target language increasingly becomes more about learning socio-pragmatic appropriateness, register-specific speech styles, and idiomatic ways of expressing something. As mentioned earlier, the three semi-fixed constructions are learned by native language learners at a rather later stage of language acquisition when they become increasingly socialized in a larger social sphere outside of individual families. In a foreign language learning, access to socially advanced authentic language use is limited. Given the availability of large, well-represented and balanced corpora, the corpus-based language materials can in principle replace the natural language experience the native language learners have. Put differently, the corpus-based materials can provide condensed authentic language exposure for the foreign language learners, and with the guidance of the teacher, they can help the learners develop nativelike intuitions about preferred idiomatic ways of speaking (see Gabrielatos 2005).

From a methodological point of view, I hope that this study has demonstrated the feasibility of empirically investigating formulaicity in Japanese language as well as the applicability of corpus-based methodologies to agglutinating languages, with the morpheme as the basic unit of analysis. The next feasible step is to conduct a more extensive investigation of semi-fixed constructions, by using carefully tagged corpora and regular expressions to search for various patterns of speech in Japanese.

## APPENDIX

### Functional classification of morphemic bundles across three registers

- \* 80-129 per million morphemes
- \*\* 130-179 per million morphemes
- \*\*\* 180-229 per million morphemes
- \*\*\*\* 230-279 per million morphemes
- \*\*\*\*\* over 280 per million morphemes

[Note: the bundles that are discussed in the main text of the dissertation are indicated in bold.]

		Conversation	Interview	Speech
<b>(1) STANCE EXPRESSIONS</b>				
<b>(A) Epistemic stance</b>				
<b><i>ta no ka na</i></b> PST NML Q PP	<b>&lt;uncertainty&gt;</b>	*		
<b><i>(doo) na n daro o</i></b> how COP:ATT NML COP AUX	<b>&lt;uncertainty&gt;</b>	*		
<i>na n daro o</i> COP:ATT NML COP AUX	<uncertainty>	*		
<i>ja nai ka na</i> COP NEG Q PP	<uncertainty>	*	**	
<b><i>(dochira) ka to iu to</i></b> which Q QT say COND	<b>&lt;relativeness&gt;</b>		*	
<b><i>kamo shire mase n</i></b> Q know POL NEG	<b>&lt;uncertainty&gt;</b>		*	*
<i>no kamo shire nai</i> NML Q know NEG	<uncertainty>		*	
<i>no ka na to</i> NML Q PP QT	<uncertainty>		*	
<i>nai ka na to</i> NEG Q PP QT	<uncertainty>		*	
<i>ja nai ka to</i> COP NEG Q QT	<uncertainty>		***	*
<b><i>n ja nai ka</i></b> NML COP NEG Q	<b>&lt;uncertainty&gt;</b>	***	***** (398)	**

		Conversation	Interview	Speech
<i>de wa nai ka</i> COP TOP NEG Q	<uncertainty>			***** (637)
<i>wa nai ka to</i> TOP NEG Q QT	<uncertainty>			***** (519)
<i>no de wa nai</i> NML COP TOP NEG	<uncertainty>			***** (488)
<i>wa nai desho o</i> TOPNEG COP AUX	<uncertainty>			*
<i>de wa nai desho</i> COP TOP NEG COP	<uncertainty>			*
<i>te iru no ka</i> TE ASP NML Q	<uncertainty>			*
<i>de aro o to</i> COP exist AUX QT	<uncertainty>			*
<hr/>				
(B) Attitudinal stance				
<i>oo ka na to</i> VOL Q PP QT	<intention>	**		
<i>te iki tai to</i> TE go want QT	<desire>			**
<i>ta hoo ga ii</i> PST way NOM good	<recommendation>	***		
<i>nakere ba nara nai</i> NEG COND must NEG	<obligation>			*
<i>te mi mashi ta</i> TE look POL PST	<attempt>			****
<hr/>				
(2) DISCOURSE ORGANIZERS				
(A) Topic introduction/elaboration				
<i>hon kenkyuu de wa</i> this research LOC TOP	‘in this research’			**
<i>na n desu ga</i> COP:ATT NML COP but	‘as for’			***** (553)
<i>na n desu keredomo</i> COP:ATT NML COP but	‘as for’			***** (386)

	Conversation	Interview	Speech
<i>to shi mashi te</i> QT do POL TE			****
<i>to itashi mashi te</i> QT do:POL POL TE			***
<i>ni tsui te wa</i> DAT adhere TE TOP			*
<i>sore ni taishi te</i> that DAT oppose TE			**
<i>te mi masu to</i> TE ASP POL COND			**
<i>na n desu kedomo</i> COP:ATT NML COP but			**
<i>te ori masu ga</i> TE ASP POL but			**
<i>shi mashi te wa</i> do POL TE TOP			**
<i>ni tsuki mashi te</i> DAT adhere POL TE			**
<i>to ii masu to</i> QT say POL COND			*
<i>te i masu ga</i> TE ASP POL but			*
<i>tsuki mashi te wa</i> adhere POL TE TOP			*
<i>ni taishi te wa</i> DAT oppose TE TOP			*
<i>ka to iu to</i> Q QT say COND			*
<i>ka to ii masu (to)</i> Q QT say POL COND			*
<i>de ari masu keredomo</i> COP exist POL but			*

		Conversation	Interview	Speech
<i>itashi mashi te wa</i> do:POL POL TE TOP	‘regarding’			*
<i>ni tsui te e:</i> DAT adhere TE INJ	‘concerning’			*
<hr/>				
(B) Cohesion				
<i>te ru n da</i> TE ASP NML COP	<cohesion>	***** (283)		
<i>te ru n desu</i> TE ASP NML COP	<cohesion>	*	***** (629)	*
<i>te ta n da/desu</i> TE PST NML COP	<cohesion>	***	***** (323)	
<i>dat ta n da/desu</i> COP PST NML COP	<cohesion>	**	***** (533)	
<i>te ki ta no</i> TE ASP PST PP	<cohesion>	*		
<i>nakat ta n da/desu</i> NEG PST NML COP	<cohesion>	*	****	
<i>shi te n no</i> do ASP NML PP	<cohesion>	*		
<i>ja nai n da/desu</i> COP NEG NML COP	<cohesion>	*	**	
<i>ki ta n da</i> ASP PST NML COP	<cohesion>	*		
<i>shi te ta no</i> do TE PST PP	<cohesion>	*		
<i>chat ta n da</i> ASP PST NML COP	<cohesion>	*		
<i>te nai n da/desu</i> TE NEG NML COP	<cohesion>	*	*	
<i>te ki ta n</i> TE ASP PST NML	<cohesion>	*	*	
<i>omot ta n da/desu</i> think PST NML COP	<cohesion>	*	*	

		Conversation	Interview	Speech
<i>tte iu n da/desu</i> QT say NML COP	<cohesion>	*	**	
<i>shi te ru n</i> do TE ASP NML	<cohesion>	*		
<i>ga aru n desu</i> NOM exist NML COP	<cohesion>		***** (297)	*
<i>at ta n desu</i> exist PST NML COP	<cohesion>		****	
<i>ni nat ta n</i> COP become PST NML	<cohesion>		***	
<i>nat ta n desu</i> become PST NML COP	<cohesion>		***	
<i>shi ta n desu</i> do PST NML COP	<cohesion>		***	
<i>i ta n desu</i> ASP PST NML COP	<cohesion>		**	
<i>wa nai n desu</i> TOP NEG NML COP	<cohesion>		**	
<i>te iu n desu</i> QT say NML COP	<cohesion>		**	
<i>ga at ta n desu</i> NOM exist NML COP	<cohesion>		**	
<i>koto na n desu</i> NML COP NML COP	<cohesion>		**	*
<i>te iu n desu</i> TE ASP NML COP	<cohesion>		**	*
<i>te i ta n</i> TE ASP PST NML	<cohesion>		**	
<i>shi te ru n</i> do TE ASP NML	<cohesion>		*	
<i>yat te ru n</i> do TE ASP NML	<cohesion>		*	

		Conversation	Interview	Speech
<i>dat ta n de</i> COP PST NML COP	<cohesion>		*	
<i>te kuru n desu</i> TE ASP NML COP	<cohesion>		*	
<i>mono na n desu</i> NML COP NML COP	<cohesion>		*	
<i>kanji na n desu</i> feeling COP NML COP	<cohesion>		*	
<i>de wa nai n</i> COP TOP NEG NML	<cohesion>		*	
<i>ni naru n desu</i> COP become NML COP	<cohesion>		*	
<i>shi te ta n</i> do TE ASP NML	<cohesion>		*	
<i>mo aru n desu</i> also exist NML COP	<cohesion>		*	
<i>te iru wake desu</i> TE ASP reason COP	<cohesion>			*
<i>te ru wake desu</i> TE ASP reason COP	<cohesion>			*
<hr/>				
(C) Background				
<i>ta n da kedo</i> PST NML COP but	<background>	***** (741)	***	
<i>ta n desu kedo</i> PST NML COP but	<background>	*	***** (830)	
<i>nai n da/desu kedo</i> NEG NML COP but	<background>	***** (280)	****	
<i>omou n da/desu kedo</i> think NML COP but	<background>	**	***** (336)	
<i>na n da/desu kedo</i> COP:ATT NML COP but	<background>	****	***** (437)	
<i>ta n desu keredomo</i> PST NML COP but	<background>		***** (424)	**

		Conversation	Interview	Speech
<i>ta n desu ga</i> PST NML COP but	<background>			***
<i>n da kedo ne</i> NML COP but PP	<background>	****		
<i>n da kedo sa</i> NML COP but PP	<background>	***		
<i>ru n da/desu kedo</i> ASP NML COP but	<background>	**	**	
<i>ii n da kedo</i> good NML COP but	<background>	**		
<i>n desu kedo ne</i> NML COP but PP	<background>	*	**	
<i>aru n da/desu kedo</i> exist NML COP but	<background>	*	**	
<i>na n desu keredomo</i> COP NML COP but	<background>		****	
<i>omou n desu keredomo</i> think NML COP but	<background>		***	
<i>ta n desu kedomo</i> PST NML COP but	<background>		**	
<i>ta n desu ga</i> PST NML COP but	<background>		**	
<i>nai n desu keredomo</i> NEG NML COP but	<background>		**	
<i>na n desu ga</i> COP NML COP but	<background>		**	
<i>nai n desu ga</i> NEG NML COP but	<background>		*	
<i>omou n desu ga</i> think NML COP but	<background>		*	*
<i>na n desu kedomo</i> COP NML COP but	<background>		*	

		Conversation	Interview	Speech
<i>ru n desu kedomo</i> ASP NML COP but	<background>		*	
<i>aru n desu keredomo</i> exist NML COP but	<background>		*	
<b>(3) REFERENTIAL EXPRESSIONS</b>				
<b>(A) Conceptualization</b>				
<i>tte iu no wa</i> QT say NML TOP	<conceptualization>	***** (337)	***** (1407)	***** (494)
<i>to iu no wa</i> QT say NML TOP	<conceptualization>		***** (420)	***** (987)
<i>to iu koto de</i> QT say NML COP	<conceptualization>		****	***** (651)
<i>to iu koto ni</i> QT say NML COP:ADV	<conceptualization>		*	***** (336)
<i>to iu koto desu</i> QT say NML COP	<conceptualization>		**	****
<i>tte iu koto desu</i> QT say NML COP	<conceptualization>		*	
<i>tte iu no ga</i> QT say NML NOM	<conceptualization>	***	***** (603)	***
<i>tte iu no o</i> QT say NML ACC	<conceptualization>		***	*
<i>ka tte iu no</i> Q QT say NML	<conceptualization>		***	*
<i>tte iu no mo</i> QT say NML also	<conceptualization>		**	
<i>tte iu koto de</i> QT say NML COP	<conceptualization>		**	*
<i>to iu no ga</i> QT say NML NOM	<conceptualization>		**	***** (561)
<i>iu no wa doo</i> say NML TOP how	<conceptualization>		*	

		Conversation	Interview	Speech
<i>te ru tte iu</i> TE ASP QT say	<conceptualization>		*	
<i>no ka tte iu</i> NML Q QT say	<conceptualization>		*	*
<i>ta tte iu no</i> PST QT say NML	<conceptualization>		*	
<i>tte iu koto wa</i> QT say NML TOP	<conceptualization>		*	
<i>to iu koto wa</i> QT say NML TOP	<conceptualization>		*	*
<i>toka tte iu no</i> such QT say NML	<conceptualization>		*	
<i>nai ka tte iu</i> NEG Q QT say	<conceptualization>		*	
<i>no wa doo iu</i> NML TOP how say	<conceptualization>		*	
<i>ta to iu koto</i> PST QT say NML	<conceptualization>		*	**
<i>iu no ga at</i> say NML NOM exist	<conceptualization>		*	
<i>to iu koto ga</i> QT say NML NOM	<conceptualization>		*	***** (640)
<i>ka na tte iu</i> Q PP QT say	<conceptualization>		*	
<i>to iu koto o</i> QT say NML ACC	<conceptualization>		*	***** (437)
<i>ka to iu koto</i> Q QT say NML	<conceptualization>			***** (457)
<i>te iru to iu</i> TE ASP QT say	<conceptualization>			***** (384)
<i>to iu no o</i> QT say NML ACC	<conceptualization>			***** (319)

	Conversation	Interview	Speech
<i>to iu mono o</i> QT say NML ACC	<conceptualization>		***** (316)
<i>nai ka to iu</i> NEG Q QT say	<conceptualization>		***** (282)
<i>to iu mono ga</i> QT say NML NOM	<conceptualization>		****
<i>koto ni yot te</i> NML COP depend TE	<conceptualization>		****
<i>de aru to iu</i> COP exist QT say	<conceptualization>		****
<i>shi te iru to</i> do TE ASP QT	<conceptualization>		***
<i>suru to iu koto</i> do QT say NML	<conceptualization>		***
<i>no ka to iu</i> NML Q QT say	<conceptualization>		***
<i>iru to iu koto</i> exist QT say NML	<conceptualization>		***
<i>iu koto ni nari</i> say NML COP become	<conceptualization>		**
<i>nai to iu koto</i> NEG QT say NML	<conceptualization>		**
<i>aru to iu koto</i> exist QT say NML	<conceptualization>		**
<i>ka to iu no</i> Q QT say NML	<conceptualization>		**
<i>te iru koto ga</i> TE ASP NML NOM	<conceptualization>		*
<i>iu no wa ma:</i> say NML TOP INJ	<conceptualization>		*
<i>ka tte iu koto</i> Q QT say NML	<conceptualization>		*

	Conversation	Interview	Speech
<i>iu koto ga wakari</i> say NML NOM understand	<conceptualization>		*
<i>te iku to iu</i> TE ASP QT say	<conceptualization>		*
<i>tte iu koto mo</i> QT say NML also	<conceptualization>		*
<i>to iu koto mo</i> QT say NML also	<conceptualization>		*
<i>iru koto ga wakari</i> exist NML NOM understand	<conceptualization>		*
<i>ga aru to iu</i> NOM exist QT say	<conceptualization>		*
<i>to iu mono wa</i> QT say NML TOP	<conceptualization>		*
<i>oo to iu koto</i> VOL QT say NML	<conceptualization>		*
<i>te ru to iu</i> TE ASP QT say	<conceptualization>		*
<i>suru koto ga deki</i> do NML NOM can	<conceptualization>		*
<i>suru koto ni yot</i> do NML COP depend	<conceptualization>		*
<i>tte iu koto ga</i> QT say NML NOM	<conceptualization>		*
<i>iu koto na n</i> say NML COP NML	<conceptualization>		*
<i>ni oi te wa</i> DAT put TE TOP	<conceptualization>		*
<i>te i ta to</i> TE ASP PST QT	<conceptualization>		*
<hr/>			
(B) Equivocation			
<i>soo iu fuu ni</i> that say way COP:ADV	<equivocation>	**	****

		Conversation	Interview	Speech
<i>soo iu no wa</i> that way NML TOP	<equivocation>	*	**	
<i>tte iu fuu ni</i> QT say way COP:ADV	<equivocation>		***** (415)	*
<i>to iu fuu ni</i> QT say way COP:ADV	<equivocation>		****	***** (792)
<i>tte iu yoo na</i> QT say like COP:ATT	<equivocation>		***** (288)	
<i>tte/to iu yoo na</i> QT say like COP:ATT	<equivocation>		**	***** (547)
<i>doo iu fuu ni</i> how say way COP:ADV	<equivocation>		****	*
<i>iu yoo na koto</i> say like COP:ATT NML	<equivocation>		**	***
<i>koo iu fuu ni</i> this say way COP:ADV	<equivocation>		**	*
<i>soo iu koto o</i> that say NML ACC	<equivocation>		*	
<i>soo iu no ga</i> that say NML NOM	<equivocation>		*	
<i>te ru yoo na</i> TE ASP like COP:ATT	<equivocation>		*	
<i>iu fuu ni omot</i> say way COP:ADV think	<equivocation>		*	
<i>soo iu imi de</i> that say meaning COP	<equivocation>		*	
<i>toka soo iu no</i> such that say NML	<equivocation>		*	
<i>soo iu koto wa</i> that say NML TOP	<equivocation>		*	
<i>wa doo iu fuu</i> TOP how say way	<equivocation>		*	

		Conversation	Interview	Speech
<i>iu fuu ni kangae</i> say way COP:ADV think	<equivocation>			***
<i>mashi ta yoo ni</i> POL PST like COP:ADV	<equivocation>			**
<i>aru to iu fuu</i> exist QT say way	<equivocation>			*
<i>it ta yoo na</i> ASP PST like COP:ATT	<equivocation>			*
<i>e kono yoo na</i> INJ this like COP:ATT	<equivocation>			*
<i>to iu yoo ni</i> QT say like COP:ATT	<equivocation>			*
<i>yoo na koto ga</i> like COP:ATT NML NOM	<equivocation>			*
<i>to iu katachi de</i> QT say NML COP	<equivocation>			*
<i>yoo na mono ga</i> like COP:ATT NML NOM	<equivocation>			*
<i>ka to iu fuu</i> Q QT say way	<equivocation>			*
<i>to it ta yoo</i> QT say PST like	<equivocation>			*
<i>yoo na koto o</i> like COP:ATT NML ACC	<equivocation>			*
<i>yoo na katachi de</i> like COP:ATT NML COP	<equivocation>			*

#### (4) SOCIO-INTERACTIONAL EXPRESSIONS

##### (A) Reporting

<i>na to omot te</i> PP QT think TE	<stance reporting>	***** (394)	*
<i>oo to omot te</i> VOL QT think TE	<stance reporting>	***	*

		Conversation	Interview	Speech
<i>tai to omoi masu</i> want QT think POL	<stance reporting>			***** (601)
<i>ka to omoi masu</i> Q QT think POL	<stance reporting>			***** (282)
<i>ka na to omot</i> Q PP QT think	<stance reporting>	***** (299)		
<i>na to omot ta</i> PP QT think PST	<stance reporting>	*		
<i>tai to omot te</i> want QT think TE	<stance reporting>		*	
<i>oo to omot ta</i> VOL QT think PST	<stance reporting>		*	
<i>ka to omou n</i> Q QT think NML	<stance reporting>		*	
<i>to omoi masu ne</i> QT think POL PP	<stance reporting>		*	
<i>to omoi masu yo</i> QT think POL PP	<stance reporting>		*	
<i>to omoi masu kedo</i> QT think POL but	<stance reporting>		*	
<i>to omot te i</i> QT think TE ASP	<stance reporting>		*	
<i>nai to omoi masu</i> NEG QT think POL	<stance reporting>		*	
<i>to omot te masu</i> QT think TE POL	<stance reporting>		*	
<i>nai to omou n</i> NEG QT think NML	<stance reporting>		*	
<i>shi tai to omoi</i> do want QT think	<stance reporting>			**
<i>iki tai to omoi</i> ASP want QT think	<stance reporting>			*

		Conversation	Interview	Speech
<i>itadaki tai to omoi</i> ASP want QT think	<stance reporting>			*
<i>nai ka to kangae</i> NEG Q QT think	<stance reporting>			*
<i>to omou n da/desu</i> QT think NML COP	<impersonal stance reporting>	**	***** (887)	***
<i>da to omoi masu</i> COP QT think POL	<impersonal stance reporting>		****	**
<i>to kangae rare masu</i> QT think POT POL	<impersonal stance reporting>			**
<i>da to omou n</i> COP QT think NML	<impersonal stance reporting>		***	
<i>to omow are masu</i> QT think PASS POL	<impersonal stance reporting>			**
<i>aru to omoi masu</i> exist QT think POL	<impersonal stance reporting>			*
<i>ta to omou n</i> PST QT think NML	<non-stance reporting>		**	
<i>aru to omou n</i> exist QT think NML	<non-stance reporting>		*	
<i>ta to omoi masu</i> PST QT think POL	<non-stance reporting>		*	
<i>to omot ta n</i> QT think PST NML	<thought reporting>	*	*	
<i>to omot te ru</i> QT think TE ASP	<thought reporting>	*		
<i>to omot te sa</i> QT think TE PP	<thought reporting>	*		

		Conversation	Interview	Speech
<i>tte it te ta</i> QT say TE PST	<utterance reporting>	*****		
<i>tte iw are te</i> QT say PASS TE	<utterance reporting>	**		
<i>toka it te ta</i> such say ASP PST	<utterance reporting>	*		
<i>ta n da tte</i> PST NML COP QT	<utterance reporting>	*		
<i>toka tte it te</i> such QT say TE	<utterance reporting>	*		
<i>tte iw are ta</i> QT say PASS PST	<utterance reporting>	*		
<i>toka iw are te</i> such say PASS TE	<utterance reporting>	*		
<i>ta n desu tte</i> PST NML COP QT	<utterance reporting>		**	
<i>ja nai ka tte</i> COP NEG Q QT	<utterance reporting>		*	
<hr/>				
(B) Interpersonal				
<i>n da yo ne</i> NML COP PP PP	<interpersonal>	***** (616)		
<i>n desu yo ne</i> NML COP PP PP	<interpersonal>	**	***** (555)	
<i>na n da yo</i> COP:ATT NML COP PP	<interpersonal>	****		
<i>na n desu yo</i> COP:ATT NML COP PP	<interpersonal>	*	***** (376)	
<i>ta n desu yo</i> PST NML COP PP	<interpersonal>	***	***** (411)	
<i>ta n da yo</i> PST NML COP PP	<interpersonal>	****		
<i>n ja nai no</i> NML COP NEG PP	<interpersonal>	***** (353)		

		Conversation	Interview	Speech
<i>ta n desu ne</i> PST NML COP PP	<interpersonal>		***** (446)	
<i>ta n ja nai</i> PST NML COP NEG	<interpersonal>	**	*	
<i>ii n ja nai</i> good NML COP NEG	<interpersonal>	**		
<i>nai n ja nai</i> NEG NML COP NEG	<interpersonal>	**		
<i>nai n da/desu yo</i> NEG NML COP PP	<interpersonal>	**	**	
<i>na n ja nai</i> COP:ATT NML COP NEG	<interpersonal>	*		
<i>it te ta yo</i> say ASP PST PP	<interpersonal>	*		
<i>n daro o ne</i> NML COP AUX PP	<interpersonal>	*		
<i>ta no yo ne</i> PST NML PP PP	<interpersonal>	*		
<i>te n no yo</i> ASP NML NML PP	<interpersonal>	*		
<i>na no yo ne</i> COP:ATT NML PP PP	<interpersonal>	*		
<i>te n ja nai</i> ASP NML COP NEG	<interpersonal>	*		
<i>na n desu ne</i> COP:ATT NML COP PP	<interpersonal>		***** (306)	
<i>ja nai desu ka</i> COP NEG COP Q	<interpersonal>		****	
<i>n desho o ne</i> NML COP AUX PP	<interpersonal>		****	
<i>n desu ka ne</i> NML COP Q PP	<interpersonal>		***	

		Conversation	Interview	Speech
<i>omou n desu ne</i> think NML COP PP	<interpersonal>		**	
<i>omou n desu yo</i> think NML COP PP	<interpersonal>		**	
<i>wake desu yo ne</i> reason COP PP PP	<interpersonal>		**	
<i>ari masu yo ne</i> exist POL PP PP	<interpersonal>		**	
<i>desho o ka ne</i> COP AUX Q PP	<interpersonal>		*	
<i>nai n desu ne</i> NEG NML COP PP	<interpersonal>		*	
<i>iu koto desu ne</i> say NML COP PP	<interpersonal>		*	
<i>ru n desu yo</i> ASP NML COP PP	<interpersonal>		*	
<i>nai desu yo ne</i> NEG COP PP PP	<interpersonal>		*	
<i>aru n desu yo</i> exist NML COP PP	<interpersonal>		*	
<hr/>				
(C) Reactive				
<i>soo soo soo soo</i> so so so so	<reactive>	***** (657)		
<i>un un un un</i> yes yes yes yes	<reactive>	***** (337)	***** (551)	
<i>soo na n da</i> so COP:ATT NML COP	<reactive>	***** (618)	***	
<i>soo na n desu</i> so COP:ATT NML COP	<reactive>		***** (878)	
<i>aa soo desu ka</i> INJ so COP Q	<reactive>		***** (1447)	
<i>soo da/desu yo ne</i> so COP PP PP	<reactive>	***** (563)	***** (328)	

		Conversation	Interview	Speech
<i>a soo na n</i> INJ so COP:ATT NML	<reactive>	***	***	
<i>sok ka sok ka</i> so Q so Q	<reactive>	***		
<i>soo ka soo ka</i> so Q so Q	<reactive>	***		
<i>a soo na no</i> INJ so COP:ATT NML	<reactive>	**		
<i>soo na no yo</i> so COP:ATT NML PP	<reactive>	**		
<i>aa soo na no</i> INJ so COP:ATT NML	<reactive>	**		
<i>un soo soo soo</i> yes so so so	<reactive>	*		
<i>aa soo na n</i> INJ so COP:ATT NML	<reactive>	*	****	
<i>a soo desu ka</i> INJ so COP Q	<reactive>		***** (433)	
<i>ee ee ee ee</i> yes yes yes yes	<reactive>		**	
<i>u:n un un un</i> INJ yes yes yes	<reactive>		*	
<i>soo desho o ne</i> so COP AUX PP	<reactive>		*	
<i>soo desu ne hai</i> so COP PP yes	<reactive>		*	
<i>hai soo desu ne</i> yes so COP PP	<reactive>		*	
<i>a soo desu ne</i> INJ so COP PP	<reactive>		*	

(D) Word/expression search

<i>nan te iu no</i> what QT say NML	<word/expression search>	*****
--	-----------------------------	-------

		Conversation	Interview	Speech
<i>te iu no ka</i> QT say NML Q	<word/expression search>	*		
<i>iu no ka na</i> say NML Q PP	<word/expression search>	*		
<i>nan te iu n</i> what QT say NML	<word/expression search>		***** (306)	
<i>te iu n desho</i> QT say NML COP	<word/expression search>		*	
<i>iu n desho o</i> say NML COP AUX	<word/expression search>		**	
(E) Inquiry				
<i>de irasshai masu ka</i> COP exist:HON POL Q	<status inquiry>		**	
<i>te rasshai masu ka</i> TE ASP:HON POL Q	<inquiry>		*	
<i>ta n desu ka</i> PST NML COP Q	<inquiry>		***** (594)	
<i>na n desu ka</i> COP:ATT NML COP Q	<inquiry>		***** (835)	
<i>n desho o ka</i> NML COP AUX Q	<inquiry>		***** (288)	
<i>iu n desu ka</i> say NML COP Q	<inquiry>		****	
<i>na n desho o</i> COP:ATT NML COP AUX	<inquiry>		****	
<i>nai n desu ka</i> NEG NML COP Q	<inquiry>		**	
<i>te rassharu n desu</i> TE ASP:HON NML COP	<inquiry>		**	

		Conversation	Interview	Speech
<i>aru n desu ka</i> exist NML COP Q	<inquiry>		**	
<i>ru n desu ka</i> ASP NML COP Q	<inquiry>		*	
<i>iu koto desu ka</i> say NML COP Q	<inquiry>		*	
<i>ta n deho o</i> PST NML COP AUX	<inquiry>		*	
<i>rassharu n desu ka</i> ASP:HON NML COP Q	<inquiry>		*	
<i>doo na n desho</i> how COP:ATT NML COP	<inquiry>		*	
<i>de rasshai masu ka</i> COP ASP:HON POL Q	<status inquiry>		*	
<i>yat te n no</i> do ASP NML PP	<inquiry>	*		
<hr/>				
(F) Others				
<i>yoroshiku onegai shi masu</i> favorably please do POL	<b>‘thank you in advance’</b>		***** (288)	
<i>arigatoo gozai mashi ta</i> thank exist:POL POL PST	<b>‘thank you’</b>		***** (284)	
<i>doomo arigatoo gozai mashi (ta)</i> much thank exist:POL POL PST	<b>‘thank you very much’</b>		*	
<i>s ase te itadaki</i> do CAUS TE receive:HUM	<announcement>			**
<i>ase te itadaki masu</i> CAUS TE ASP:HUM POL	<announcement>			**
<i>n de wa nai</i> NML COP TOP NEG	<proposal>			*
<i>nai desho o ka</i> NEG COP AUX Q	<proposal>			**

		Conversation	Interview	Speech
<i>to omoi masu ga</i> QT think POL but	<disclaimer>			*
(5) OBJECTIVE STATEMENTS				
<i>kaet te ki ta</i> return TE ASP PST	<statement>	*		
<i>kaet te ki te</i> return TE ASP TE	<statement>	*		
<i>ni nat te ru</i> COP:ADV become TE ASP	<statement>	*	*	
<i>ni nari mashi ta</i> COP:ADV become POL PST	<statement>		*	*
<i>ni nat te ki</i> COP:ADV become TE ASP	<statement>		*	
<i>yoo ni nat te</i> like COP:ADV become TE	<statement>		*	*
<i>nat te i masu</i> become TE ASP POL	<statement>			***** (400)
<i>ni nat te i</i> COP:ADV become TE ASP	<statement>			****
<i>koto ni nari masu</i> NML COP:ATT become POL	<statement>			***
<i>nat te ori masu</i> become TE ASP POL	<statement>			***
<i>ni nat te ori</i> COP:ADV become TE ASP	<statement>			***
<i>to nat te i</i> QT become TE ASP	<statement>			**
<i>ni nat te iru</i> COP:ADV become TE ASP	<statement>			*
<i>shi te i masu</i> do TE ASP POL	<statement>			***** (392)
<i>shi te ori masu</i> do TE ASP POL	<statement>			****

		Conversation	Interview	Speech
<i>o shi te iru</i> ACC do TE ASP	<statement>			*
<i>shigoto o shi te</i> work ACC do TE	<statement>		*	
<i>o shi te ru</i> ACC do TE ASP	<statement>		*	
<i>o shi te i</i> ACC do TE ASP	<statement>		*	
<i>ga ari mashi te</i> NOM exist POL TE	<statement>		*	*
<i>ga ari mashi ta</i> NOM exist POL PST	<statement>			**
<i>wa ari mase n</i> TOP exist POL NEG	<statement>			**
<b><i>o okonai mashi ta</i></b> ACC conduct POL PST	<statement>			****
<i>kai te at ta</i> write TE ASP PST	<statement>	*		
<i>kai te ari masu</i> write TE ASP POL	<statement>			*
<i>ga de te ki</i> NOM go.out TE ASP	<statement>			*
<i>o arawashi te i</i> ACC express TE ASP	<statement>			*
<i>o mot te iru</i> ACC have TE ASP	<statement>			*
<i>koto ga wakari masu</i> NML NOM understand POL	<statement>			***
<i>ga wakari mashi ta</i> NOM understand POL PST	<statement>			**
<i>koto ga wakari mashi</i> NML NOM understand POL	<statement>			*

		Conversation	Interview	Speech
<i>koto ga deki masu</i> NML NOM can POL	<statement>			**
<i>mase n deshi ta</i> POL NML COP PST	<statement>		**	***
<i>de wa naku te</i> COP TOP NEG TE	<statement>			****
<i>te ori mashi te</i> TE ASP POL TE	<statement>			***
<i>te i mashi ta</i> TE ASP POL PST	<statement>			**
<i>te ki mashi ta</i> TE ASP POL PST	<statement>			*
<i>te ori mashi ta</i> TE ASP POL PST	<statement>			*
<i>te i masse n</i> TE ASP POL NEG	<statement>			*
<i>de wa ari mase</i> COP TOP ASP POL	<statement>			*
<b><i>are te i masu</i></b> PASS TE ASP POL	<statement>			***** (417)
<b><i>s are te i</i></b> do PASS TE ASP	<statement>			***** (296)
<i>are te ori masu</i> PASS TE ASP POL	<statement>			****
<i>s are te iru</i> do PASS TE ASP	<statement>			****
<i>s are te ori</i> do PASS TE ASP	<statement>			**
<i>s are mashi ta</i> do PASS POL PST	<statement>			*
<i>rare te i masu</i> PASS TE ASP POL	<statement>			*

		Conversation	Interview	Speech
<i>mir are mashi ta</i> see PASS POL PST	<statement>			**
<i>ga mir are masu</i> NOM see PASS POL	<statement>			*
<i>ga mir are mashi</i> NOM see PASS POL	<statement>			*

## REFERENCES

- Allen, D. (2009). Lexical bundles in learner writing: an analysis of formulaic language in the ALESS learner corpus. *Komaba Journal of English Education, 1*, 105-127.
- Altenberg, B. (1991). The London-Lund corpus: research and applications. In *Using corpora* (pp. 71-83). Oxford: UW Centre for the New OED.
- Altenberg, B. (1998). On the phraseology of spoken English: the evidence of recurrent word combinations. In A. Cowie (Ed.), *Phraseology: theory, analysis, and applications* (pp. 101-122). Oxford: Oxford University Press.
- American Council on the Teaching of Foreign Languages. (2012) ACTFL Proficiency Guidelines 2012. Retrieved from <http://www.actfl.org>
- Aoki, H., & Okamoto, S. (1988). *Rules for conversational rituals in Japanese*. Tokyo: Taishukan Shoten.
- Austin, J. L. (1962). *How to do things with words*. Cambridge, MA: Harvard University Press.
- Bach, K. (1998). Speech acts. In E. Craig (Ed.), *Routledge encyclopedia of philosophy* (pp. 81-87). London: Routledge.
- Bannard, C. & Lieven, E. (2009). Repetition and reuse in child language learning. In R. Corrigan, E. A. Moravcsik, H. Quali, & K. M. Wheatley (Eds.), *Formulaic language: Volume II: Acquisition, loss, psychological reality, and functional explanations*. (pp. 299-321). Philadelphia: John Benjamins.
- Banno, E., Ikeda, Y., Ohno, Y., Shinagawa, C., & Tokashiki, K. (2011). *Genki: An integrated course in elementary Japanese I*. Tokyo: Japan Times.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber, D. & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes, 26*, 263-286.
- Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgard, & S. Oksefjell (Eds.), *Out of corpora: Studies in honor of Stig Johansson* (pp. 181-189). Amsterdam: Rodopi.
- Biber, D., & Finegan, E. (1994). *Sociolinguistic perspectives on register*. New York/Oxford: Oxford University Press.
- Biber, D., Conrad, S., & Cortes, V. (2003). Lexical bundles in speech and writing: an initial taxonomy. In A. Wilson, P. Rayson, & T. McEnery (Eds.), *Corpus linguistics by the lune: A festschrift for Geoffrey Leech* (pp. 71-92). Frankfurt/Main: Peter Lang.

- Biber, D., S. Conrad, & V. Cortes. (2004). *If you look at... : Lexical bundles in university teaching and textbooks*. *Applied Linguistics*, 25, 371-405.
- Biber, D., Reppen, R., & Conrad, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Biber, D., Kim, Y-J., & Tracy-Ventura, N. (2009). A Corpus-driven approach to comparative phraseology: lexical bundles in English, Spanish, and Korean. In S. Iwasaki, H. Hoji, P. M. Clancy, & S-O. Sohn (Eds.), *Japanese/Korean Linguistics, Volume 17* (pp. 75-94). Stanford, CA: CSLI Publication.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999) *Longman grammar of spoken and written English*. London: Longman.
- Bolinger, D. (1976). Meaning and memory. *Forum Linguisticum*, 1, 1-14.
- Bybee, J. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.
- Bybee, J. (2003). Cognitive processes in grammaticalization. In M. Tomasello (Ed.), *The new psychology of language, vol. 2* (pp. 145-167). Mahwah, NJ: Lawrence Erlbaum.
- Bybee, J. (2006). From usage to grammar: the mind's response to repetition. *Language*, 82(4), 711-733.
- Bybee, J., & Cacoullos, R. T. (2009). The role of prefabs in grammaticization: how the particular and the general interact in language change. In R. Corrigan, E. A. Moravcsik, H. Quali, & K. M. Wheatley (Eds.), *Formulaic language: Volume I: Structure, distribution and historical change* (pp. 187-217). Philadelphia: John Benjamins.
- Chafe, W. (1982). Integration and involvement in speaking, writing, and oral literature. In D. Tannen (Ed.), *Spoken and written language: Exploring orality and literacy* (pp. 34-53). Norwood: Ablex.
- Chafe, W. (1994). *Discourse, consciousness, and time*. Chicago/London: The University of Chicago Press.
- Chafe, W., & Danielewicz, J. (1987). Properties of spoken and written language. In R. Horowitz, & S. Samuels (Eds.), *Comprehending oral and written language* (pp. 83-113). San Diego, CA: Academic Press.
- Chen, Y-H., & Baker, P. (2010) Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, 14(2), 30-49.
- Clancy, P. M. (1986). The acquisition of communicative style in Japanese. In B. B. Schieffelin, & E. Ochs (Eds.), *Language socialization across cultures* (pp. 213-250). Cambridge: Cambridge University Press.

- Clancy, P. M. (2009). Usage-based linguistics: theory and methodology. In S. Iwasaki, H. Hoji, P. M. Clancy, & S-O. Sohn (Eds.), *Japanese/Korean Linguistics, Volume 17* (pp. 135-147). Stanford, CA: CSLI Publication.
- Cook, H. M. (2008). Style shifts in Japanese academic consultations. In K. Jones, & T. Ono (Eds.), *Style shifting in Japanese* (pp. 9-38). Amsterdam: John Benjamins.
- Cowie, A. (1981). The treatment of collocations and idioms in learners' dictionaries. *Applied Linguistics, 2/3*, 223-35.
- Cowie, A. (1998). *Phraseology: theory, analysis, and applications*. Oxford: Oxford University Press.
- Dixon, R. M. W., & Aikhenvald, A. (Eds.). (2002). *Word: A cross-linguistic typology*. Cambridge: Cambridge University Press.
- Dorgeloh, H., & Wanner, A. (2009). Formulaic argumentation in scientific discourse. In R. Corrigan, E. A. Moravcsik, H. Quali, & K. M. Wheatley (Eds.), *Formulaic language: Volume II: Acquisition, loss, psychological reality, and functional explanations* (pp. 523-544). Philadelphia: John Benjamins.
- Du Bois, J. (1985). Competing motivations. In J. Haiman (Ed.), *Iconicity in syntax* (pp. 343-365). Amsterdam/Philadelphia: John Benjamins.
- Du Bois, J. (1987). The discourse basis of ergativity. *Language, 63*, 805-55.
- Ellis, N. C., & Frey, E. (2009). The psycholinguistic reality of collocation and semantic prosody (2): affective priming. In R. Corrigan, E. A. Moravcsik, H. Quali, & K. M. Wheatley (Eds.), *Formulaic Language: Volume II: Acquisition, loss, psychological reality, and functional explanations* (pp. 117-145). Amsterdam/Philadelphia: John Benjamins.
- Erman, B. (2009). Formulaic language from a learner perspective: What the learner needs to know. In R. Corrigan, E. A. Moravcsik, H. Quali, & K. M. Wheatley (Eds.), *Formulaic language: Volume II: Acquisition, loss, psychological reality, and functional explanations* (pp. 323-346). Amsterdam/Philadelphia: John Benjamins.
- Erman, B., & Warren, B. (2000). The idiom principle and the open-choice principle. *Text, 20*, 29-62.
- Ferguson, C. (1983). Sports announcer talk: syntactic aspects of register variation. *Language in Society, 12*, 153-72.
- Fillmore, C. J. (1979). On fluency. In C. J. Fillmore, D. Kempler, & W. S-Y. Wang (Eds.), *Individual differences in language ability and language behavior* (pp. 85-101). New York: Academic Press.

- Fillmore, C. J. (1988). The mechanisms of 'Construction Grammar.' *BLS*, 14, 35-55.
- Fillmore, C. J., Kay, P., & O'Connor, M. C. (1988). Regularity and idiomaticity in English grammar: the case of let alone. *Language*, 64(3), 501-538.
- Ford, C. E., & Thompson, S. A. (1996). Interactional units in conversation: syntactic, intonational, and pragmatic resources for the management of turns. In E. Ochs, E. A. Schegloff, & S. A. Thompson (Eds.), *Interaction and grammar* (pp. 134-84). Cambridge: Cambridge University Press.
- Fox, B. (2001). An exploration of prosody and turn projection in English conversation. In E. Couper-Kuhlen, & M. Selting (Eds.), *Interactional linguistics* (pp. 287-315). Amsterdam: John Benjamins.
- Fujii, N. (1991). *Historical Discourse Analysis: Grammatical Subject in Japanese*. Berlin: Mouton de Gruyter.
- Fujii, N., & Ono, T. (2000). The occurrence and non-occurrence of the Japanese direct object marker "o" in conversation. *Studies in Language*, 24(1), 1-39.
- Gabrielatos, C. (2005). Corpora and language teaching: just a fling or wedding bells? *TESL-EJ*, 8(4), A1, 1-37.
- Gavioli, L., & Aston, G. (2001). Enriching reality: language corpora in language pedagogy. *ELT Journal*, 55(3), 238-246.
- Givón, T. (1989). *Mind, code and context*. Hillsdale, NJ: Lawrence Erlbaum.
- Gläser, R. (1986). The plea for phraseo-stylistics. In D. Kastovsky, & A. Szwedek (Eds.), *Linguistics across historical and geographical boundaries* (pp. 41-52). Berlin: Mouton de Gruyter.
- Gläser, R. (1988). The grading idiomaticity as a presupposition for a taxonomy of idioms. In W. Hüllen, & R. Schulze (Eds.), *Understanding the lexicon: meaning, sense and world knowledge in lexical semantics* (pp. 264-279). Tübingen: Max Niemeyer.
- Goldberg, A. (1995). *A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
- Haiman, J. (1994). Ritualization and the development of language. In W. Pagliuca (Ed.), *Perspectives on grammaticalization* (pp. 3-28). Amsterdam: John Benjamins.
- Hasegawa, Y. (2010). *Soliloquy in Japanese and English*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

- Hatasa, Y.A., Hatasa, K., & Makino, S. (2010). *Nakama 1: Introductory Japanese: Communication, culture, context*. Florence, KY: Cengage Learning.
- Haugh, M. (2008). Utterance-final conjunctive particles and implicature in Japanese conversation. *Pragmatics*, 18(3), 425-451.
- Hayashi, M. (2004). Discourse within a sentence: An exploration of postpositions in Japanese as an interactional resource. *Language in Society*, 33, 343-376.
- Hirose, Y. (1995). Direct and indirect speech as quotations of public and private expression. *Lingua*, 95, 223-238.
- Hopper, P. (1987). Emergent grammar. *Berkeley Linguistics Conference (BLS)*, 13, 139-157.
- Hopper, P. (1998). Emergent Grammar. In M. Tomasello (Ed.), *The New Psychology of Language: Cognitive and Functional Approaches to Language Structure* (pp. 155-175). New Jersey: Lawrence Erlbaum Associates.
- Horie, K. (1993). *A cross-linguistic study of perception and cognition verb complements: A cognitive perspective*. University of Southern California, Ph.D. dissertation.
- Hornby, A. S., Gatenby, E. V., & Wakefield, H. (1942). *Idiomatic and syntactic English dictionary*. Tokyo: Kaitakusha.
- Howarth, P. (1996). *Phraseology in English academic writing: some implications for language learning and dictionary making*. Tübingen: Max Niemeyer.
- Howarth, P. (1998). The phraseology of learners' academic writing. In A. P. Cowie (Ed.), *Phraseology: theory, analysis, and applications* (pp. 161-186). Oxford: Clarendon Press.
- Hudson, J., & Wiktorsson, M. (2009). Formulaic language and the relater category - the case of *about*. In R. Corrigan, E. A. Moravcsik, H. Quali, & K. M. Wheatley (Eds.), *Formulaic language: Volume I: Structure, distribution and historical change* (pp. 77-95). Philadelphia: John Benjamins.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Hymes, D. H. (1968). The ethnography of speaking. In J. A. Fishman (Ed.), *Readings in sociology of language* (pp. 99-138). The Hague/Paris: Mouton.
- Hymes, D. H. (1984). Sociolinguistics: stability and consolidation. *International Journal of the Sociology of Language*, 45, 39-45.
- Ikeda, K. (2007). The change-of-state token in Japanese language proficiency interviews. In *Proceedings of the 6<sup>th</sup> Annual JALT Pan-SIG Conference* (pp. 56-64). Sendai: Tohoku Bunka Gakuen University.

- Ikegami, Y. (1981). *SURU to NARU no gengogaku: Gengo to bunka no taiporojii e no shiron* [Linguistics of DOING and BECOMING: An essay towards a typology of language and culture]. Tokyo: Taishukan Shoten.
- Iwasaki, S. (1985). Cohesion, challengeability, and *-n desu* clauses in spoken discourse. *Journal of Asian Culture*, 9, 125-142.
- Iwasaki, S. (2002). *Japanese*. Amsterdam/Philadelphia: John Benjamins.
- Jimbo, K. (1948). *kokugo onseegaku gaiyoo* [Outline of Japanese phonetics]. Tokyo: Meiji Tosho.
- Johnson, K. (1997). Speech perception without speaker normalization. In K. Johnson, & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 145-65). San Diego: Academic Press.
- Jones, K., & Ono, T. (Eds.). (2008). *Style shifting in Japanese*. Amsterdam: John Benjamins.
- Kaiser, S., Ichikawa, Y., Kobayashi, N., & Yamamoto, H. (2001). *Japanese: A comprehensive grammar*. New York: Routledge.
- Kamio, A. (1997). *Territory of information*. Amsterdam/Philadelphia: John Benjamins.
- Kay, P. (1997). *Words and the grammar of context*. Stanford, CA: CSLI Publication.
- Kay, P., & Fillmore, C. J. (1999). Grammatical constructions and linguistic generalizations: the *What's X Doing Y?* construction. *Language*, 75, 1-33.
- Kerz, E., & Haas, F. (2009). The aim is to analyze NP: The function of prefabricated chunks in academic texts. In R. Corrigan, E. Moravcsik, H. Quali, & K. Wheatley. (Eds.), *Formulaic language: Volume I: Structure, distribution and historical change* (pp. 97-115). Philadelphia: John Benjamins.
- Kim, Y-J. (2009). Korean lexical bundles in conversation and academic texts. *Corpora*, 4(2), 135-165.
- Kunihiro, T. (1985). *kanyookuron* [phraseology]. *Nihongogaku*, 4(1), 4-14.
- Kuno, S. (1973). *The Structure of the Japanese language*. Cambridge, MA: MIT Press.
- Langacker, R. (1988). A view of linguistics semantics. In B. Rudzka-Ostyn (Ed.), *Topics in cognitive linguistics* (pp. 50-90). Amsterdam/Philadelphia: John Benjamins.
- Langacker, R. W. (1987). *Foundations of cognitive grammar, Volume I: Theoretical prerequisites*. Stanford, CA: Stanford University Press.

- Langacker, R. W. (1991). *Foundations of cognitive grammar, Volume II: Descriptive application*. Stanford, CA: Stanford University Press.
- Langacker, R. W. (1998). Conceptualization, symbolization, and grammar. In M. Tomasello (Ed.), *The new psychology of language* (pp. 1-39). Mahwah, NJ: Lawrence Erlbaum Associates.
- Lenk, U. (1997). Discourse markers. In J. Verschueren, J. O. Östman, J. Blommaert, & C. Bulcaen (Eds.), *Handbook of pragmatics* (pp. 1-17). Amsterdam: John Benjamins.
- Locke, J. (1993). *The child's path to spoken language*. Cambridge, MA: Harvard University Press.
- Locke, J. (1995). Development of the capacity for spoken language. In P. Fletcher, & B. MacWhinney (Eds.), *The handbook of child language* (pp. 278-302). Oxford: Blackwell.
- Locke, J. (1997). A theory of neurolinguistics development. *Brain and Language*, 58, 265-326.
- Lyons, J. (1968). *Introduction to theoretical linguistics*. Cambridge: University Press.
- Matsui, T. (2002). Semantics and pragmatics of a Japanese discourse marker *dakara*: a unitary account. *Journal of Pragmatics*, 34, 867-891.
- Matsumoto, Y. (1998). Semantic change in the grammaticalization of verbs into postpositions in Japanese. In T. Ohori (Ed.), *Studies in Japanese grammaticalization: Cognitive and discourse perspectives* (pp. 25-60). Tokyo: Kuroshio Shuppan.
- Maynard, S. K. (1989). Functions of the discourse marker “*dakara*” in Japanese conversation. *Text*, 9(4), 389-414.
- Maynard, S. K. (1992). Cognitive and pragmatic messages of a syntactic choice: The case of the Japanese commentary predicate *n(o) da*. *Text*, 12(4): 563-613.
- Maynard, S. K. (1993). *Discourse modality: Subjectivity, emotion, and voice in the Japanese language*. Amsterdam: John Benjamins Publishing Company.
- Maynard, S. K. (1996). Multivoicedness in speech and thought representation: the case of self-quotation in Japanese. *Journal of Pragmatics*, 25, 207-226.
- Maynard, S. K. (2002). *Linguistic emotivity: Centrality of place, the topic-comment dynamic, and an ideology of pathos in Japanese discourse*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Maynard, S. K. (2005). Another conversation: Expressivity of *mitaina* and inserted speech in Japanese discourse. *Journal of Pragmatics*, 37(6), 837-869.

- Maynard, S. K. (2007). *Linguistic creativity in Japanese discourse: Exploring the multiplicity of self, perspective, and voice*. Amsterdam: John Benjamins.
- McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. London: Routledge.
- Miyaji, Y. (1982). *kanyooku no imi to yoohoo* [meanings and functions of idioms]. Tokyo: Meiji Shoin.
- Miyaji, Y. (1985). *kanyooku no shuuhen—rengo, kotowaza, fukugoogo* [vicinity of idioms—collocations, proverbs, compounds]. *Nihongogaku*, 4(1), 62-75.
- Moon, R. (1998). Frequencies and forms of phrasal lexemes in English. In A. P. Cowie (Ed.), *Phraseology: theory, analysis, and applications* (pp. 79-100). Oxford: Clarendon Press.
- Mori, J. (1999). *Negotiating agreement and disagreement in Japanese*. Amsterdam: John Benjamins Publishing Company.
- Müller, S. (2005). *Discourse markers in native and non-native English discourse. Pragmatics & Beyond New Series 138*. Amsterdam: John Benjamins.
- Najima, Y. (2007). *Noda no imi/kinoo: Kanrensei riron no kanten kara* [The meaning and function of *noda*: From the perspective of relevance theory]. Tokyo: Kuroshio Shuppan.
- Nakayama, T., & Ichihashi-Nakayama, K. (1997). Japanese *kedo*: discourse genre and grammaticization. In H. Sohn, & J. Haig (Eds.), *Japanese/Korean Linguistics, Volume 6* (pp. 607-618). Stanford, CA: CSLI Publications.
- Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.
- Noda, H. (1997). '*No (da)*' *no kinoo* [Functions of *no (da)*]. Tokyo: Kuroshio Shuppan.
- Ochs, E., Schegloff, E. A., & Thompson, S. A. (Eds.). (1996). *Interaction and grammar*. Cambridge: Cambridge University Press.
- Ohori, T. (1997). Framing effects in Japanese non-final clauses: toward an optimal grammar-pragmatics interface. *BLS*, 23, 471-480.
- Ono, T. (2006). An emotively motivated post-predicate constituent order in a 'strict predicate final' language: emotion and grammar meet in Japanese everyday talk. In S. Suzuki (Ed.), *Emotive Communication* (pp. 139-153). Amsterdam: Benjamins.
- Ono, T., & Suzuki, R. (1992). Word order variability in Japanese conversation: motivations and grammaticization. *Text*, 12, 429-445.

- Ono, T., & Thompson, S. A. (2003). Japanese (*w*)*atashilore/boku* 'I': They're not just pronouns. *Cognitive Linguistics*, 14(4), 321-347.
- Ono, T., & Thompson, S. A. (2009). Fixedness in Japanese adjectives in conversation: toward a new understanding of a lexical ('part-of-speech') category. In R. Corrigan, E. A. Moravcsik, H. Quali, & K. M. Wheatley (Eds.), *Formulaic Language: Volume I: Structure, distribution and historical change* (pp. 117-145). Philadelphia: John Benjamins.
- Ono, T., Thompson, S. A., & Suzuki, R. (2000). The pragmatic nature of the so-called subject marker "ga" in Japanese: evidence from conversation. *Discourse Studies* 2(1), 55-84.
- Onodera, N. O. (2004). *Japanese discourse markers: Synchronic and diachronic discourse analysis*. Philadelphia/Amsterdam: John Benjamins.
- Palmer, H. E. (1933). *Second interim report on English collocations*. Tokyo: Kaitakusha.
- Pawley, A. (2007). Developments in the study of formulaic language since 1970: A personal view. In P. Skandera (Ed.), *Phraseology and culture in English* (pp. 3-34). Berlin: Mouton de Gruyter.
- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In: J. Richards, R. Schmidt (Eds.), *Language and communication* (pp. 191-225). London: Longman.
- Perkins, M. R. (1999). Productivity and formulaicity in language development. In M. Garman, C. Letts, B. Richards, C. Schelletter, & S. Edwards (Eds.), *Issues in normal & disordered child language: from phonology to narrative. Special issue of the new bulmershe papers* (pp. 51-67). Reading: University of Reading.
- Peters, A. (1983). *The units of language acquisition*. Cambridge: Cambridge University Press.
- Peters, A. (2009). Connecting the dots to unpack the language. In R. Corrigan, E. A. Moravcsik, H. Quali, & K. M. Wheatley (Eds.), *Formulaic language: Volume II: Acquisition, loss, psychological reality, and functional explanations* (pp. 387-404). Philadelphia: John Benjamins.
- Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee, & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 137-58). Amsterdam: John Benjamins.
- Pierrehumbert, J. (2002). Word-specific phonetics. In C. Gussenhoven, & N. Warner (Eds.), *Laboratory phonology 7* (pp. 101-39). Berlin: Mouton de Gruyter.

- Raupach, M. (1984). Formulae in second language speech production. In H. Dechert, D. Möhle, & M. Raupach (Eds.), *Second language productions* (pp. 114-137). Tübingen: Gunter Narr.
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language*, 50, 696-735.
- Saft, S. (2011). Displays of concession in university faculty meeting: culture and interaction in Japanese. *Pragmatics*, 11(3), 223-262.
- Sams, J. (2009). Genre-controlled construction in written language quotatives. In R. Corrigan, E. A. Moravcsik, H. Quali, & K. M. Wheatley (Eds.), *Formulaic language: Volume I: Structure, distribution and historical change* (pp. 147-169). Philadelphia: John Benjamins.
- Schegloff, E. A. (1979). The relevance of repair to syntax-for-conversation. In T. Givón (Ed.), *Syntax and semantics (vol. 12. Discourse and syntax)* (pp. 261-286). New York: Academic Press.
- Schegloff, E. A. (1987). Analyzing single episodes of interaction: an exercise in conversation analysis. *Social Psychology Quarterly*, 50, 101-114.
- Schegloff, E. A. (1996). Issues of relevance for discourse analysis: Contingency in action, interaction and co-participant context. In E. H. Hovy, & D. Scott (Eds.), *Computational and conversational discourse: Burning issues - an interdisciplinary account* (pp. 3-38). Heidelberg: Springer Verlag.
- Schegloff, E. A. (1999). Discourse, pragmatics, conversation, analysis. *Discourse Studies*, 1(4), 405-435.
- Schiffrin, D. (1987). *Discourse markers*. Cambridge: Cambridge University Press.
- Searle, J. (1969) *Speech acts: An essay in the philosophy of language*, Cambridge: Cambridge University Press.
- Segalowitz, N. (2001). On the evolving connections between psychology and linguistics. *Annual Review of Applied Linguistics*, 21, 3-22.
- Shibatani, M. (1990). *The Languages of Japan*. Cambridge: Cambridge University Press.
- Shiraishi, D. (1950). *Nihongo no idiomu* [Japanese idioms]. Tokyo: Sanseido.
- Shoji, K. (1999). *Japanese core words and phrases: Things you can't find in a dictionary*. Tokyo: Kodansha International.

- Sidtis, D. (2009). Formulaic and novel language in a ‘dual process’ model of language competence. In R. Corrigan, E. A. Moravcsik, H. Quali, & K. M. Wheatley (Eds.), *Formulaic language: Volume II: Acquisition, loss, psychological reality, and functional explanations* (pp. 445-470). Philadelphia: John Benjamins.
- Sinclair, J. (1987). Collocation: a progress report. In R. Steele, & T. Threadgold (Eds.), *Language topics, essays in honor of Michael Halliday* (pp. 319-331). Amsterdam: John Benjamins.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Strawson, P. F. (1964). Intention and convention in speech acts. *Philosophical Review*, 73, 439-60.
- Sohn, S-O., & Nam, K. (2011). The formulaic expressions of ‘thing’ in spoken and written registers: a corpus study of bound nouns *kes* and *ke* in Korean. In B. Frellesvig & P. Sells (Eds.), *Japanese/Korean Linguistics, Volume 20*. Stanford, CA: CSLI Publications.
- Stubbs, M. (1995). Collocations and semantic profile: on the cause of the trouble with quantitative methods. *Functions of Language*, 2, 1-33.
- Suzuki, S. (1995). A study of the sentence-final *mitai na*. *Journal of the Association of Teachers of Japanese*, 29(2), 55-78.
- Suzuki, S. (1996a). An analysis of *tte-* and *to yuu no wa-* marked phrases in Japanese discourse. *Word*, 47(3), 321-337.
- Suzuki, S. (1996b). Incorporation of information and complementizers in Japanese. *Pragmatics*, 6(4), 511-551.
- Suzuki, R. (2006). How does ‘reason’ become less and less reasonable? Pragmatics of the utterance-final *wake* in conversational discourse. In S. Suzuki (Ed.), *Emotive communication in Japanese* (pp. 35-51). Amsterdam: John Benjamins.
- Takubo, Y., & Kinsui, S. (1997). *Ootooshi, kandooshi no danwateki kinoo* [The discourse management function of fillers in Japanese]. In Spoken Language Working Group (Ed.), *Bunpoo to Onsei* [Speech and grammar] (pp. 257-279). Tokyo: Kuroshio Shuppan.
- Tanaka, H. (1999). *Turn-taking in Japanese conversation: A study in grammar in interaction*. Amsterdam: John Benjamins.
- Tanaka, H. (2005). Grammar and the “timing” of social action: word order and preference organization in Japanese. *Language in Society*, 34, 389-430.

- Tao, H. (2001). Discovering the usual with corpora: The case of 'remember'. In R. Simpson, & J. Swales (Eds.), *Corpus linguistics in North America: Selections from the 1999 symposium* (pp. 116-144). Ann Arbor: University of Michigan Press.
- Tao, H. (2003). A usage-based approach to argument structure: 'remember' and 'forget' in spoken English. *International Journal of Corpus Linguistics*, 8(1), 75-95.
- Taylor, Y. (2010). Functions of Japanese exemplifying particles in spoken and written discourse. University of California, Los Angeles, Ph.D. dissertation.
- Thompson, S. A., & Couper-Kuhlen, E. (2005). The clause as a locus of grammar and interaction. *Discourse Studies* 7.4/5, 481-505.
- Thompson, S. A., & Ono, T. (2009). Toward a conversation-centered understanding of human language: Goals, assumptions, recent findings, and future directions. In S. Iwasaki, H. Hoji, P. M. Clancy, & S-O. Sohn (Eds.), *Japanese/Korean Linguistics, Volume 17* (pp. 95-115). Stanford, CA: CSLI Publications.
- Tohsaku, Y-H. (2006). *Yookoso! An invitation to contemporary Japanese*. New York: McGraw-Hill.
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Tracy-Ventura, N., Biber, D., & Cortes, V. (2007). Lexical bundles in Spanish speech and writing. In G. Parodi (Ed.), *Working with Spanish corpora* (pp. 217-231). London: Continuum.
- Van Lancker, D. (1987). Nonpropositional speech: neurolinguistic studies. In A.W. Ellis (Ed.), *Progress in the psychology of language, vol.3* (pp. 49-118). Hillsdale, NJ: Lawrence Erlbaum.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge, Cambridge University Press.
- Wray, A. (2008). *Formulaic language: pushing the boundaries*. Oxford: Oxford University Press.
- Wray, A., & Grace, G. W. (2007). The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua*, 117(3), 543-578.
- Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: an integrated model. *Language & Communication*, 20(1), 1-28.

## DATA SOURCES

- Usami, M. (Ed.) (2007). *Multilingual corpus of spoken language by basic transcription system (BTS) — Japanese conversation*. Tokyo University of Foreign Studies.
- Aoki, H., Matsugu, Y., Miyashita, M., Ono, T. and Sadler, M. (to appear) *Japan Corpus*. Department of East Asian Studies, University of Arizona.
- Gendai Nihongo Kenkyukai. (1997). *Josei no kotoba: shokuba hen* [Female corpus: in workplace]. Tokyo: Hitsuji Shobo.
- Gendai Nihongo Kenkyukai. (2002). *Dansei no kotoba: shokuba hen* [Male corpus: in workplace]. Tokyo: Hitsuji Shobo.
- Miyata, S., Banno, K., Konishi, S., Matsui, A., Matsumoto, S., Ooki, R., Takahashi, A. and Murai, K. (2010). *Sakura Corpus*. Talkbank.
- Wheatley, B., Kaneko M. and Kobayashi, M. (1996). *CALLHOME Japanese Transcripts*. Linguistic Data Consortium, Philadelphia.
- Canavan, A. and Zipperlen, G. (1996). *CALLHOME Japanese Speech*. Linguistic Data Consortium, Philadelphia.
- Furui, S., Maekawa, K. and Isahara, H. (2004). *The Corpus of Spontaneous Japanese*. National Institute for Japanese Language and National Institute of Information and Communications Technology.
- Uemura, R. (1997). *Hypermedia Corpus of Spoken Japanese*. Fukuoka Institute of Technology. Retrieved September 19, 2011, from <http://www.env.kitakyu-u.ac.jp/corpus/docs/index.html>.
- Fujii, T. (General Producer). *Tetsuko no heya* [Tetsuko's room]. TV Asahi Corporation. August, 2007–September, 2010.

## SOFTWARE PROGRAMS

- ChaSen for Windows version 2.0 (1999). Computational Linguistics Laboratory, Graduate School of Information Science, NAIST (Nara Institute of Science and Technology).
- Anthony, L. (2011). AntConc (Version 3.2.2) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.antlab.sci.waseda.ac.jp/>