

VARIATION IN JAPANESE NOMINAL PARTICLE OMISSION: TOWARDS A  
CORPUS-BASED SYNTACTIC ANALYSIS

BY

RACHEL ABIGAIL ANKIRSKIY

(Under the Direction of Vera Lee-Schoenfeld)

ABSTRACT

The numerous Japanese dialects exhibit phonological, lexical and morphosyntactic variation. The dialects can vary in frequency of post-nominal particle omission, a phenomenon conditioned by many extra-linguistic and linguistic factors. In this thesis, a quantitative analysis of particle omission in transcribed natural conversation data from the COJADS corpus (NINJAL 2021) will inform a syntactic analysis of the phenomenon. Data from four prefectures (Hokkaido, Tokyo, Osaka, and Fukuoka) will be examined. I address two key research questions: Does evidence in the COJADS data uphold prior conclusions that particle omission varies by dialect and linguistic and extra-linguistic factors? How can we explain the phenomenon via traditional generative analysis? The quantitative analysis confirms dialectal variation and variation based on extra-linguistic factors. These results inform an analysis of particle omission in the Minimalist framework, and it is suggested that the particle omission phenomenon can be explained in terms of phases and Merge.

INDEX WORDS: Dialectal variation; Japanese dialects; Syntax; Generative linguistics; Morphology; Minimalism; Phases; Corpus linguistics; Sociolinguistic variation; Particle ellipsis

VARIATION IN JAPANESE NOMINAL PARTICLE OMISSION: TOWARDS A  
CORPUS-BASED SYNTACTIC ANALYSIS

By

RACHEL ABIGAIL ANKIRSKIY

B.A., University of Georgia, 2016

A Thesis Submitted to the Graduate Faculty of the University of Georgia in Partial Fulfillment of  
the Requirements for the Degree

MASTER OF ARTS

ATHENS, GEORGIA

2022

© 2022

Rachel Abigail Ankirskiy

All Rights Reserved

VARIATION IN JAPANESE NOMINAL PARTICLE OMISSION: TOWARDS A  
CORPUS-BASED SYNTACTIC ANALYSIS

By

RACHEL ABIGAIL ANKIRSKIY

Major Professor: Vera Lee-Schoenfeld

Committee: Keith Langston  
Jon Forrest  
Pilar Chamorro Fernández

Electronic Version Approved:

Ron Walcott  
Vice Provost for Graduate Education and Dean of the Graduate School  
The University of Georgia  
May 2022

**DEDICATION**

In memory of Grandpa, Robert R Nason.

You must have given me a genetic disposition to

love languages, especially Japanese.

&

To my family

## **ACKNOWLEDGEMENTS**

I would like to thank my major professor, Dr. Vera Lee-Schoenfeld, for her help and mentorship during my graduate studies, and for introducing me to the wonderful world of syntax during my time as an undergraduate student. I would also like to thank my committee, Drs. Pilar Chamorro, Jon Forrest, and Keith Langston, for their help and support as instructors and committee members. Each of their courses inspired me towards the topic of this thesis. Without the past research conducted by NINJAL, this project would not have been possible, and I thank them for making such valuable data as is contained in the COJADS corpus available to enable vital research on Japanese dialects. It is not an easy task to pursue a higher degree while holding a full-time job, but my supervisor Dr. Jessie Kissinger made this possible; I would like to thank her for her support and encouragement over the years. I am also grateful to my co-workers, especially Rodrigo, Rui, and Jeremy, for providing advice even though their own fields of research are vastly different from my own. I cannot forget to thank my colleague Kihoon Kim, who pointed me in the right direction with regards to Japanese linguistics. I would also like to thank Becky Sexton, for her friendship and advice. Finally, I am most grateful to my family for their love and support; to Elena and Viktor who supported and believed in me from afar; to my parents Robin and Jeff, for nurturing and supporting my curiosity in all things science and language, and for always believing in me; to my brother Joshua, for his friendship; and to Bonzo, for his feline companionship and constant antics. Most of all, I am grateful to my husband Alex; without his love, support and patience, this entire endeavor would have been impossible.

## **TABLE OF CONTENTS**

ACKNOWLEDGEMENTS.....	v
CHAPTER 1 – INTRODUCTION.....	1
1.1 Background.....	1
1.2 Questions & Objectives.....	5
1.3 Thesis Structure Overview.....	6
CHAPTER 2 – BACKGROUND INFORMATION.....	7
2.1 The Japanese Language.....	7
2.2 Japanese Dialectology and Dialectal Variation.....	7
2.3 Formality & Sociolinguistic Factors.....	14
2.4 Nominal Particles.....	15
2.5 Particle Omission.....	20
2.6 Labeling, Merge and Phases.....	36
CHAPTER 3 – QUANTITATIVE ANALYSIS OF THE COJADS DATA.....	40
3.1 Hypotheses.....	40
3.2 The Data & Data Preparation.....	41
3.3 The Variables.....	44
3.4 Data Visualization.....	49
3.5 The Statistical (ANOVA) Tests.....	66
3.7 ANOVA Results Discussion.....	71
CHAPTER 4 – THE SYNTACTIC ANALYSIS.....	73
4.1 Overview of Chapter 4.....	73
4.2 Summary of the Particle Omission Puzzle.....	74
4.3 Towards a Syntactic Analysis of Particle Omission.....	75
4.4 Complications to the Proposal.....	89
4.5 Revised Proposal & Recap of Outstanding Questions.....	97
CHAPTER 5 - CONCLUSION.....	99
REFERENCES.....	104
APPENDICES.....	112

Appendix A. Abbreviations and Transcription System .....	112
Appendix B. Sample Datasets .....	113
Appendix C. R Code .....	115

## CHAPTER 1 – INTRODUCTION

### **1.1 Background**

Japanese is most commonly considered a language isolate, although it consists of a large number of dialects (Shibatani 1990). Japanese is an SOV language (Shibatani 1990), which has fairly free word order (typically explained as “scrambling”; cf. Oku 1998, Saito 2016). As Nariyama (2003: 3) explains, “there is no grammatical requirement for nominal arguments, such as subject and object, to be overt in Japanese...” Nariyama goes on to explain that sentences can consist only of a single verb, yet Japanese lacks verbal agreement inflection that is typical of European “pro-drop” languages (such as Spanish) which generally permit subject omission because verbal inflection provides the necessary information (e.g. person and number). Japanese, while lacking such verbal inflectional agreement, allows not only subject but also object omission. The phenomenon of omitting these nominal arguments is typically referred to as ‘ellipsis’ and a lengthy treatment of the phenomenon can be found in Nariyama 2003. Recent investigations of the phenomenon refer to it as ‘argument ellipsis’ (cf. Saito 2016 , Saito 2017) a term which incorporates Sluicing, subject and object ellipsis, and other types of ellipsis phenomena.

Nominal arguments are typically marked with post-nominal particles (see Section 2.4 for details), which encode grammatical relations (including case and topic relations) (Kishimoto 2017). Some particles are typically treated as Case markers (*-ga* NOM and *-o* ACC), while others (e.g. *-de* INS or ‘in/at’ and *-e* ‘to a place’) are treated as postpositions; however, all are morphologically similar. These particles are frequently omitted in spoken language (Nariyama 2003; Shibatani 1990; Shimojo 2006; Takano 1998). Example (1) below provides a basic example of nominal

particle omission. All four utterances have the same referential meaning, but in (1b), (1c) and (1d), different particles have been omitted (# indicates an omitted particle).

(1) a. boku=wa kono hon=o katta

I=TOP this book=ACC bought

‘I bought this book.’

b. boku=# kono hon=# katta

I=# this book=# bought

‘I bought this book.’

c. boku=# kono hon=o katta

I=# this book=ACC bought

‘I bought this book.’

d. boku=wa kono hon=# katta

I=TOP this book=# bought

‘I bought this book.’

(Lit. ‘As for me, (I) bought this book.’)

*(Adapted from Lee 2002b: 684, Examples 1a and 2a.)*

These ellipsis phenomena present challenges for generative grammar, which traditionally (e.g. Chomsky 2000, 2001, 2013) requires grammatical relations to be explained in terms of Case and phi-feature agreement between the verb and nominal argument; if ellipsis of any nominal argument is possible, as is the omission of any post-nominal particle, how do speakers understand utterances? While there have been many attempts to explain both phenomena in various

frameworks and sub-fields of linguistics, a suitable explanation of particle omission in particular appears to be non-existent (at least in the literature currently available in English). While there are numerous attempts at explaining argument ellipsis (cf. Nariyama 2003, Saito 2016), there are far fewer attempts at explaining particle omission. In this thesis, I use the terms “particle omission” and “nominal particle omission” to refer to two phenomena which both result in a “missing” (not pronounced) particle: “particle ellipsis,” whereby a particle is elided, and the “zero particle,” which is argued to be a separate grammatical entity that cannot be traced back to a specific particle. There are many arguments surrounding these two phenomena, and the distinction will be further discussed in Section 2.5.2. However, I will take no position on the matter due to the limited scope of this thesis. Particle omission as a whole is an important phenomenon for understanding the structure of the Japanese language and may have wider implications in generative linguistics. Many recent studies, including Fry 2003, Shimojo 2006, and Takano 1998 have attempted to explain the phenomenon, but as Fry (2003: 119) concluded, “...theories that rely on a single categorical explanation...are doomed to inadequacy.”

As mentioned above, Japanese exhibits a large amount of dialectal variation (Koiso et al. 2019; Onishi 2017; Shibatani 1990; Tsujimura 2014), and the dialects are typically grouped into Eastern and Western varieties based on variation in accent and lexical items (Onishi 2017; Shibatani 1990; Tsujimura 2014). Regional dialects tend to be used primarily in informal or colloquial situations in local community groups and families, while Standard Japanese is used in other contexts such as school and the workplace where interactions with speakers of other dialects may occur. Despite a growing body of research into the dialectal diversity in Japan, as described by Onishi (2017), the lack of proper documentation of endangered dialects in Japan is still being

lamented along with the overall lack of consideration for sociolinguistic factors in most of the studies Onishi cites to date.

The National Institute for Japanese Language and Linguistics (NINJAL) began preparing the Corpus of Japanese Dialects (COJADS)<sup>1</sup> in recent years; this corpus addresses some of the problems cited by Onishi (2017), as it provides a very large database of natural conversation data of speakers of many dialects of Japan. COJADS consists primarily of data from the “Urgent Survey of Japan’s Regional Dialects,” a language documentation project which was conducted between 1977 and 1985 by the Agency for Cultural Affairs to collect natural conversation data for various dialects (Koiso et al. 2019; Nakagawa et al. 2019). The corpus consists of audio recordings, transcriptions, and translations into Standard Japanese, and the data is currently available for public download and search (for academic use, provided a free account is created) via COJADS or the accompanying website where the raw CSV data files are published.<sup>2</sup>

One way in which the dialects are thought to vary is in frequency of nominal particle omission in casual conversation, as was noted recently by Kibe et al. (2020). Because dialects and particle omission are both more commonly seen in informal conversation settings, it is likely that the two are interrelated and the recent study by Kibe et al. (2020) showed that this was in fact the case for the ACC *-o* omission they observed in the COJADS data. Variation based on other sociolinguistic factors as well as extra-linguistic factors has also been observed (cf. Fry 2003, Takano 1998, Lee 2002a).

In the hopes of addressing Fry’s (2003) conclusion that particle omission needs to be examined from many angles, as well as Onishi’s (2017) plea for consideration of sociolinguistic factors in Japanese dialectology, this thesis will examine the phenomenon of particle omission

---

<sup>1</sup> <https://chunagon.ninjal.ac.jp>

<sup>2</sup> <https://www2.ninjal.ac.jp/cojads/index.html?targ=data>

from an inter-disciplinary approach which combines syntax, sociolinguistics, corpus analysis, and to some extent pragmatics. Nariyama's belief that "...Japanese must contain concrete mechanisms with which to track referents without expressing them overtly" (Nariyama (2003: x), parallels my own intuition, and the idea was influential in the design and implementation of this thesis.

To this end, this thesis presents a two-part study of the nominal particle omission phenomenon; the first portion consists of a corpus analysis of the phenomenon in natural conversation data from native speakers of four Japanese dialects (obtained from raw data from the COJADS corpus), while the second portion consists of a proposed syntactic treatment of the phenomenon. Several questions will be addressed, which are expounded upon in Section 1.2 but can be condensed into two broad questions. First, does nominal particle omission vary in frequency between the four dialects examined, and is there variation based on other factors (both sociolinguistic and linguistic)? Second, how can we explain the phenomenon, and any observed variation, in generative terms? In order to answer the first question posed above, a quantitative analysis in R was conducted on a subset of data from the COJADS corpus, followed by a syntactic treatment of the phenomenon in the Minimalist framework (Chomsky 1995, 2000, 2001, etc.).

## **1.2 Questions & Objectives**

This thesis addresses two main goals. The first is to determine whether nominal particle omission varies in frequency between particle type and dialect of Japanese, as well to determine whether the sociolinguistic factors (age and gender) and other linguistic factors (sentence-final particle use and honorific prefix use) impact the frequency of omission.

The second goal is to explain the particle omission phenomenon in generative terms, while taking into consideration observations from the COJADS data analysis.

The specific research questions are as follows:

- 1) Does the frequency of nominal particle omission vary by dialect?
- 2) Are certain nominal particles omitted more than others, and is there any variation observed between dialects? If so, is this variation statistically significant?
- 3) Is there variation based on other extralinguistic factors, specifically speaker gender and speaker age? Is there variation observed based on other linguistic factors, namely use of honorific prefixes and sentence-final particles? If so, is this variation statistically significant?
- 4) How can particle omission be explained syntactically, in keeping with the Minimalist framework?

### **1.3 Thesis Structure Overview**

This thesis is made up of five chapters. This introduction is the first. Chapter 2 introduces background information on the main topics of the thesis: an overview of relevant features of the Japanese language, background on the Japanese dialects and history of dialectology in Japan, an introduction to the Corpus of Japanese Dialects (COJADS) and data used, sociolinguistic factors observed to impact Japanese language use, an introduction to nominal particles and the nominal particle omission problem, and a brief introduction to three key Minimalist concepts (Phases, Labeling and Merge). Chapter 3 describes the hypotheses posited for the quantitative analysis of the COJADS data, detail on the methods used for this portion of the analysis, and the results of the quantitative analysis. Chapter 4 presents a proposal for explaining the particle omission phenomenon in generative (Minimalist) terms that is informed by the quantitative analysis results. Chapter 5 wraps up the thesis with a conclusion and direction for future study. The bibliography follows, along with Appendices which include a brief description of the transcription system and abbreviations used, samples of the transcribed and formatted data, and the R script used.

## **CHAPTER 2 – BACKGROUND INFORMATION.**

### **2.1 The Japanese Language**

Japanese is most commonly considered a language isolate. While it shares features with Korean, Turkish and the Ryukyuan languages, the only languages it has been conclusively proven to be related to are the Ryukyuan languages (spoken in the islands of Okinawa in southern Japan) which are often considered to simply be very distant dialects of Japanese (Shibatani 1990; Tsujimura 2014). Japanese is an SOV language (Hasegawa 2018a; Shibatani 1990) with fairly free word order, and is said to allow extensive “scrambling” (cf. Oku 1998, Saito 2016). However, Japanese lacks the subject-verb agreement that is typical of European “pro-drop” languages (person and number), while allowing not only subject but also object omission. Japanese nominal arguments are typically marked with post-nominal particles (see Section 2.5), referred to as “nominal particles” or “particles” in this thesis, which encode grammatical relations (Kishimoto 2017) and which are frequently omitted (not pronounced) in spoken language (Nariyama 2003; Shibatani 1990; Shimojo 2006; Takano 1998). In addition to the nominal particles, sentence/clause-final particles are employed to convey speaker intent (for example the question particle *-no*, or the particle *-yo* used on emphatic sentences). Japanese also has a complex honorifics and formality system (see Section 2.3).

### **2.2 Japanese Dialectology and Dialectal Variation**

Japan displays a large amount of dialectal variation, and regional dialects are used primarily in informal or colloquial situations in local community groups or with friends and family who speak the same dialect, while Standard Japanese tends to be used in more formal contexts

including school and the workplace and settings with strangers who may not speak the same dialect. Phonological, lexical and morpho-syntactic variation between dialects and dialect groups has been attested, and the dialects are typically grouped into Eastern and Western varieties based on variation in accent and lexical items (Tsuji-mura 2014). The map in Figure 1 (adapted by Sum 2008 from the two dialect grouping maps in Shibatani 1990) shows these major groupings. The Eastern dialects include Tokyo, Tohoku, and Hokkaido dialects, while the Western dialects include Kansai, Shikoku, and Kyushu dialects. The Ryukyuan languages are typically treated separately from these two groupings.

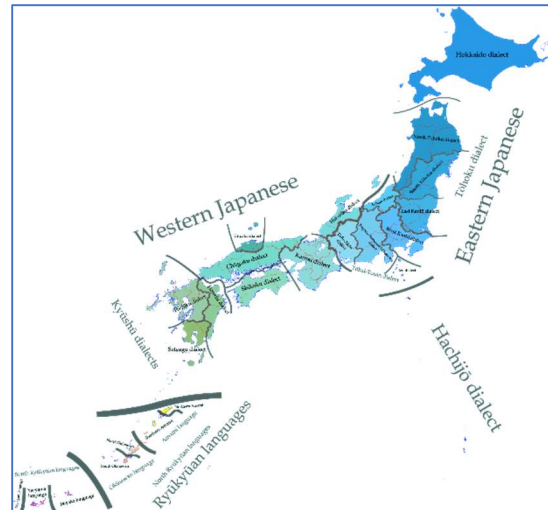


Figure 1. Major Dialect Groupings

Phonological variation has been attested, including the /i/ and /u/ high vowel merger in the Northern Tohoku dialects and the voicing of voiceless stops /k/ and /t/ in the Tohoku dialects, as well as pitch-accent variation (Tsuji-mura 2014). Lexical differences can vary from differences in common phrases to individual words (Tsuji-mura 2014). The most studied morphosyntactic variation is that of verbal conjugation according to Tsuji-mura (2014); however, variation in nominal particles has also been observed. In particular, there is evidence that nominal particle omission frequency varies between dialects (see for example Kibe et al. 2020).

Speakers of Japanese are known to code-switch between Standard Japanese and their dialect based on the situation and style or register that is called for (Okamoto 2008). Regional

dialects are more likely to be used when speaking with close friends or family, as indicated by Igarashi (2007). Despite this, dialect leveling towards Standard Japanese has been occurring for many reasons (some of which are discussed briefly in Section 2.2.1). Variation in regional dialect use based on speaker gender has also been suggested (cf. Shibamoto 1985); however, more recent studies, such as Didi-Ogren 2020, have indicated that this may not be the case.

### 2.2.1 Overview of Past Dialectology Studies

The use of regional dialects has been historically discouraged since the push to unify the country in the Meiji Restoration (1867), when the concept of *hyoujungo* ‘standard language’ was introduced by the Meiji government (Shibatani 1990). According to Shibatani (1990), *hyoujungo* was based on the Edo (Tokyo) dialect, and one of the goals of the Meiji government was to have a “unified common language.” Shibatani explains that with compulsory education came additional enforcement of the Tokyo dialect as the standard, and this resulted in dialects being viewed as inferior to the standard. After the end of World War II, the idea of a “common language” or *kyoutsugo* was introduced; as Shibatani (1990) explains, this variety was influenced by the original *hyoujungo*, but allowed for regional variation (such as accent features) while still retaining enough standard features to enable mutual intelligibility. This enabled more recent movements towards acceptance of dialectal differences, although as Shibatani (1990) explains, textbooks are still written with the “standard” language in mind. He mentions that there is still a leveling towards “variations” of *kyoutsugo* and that this poses a risk to dialectal variation in Japan. Throughout this thesis, references made to “Standard Japanese” are generally referring to the most standard version of *kyoutsugo* that is taught in textbooks such as for second-language learners of Japanese or grade-school students in Japan.

With the implementation of the more modern idea of *kyoutsugo* in the mid-20<sup>th</sup> century, the Japanese government began to acknowledge the importance of regional dialect variation. In 1949, the *Kokuritsu Kokugo Kenkyujo* ‘National Language Research Institute’ or ‘NLRI’ (currently, *Nihonhogen-no Kijyutsuteki Kenkyu* ‘National Institute for Japanese Language and Linguistics’, or ‘NINJAL’) was founded (Onishi 2017). NLRI/NINJAL has done extensive work in Japanese dialectology since it was founded, including descriptive studies of the dialects, a 6-volume *Grammar Atlas of Japanese Dialects*, and *The Linguistic Atlas of Japan* (Onishi 2017).

Recently, NINJAL began preparing the Corpus of Japanese Dialects (COJADS)<sup>3</sup>, which is currently (as of 2019) the largest corpus of Japanese dialects available according to Nakagawa et al. (2019). This corpus consists primarily of data from the “Urgent Survey of Japan’s Regional Dialects,” a language documentation project conducted between 1977 and 1985 by the Agency for Cultural Affairs to collect natural conversation data between native speakers of various dialects (Koiso et al. 2019; Nakagawa et al. 2019). Prior to the ongoing corpus preparation effort, only part of the data from the “Urgent Survey of Japan’s Regional Dialects” had been published (via the series of volumes titled *Zenkoku hogen danwa deetabeesu – Nihon no furusata kotoba shuusei*), but once the COJADS corpus is complete all the data collected via the “Urgent Survey of Japan’s Regional Dialects” will become easily searchable (Koiso et al. 2019; Nakagawa et al. 2019).

In addition to the increasing focus on documenting and studying the dialects of Japan, there have also been a number of sociolinguistics studies performed in relation to the dialects of Japan. For instance, Didi-Ogren 2020 examined regional dialect use by women, and King et al. 2021 performed a dialect perception study of the Osaka dialect in relation to the concept of masculinity

---

<sup>3</sup> <https://chunagon.ninjal.ac.jp>

and fatherhood. Despite these efforts, Onishi (2017) stresses the need for ongoing documentation efforts, as well as the need to perform not only dialectology studies but also sociolinguistic ones.

Japanese itself is by no means an endangered language; it had over 127 million native speakers in Japan as of 2015 and millions of second-language learners worldwide (Hasegawa 2018b). However, according to the Endangered Languages Database (Celik & Kibe 2019), many Japanese dialects are endangered, in addition to 8 languages spoken in Japan (Ainu, Hachijo, Amami, Kunigami, Okinawa, Miyako, Yaeyama and Yonaguni). The historical suppression of regional dialect use, combined with the continued use of Standard Japanese in schools and workplace settings and a population shift towards cities and away from traditional occupations has caused a leveling towards Standard Japanese; as Okamoto (2008: 132) states, "...dialect leveling toward 'Standard Japanese', ... or *hyoujungo*, has been steadily advancing in modern Japan." Onishi (2017) also mentions that there is a "general tendency" towards standard Japanese, and although he mentions that people in cities still learn and use local dialects. In fact, Nakagawa et al. (2019: 371) state that "under the influence of the dominant language, so-called Standard Japanese, most, if not all, of such local languages are in danger of extinction." Efforts should therefore be made to continue the study of the many dialects of Japan, as well as to document them, such as the Endangered Languages Database team has been doing since 2016 (Celik & Kibe 2019). While this thesis does not contribute to the documentation effort, I hope that it provides a basis for future study on dialectal variation in Japan and presents questions that will invite further research.

### 2.2.2 *The COJADS Corpus*

The Corpus of Japanese Dialects (COJADS) is currently the largest corpus of Japanese dialects available according to Nakagawa et al. (2019). As mentioned in Section 2.2.1, the corpus consists of data from the "Urgent Survey of Japan's Regional Dialects" and the "Hometown

Language Collection” (surveys which were conducted between 1977 and 1985 to collect natural conversation data between native speakers of various dialects) (Koiso et al. 2019; Nakagawa et al. 2019). The original data consists of recording tapes of dialect conversations amounting to about 4000 hours and representing around 200 locations in 47 prefectures of Japan. Although not all of this original data is available in COJADS yet, an extensive amount of data is already available and the full corpus consists of audio recordings, dialect text transcriptions, and translations into Standard Japanese. Transcriptions are presented in *katakana* script, with translations in Standard Japanese provided in a combination of *kana* and *kanji* scripts. These transcriptions are currently available for public download and search (for academic use, provided a free account is set up) via the COJADS corpus or the accompanying website where the raw CSV data files are published.<sup>4</sup> The data used for the analysis presented in Chapter 3 was downloaded in the form of raw CSV data files from the accompanying COJADS website, after a free user account was set up for the full COJADS corpus.

The raw CSV files include extensive metadata, such as speaker age at time of recording, gender, birth year, topic of conversation and location of conversation. Speakers are kept anonymous, as are any references made to other people in each conversation. The NINJAL researchers also marked where particles were missing in the speech by marking particle omission using brackets in the Standard translation text (Koiso et al. 2019), and these annotations were vital to this thesis project. Further detail on the data used in this thesis is provided in Chapter 3.

### 2.2.3 *The Four Dialects*

The analysis of the COJADS data presented in Chapter 3 focuses on four dialects (named by their prefecture throughout this thesis): Kitakyushu Dialect (Fukuoka Prefecture), Osaka

---

<sup>4</sup> <https://www2.ninjal.ac.jp/cojads/index.html?targ=data>

Dialect/Prefecture, Tokyo Dialect/Prefecture, and Hokkaido Dialect/Prefecture. The locations of each are roughly marked with red arrows in Figure 2. These four dialects were chosen to represent each major dialect grouping of Japanese (Eastern and Western).

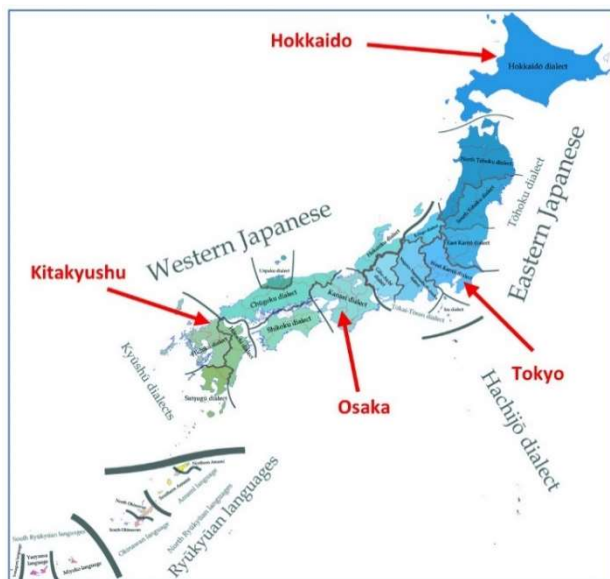


Figure 2. The Four Dialects

The Kitakyushu dialect is spoken on the southern island of Kyushu in the northern portion of Fukuoka Prefecture and belongs to the Western dialect grouping as well as to the historic Hichiku dialect family (Onishi 2017). The Osaka dialect is part of the Kansai dialect and is spoken in the Kansai region (also referred to as the Kinki region, for instance in Onishi 2017) of the main island. It represents the eastern-most portion of the Western dialect grouping.

The Tokyo dialect is spoken in Tokyo Prefecture in the Kanto Region and most closely resembles Standard Japanese; in fact, Standard Japanese was based originally on the Edo (Tokyo) Dialect (Onishi 2017; Shibatani 1990). The Tokyo dialect is also representative of the Kanto dialect, spoken in the central area of the island of Honshu, and belongs to the Eastern dialect

grouping. The Hokkaido dialect, spoken on the northern-most of the three main islands, represents the Northern-most dialect and the Eastern-most dialect of the Eastern dialect grouping.

### 2.3 Formality & Sociolinguistic Factors

Japanese is known for its nuanced formality and honorifics system, which encodes many styles and levels of speech (Tsujimura 2014). Nominals can be marked with an honorific prefix (*go-* or *o-*) and verbal elements can be marked with referent or addressee honorifics (via root changes and suffixes), and an even more nuanced form of referent honorifics is observed (respectful and humble forms) (Tsujimura 2014). In addition to honorification of verbs, there are many ways of referring to a person that denote their relationship to the speaker (in age, closeness of relationship, and so forth), for example by using suffixes that occur after a person's name (e.g. *-san* 'Mr./Mrs.')

or by referring to a person by their given name (less formal) or family name (more formal). In addition to honorifics, the language also exhibits varying formality levels, typically referred to as *-desu/-masu* form (formal speech) and "plain form" (informal speech). Finally, a nominal can be turned into an honorific form by adding the prefix *o-* or *go-*; *o-* is attached to native Japanese words, and *go-* is prefixed to Sino-Japanese words (Tsujimura 2014).

As Tsujimura (2014) explains, while the expectation is that formal speech forms are used in more formal situations (such as when speaking to a stranger who is older, or when assisting a customer/client) and informal speech forms are used in informal settings (such as when speaking with close friends or family), this is not always true. It is true that the age and relationship of the speaker, addressee and sometimes even audience matters greatly, and that social status and gender composition of the interlocutors can also play into formality choices; however, as Tsujimura (2014) explains, this does not capture nuanced interactions. Instead, Tsujimura (2014: 414) argues that "the decision by a speaker to use a particular form of honorification [or formality level] is not

so easily reduced to an ideological norm, but is controlled by the interaction of numerous factors, of which linguistic, social and cultural ones are just a few.”

In addition to the variation observed with formality and style, Japanese also exhibits gender differences in speech. Tsujimura (2014) lists a few of these, namely different personal pronouns, differences in use of sentence-final particles, and possibly also differences in slang and a hypercorrect use of honorific prefixes (*o/go-*). However, as with Tsujimura’s (2014) conclusions regarding honorific and formality levels being overly ideological, it is likely that these generalizations regarding gender variation in Japanese speech are also ideological. These generalizations also appear to be grounded in much older concepts of language and gender, such as that of Trudgill (1972); the reality is likely much more nuanced than this, as suggested by Tsujimura (2014) and demonstrated by studies such as Takano 1998 and Did-Ogren 2020. This is what we would expect, based on recent developments in language and gender studies, which all point towards a much more complex picture of the interaction between language and gender (cf. Eckert & McConnell-Ginet 1992, Levon 2014).

## 2.4 Nominal Particles

Japanese nominal arguments are marked with post-nominal particles which occur immediately following a noun. Several nominal particles (*-wa*, *-ga*, and *-ni*) can be seen in (2), marked in bold font.

(2) Tarou=**wa** neko=**ga** daisuki dakara, itsumo neko kafe=**ni** iku.

Tarou=TOP cat=NOM love because always cat café=GOAL go

‘Because Tarou loves cats, he always goes to the cat café.’

*(Lit. ‘Because cats are loved by Tarou, he always goes to the cat café.’)*

The nominative *-ga* and accusative *-o* are often considered to be Case markers and are typically referred to as Case particles. Other particles (such as *-de* ‘in/at’ and *-e* ‘to’) are often referred to as Postpositions or Postpositional particles, rather than Case markers. The particle *-ni* is often glossed as either a DAT Case particle or simply ‘to’, but it patterns more closely with *-de* and *-e* which are considered Postpositions, and which are typically glossed with the English prepositions ‘at/by’ and ‘to’ (respectively). The Topic Marker *-wa* is typically considered a discourse marker, but sometimes patterns with the Case particles *-ga* and *-o*, and sometimes with the discourse marker *-mo* ‘too’ which can occur with or to replace nominal particles (see Section 4.3.4 for a brief discussion). In this thesis, I will refer to all particles occurring at the end of a nominal constituent (both Case and Postpositional particles) as “nominal particles.” I will also use the terms “Case particle” and “Postpositional particle” when it is convenient to refer to each group of particles. However, use of these two terms is for convenience and exposition purposes only; while it is clear there is some semantic distinction between the two groups, there may not in fact be a syntactic distinction (see Chapter 4 for a discussion of this).

Nominal particles are often treated as suffixes. However, it could be argued that they function as clitics as suggested by Nakamura (2018). Nakamura explained that although nominal particles are bound phonologically to the preceding word, in some instances an element can occur between the nominal and particle. For example, *dake* ‘only’ can intervene between the noun and particle as seen in Example 3.

- (3) Tarou=to Hanako-dake=ga eki=kara arui-ta.  
 Tarou=CONJ Hanako-only=NOM train.station=ABL walk-PST  
 ‘Only Taro and Hanako walked from the train station.’  
*(Adapted from Nakamura 2018, page 249, example 1.)*

Nakamura (2018) also explains that the particle’s scope can include not only the nominal it immediately attaches to, but also any coordinated NPs (as in Examples 4-5). In addition, nominal particles must always occur at the end of the nominal phrase which can include relative clauses and take scope over the whole constituent, not just the noun itself (as in Examples 6-7). Based on the arguments presented by Nakamura (2018), I will assume in this thesis that particles are better considered clitics than true nominal suffixes.

- (4) mukashimukashi aru tokoro=ni [ojisan=to obaasan]=ga sundeimashita.  
 old.days certain place=LOC old.man=CONJ old.woman=NOM were.living  
 ‘A long, long time ago, there lived an old man and an old woman.’  
*(Adapted from Nakamura 2018, page 282, example 16; brackets added.)*

- (5) watashi=wa kesa [gakkou=to toshokan]=ni ikimashita.  
 I=TOP this.morning [school=CONJ library]=GOAL went  
 ‘I went to school and to the library this morning.’  
*(Example, brackets, and judgement are my own.)*

(6) [watashi=ga kinou mita] eiga=wa omoshirokunakatta.

I=NOM yesterday saw movie=TOP was.not.interesting

‘The movie I saw yesterday was not interesting.’

*(Adapted from Nakamura 2018, page 288, example 28; brackets added.)*

(7) [[sono hon=o yonda] hito]=wa Tarou desu.

that book=ACC read.PST person=TOP Tarou is

‘The person who read that book is Taro.’

*(Adapted from Fukuda 1993, page 169, example 2b.)<sup>5</sup>*

In this thesis, only the particles *-ga*, *-o*, *-ni*, *-de*, *-e*, *-no*, and *-wa* will be examined; other particles, such as *-kara* (ablative) and *-made* (terminative), will not be examined because of the limited scope of this thesis. Table 1 presents each particle to be examined in Standard Japanese, followed by the phonological realization in each of the four dialects being examined. The information presented in Table 1 in the four dialect columns is from the COJADS corpus data examined in Chapter 3; note that this information is based on my own observations of the transcriptions provided in the dataset and has not been verified in the literature. As we can see from Table 1, in most dialects the realization of the particles is identical to that of Standard Japanese; however, the first consonant of the NOM particle is devoiced in the Hokkaido and Tokyo dialects.

---

<sup>5</sup> The brackets around the NP are Fukuda’s; I added the topic-marker, predicate, and full sentence translation.

Table 1. Particle Realization per Dialect <sup>6</sup>

	Standard	Hokkaido	Tokyo	Osaka	Kitakyushu
<b>NOM</b>	-ga (が)	-ka (カ)	-ka (カ)	-ga (ガ)	-ga (ガ)
<b>ACC</b>	-o (を)	-o (オ)	-o (オ)	-o (オ)	-o (オ)
<b>DAT / 'to'</b>	-ni (に)	-ni (ニ)	-ni (ニ)	-ni (ニ)	-ni (ニ)
<b>'to'</b>	-e (へ)	-e (エ)	-e (エ)	-e (エ)	N/A
<b>LOC / INS</b>	-de (で)	-de (デ)	-de (デ)	-de (デ)	-de (デ)
<b>GEN</b>	-no (の)	-no (ノ)	-no (ノ)	-no (ノ)	-no (ノ)
<b>TOP</b>	-wa (は)	-wa (ワ)	-wa (ワ)	-wa (ワ)	-wa (ワ)

It should also be noted that phonological reduction of some particles was observed in the Tokyo, Osaka and Kitakyushu dialects in certain environments; this is not represented in the table due to space constraints. In the Tokyo dialect, the particle *-ni* is reduced to *-n* when the next syllable begins with [n]. A word-final [e] sometimes appears to lengthen in the Tokyo dialect before the particles *-e* and *-de* which appear to be omitted; this could indicate either that the final [e] is lengthened and is followed by an omitted, or it could be that the particle merges with the preceding word (in the case of *-de* this would involve a loss of the initial [d]). In the Osaka dialect, the only observed phonological reduction was that of *sore-wa* ‘that-TOP’ becoming *sora*. In the Kitakyushu dialect, the particle *-wa* was observed to cause the final vowel to lengthen, apparently by merging with the final syllable of the noun it attaches to. For example, *kengyounouka-wa* ‘part.time.farmer-NOM’ became *kengyounoukaa* (Kyushu 40\_a\_002-1.csv, lines 122 and 125). Note that there were no tokens for *-e* in the data for Kitakyushu. It is possible that this allomorphy could support claims

<sup>6</sup> Note that although the standard method of writing particles in Japanese is to use *hiragana* script, the *katakana* script is used in Table 1 for particles in the dialect columns to match how the particles appeared in the original COJADS data. As previously mentioned, in the original files the *katakana* script was used for the dialect transcriptions, while a mix of *kana* and *kanji* was used for the Standard Japanese translation column.

that the particles are suffixes and not clitics; however, Nakamura 2018 (who treats the particles as clitics, as mentioned above) does mention that the particles are phonologically bound to the preceding word and yet still favors the clitic argument based on syntactic evidence. While more study is needed into this topic, it is beyond the scope of my thesis, and I will follow Nakamura 2018 in considering the particles as clitics.

Case marking in Japanese can display other patterns, where a particular particle can be used to express another grammatical relation. For example, in some cases *-ga* or *-ni* are allowed to mark a subject (Nakamura 2018). However, this is outside the scope of this thesis, and for the purposes of this study each particle observed in the data will be treated according to the information in Table 1 above. Further research is needed, in consultation with native speakers of each dialect, in order to further annotate the data (presented in Chapter 3) for Case and examine any patterns of case marking that may differ from the standard.

## 2.5 Particle Omission

Nominal particles are frequently omitted in spoken language (Nariyama 2003; Shibatani 1990; Shimojo 2006; Takano 1998), and there is some evidence that the frequency and nature of particle omission varies between dialects (cf. Kibe et al. 2020). Claims have been made that only Case particles can be omitted, but not Postpositional particles (cf. Nakamura 2018, Tsujimura 2014); this is illustrated by Examples 8-9.

(8) a. kono ringo=**o** tabe-ta?

this apple=ACC eat-pst

‘Did you eat this apple?’

b. kono ringo=# tabe-ta

this apple=# eat-pst

‘Did you eat this apple?’

*(Adapted from Tsujimura 2014, page 136, Example 28b.)*

(9) a. Tarou=ga toshyokan=**de** hon=o yondeiru.

Tarou=NOM library=LOC book=ACC is.reading

‘Tarou is reading a book at the library.’

b. \*Tarou=ga toshyokan=# hon=o yondeiru.

Tarou=NOM library=# book=ACC is.reading

Intended: ‘Tarou is reading a book at the library.’

*(Adapted from Tsujimura 2014, page 136, Example 29a.)*

However, Takano (1998) argued that the topic and nominative markers are commonly omitted, and it is clear that other particles can also be omitted (cf. Kibe et al. 2020, Shimojo 2006, Fry 2003). Nariyama (2003) confirms that any particle, including the topic marker *-wa* and the Postpositional particles, can be omitted.

Examples 10-12 further confirm that other particles can be omitted in colloquial speech. Example 10 illustrates omission of the ACC particle *-o* (commonly considered a “Case particle”) in the Hokkaido dialect, while examples 11-12 present examples of “Postpositional particle” omission in the Osaka and Kyushu dialects. As these examples are from the COJADS data analyzed in Chapter 4, corresponding translations in Standard Japanese are provided in the (b) example for each. Example 1 in Section 1.1 can be referred to for examples of omission of the nominative *-ga* and topic *-wa* particles.

(10) a. Hokkaido dialect text:

oshyookatsu=[o] mukaen=no [tame=no] junbi=ka saa.

New-Year=[ACC] approach=NOMINALIZER [sake=GEN] preparations=NOM right

”It’s the preparations for the (sake of) the approaching New Year, right?”<sup>7</sup>

b. Standard Japanese translation text:

oshoogatsu=[o] mukaeru tame=no junbi ga ne .

New Year=[ACC] approach sake=GEN preparations NOM right?

“It’s the preparations for the (sake of) the approaching New Year, right?”

*(Adapted from COJADS (NINJAL 2021), Hokkaido file 01\_b\_099.csv, line 17)*<sup>8</sup>

(11) a. Osaka dialect text:

soto=[ni] deru koto=[ga] arimasennen.

outside=[GOAL] exit nominalizer=[NOM] do.not.have

‘(I) don’t go outside.’

b. Standard Japanese translation text:

soto=[ni] deru koto=[ga] naindesu.

outside=[GOAL] exit nominalizer=[NOM] do.not.have

‘(I) don’t go outside.’

*(Adapted from COJADS (NINJAL 2021), Osaka file 27\_b\_099.csv , line 870.)*

---

<sup>7</sup> I added *tame=no* ‘sake=GEN’ in brackets to illustrate where this phrase would appear in Standard Japanese; this is further illustrated in Example 10b.

<sup>8</sup> Transcription into Roman characters, annotation, and translation added for examples 10-12. Note that brackets are used to indicate the missing particle to illustrate how this appears in the corpus; this is how the missing particles were marked in the Standard Translation text which can be seen in the (b) examples in 10-12.

(12) a. Kitakyushu dialect text:

uchi=[no] gata madaa

1sg=GEN house still

‘My house is still ...’ (*intended predicate not clear*)

b. Standard Japanese translation text:

watashi=[no] ie mada

1sg=[GEN] house still

‘My house is still ...’ (*intended predicate not clear*)

(Adapted from COJADS (NINJAL 2021), Kitakyushu 40\_a\_002-1.csv, line 407)

### 2.5.1 Phrase Structure of Nominal Particles

One of the more traditional syntactic analyses of particles (as described in Tsujimura 2014) is that particles are either assumed to be case-marking suffixes, affixed directly to the N, or they head a PP. In Figure 4, we can see a particle heading a PP (Postpositional Phrase), while in Figure 3 we can see a particle simply suffixed to the N.<sup>9</sup>

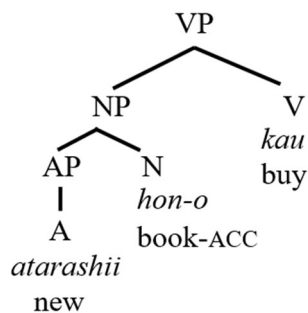


Figure 3. Structure of NP with Particle Suffixed to N

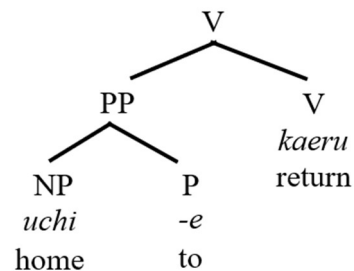


Figure 4. Structure of NP with Postpositional Phrase

<sup>9</sup> Figures 3 and 4 are adapted from Tsujimura 2014, page 232, example 7.

This analysis presents two problems. First, by attaching the Case Particle *-o* ACC in Figure 3 directly to N as a suffix, we lose the intuition that the particles have scope over the entire nominal phrase (which could include a relative clause). The treatment of Postpositional particles as P in Figure 4 appears to allow this more readily. However, by positing the two particle types as having separate syntactic structures, we lose the idea that all nominal particles encode the grammatical relationship of the entire nominal constituent (as described by Nakamura 2018 and presented in Examples 3-7 in Section 2.4); it is not just the so-called Postpositional particles that have scope over the entire nominal phrase.

This treatment also seems to simply parallel the structure of English NPs and PPs. Additionally, how can we elide a Case suffix that attaches directly to N, yet not elide the Postpositional particle which is described as head P in Figure 4 (see 2.4 for more details)? In both DPs and PPs in English, for example, the head D (e.g. *the*) or the head P (e.g. *to*) could (typically) not be elided, leaving the argument NP/DP behind; English allows preposition stranding, where the DP is moved but not elided, and the preposition itself certainly is not. For the DP, English does not allow the D to be elided under most circumstances (it can, for instance, be elided in newsprint headlines). Finally, both structures fail to capture the fact that particles have scope over the nominal constituent as explained by Nakamura (2018).

Fukuda (1993) proposed that treating Case particles (which he took to be only *-ga* and *-o*) as head K in a “Kase Phrase” (KP) could solve this problem. He argued that if Case particles are located in K, they can elide under the ECP (Empty Category Principle). Specifically, with Case particles heading KP as in Example 13, a K inside VP is properly head-governed by V allowing an empty category (omitted particle) in K (as in Example 14). He then argues that the presence of sentence-final particles allows subject particles to drop because they fill C and properly govern the

subject-position K. This can be seen in (15a) with the emphatic sentence-final particle *yo* or *zo* and in (15b) with the sentence-final question particle *no*; in both sentences, the sentence-final particle (*yo*, *zo*, or *no*) occurs in C and governs the subject as illustrated in (15c). Although sentence-final particle use does not seem to correlate with nominal particle omission (see Chapter 3), Fukuda's argument seems fairly sound in the Government & Binding framework (Chomsky 1981) at least as far as the accusative (non-subject) K goes.

(13) [<sub>KP</sub> [<sub>K'</sub> NP [<sub>K</sub> *ga/o*]]]

*(From Fukuda 1993, page 170.)*

(14) a. *jon=ga sono hon=# yon-da*

John=NOM that book=# read-PST

'John read that book.'

*(Adapted from Fukuda 1993, page 169, example 1)*

b. [<sub>IP</sub> [<sub>KP</sub> [<sub>K'</sub> *jon* [<sub>K</sub> *ga*]]] [<sub>I'</sub> [<sub>VP</sub> [<sub>KP</sub> [<sub>K'</sub> *sono hon* [<sub>K</sub> #]]] [<sub>V</sub> *yon*]] [<sub>I</sub> *da*]]]

*(Adapted from Fukuda 1993, page 170, example 8.)*

(15) a. *jon=# sono hon=o yon-da yo/zo.*

John=# that book=ACC read-PST EMPH

'John read that book.'

*(Adapted from Fukuda 1993, page 169, example 4.)*

b. *jon=# sono hon=o yon-da no?*

John=# that book=ACC read-PST Q

'Did John read that book?'

*(Adapted from Fukuda 1993, page 169, example 3.)*

- c. [CP [IP [KP [K' *jon* [K #]]] [I' [VP [KP [K' *sono hon* [K -o]]] [v *yon*]] [I *da*]] [C *yo/zo/no*]]]  
*(Adapted from Fukuda 1993, page 171, example 9.)*

Although Fukuda’s argument about how the particles can elide is no longer viable under the Minimalist program (Chomsky 1995, 2000, 2001, 2013, etc.), he did demonstrate that particles can be treated syntactically (as opposed to simply being counted as suffixes and therefore handled primarily by morphology). Fukuda’s KP structure (as shown in Example 13 and illustrated in Figure 5) does provide a potential alternative structure that could still be functional in the Minimalist program.

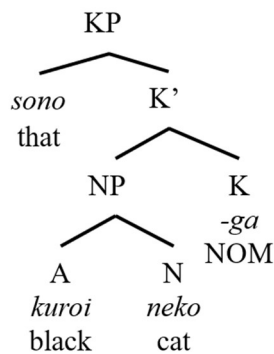


Figure 5. Illustration of Fukuda’s (1993) KP Structure<sup>10</sup>

An alternative could be to combine the KP and DP, as seen in Figure 6. This would capture possible similarities of demonstratives in Japanese to determiners of other languages, while still ensuring that the particles (here positioned in K) have scope over the nominal phrase.

<sup>10</sup> The example text “that black cat” in Figure 5 is my own; the structure presented is adapted from the basic KP structure suggested by Fukuda 1993 on page 170, presented here in Example 4. Note that in Figure 5, the demonstrative is tentatively placed into Spec K, but no category is listed. As briefly discussed in relation to Figure 7, demonstratives could be types of noun phrases which also have their own particle (cf. Shibatani 1990, Hoji et al. 2003); it is for this reason I place them tentatively in Spec K in Figure 5. However, I do not give them a category in the KP tree as that option is not the one I proceed to use in my analysis and I am making no claims about the location of nominal modifiers in the KP structure presented in Figure 6.

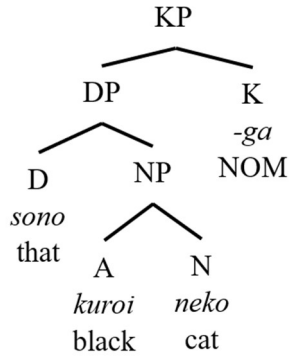


Figure 6. Combined KP and DP Option

The third possible alternative is to simply treat the particles as head D of a DP, as shown in Figure 7. Japanese does not have determiners (such as ‘a’ or ‘the’ in English) (cf. Nishiguchi 2009, Shibatani 1990), and therefore this slot appears to be open; a possibility is to consider the particles as D.

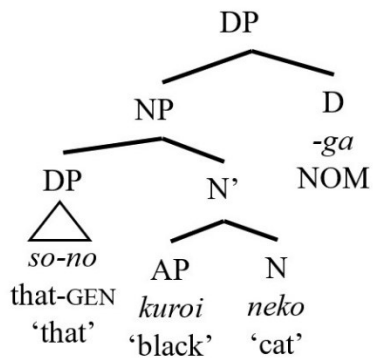


Figure 7. DP Structure, with Particles in D

While the structure in Figure 7 looks quite similar to the KP structure illustrated in Figure 5, it has some advantages over the KP structure (Fig. 5) and the combined KP-DP structure (Figure 6). First, the combined DP and KP option (Fig. 6), where the demonstrative *sono* ‘that’ is treated

as D, results in a left-headed DP. Japanese is a right-headed language, and this would be (to my knowledge) the only left-headed phrase.

In addition, the demonstratives of Japanese do not always act in a similar manner to demonstratives in other languages such as English (Shibatani 1990). In particular, both Hoji et al. 2003 and Shibatani 1990 treat the demonstratives, as well as quantifiers and classifiers, as nominal modifiers. They treat the deictic forms (e.g. *sono* ‘that’ and *kono* ‘this’) as locative nouns *so* and *ko* followed by the genitive particle *-no*. This parallels the other demonstratives which act similarly to nouns in that they can take a particle (such as *sore* ‘that thing’, *kore* ‘this thing’, *soko* ‘that place’, *koko* ‘this place’, etc.). Shibatani (1990:258) also mentions that “these nominal modifiers, including the ones involving numerals (and classifiers), do not terminate the expansion of the noun phrase...” Example 13 illustrates Shibatani’s (1990) point that the demonstratives, such as *sono* ‘that’, do not block subsequent modification; here we can see the adjective *ookina* ‘big’ preceding the demonstrative and the quantifier *issatsu* ‘one (flat object).’ This example also makes clear the parallel between demonstratives and quantifiers; they can both take the genitive *-no*. Examples 14-17 are adapted from Hoji et al. 2003, and further demonstrate that the demonstratives (following the treatment by Hoji et al. 2003) can be considered nominals which take a particle.

(13) *ookina so=no issatsu=no Tarou=no hon*  
big that=GEN one.CL=GEN Tarou=GEN book  
‘That big book of Tarou’s’

*(Adapted from Shibatani 1990, page 258)*

(14) [Pointing to someone standing 10 meters away.]

a=no hito=wa amerikajin desu.

that-GEN person=TOP American is

‘That person is an American.’

*(Adapted from Hoji et al. 2003, page 106, example 37a.)<sup>11</sup>*

(15) [Pointing to someone the speaker has his/her arm around.]

ko=no hito=wa amerikajin desu

this=GEN person=TOP American is

‘This person is an American.’

*(Adapted from Hoji et al. 2003, page 106, example 37b.)*

(16) do=no jidoushya-gaishya=ga [so-ko=no ko-gaishya]=o

which=GEN automobile-company=NOM [**that-place=GEN** child-company]=ACC

suisenshita no?

recommended Q

‘Which automobile company recommended [its subsidiary]?’

*(Adapted from Hoji et al. 2003, page 100, example 15c. Brackets were in original.)*

(17) [hon=o hiraita hito]=wa minna so-re=o ka-wanakerebanaranai.

book=ACC opened person=TOP everyone **that-thing=ACC** buy-must

‘Everyone who has opened a book must buy it.’

*(Adapted from Hoji et al. 2003, page 101, example 15d. Brackets were in original.)*

---

<sup>11</sup> Note that in the original example by Hoji et al. 2003, the demonstrative was shown as a-no ‘that-GEN’; I have substituted the dashes for the equals sign following my assumption that the nominal particles in Japanese are clitics. This is the same for Examples 14-17 which were also adapted from Hoji et al. 2003.

It therefore follows that demonstratives (as well as quantifiers) should not be placed in D head of DP in Japanese and instead it may be more suitable to place them in a Specifier position, either Spec NP, as illustrated in Figure 7, or Spec DP (as discussed in Chapter 4 in relation to possessive phrases). They likely also head their own DP, as they are also made up of an NP and particle (D). This idea is further supported by Shibatani's (1990:258) statement that "...Japanese has no functional category of determiners comparable to the English determiner system, which, consisting of demonstratives articles, and genitives, terminates the expansion of a noun phrase." Based on Shibatani's analysis of the Japanese demonstratives and quantifiers, as well as the fact that Japanese does not have articles, it is unlikely that any of these categories, which typically occur in D in English, would fall under D in Japanese. The exact placement of these modifiers will be discussed in Chapter 4, Section 4.4.1, along with the placement of possessives.

Finally, genitives in some languages, such as the English genitive clitic 's, are spelled out in the D slot attaching to the possessor in Spec DP when D bears a genitive Case feature. It follows that other Case markers could occur in D, especially ones that function as clitics more so than suffixes. For these reasons, I will assume in this thesis that nominal particles head D in a DP as seen in Figure 7. While this does not address the problem of how particles, as head of a phrase, can elide, it does appear to better account for the scope they have over the whole DP (as seen in Examples 4-7).

Further investigation is needed to explain the particle omission phenomenon syntactically. In particular the following questions should be addressed: Should the same structure be used for all particles, or should Postpositional particles and Case particles be treated differently? Can one of these options better account for the ability for the discourse particles to stand in for, or occur with, certain particles? These questions will be discussed in Chapter 4.

### 2.5.2 *Zero Particles vs. Particle Ellipsis?*

A question, which has been addressed in the literature from many perspectives, arises regarding particle omission; are the missing particles in fact null “pro-form” particles (“zero particles”), or are they a case of ellipsis (“particle ellipsis”)? This section provides a brief account of the two main proposals that have been made in answer to this question, as well as a brief discussion of some possible evidence in support of each proposal.

In his quantitative study, Takano (1998) treated the phenomenon as ellipsis, and focused on sociolinguistic variation in the occurrence of particle ellipsis of *-wa* and *-ga*. Although his focus was on sociolinguistic variation, particularly on testing claims made by Shibamoto (1985) regarding particle ellipsis being a feature of women’s speech, Takano made the assumption that the phenomenon is ellipsis based on prior studies, including one by Shibatani (1990), who treated it as particle drop or ellipsis but did not go into great detail on why it was considered ellipsis. Fry (2003: 96) also treats it as ellipsis, stating that in cases where the particle is missing, “the meaning of the sentence is nonetheless clear from context, as is typically the case with ellipsis in general.”

Lee (2002a) explains that some examples can be explained more readily in terms of a “zero particle”, as opposed to ellipsis. In some cases, although several particles can optionally appear with a given nominal, a different nuance is obtained depending on which particle is used, including the zero particle. According to Lee (2002a), examples 18-20 (adapted from Lee 2002a: 646 examples 2-3) include the same lexical content, but exhibit different nuances which can only be explained by the different particles used. Lee (2002a: 646) explains that (18) “would typically be used in a situation where the speaker is upset for some reason and declares that he/she will not eat dinner.” Example 19, according to Lee 2002a, is a statement that means the speaker does not want

to eat; the scenario he provides is that the speaker has been sick, and is not eating these days; it is just a fact. Finally, example 20 implies that the speaker doesn't eat dinner, but may eat other meals.

(18) gohan=# tabe-nai!

meal=# eat-NEG

'I won't eat dinner!'

(19) gohan=0 tabe-nai!

meal=ACC eat-NEG

'I will not eat dinner'

(20) gohan=**wa** tabe-nai!

meal=TOP eat-NEG

'I do not eat dinner' (lit. 'As for dinner, I do not eat it.')

*(Examples 18-20 adapted from Lee 2002a, page 646, examples 2-3.)*

Lee (2002a) argues that examples such as these point to the need for a zero particle with its own grammatical role/meaning to capture this effect. Shimojo (2005) also includes the zero particle as one of the argument encoding options in Japanese. Some additional examples of the zero particle are provided in Examples 21-23.

(21) terebi=# miterassharu n desu ne

TV=# is.watching NOMINALIZER is FP

'He is watching TV.'<sup>12</sup>

---

<sup>12</sup> The examples and judgements in (21) are my own.

(22) hon=# ippai aru tokoro=ni ikitai

Book=# lots exist place=GOAL want.to.go

‘I want to go to the place which has lots of books.’

*(Examples 21-22 adapted from Lee 2002a, page 649, example 6.)*

(23) soko=# boku=no hon aru?

there=# I=GEN book exist

‘Is my book there?’

*(Example 23 adapted from Lee 2002a, page 650, example 7.)*

However, the zero particle analysis does not seem to satisfactorily account for all instances of particle omission; in many cases, it is clear which particle should have been pronounced. For example, in (21b) it is clear that *-o* (ACC) was omitted, and no different nuance is obtained in (21a) versus (21b). It is possible to substitute the particle *-o* with the particle *-mo* or *-wa*, but as these are discourse particles these do change the nuance (as observed with example 20 versus 18 and 19).

(21) a. ken=ga shukudai=**o** shi-ta

Ken=NOM homework=ACC do-PST

‘Ken did his homework.’

b. ken=ga shukudai=# shi-ta

Ken=NOM homework=# do-PST

‘Ken did his homework.’

c. \*ken=ga shukudai=**ni** shi-ta

Ken=NOM homework=**GOAL** do-PST

Intended: ‘Ken did his homework.’

While the zero particle analysis does not appear to fully capture the phenomenon, it also does not seem appropriate to dub every instance of a missing particle “particle ellipsis” when the phenomenon does not always appear to be true ellipsis. However, Fry (2003) does describe the meaning of the sentence as clear even without the overt particles. Shimojo (2005) briefly mentions that several studies allowed for both types of omitted particles – ones that are not obligatory (can be replaced by an overtly pronounced particle) and ones that are (cannot be replaced by another particle; only the zero particle can occur without a change in meaning) – however, they are still both referred to as “zero particle” by Shimojo (2005, 2006). It is beyond the scope of this thesis to take a stance on whether both types of omission exist; it is for this reason that, as previously mentioned, I will refer to all instances where a particle is expected but where no overt particle exists as “particle omission.” In some cases I will refer to each of the two possible treatments separately (particle ellipsis and zero particle), but only for illustrative purposes; I am not taking a position on this matter due to the limited scope of this thesis.

The problem has also been analyzed in terms of semantics rather than syntax; for instance, in Lee 2002a and Shimojo 2006. Still others have attempted a more holistic analysis, such as is attempted here; these include Fry’s (2003) and Takano’s (1998) analyses. Both use corpus work to explain the phenomenon.

#### 2.5.4 Variation in Particle Omission

As Podesva (2007) explains, people speak languages in social and situational contexts in everyday life, and not in laboratory settings – it follows that a complex phenomenon such as particle omission, would be affected by sociolinguistic factors. In fact, particle omission has been argued to vary based on numerous sociolinguistic and pragmatic factors, including dialect, gender, and situational context.

Takano (1998), looking at a small corpus of natural conversation data, concluded that, while linguistic constraints on particle omission are stronger than sociolinguistic and functional considerations, sociolinguistic factors (in particular gender composition) do matter. He found that examining intraspeaker variation is important. Fry's (2003) data was more extensive (involving a corpus of phone calls), but one of his conclusions was that a more holistic perspective is needed in order to ensure the whole picture of particle omission is captured. Shimojo (2006) and Lee (2002a) focused on investigating the phenomenon from a pragmatics point of view, and demonstrate that context should be considered when investigating particle omission.

Particle omission has also been observed to vary based on dialect, although there are fewer studies on dialectal variation than particle omission in Standard Japanese. One very recent study (Kibe et al. 2020) was conducted on COJADS data (looking at a different set of dialects, and a much larger dataset than is used in this thesis), and found that omission of the ACC particle *-o* varied by region (dialect). In addition to attested dialectal variation, claims have been made that particle omission varies by speaker gender (cf. Shibamoto 1985). It has also been suggested that grammatical elements such as sentence-final particle use impacts omission of the nominative particle *-ga* (see for example Fukuda 1993, who argued that the presence of sentence-final particles

enabled omission of particles marking subject NPs). However, there are arguments against this, such as in Lee 2002b.

Based on these prior studies, it seems likely that particle omission is affected by a great variety of factors, and I hope to examine some of them in this thesis. Although examining every possible factor is of course beyond the scope of this thesis, I will examine (in Chapter 3) grammatical, sociolinguistic, and, to some extent pragmatic, factors; in particular, the impact of dialect, speaker age and gender, sentence-final particle use, and honorific prefix use on the frequency of particle omission. I hope to shed light on the phenomenon in the context of dialectal variation in particular while revisiting some factors that have been previously argued to matter for particle omission.

## **2.6 Labeling, Merge and Phases**

Three mechanisms of the Minimalist framework (Chomsky 1995, 2000, 2001, 2013, etc.) are important for the syntactic analysis presented in Chapter 4; Merge, the Labeling Algorithm, and Phases. These will be briefly discussed in this section.

### *2.6.1 Merge & The Labeling Algorithm*

In the Minimalist framework (Chomsky 1995, 2000, 2001, 2013, etc.), Merge is the most basic operation for building syntactic structure, and according to Hornstein (2009):, “all grammatical relations are grammatically executed under Merge”. Merge occurs when two distinct syntactic elements are combined to make a new syntactic element. Collins & Stabler 2016 provides the definition of Merge in terms of the notion of a “syntactic object” which is either a lexical item or a set of syntactic objects (e.g. phrases). These syntactic objects are the building blocks of syntactic structure under Merge. An additional important notion is that by the definition of Merge as an operation that combines two syntactic objects (which consist of lexical items), only what is

necessary should be merged into the derivation. For instance, there should not be empty nodes merged into the derivation, because only lexical items (or semantically meaningful pieces of morphology) and syntactic objects can be merged.

The Labeling Algorithm as described by Chomsky (2013) explains how phrases are labeled after Merge occurs. When two syntactic objects (phrases or heads) undergo Merge, the Labeling Algorithm explains how the resulting phrase is labeled. For instance, if a phrase (XP) and a head (Y) are combined, the head provides the label (resulting in the new phrase YP). Problems arise when the two elements being merged are of the same type (e.g. {XP, YP} or {X, Y}), and Chomsky 2013 provides some solutions for these; however, only the basic notion of Labeling as described here (where a phrase and lexical head are merged) is important for the syntactic analysis presented in Chapter 4.

### 2.6.2 Phases

In the Minimalist framework, phases are syntactic domains of a derivation which are completed before higher parts of the derivation (Chomsky 2000, 2001). Phases are typically considered semantically “saturated” domains, meaning that they contain all the necessary semantic elements (cf. Lee-Schoenfeld 2008); as such vP and CP are typically considered phases (Chomsky 2000, 2001). Phases explain how syntactic derivations are transferred to the interfaces; as a phase is completed, its complement is transferred to the interfaces, making that chunk inaccessible to future parts of the derivation (Chomsky 2000; Chomsky 2001;

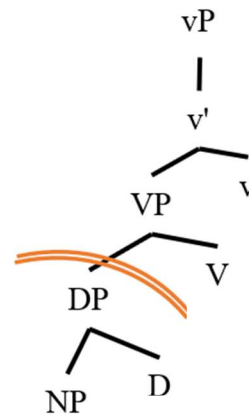


Figure 8. Illustration of a DP Phase

Sakamoto 2016). For example (following the explanation provided by Sakamoto 2016), if vP is a

phase then when it is completed, the complement (VP) is transferred but not the whole phase including vP. This leaves the head v behind and v is therefore available to ‘search’ in future parts of the derivation (for example it is then able to move to T to get Tense features). Figure 8 illustrates this with respect to a DP phase; in Figure 8 the double curved line illustrates the phase boundary of a DP phase. When the phase is complete, NP (the complement) would transfer to the interfaces, but the head D would be left behind. Because the topmost elements (the specifier(s) and head) of a phase are not transferred immediately to the interfaces (VP in the case of a vP phase or NP in the case of a DP phase), this leaves them available for interactions with the next higher phase. The complement of the phase-head, which is transferred to the interfaces, becomes inaccessible to higher operations of the derivation, which reduces the search space, as search can then only go as far as the remaining head at the phase edge. A head located at the phase edge is therefore accessible to the next higher phase, while elements lower in the derivation are not.

Under the proposal made by Fukuda (1993), particles (which he argued headed KP) could elide when properly head-governed under the Empty Category Principal (ECP) (see Section 2.5.1). Although the ECP is no longer part of the modern Minimalist framework (Chomsky 1995, 2000, 2001, 2013, etc.), an extension of the ECP to Minimalist terms is that a non-pronominal empty category (e.g. an omitted particle) must be properly licensed in its Phase, or as part of the next higher phase if it is at the phase edge. Phases also define binding and locality, and explain movement and anaphoric relations according to Lee-Schoenfeld (2008). They can also explain Japanese argument ellipsis (cf. Sakamoto 2016).

Although Chomsky (2000, 2001) considers CP and vP to be phases, other phrases have also been argued to be phases. In particular, DP is considered to be a phase by Svenonius (2004) and McCloskey (2000) also mentions the possibility that DPs could be phases in a footnote. If we

assume that DPs are phases, and that phases explain some ECP effects, then phases could explain how particles in head D are able to be empty (whether elided or null) (this will be discussed in more detail in Chapter 4).

## **CHAPTER 3 – QUANTITATIVE ANALYSIS OF THE COJADS DATA**

### **3.1 Hypotheses**

A quantitative analysis was conducted to determine whether there is dialectal variation in the frequency of particle omission. A secondary goal was to determine whether certain social factors (speaker age and gender) and linguistic factors (use of sentence-final particles and honorific prefixes) impact the frequency of particle omission. My hypotheses with respect to the data examined are specifically that:

- 1) The frequency of nominal particle omission will vary from dialect to dialect, with some particles exhibiting more omission than others.
- 2) The social factors (speaker age and gender) will affect the frequency of nominal particle omission and may also interact with any dialectal variation.
- 3) The linguistic factors (sentence-final particle use and honorific prefix use) will not impact frequency of nominal particle omission due to the overall more informal style of the conversation data.

To test my hypotheses, I examined natural conversation data from the COJADS corpus (NINJAL 2021) representing four different dialects of Japanese (the Kitakyushu, Osaka, Tokyo, and Hokkaido dialects). The data was first processed and visualized, and then a mixed effects model in R (R Core Team 2020) was used to test for statistical significance.

### 3.2 The Data & Data Preparation

One CSV file per dialect was downloaded from the website<sup>13</sup> where the Corpus of Japanese Dialects (NINJAL 2021) raw CSV data files are published (see 2.2.3 for a discussion of dialect choices, and 2.2.2 for details about COJADS and the website). The specific files downloaded from the database webpage were as follows: Hokkaido 01\_b\_099.csv, Tokyo 13\_a\_028.csv, Osaka 27\_b\_099.csv, and Kitakyushu 40\_a\_002-1.csv. These files contain data from both the “Urgent Survey of Japan’s Regional Dialects” and the “Hometown Language Collection” (see Section 2.2.2 for details). These files are publicly available for download and search for academic use with a free user account, and the webpage link is provided in the References section and in several footnotes. Note that the files used in this project were downloaded in March 2021. Only the text data was used in this project; no audio data was accessed.

As described in 2.2.2, each file contains transcriptions and corresponding metadata from recorded conversations held between speakers of the same dialect. The original recordings for the four spreadsheets used were approximately 20-37 minutes long, and each produced between 700 and 1,500 lines of data when transcribed by the NINJAL researchers. The conversations were recorded in the 1970s and 80s, and the speakers were primarily above the age of 60 at the time of recording (with dates of birth starting in the late 1890s to early 1900s). One speaker was an outlier, at age 47.

Only transcripts of conversations between mixed-gender groups were chosen (with varying numbers of speakers involved), as there is some evidence that speakers use different linguistic features when speaking in same-gender groups versus mixed-gender groups (cf. Takano 1998). Otherwise, the files were chosen at random from the files available for download at the time

---

<sup>13</sup> <https://www2.ninjal.ac.jp/cojads/index.html?targ=data>

(March 2021). Each file includes transcriptions of the conversation in katakana script, a translation into Standard Japanese, and metadata (speaker ID, age, sex and date of birth, topic of conversation, location and region of recording, and dialect). As previously mentioned, speakers are kept anonymous, as are any references made to other people in each conversation. The NINJAL researchers also marked instances of particle omission using brackets in the Standard translation text column (Koiso et al. 2019), and these annotations were vital to this thesis project.

Each file was originally entirely in Japanese characters and all columns were converted to *romaji* (Roman characters) and English using a combination of my own knowledge of Japanese and several online tools to prepare it for analysis in R (R Core Team 2020) (R is unable to handle Japanese characters and certain special symbols). For the transcriptions of the data into roman characters (both standard and dialect columns), I used a modified Hepburn orthography system (in particular the long ‘o’ sound was transcribed as ‘oo’ rather than the more standard ‘ou’ or ‘ō’ in order to ensure readability in R). The orthography system used in the dataset and in this thesis in general is described in further detail in Appendix A.

The dialect text column in each file was converted to roman characters using the [lexilogos.com](http://lexilogos.com) *katakana* conversion tool (Negre 2021), and the standard language translation column (which included a combination of kana and kanji) was converted using the translator at [romajidesu.com](http://romajidesu.com) (Haibuihoang 2015). I verified the accuracy of these automated transcriptions by first inserting a new column beside the one being transcribed and working in batches of 25-50 rows at a time, spot-checking for accuracy in each batch by comparing the original and transcribed columns. The metadata columns (e.g. speaker age, location, prefecture) were converted to English using a combination of my own knowledge and the [jisho.org](http://jisho.org) dictionary (Ahlström et al. 2022). I then used simple substitution regular expressions using [Regex101.com](http://Regex101.com) (Dib 2021) and the Find-

and-Replace feature in the software LibreOffice Calc and Microsoft Excel to replace all remaining Japanese characters and other special characters to ensure readability in R. A sample of the translated data file is provided in Appendix B, along with a sample of the long-format version of the data generated in R.

Once the original data and metadata were formatted as CSV files and readable in R, I annotated the data for the variables which were not already included in the metadata. One column per nominal particle under examination (*-wa*, *-ga*, *-o*, *-ni*, *-de*, *-e*, *-no*) was added, and each line was marked PRESENT/OMITTED/NA based on whether that particle occurred in it. Similar columns were added for sentence-final particles (eg. *-ne*, *-yo*, *-ka*), and honorific nominal prefixes (*go-* and *o-*), although these were grouped together into one column rather than broken out by individual particle/prefix. Detail on all variables considered and the annotation method used for each is provided in Section 3.3.

Finally, I added the first letter of the prefecture name before each speaker ID and line ID; each file had started the speaker ID at “A” and line ID at “1”, meaning there were four speaker A’s and four Line 1’s. As all spreadsheets needed to be combined into one single spreadsheet, this ensured I could identify which speaker belonged to which dialect. For example, speaker “A” became speaker “H\_A” standing for “Hokkaido speaker A.” After combining the files into one large CSV file, I loaded it into RStudio (Version 1.4.1103) (RStudio Team 2021), software that is designed for working in the R programming language (Team 2020). I then converted the data to long format using the `gather()` function from the `tidyverse` package (Wickham et al. 2019), omitted the NA values using the `na.omit()` function in base R (R Core Team 2020), and re-ordered the columns. A sample of this re-ordered long format data can be found in Appendix C. Note that in the data, the Kitakyushu dialect data is referred to as “Fukuoka”, the prefecture in

which the dialect is spoken as all dialects are referred to by their prefecture in the data; “Fukuoka” was therefore used in the figures and tables in this chapter.

### 3.3 The Variables

Table 2 below shows a brief summary of the variables that were annotated for, and the following subsections (3.3.1 - 3.3.6) describe the annotation methods used for each variable.

*Table 2. Summary of Variables Examined*

Dependent Variable:	Post-nominal particles (- <i>wa</i> , - <i>ga</i> , - <i>o</i> , - <i>ni</i> , - <i>de</i> , - <i>e</i> , - <i>no</i> )
Independent Variables:	Sentence/Clause-final particles
	Honorific nominal prefixes
	Speaker gender (biological sex)
	Speaker age (at time of recording) & year of birth
	Dialect (labeled as “Prefecture”)

#### 3.3.1 Post-Nominal Particles

One column per post-nominal particle was added, and each particle was marked as “PRESENT”, “OMITTED”, or “NA” for each sentence where post-nominal particles were possible (sentences with null arguments were not considered as having overt particles possible due to the fact that the entire nominal argument is elided in argument ellipsis, including any particles). I also added columns indicating whether each utterance could have particles or not, and whether particle omission was present at least once in the utterance.

I examined the presence/absence of most mono-syllabic nominal particles, and these are listed in Table 3. Instances of -*no* as a nominalizer were not included, as these are instances where a particle follows a verbal element, rather than a nominal element. Instances where particles can

be used consecutively (for instance *-ni* + *-wa*) were treated as separate particles in the annotation of each particle, rather than as a separate phenomenon; however, instances of these double-particles were marked in a separate column and reviewed visually (this column was not used in the R analysis and was therefore removed before the file was finalized for R).

*Table 3. Nominal Particles Examined in R<sup>14</sup>*

<b><u>Particle</u></b>	<b><u>Use</u></b>
<i>-wa</i>	Topic (TOP)
<i>-ga</i>	Nominative (NOM)
<i>-o</i>	Accusative (ACC)
<i>-ni</i>	Goal
<i>-de</i>	Locative (LOC)/Instrumental (INS)
<i>-e</i>	To a place
<i>-no</i>	Genitive (GEN)

A shortcoming of relying solely on the annotation provided originally by the NINJAL researchers should also be mentioned: it is not clear how they accounted for the discourse particles which can stand in for or occur with other particles (for instance *-wa* and *-mo*) when coding for the omitted particles. These particles can replace other particles, and in some cases it could be difficult to determine which particle was truly intended. Additionally, there are some instances where either *-ga* or *-o* would be appropriate; it is not clear how the researchers made the determination of which particle was omitted in every instance. For the purposes of this thesis, I have assumed that the researchers relied on their native speaker intuitions to indicate which particle was missing. However, it should be acknowledged that this is a shortcoming, especially in light of arguments that there are two types of missing particles (a zero particle and an elided particle; see

---

<sup>14</sup> Note that the particles *-ni* and *-e* are quite similar; however, I have labeled *-ni* as a “goal” marker to account for the fact that it acts as a goal in the sense of going to a place as well as giving an item to someone. The particle *-e* on the other hand is only used to indicate movement towards a location, and cannot be used to mark the recipient of an item for instance.

section 2.5.2 for a brief discussion). It is not clear from the way the researchers annotated the data which of the particles they marked as missing were instances of the zero particle, and which were instances of particle ellipsis (if we take both analyses as viable). For the purposes of this chapter, references made to “particle omission” are all referring to particle morphemes that were marked as “missing” by the researchers, and the two phenomena are treated together as one. Note also that I did not distinguish between each use of particles which can have more than one use (such as *-de* which can be instrumental or locative); properly annotating each case would ideally involve assistance from a native speaker of each dialect who has linguistic training – and this was beyond the limited scope of this thesis. Future research should certainly involve annotating each case.

### 3.3.2 *Nominal Honorific Prefixes*

In order to examine whether there is any correlation between the presence of honorific prefixes in an utterance and particle omission, I included a column to indicate whether honorific prefixes were used in each line. These were coded as YES/NO/NH: YES indicates that one or more honorific prefixes (either *go-* or *o-*) were used in that line; NO indicates that no honorific prefixes were observed, but that one was expected (for instance a noun that is typically pronounced with the honorific prefix attached, such as *o-shougatsu* ‘HON-new year’); NH indicates that an honorific prefix was not possible (e.g. the utterance was some form of backchanneling such as *un* ‘yes’), or the noun does not typically occur with an honorific prefix. The goal of including this variable was to look for correlations between formality levels and particle omission; due to the limited scope of the project and complexity of the Japanese honorifics system, it was more feasible to identify honorific prefixes than to annotate the data based on overall sentence formality.

### 3.3.4 Sentence/Clause-final Particles

Each line was coded for the presence or absence of sentence/clause-final particles. Each line was manually checked for the sentence/clause-final particles contained in Table 4 (the definitions contained in this table are brief summaries of those discussed by Endo (2007)). Note that some of these sentence/clause-final particles can occur at the end of a clause, but all can occur at the end of a sentence; therefore, they are referred to as “sentence-final” particles in this thesis.

*Table 4. Sentence/Clause-final Particles Examined*

WA	Indicates “mildness” of the statement, and is typically used by females (or by males to sound “softer”)
ZO	Emphatic, similar to YO but indicating the speaker has higher social status than the addressee
YO	Speaker’s emphasis; neutral social status
SA	Reports speaker’s familiarity about the proposition (the sentence should be taken as a matter-of-fact)
NE, NA, NEE, NAA	Weak confirmation, similar to English tag question ‘right?’
KA, NO	Question marker

Lines where no sentence-final particle was possible (for instance short affirmations such as *un* ‘yes’ or *hai* ‘yes’) were labeled “NP” for “not possible”; lines with a sentence-final particle were labeled “YES”; lines where sentence-final particles were possible but where none was used were labeled “NO.” It should be noted that when these particles occurred at the end of a clause, they were still coded as “YES.” This variable was included due to claims by Fukuda (1993) and

Endo (2007), and others that sentence-final particles condition particle omission, as well as claims to the opposite (that they do not condition particle omission) such as that made by Lee (2002b).

### *3.3.5 Speaker Gender & Age*

The gender, age at time of recording, and year of birth for each speaker was provided by the NINJAL researchers in the original datasets (as previously mentioned, all data was already de-identified, with speakers being labeled by de-identified letter ID). The original coding for gender, age and year of birth was not changed, beyond translation to English.

It should be noted that what is labeled “gender” in the data likely corresponds to biological sex, as opposed to gender in the modern sense (such as in Levon 2014 or West & Zimmerman 1987). It was not clear what determination the researchers used in making this distinction (perhaps using visual or auditory cues), nor was it clear whether the determination was made by the researchers who made the recordings or who later transcribed them. As such, this information may not be accurate, or up to modern standards in language and gender studies. However, short of contacting the speakers (which is impossible as the data was de-identified), it is impossible to determine a better coding method for speaker gender. Additionally, due to Japanese cultural norms, it may not be appropriate (or would prove difficult) to determine whether speakers identified as any gender other than the binary male/female options. Therefore, I will assume that the column for “gender” is really that of biological sex, and was assigned based on researcher observations; however, I will still use the term “gender” in figures and discussion presented in this section in keeping with the terminology employed in the original metadata.

Although it would be ideal to be able to analyze the gender and age of all interlocutors involved in each speech act (as studies such as Coupland 1980, Sharma 2018, and Bilous & Krauss 1988 indicate, conversation composition matters, and Takano 1998 indicates that it matters for

Japanese), only speaker gender and age were examined for this thesis. The reason for this is that it would be difficult (or even impossible) to determine who was being addressed in each line (as well as who was in the audience). This may be possible if recordings were listened to, but this was beyond the limited scope of this project and would be a point for future study.

### *3.3.6 Dialect (prefecture)*

This was coded by the NINJAL speakers and contained in the original data files. One column was provided for the prefecture name, and one for the city in which the recording took place. I used the prefecture column as the basis for naming each dialect. The original coding of dialect and recording location was not changed from the originally downloaded files (beyond translating them into English).

## **3.4 Data Visualization**

Once the data was in R and had been formatted and re-ordered, tables and charts were generated to visualize possible interactions between the factors described in Section 3.3 and particle omission. Each subsection below includes plots, tables, and a brief discussion to help visualize the data. Each plot in Sections 3.5.1 – 3.5.7 below present data for all particles listed in Table 3 in Section 3.3.1.

### *3.4.1 Overall Particle Omission Observed in the Data*

Table 5 presents a simple count of tokens of omitted and present particles. We can see from this table that *-wa* had the most tokens overall (of both omitted and present tokens), followed by the particles *-no*, *-ni*, *-ga* and *-o*. We can also see that *-o* had the most tokens of omitted particles.

Table 5. Count of Tokens per Particle

	<b>Tokens Omitted</b>	<b>Tokens Present</b>	<b>Total Particles</b>
<b>DE</b>	13	216	229
<b>E</b>	13	53	66
<b>GA</b>	133	325	458
<b>NI</b>	74	457	531
<b>NO</b>	46	514	560
<b>O</b>	264	165	429
<b>WA</b>	179	591	770
<b>Total:</b>	<b>722</b>	<b>2,321</b>	<b>3,043</b>

Table 6 and Figure 9 present the same data as in Table 5 above, but in percentage format. From these, we can see that *-o* led by far in percentage of omission, followed by *-ga*, *-wa*, *-e* and *-ni*. The particle with the lowest rate of omission was *-de*.

Table 6. Percentage of Particle Omission by Particle

	<b>% Particle Omitted</b>	<b>% Particle Present</b>
<b>DE</b>	5.68%	94.32%
<b>E</b>	19.70%	80.30%
<b>GA</b>	<b>29.04%</b>	<b>70.96%</b>
<b>NI</b>	13.94%	86.06%
<b>NO</b>	8.21%	91.79%
<b>O</b>	<b>61.54%</b>	<b>38.46%</b>
<b>WA</b>	<b>23.25%</b>	<b>76.75%</b>

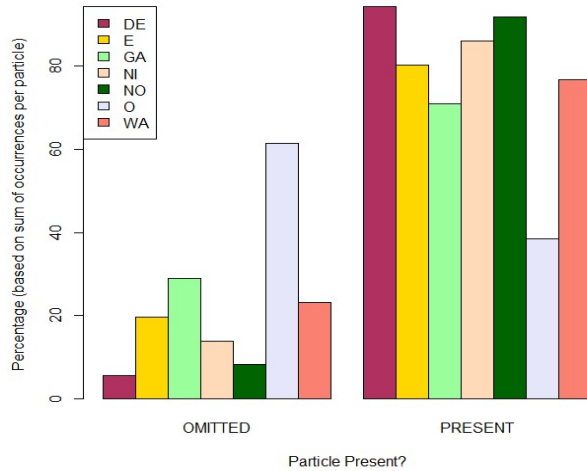


Figure 9. Percentage of Particle Omission by Particle

### 3.4.2 Particle Omission by Dialect

Dialectal variation in particle omission does appear to be present in the data. Table 7 presents an overview of tokens of particle omission. From Table 7, we can see that there was an uneven number of utterances with the possibility of having a nominal particle (lines where a nominal particle was not possible, such as *hai* ‘yes’, were filtered out before tables and charts were created). We can also see that the Hokkaido dialect had the most utterances, as well as the most instances of particle omission.

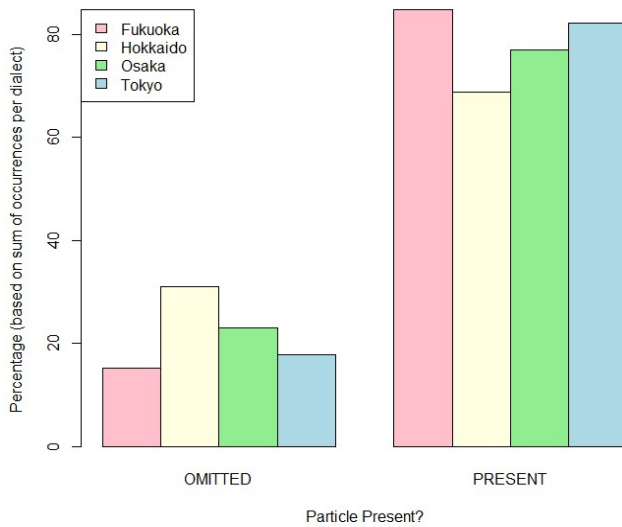
Table 7. Tokens of Particle Omission by Dialect

	<b>Tokens Omitted</b>	<b>Tokens Present</b>
<b>Fukuoka</b>	72	400
<b>Hokkaido</b>	355	786
<b>Osaka</b>	178	596
<b>Tokyo</b>	117	539

Figure 10 and Table 8 present the data in percentage format; here, we can more clearly see that Hokkaido does in fact exhibit the most frequent particle omission, while Fukuoka exhibits the least. This is consistent with the findings of (Kibe et al. 2020) who found a similar pattern in their examination of the *-o* ACC marker in a larger dataset from the COJADS database.

*Table 8. Percentage of Particle Omission by Dialect*

	<b>% Particle Omitted</b>	<b>% Particle Present</b>	
<b>Fukuoka</b>	15.25%	84.75%	100%
<b>Hokkaido</b>	31.11%	68.89%	100%
<b>Osaka</b>	23.00%	77.00%	100%
<b>Tokyo</b>	17.84%	82.16%	100%



*Figure 10. Percentage of Particle Omission by Dialect*

Figure 11 demonstrates that dialectal variation also occurs between individual particles; some dialects exhibit more omission of certain particles than others. In particular, Tokyo has more omission of *-e* ‘to’ than other dialects including Hokkaido which has the most omission for all other particles. This chart also confirms that certain particles are omitted more readily than others; in particular *-wa*, *-ga*, *-o*, and *-ni*, as seen in Section 3.4.1.

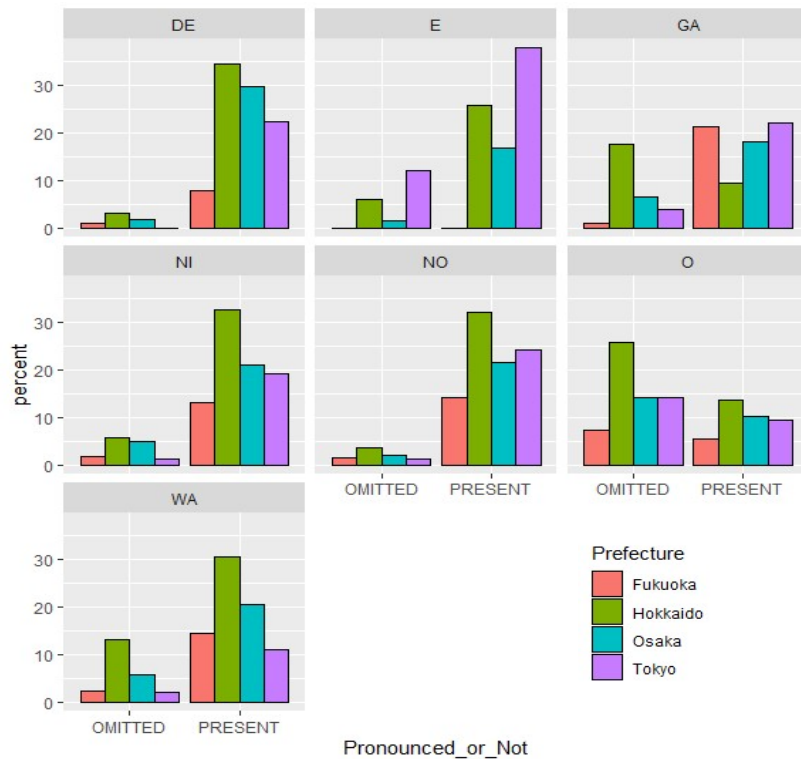


Figure 11. Percentage of Particle Omission by Prefecture (Dialect) and Particle

### 3.4.3 Particle Omission by Speaker Gender

Male speakers had more tokens overall (of both present and omitted particles) as compared to the female speakers; as a result, men had more tokens of particle omission than women (as seen

in Table 9). However, when we examine the percentages of particle omission by gender (see Table 10), we see that women omitted particles slightly more than men (by 5.51%). Although there is a slight difference in frequency, this does not seem to be a large enough difference to be statistically significant (see Section 3.5 for a discussion of the statistical tests performed to check for statistical significance).

*Table 9. Tokens of Particle Omission by Speaker Gender*

<b>Speaker Sex</b>	<b>Tokens Particles Omitted</b>	<b>Tokens Particles Present</b>
<b>Female</b>	280	735
<b>Male</b>	438	1546
<b>Unknown</b>	4	40

*Table 10. Percentage of Particle Omission by Speaker Gender*

<b>Speaker Sex</b>	<b>% Particles Omitted</b>	<b>% Particles Present</b>
<b>Female</b>	27.59%	72.41%
<b>Male</b>	22.08%	77.92%
<b>Unknown</b>	9.09%	90.91%

As mentioned above, it is possible that the terms used refer to the speakers’ biological sex rather than their gender identity. It should be noted that the speaker genders listed here are based on the metadata provided by the NINJAL researchers who transcribed the conversations; the three words used in the ‘speaker sex’ column translate to “female,” “male,” and “unknown.” It is not clear from the data how the determination of a speaker’s gender was made; however, it seems likely that the determination was made based on the sound of the speaker’s voice as well as other

cues (such as use of certain sentence-final particles or other gendered speech). It is still possible that the coding does represent gender, if speakers present their speech according to their gender identity. It is also possible that the speakers were provided a questionnaire at the time of recording.

In Figure 12, we can see that there is also slight variation in specific particles based on gender. Each plot within the figure presents the percentage of omitted versus overt particles for each gender. The percentages are calculated based on all utterances of each particle (both omitted or present). For most particles, men exhibit more particle omission than women, except for *-e* where women omit the particle more than men. We can also see that, consistent with the findings presented in Tables 5 and 6 in Section 3.4.1, the particle *-o* was omitted more than it was pronounced; in fact, both men and women omitted this particle more than pronounced it.

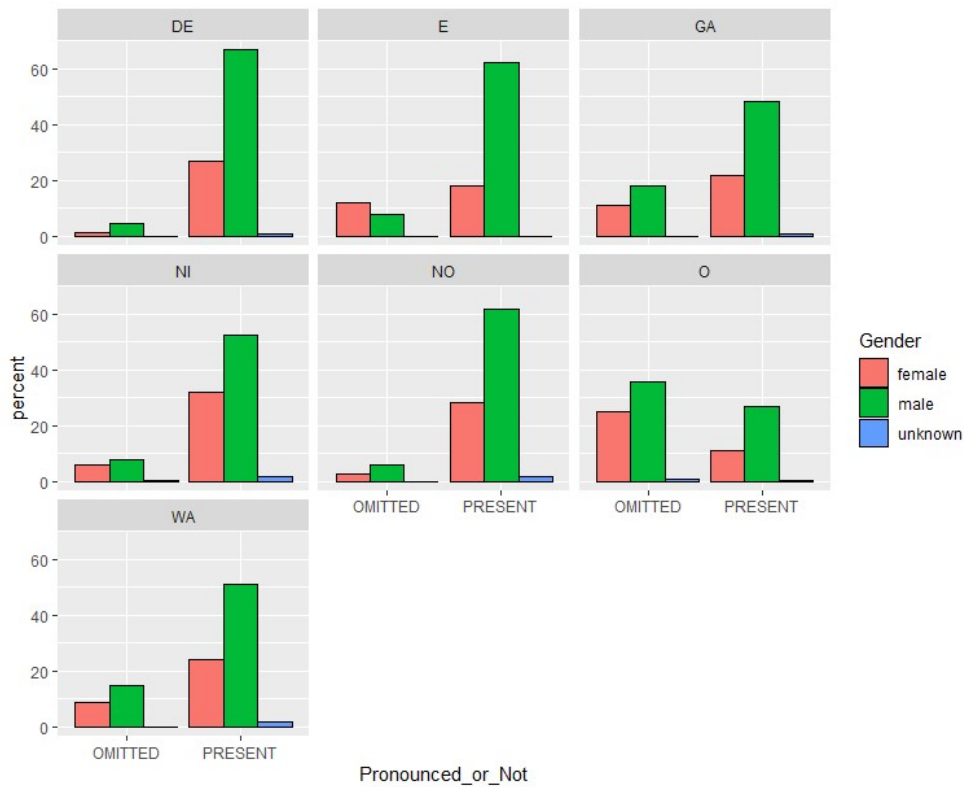


Figure 12. Percent of Particle Omission by Gender and Particle Type

### 3.4.4 Particle Omission by Speaker Age and Birth Year

Table 11 shows the individual ages of the speakers (at the time of recording). All but one speaker was over the age of 60 at the time of recording, with most speakers being in their 70s. Table 11 reveals that speakers of some age groups had more tokens of particle omission than others, but that overall, the phenomenon is varied by age. Figure 13 shows the percentages of omission by speaker age; here we can see some correlation may be emerging, with younger speakers omitting more particles (with the exception of the outlier 47-year old, who pronounces all particles). However, the ages are so close together that it is difficult to determine any correlation based on age at time of recording.

Table 11. Tokens of Nominal Particle Omission by Speaker Age

Speaker Age:	47	63	65	66	70	71	72	73	74	76	77	78	79	unknown
Omitted	0	73	9	113	48	194	32	41	70	7	7	47	77	4
Present	3	149	38	200	127	459	101	235	223	20	43	313	370	40

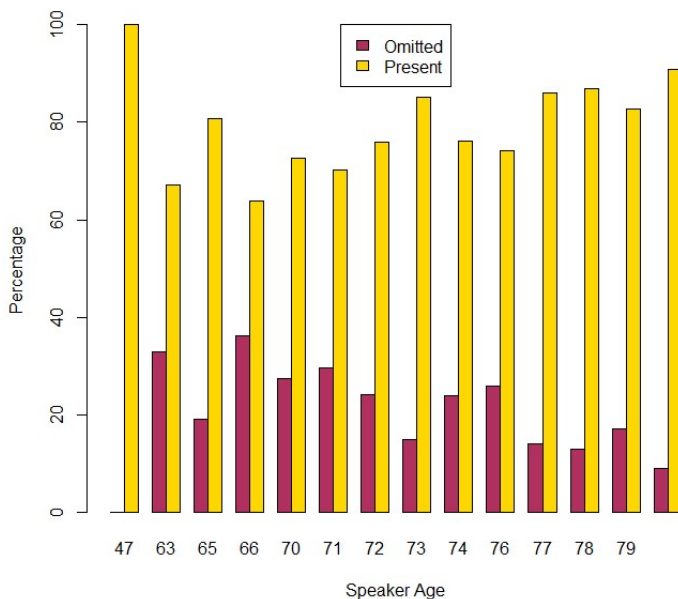


Figure 13. Percentage of Particle Omission by Speaker Age

Another way of examining the impact of speaker age is by looking at speaker birth year, as can be seen in Figure 14. The color scale here allows us to see that speakers who were born between 1904 - 1912 were generally more likely to omit particles than speakers born in other years. However, there is simply not a wide enough range of ages represented, and no generalization can readily be made.

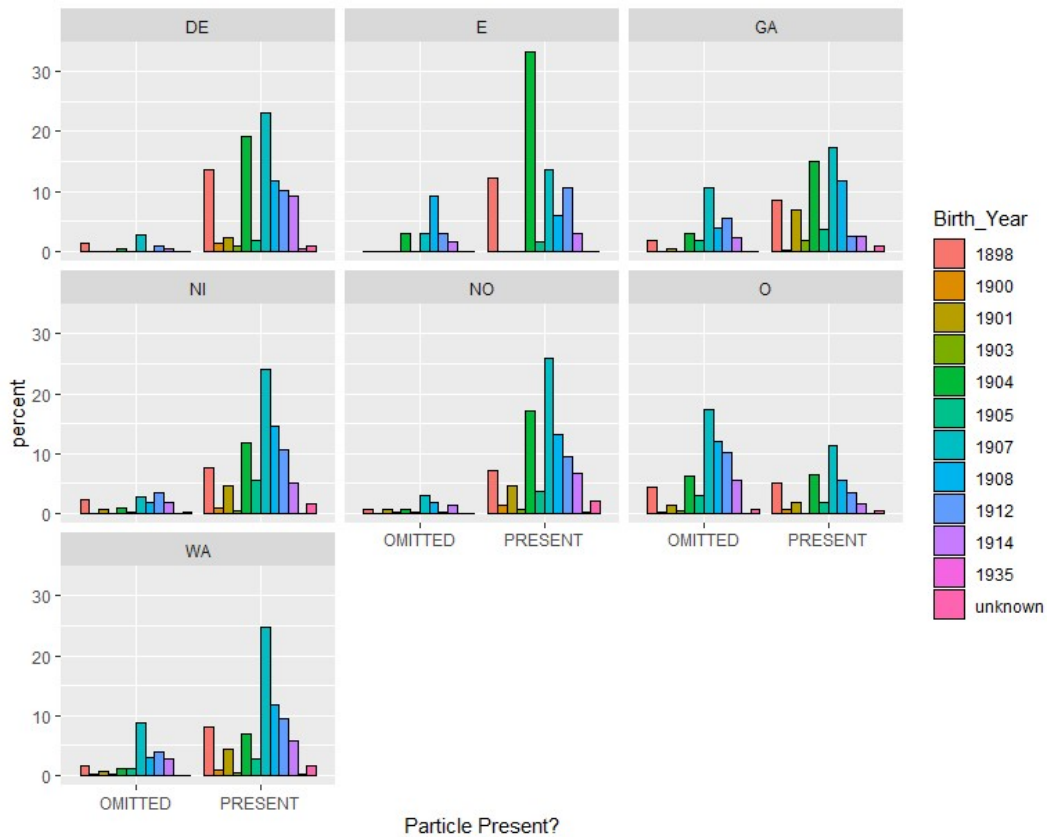


Figure 14. Percentage of Particle Omission by Speaker Birth Year & Particle Type

### 3.4.5 Particle Omission and Honorific Prefixes

Based on Figure 15, there appears to be little to no correlation between particle omission and the use of nominal honorific prefixes (*go-* and *o-*). While some particles are omitted more than

others when an honorific prefix is present (for example, *-o* and *-de* in Figure 16), there are simply not enough tokens of honorific prefixes present in the data to make any generalizations.

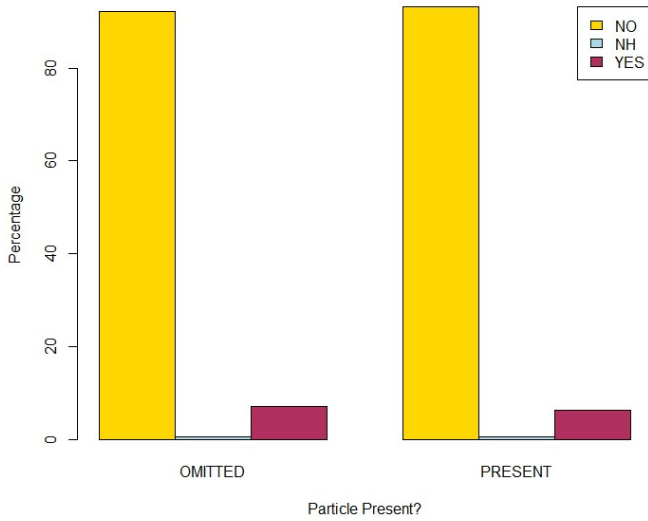


Figure 15. Percentage of Particle Omission by Nominal Honorific Prefix Use

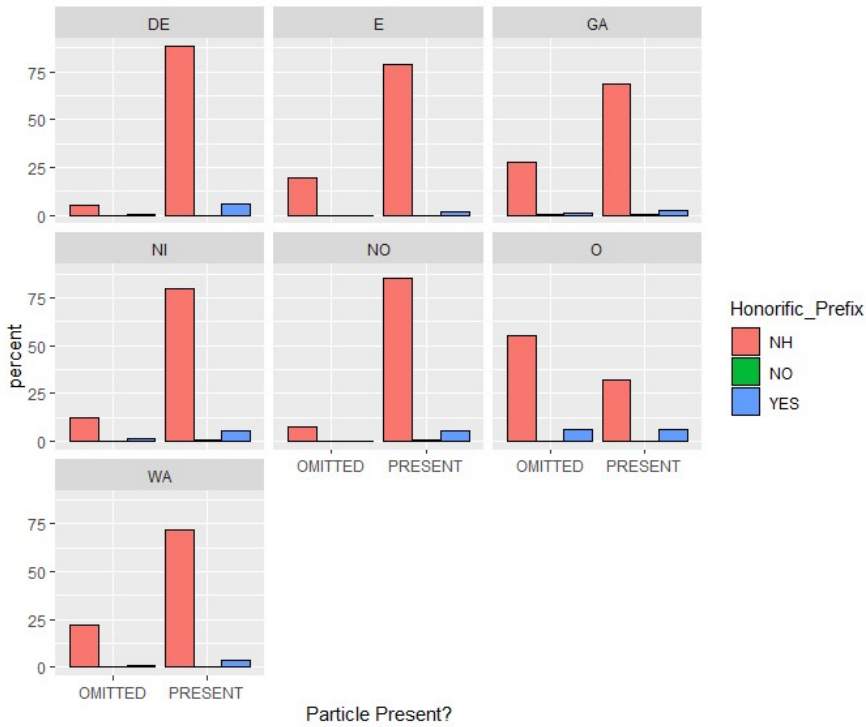


Figure 16. Percentage of Particle Omission by Presence of Nominal Honorific Prefix

### 3.4.6 Particle Omission and Sentence-final Particles

There also does not appear to be a correlation between sentence-final particle use and nominal particle omission. Figure 17 shows that the percentages of omitted and overt nominal particles are almost identical between utterances with sentence-final particles and those without, and the difference is likely not statistically significant.

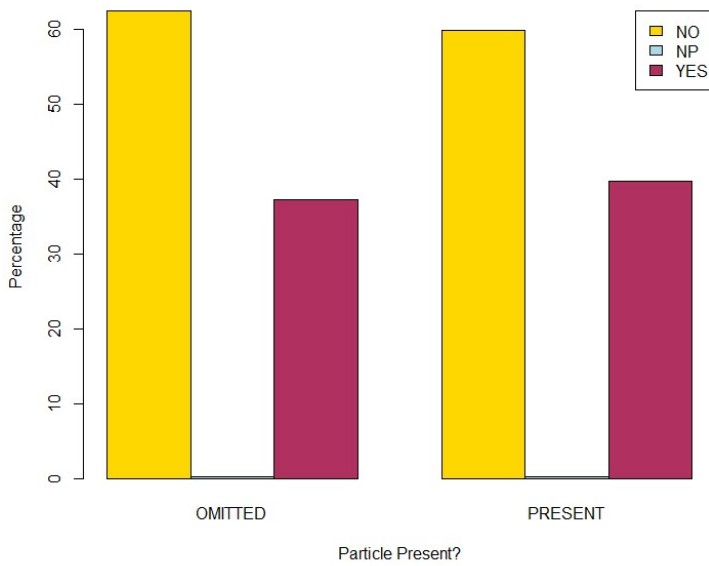


Figure 17. Percentage of Particle Omission by Sentence-Final Particle Use

Figure 18 demonstrates that there is variation in which particles are more frequently omitted. However, overall, no generalization can be made regarding any correlation between sentence-final particle use and nominal particle omission.

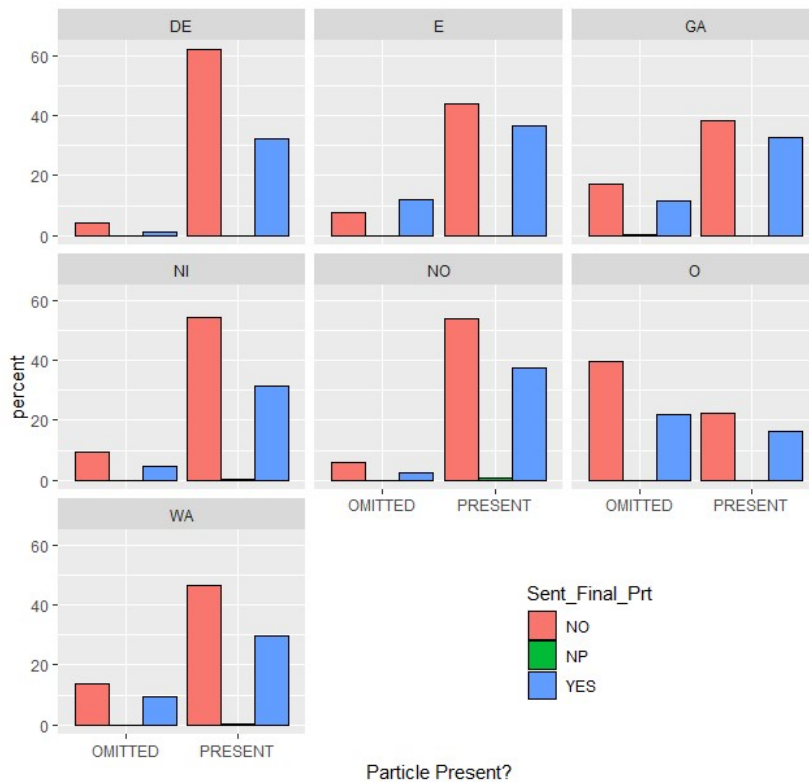


Figure 18. Percentage of Particle Omission by Particle and Use of Sentence-Final Particles

### 3.4.7 Particle Omission by Speaker

The plot in Figure 19 shows that some speakers had more utterances than others. It also shows that all speakers pronounce more particles than they omit. As previously mentioned, each speaker ID starts with the first letter of the prefecture of the recording, which I added to the original speaker ID letter. For instance, H\_A corresponds to Hokkaido Speaker A. Because of this annotation method, we can also look at possible interaction between speaker and dialect in this plot, and indeed the three Hokkaido speakers appear to have higher frequencies of omitted particles than speakers of other dialects (as expected based on previously presented tables).

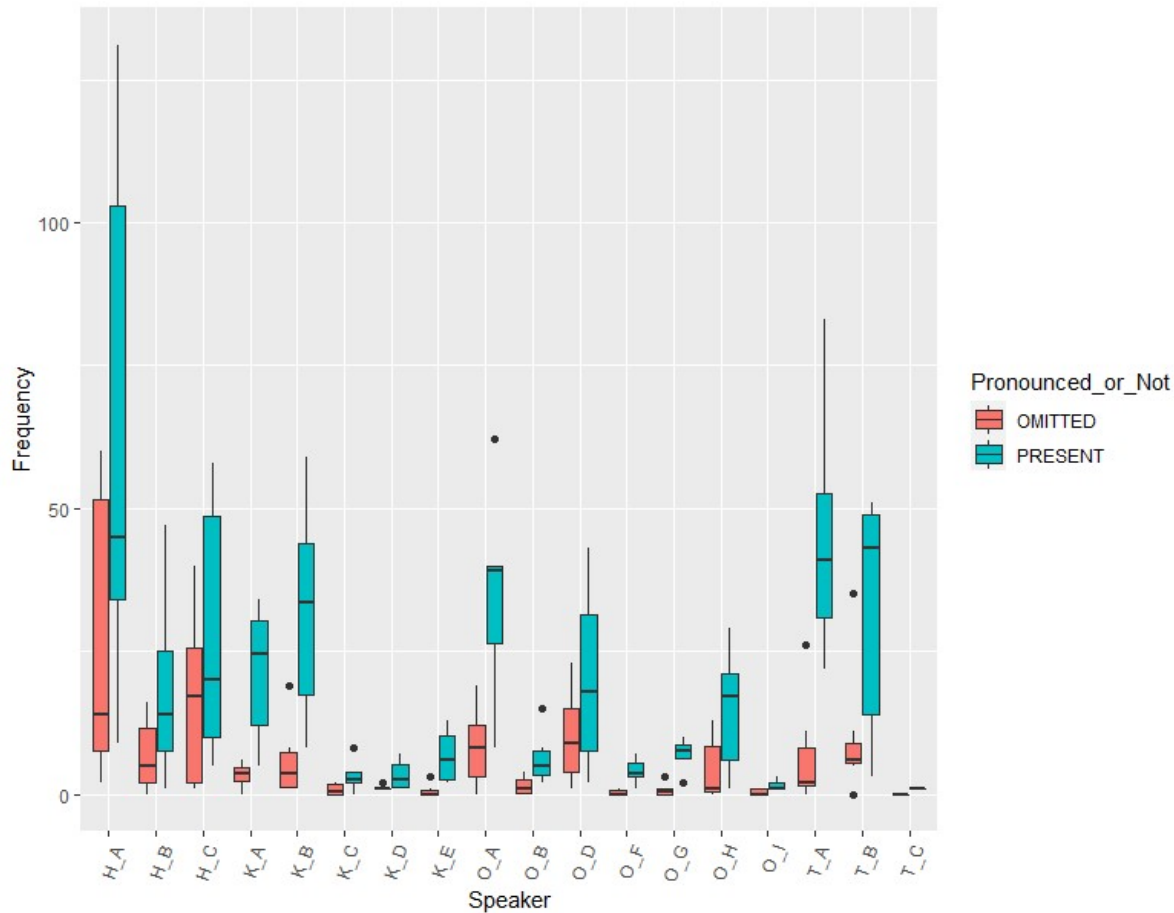


Figure 19. Frequency Plot Showing Particle Omission by Individual Speaker

### 3.4.8 Correlation Between Multiple Factors

Although it is difficult to visualize correlation between more than a few factors, the plots contained in Figure 20 and 21 are an attempt at visualizing the effects of each variable on particle omission; age, dialect (prefecture), and particle type are shown in Figure 20, and age, gender, and particle type are shown in Figure 21. The most interesting point of note in Figure 20 is that the ages of speakers were distributed fairly randomly between dialects. However, overall, it is difficult to visualize any correlation between the three factors presented in Figure 19. In Figure 21, the result is the same, and correlations are difficult to see. Although correlations are not obvious from

these multi-factor plots, a statistical test (ANOVA) was run in order to examine the effects of particular variables on particle omission; this is discussed in more detail in Section 3.5.

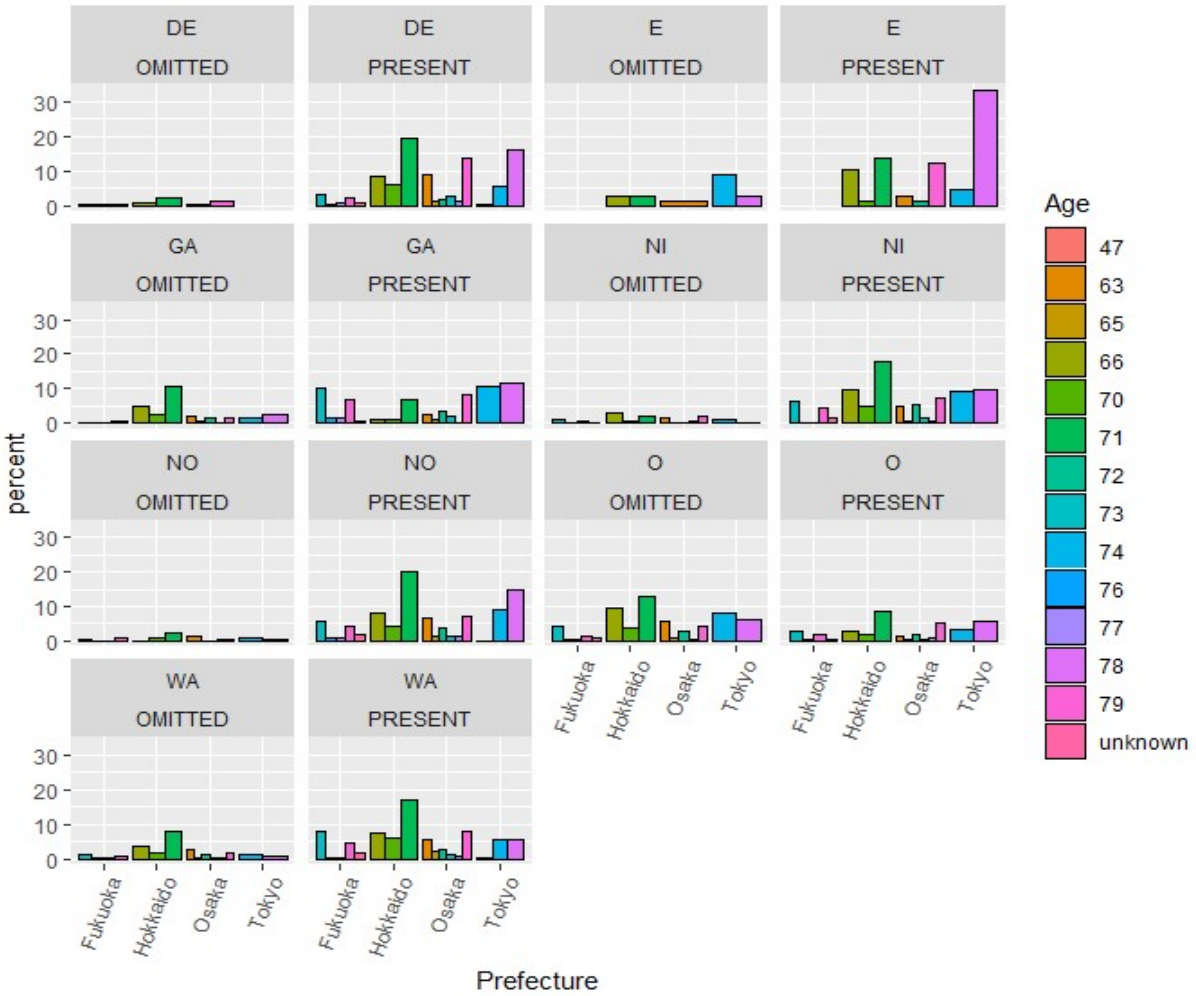


Figure 20. Percentage of Particle Omission by Speaker Age, Prefecture (Dialect) and Particle

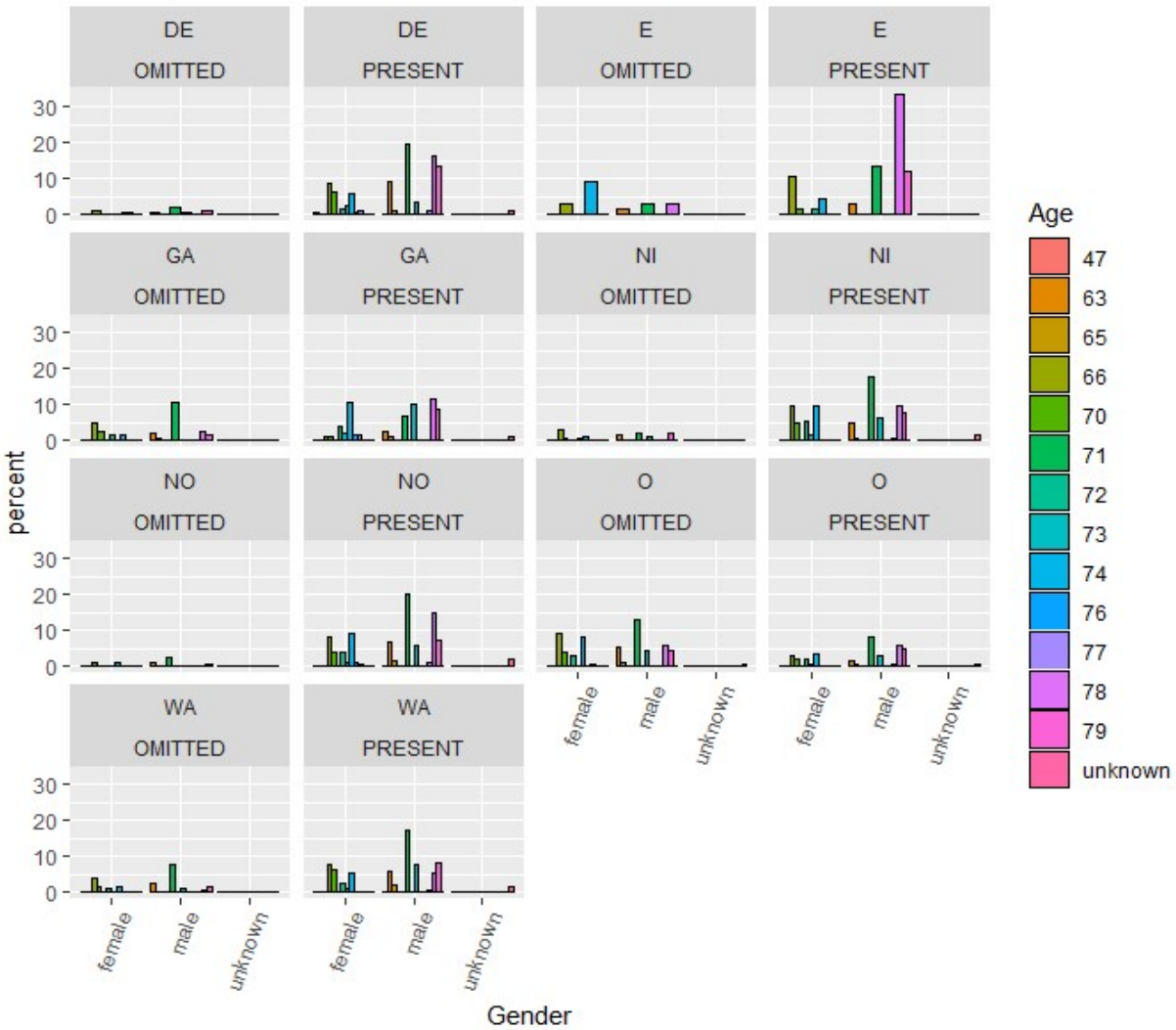


Figure 21. Percentage of Particle Omission by Speaker Age, Speaker Gender, and Particle Type

### 3.4.9 Plotting the Three Case Particles (-ga, -o, -ni)

Due to the treatment of the so-called “Case particles” separately from the so-called “Postpositional particles” in much of the literature (as discussed briefly in Section 2.4), these three particles were also examined independently of the other particles. In order to do this, a new spreadsheet was created with the columns for the other nominal particles removed. This file was then run through R in the same manner as the original file (as described in Section 3.2). This

section presents a few of the more interesting plots that resulted from looking only at these three particles. The plots presented here are similar to those presented in the previous sub-sections but contain only the three Case particles. Detail by particle was already provided in the multi-plot figures in the previous subsections.

When we examine the frequency of omitted versus overt particles by dialect with only the three Case-particles (shown in Figure 22), the resulting plot is quite similar to that observed in Figure 9, with Hokkaido exhibiting the most zero particle use and Fukuoka (Kitakyushu) the least.

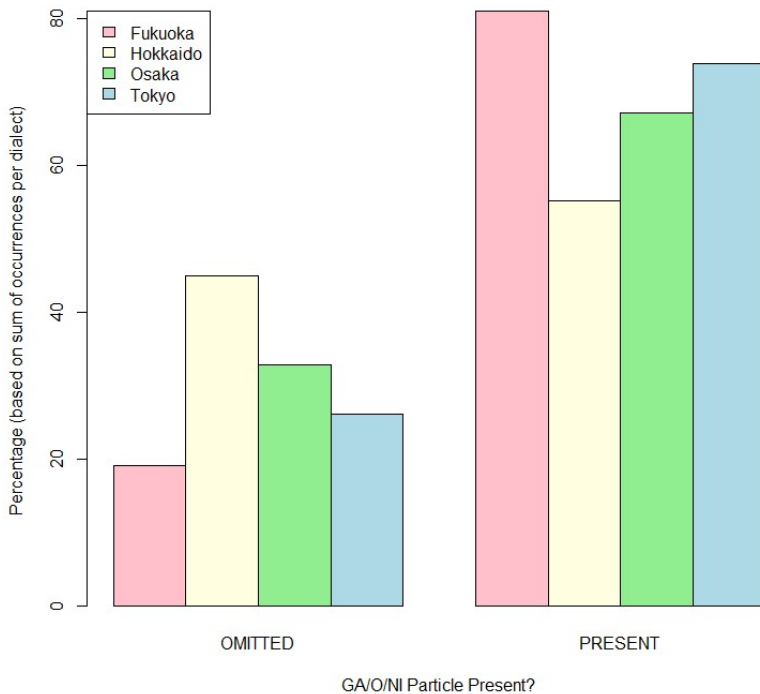
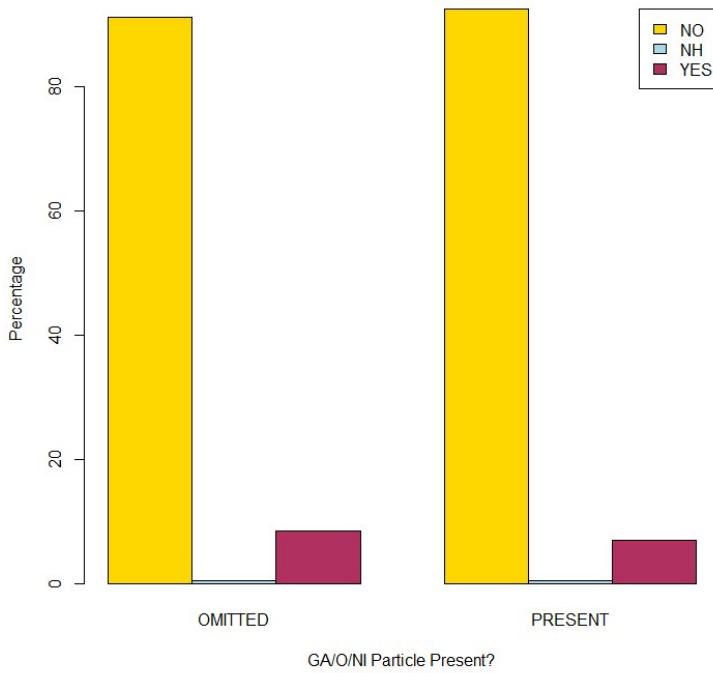


Figure 22. Percentage of Particle Omission of GA / O / NI by Dialect

We can also see that, even just with the Case particles, there is little to no correlation between honorific prefix or clause (sentence)-final particle use (Figure 23 and Figure 24, respectively). However, we can see that there is slightly more particle omission of the three case

particles when honorific prefixes are present (which was not obvious from Figure 16 above). This difference is quite slight, and may not be significant as the variation is less than 5%. These observations could suggest that the Case particles may not act differently than the other post-nominal particles. Further investigation is necessary due to the slight difference observed, but this will be left for future research.



*Figure 23. Percentage of Particle Omission of GA/O/NI by Nominal Honorific Prefix Use*

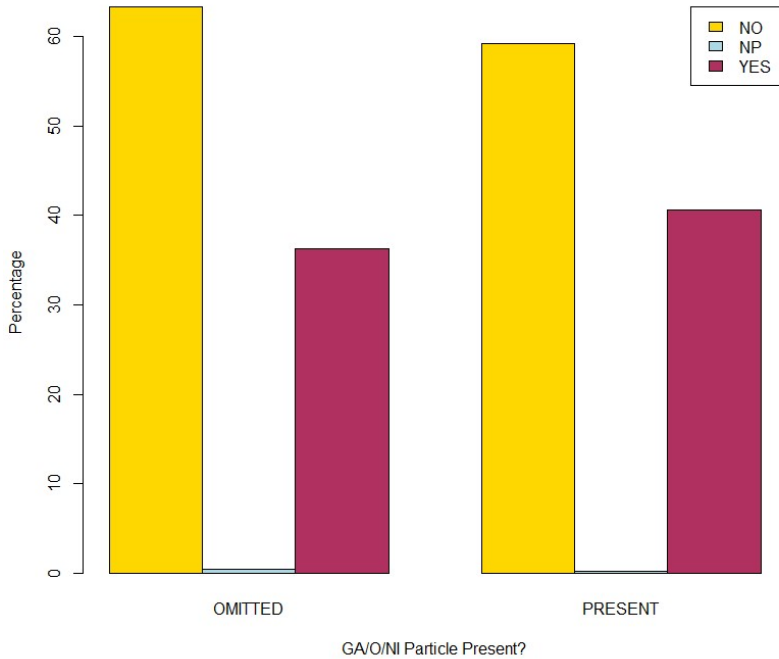


Figure 24. Percentage of Particle Omission of GA/O/NI by Sentence-Final Particle Use

### 3.5 The Statistical (ANOVA) Tests

After visually confirming that there is a possible correlation with at least dialect, speaker age, and particle type, I performed a multivariate analysis in R (R Core Team 2020) in a stepwise manner, similarly to the method used by Takano (1998) in his analysis of the effect of gender and context on *-wa* and *-ga* omission. I used a linear mixed-effects regression (mixed ANOVA), fitted with the help of the `lme4` and `lmerTest` packages in R, because the data contains fixed effects and the random effect of speaker. In order to ensure the test would work correctly, I first converted the “Omitted” and “Present” in the `Pronounced_or_Not` column to numbers: 0=Omitted, 1=Present. The code used to prepare the data for the ANOVA analysis, along with the full R code used for this analysis, is included in Appendix C.

I then used the `lmer()` function from the `lme4` package (Bates et al. 2015) to create a baseline model, with `speaker` as a random variable. I also used the package `lmerTest` (Kuznetsova et al. 2017) to add p-values to the results.

```
> ZeroPrt_LMER_0 <- lmer(Pronounced_or_Not ~ 1 + (1 | Speaker), data =
LongFourDialects_Thesis.NAomitReordered_Binary_BirthYear.df, REML =
FALSE)
> summary(ZeroPrt_LMER_0)

Linear mixed model fit by maximum likelihood . t-tests use
Satterthwaite's method ['lmerModLmerTest']
Formula: Pronounced_or_Not ~ 1 + (1 | Speaker)
Data: LongFourDialects_Thesis.NAomitReordered_Binary_BirthYear.df

      AIC      BIC   logLik deviance df.resid
 3370.7   3388.8 -1682.4   3364.7     3040

Scaled residuals:
   Min       1Q   Median       3Q      Max
-2.0598  0.3282  0.4605  0.6999  0.8286

Random effects:
 Groups   Name      Variance Std.Dev.
 Speaker (Intercept) 0.005357 0.07319
 Residual                0.175369 0.41877
Number of obs: 3043, groups: Speaker, 18

Fixed effects:
              Estimate Std. Error  df      t value  Pr(>|t|)
(Intercept)  0.78694     0.02098 17.22085    37.5    <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

I then added the additional predictors (particle, prefecture, speaker age, speaker gender, sentence(clause)-final particle use, and nominal honorific prefix use) sequentially using the `update()` function (see R script in Appendix D for details). Finally, the models were compared using ANOVA:

```
> anova(ZeroPrt_LMER_0, ZeroPrt_LMER_1, ZeroPrt_LMER_2, ZeroPrt_LMER_3,
ZeroPrt_LMER_4, ZeroPrt_LMER_5, ZeroPrt_LMER_6)
```

Data: LongFourDialects\_Thesis.NAomitReordered\_Binary\_BirthYear.df

Models:

ZeroPrt\_LMER\_0: Pronounced\_or\_Not ~ 1 + (1 | Speaker)

ZeroPrt\_LMER\_1: Pronounced\_or\_Not ~ Particle + (1 | Speaker)

ZeroPrt\_LMER\_2: Pronounced\_or\_Not ~ Particle + (1 | Speaker) + Prefecture

ZeroPrt\_LMER\_3: Pronounced\_or\_Not ~ Particle + (1 | Speaker) + Prefecture +  
scale(Birth\_Year)

ZeroPrt\_LMER\_4: Pronounced\_or\_Not ~ Particle + (1 | Speaker) + Prefecture +  
scale(Birth\_Year) + Sex

ZeroPrt\_LMER\_5: Pronounced\_or\_Not ~ Particle + (1 | Speaker) + Prefecture +  
scale(Birth\_Year) + Sex + Sentence\_Final\_Particle\_Present

ZeroPrt\_LMER\_6: Pronounced\_or\_Not ~ Particle + (1 | Speaker) + Prefecture +  
scale(Birth\_Year) + Sex + Sentence\_Final\_Particle\_Present +  
Honorific\_Prefix\_Present

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
ZeroPrt_LMER_0	3	3370.7	3388.8	-1682.3	3364.7			
ZeroPrt_LMER_1	9	2833.6	2887.8	-1407.8	2815.6	549.1080	6	<b>&lt; 2e-16 ***</b>
ZeroPrt_LMER_2	12	2829.0	2901.3	-1402.5	2805.0	10.5753	3	<b>0.01426 *</b>
ZeroPrt_LMER_3	13	2826.5	2904.7	-1400.2	2800.5	4.5411	1	<b>0.03309 *</b>
ZeroPrt_LMER_4	15	2825.8	2916.1	-1397.9	2795.8	4.7262	2	0.09413 .
ZeroPrt_LMER_5	17	2828.7	2931.1	-1397.3	2794.7	1.0421	2	0.59391
ZeroPrt_LMER_6	19	2829.8	2944.2	-1395.9	2791.8	2.8710	2	0.23800

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Three of the models, LMER\_1 (particle), LMER\_2 (prefecture), and LMER\_3 (speaker birth year), were promising as they had significant p-values (less than 0.05), with low AIC and BIC. Models LMER\_4 (speaker gender), LMER\_5 (sentence-final particle), and LMER\_6 (honorific prefix) did not have significant p-values, indicating that there was no significant correlation between particle omission and speaker gender, sentence-final particle use, or nominal honorific prefix use. Because the visuals (presented in Section 3.4) also indicated these factors did not have much impact on omission, I chose to include only particle, prefecture, and birth year as predictor variables in the final model. A summary of the final model is as follows:

```
> summary(ZeroPrt_LMER_3)
Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
method ['lmerModLmerTest']
Formula: Pronounced_or_Not ~ Particle + (1 | Speaker) + Prefecture +
scale(Birth_Year)
Data: LongFourDialects_Thesis.NAomitReordered_Binary_BirthYear.df

      AIC      BIC   logLik deviance df.resid
2826.5  2904.7 -1400.2  2800.5     3030

Scaled residuals:
      Min       1Q   Median       3Q      Max
-2.69295 -0.08285  0.34703  0.58251  1.85554

Random effects:
Groups   Name             Variance Std.Dev.
Speaker (Intercept) 0.0003354 0.01832
Residual                0.1467072 0.38302
```

Number of obs: 3043, groups: Speaker, 18

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(> t )	
(Intercept)	1.023e+00	3.266e-02	5.500e+01	31.306	< 2e-16	***
<b>ParticleE</b>	<b>-1.559e-01</b>	<b>5.376e-02</b>	<b>3.029e+03</b>	<b>-2.900</b>	<b>0.00376</b>	<b>**</b>
<b>ParticleGA</b>	<b>-2.546e-01</b>	<b>3.116e-02</b>	<b>3.043e+03</b>	<b>-8.169</b>	<b>4.52e-16</b>	<b>***</b>
<b>ParticleNI</b>	<b>-8.119e-02</b>	<b>3.036e-02</b>	<b>3.043e+03</b>	<b>-2.674</b>	<b>0.00752</b>	<b>**</b>
ParticleNO	-3.028e-02	3.012e-02	3.034e+03	-1.005	0.31483	
<b>ParticleO</b>	<b>-5.559e-01</b>	<b>3.141e-02</b>	<b>3.041e+03</b>	<b>-17.697</b>	<b>&lt; 2e-16</b>	<b>***</b>
<b>ParticleWA</b>	<b>-1.667e-01</b>	<b>2.896e-02</b>	<b>3.039e+03</b>	<b>-5.755</b>	<b>9.50e-09</b>	<b>***</b>
<b>PrefectureHokkaido</b>	<b>-1.375e-01</b>	<b>2.701e-02</b>	<b>4.669e+00</b>	<b>-5.093</b>	<b>0.00460</b>	<b>**</b>
<b>PrefectureOsaka</b>	<b>-8.257e-02</b>	<b>2.642e-02</b>	<b>7.791e+00</b>	<b>-3.125</b>	<b>0.01460</b>	<b>*</b>
PrefectureTokyo	-1.783e-02	2.871e-02	4.754e+00	-0.621	0.56316	
<b>scale(Birth_Year)</b>	<b>-2.863e-02</b>	<b>9.183e-03</b>	<b>5.661e+00</b>	<b>-3.117</b>	<b>0.02231</b>	<b>*</b>

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:

	(Intr)	PrtcE	PrtcGA	PrtcNI	PrtcNO	PrtcO	PrtcWA	PrfctH	PrfctO	PrfctT
ParticleE	-0.347									
ParticleGA	-0.675	0.383								
ParticleNI	-0.670	0.391	0.682							
ParticleNO	-0.678	0.397	0.688	0.705						
ParticleO	-0.640	0.380	0.658	0.676	0.681					
ParticleWA	-0.709	0.406	0.714	0.734	0.738	0.709				
PrefctrHkkd	-0.523	-0.017	0.067	0.025	0.036	0.015	0.027			
PrefctrOsk	-0.522	0.001	0.065	0.034	0.044	0.027	0.041	0.578		
PrefctrTky	-0.482	-0.059	0.037	0.027	0.018	0.010	0.055	0.563	0.555	
scl(Brth_Y)	0.075	0.011	0.002	-0.021	-0.020	-0.024	-0.030	-0.249	0.068	-0.063

In the final model, we can see that the particles *-e*, *-ga*, *-ni*, *-o*, and *-wa* have the most significance with significant p-values. This is consistent with what we saw in Section 3.4.1, and as expected based on the graphs, we can see that the particles *-o*, *-wa*, and *-ga* are the most significant. The particle *-no* had the highest p-value (and least significance). This summary also confirms that Hokkaido was the most significant of the dialects, which matches what was observed in Section 3.4.2. There also is some significance in speaker birth year, although not as much as other variables.

### 3.7 ANOVA Results Discussion

My first hypothesis presented in 3.1, that particle omission varies by dialect and that some particles would exhibit more omission than others, was upheld; both the visual plots and the mixed ANOVA test confirm there is statistically significant correlation with dialect and particle. The hypothesis that both the factors of gender and age would also influence omission was not upheld; in fact, gender had little difference visible in the plots as well as no significant interaction in the `lmer()` stepwise test. This is not particularly surprising, given more recent studies in language and gender that indicate there is not as much difference between the language used by men and women and in fact many other factors are highly relevant (c.f. Didi-Ogren 2020). The age variable did have some correlation however, upholding part of the hypothesis. The correlation with age was not as obvious from the charts presented in Section 3.4 (likely due to the large range of ages represented), but age did exhibit a significant p-value. This outcome makes sense; older speakers are often in a position to speak more informally in more situations than younger speakers, which could result in more particle omission. (In Japan, the age of the addressee has a large influence on speech formality choices, with more formal speech being required of young speakers speaking to older speakers). The finding that certain particles were more likely to be omitted than others seems

to hold with Takano's (1998) analysis which only looked at *-wa* and *-ga* due to the frequency with which they are omitted as compared with other particles. Finally, my hypothesis that the grammatical factors (sentence-final particle use and nominal honorific prefix use) would not impact particle omission was upheld; indeed, both in the plots presented in Section 3.4 and the results of the statistical analysis demonstrate no correlation between these factors and frequency of particle omission.

This analysis produced some interesting results, and it is my hope that it will provide a point for further research in Japanese dialectal variation. For instance, it would be interesting to further annotate the data used in this thesis (or other similar data) to include information about the use of each particle, such as the discourse context, intent of the utterance, and so forth. The broad level of annotation used in this thesis presents an interesting overview of the correlations of each variable with particle omission, but individual variables should be more closely analyzed. It would also be interesting to apply the results of this analysis to a larger study which contains data from more dialects, or to apply similar methods to newly collected data. The results presented in this chapter inform the syntactic analysis and discussion presented in Chapter 4.

## CHAPTER 4 – THE SYNTACTIC ANALYSIS

### 4.1 Overview of Chapter 4

While there have been numerous attempts at explaining argument ellipsis (cf. Nariyama 2003, Saito 2016), there has been far less work done on nominal particle omission. However, this is an important phenomenon for understanding the structure of the Japanese language. Many recent studies, including Fry 2003, Shimojo 2006, and Takano 1998, have attempted to explain it. However, as Fry (2003: 119) concluded, “...theories that rely on a single categorical explanation...are doomed to inadequacy,” and it seems that a holistic explanation of the phenomenon has not yet been found. Some of the more comprehensive studies have been corpus-based. Takano (1998) examined the ellipsis of *-wa* and *-ga* using a small corpus of conversation data); Fry (2003) examined various ellipsis phenomena using a corpus of telephone conversations; and Kibe et al. (2020) examined *-o* ellipsis with a large amount of COJADS data. The findings from the quantitative analysis presented in Chapter 3 will be used to inform the preliminary syntactic analysis presented in this chapter. A brief summary of the findings from the quantitative analysis are as follows:

- 1) Speaker gender does not impact the frequency of nominal particle omission;
- 2) Speaker age does impact the frequency of nominal particle omission;
- 3) Dialectal variation in particle omission is observed;
- 4) Some particles exhibit more frequent omission than others, but all those examined exhibited at least some omission;

- 5) The presence of sentence-final particles and honorific prefixes does not impact particle omission.

These findings, along with the topics discussed in Chapter 2, demonstrate the complex puzzle that the nominal particle omission presents. In order to move towards a solution to this puzzle, the following questions will be addressed in this chapter:

- 1) What is the best way to treat nominal particle omission syntactically?
- 2) Should Case particles and Postpositional particles be treated separately in the syntax?

These questions will be addressed in sections 4.2 to 4.3, and ideas for further study are presented in Section 4.4.

## **4.2 Summary of the Particle Omission Puzzle**

Japanese nominal arguments are marked with post-nominal particles which can be considered clitics (cf. Nakamura 2018). Two of these are typically considered to be Case markers; *-ga* (NOM) and *-o* (ACC). Other particles (such as *-de* ‘in/at’ and *-e* ‘to’) are often considered to be postpositions, rather than Case markers. The particle *-ni* is often glossed as a DAT Case particle, but it patterns more closely with *-de* and *-e* which are considered to be Postpositional particles. The Topic Marker *-wa* is considered a discourse marker, but sometimes patterns with the NOM and ACC Case particles *-ga* and *-o* and sometimes with the discourse marker *-mo* ‘too’. (See Section 4.3.4 for additional details.)

As demonstrated by the analysis in Chapter 3, all of the nominal particles can be omitted in colloquial Japanese (contrary to prior claims that only Case particles can be omitted, as

discussed in 2.5), and this presents a problem for generative analyses of Japanese nominals. How can we account for nominal particles when we know any can be omitted, not just Case particles? Should the two types of particles be treated in the same manner in the syntax, or separately? Ultimately, how can we explain particle omission syntactically? In addition, the results presented in Chapter 3 indicate that variation in the frequency of particle omission occurs based on dialect and speaker age, and some particles are omitted more frequently than others. While this last point cannot be explained via the syntax, the two questions presented above will be answered in the following sections.

### 4.3 Towards a Syntactic Analysis of Particle Omission

This section will address the two questions presented in 4.1 and 4.2, looking at examples from Standard Japanese (SJ). A preliminary proposal for treating the phenomenon syntactically (presented in the form of a discussion of several possible solutions), as well as whether to treat the two types of particles (Case and Postpositional) together, will be presented in this section.

#### 4.3.1 Option 1: *Particles are the D head of DP*

Based on the discussion presented in Section 2.5.1, I propose that particles should be treated as the head D of a DP, as seen in Figure 7 (copied here as Figure 25)<sup>15</sup>. As discussed in Section 2.5.1, this structure has some advantages over the KP structure proposed by Fukuda (1993), as well as advantages over the possible combined KP and DP structure suggested in Section 2.5.1. The combined DP and KP option (Fig. 6) results in a left-headed DP; as Japanese is a right-headed language, it would not be appropriate to have a left-headed DP when all other phrases (to my knowledge) are right-headed. Japanese demonstratives act less like determiners in other languages

---

<sup>15</sup> Note that in Figures 7 and 25 (which are identical), I do not intend for A to be a true complement to N. APs are likely adjuncts; however, AP must adjoin lower than the demonstratives (such as *sono* 'that') and closer to N. Further research is needed in order to definitively posit a structure for adjuncts such as APs.

and may in fact be nominal phrases (as discussed in Section 2.5.1), so it follows that they should not be placed in D (see Section 4.4.1 for further discussion). The final motivation is that genitives in some languages (e.g., the English genitive clitic ‘s), are spelled out in the D slot and attach to the possessor in Spec DP when D bears a genitive Case feature. It follows that other Case markers could occur in D, especially ones that function as clitics more so than suffixes (cf. Nakamura 2018).

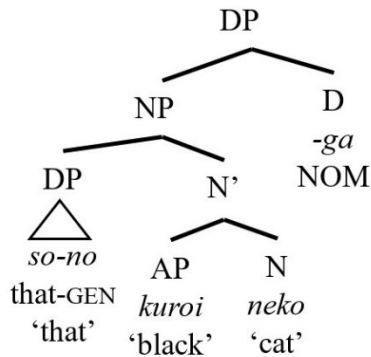


Figure 25. Proposed DP Structure

If we then assume that post-nominal particles occur in D, we next must explain how these particles in head D can be omitted; as seen in Chapter 3, all particles can optionally be omitted. Under the proposal made by Fukuda (1993), particles (which he argued headed KP) could elide when properly head-governed under the Empty Category Principal (ECP) (see Section 2.5.1 for a more detailed discussion of Fukuda’s proposal). Although the ECP is no longer part of the modern Minimalist framework (Chomsky 1995, 2000, 2001, 2013, etc.), the analysis Fukuda presented was concise and worked well in the Government & Binding framework (Chomsky 1981).

In Minimalism, phases (see Section 2.6 for an introduction to phases) seem to provide the best possible solution in line with Fukuda’s analysis. As discussed in Section 2.6, we could consider DP to be phase-defining in Japanese following the treatment of DPs as phases by

Svenonius (2004). Additionally, an extension of the ECP to Minimalist terms is that a non-pronominal empty category must be properly licensed in its phase, or as part of the next higher phase if it is at the phase edge (Chomsky 2001). If we additionally assume that phases (which can be used as a locality constraint for both movement and anaphoric relations according to Lee-Schoenfeld 2008, and explain argument ellipsis according to Sakamoto 2016) explain how empty categories are licensed, then we can also assume that particles (which occur at their phase edge as the D) would be properly licensed empty categories.

For non-subjects we could therefore consider the empty category in D to be properly licensed at the edge of its phase by the lexical head that subcategorizes for the DP; D is at the edge of its phase, with no phase-boundary between the lexical head and D which therefore allows the D to elide. This would mean that DPs in the predicate would be licensed by the lexical head V (and would be licensed prior to moving under any application of scrambling), similarly to the proposal made by Fukuda (1993). Example 22 and Figure 4 demonstrate how this might work for a non-subject DP. In this example, the lexical head V (*iku* ‘go’) first licenses D, because D is at the edge of the DP phase allowing it to be visible to V. Because of this, the D can optionally be an empty category, as it is properly head-licensed by the higher lexical head V (see Figure 26 for an illustration).

- (22) gakkoo=e/# itta yo  
 School=to went FP  
 ‘(They) went to school.’

*(Example 22 is adapted from Lee 2002a, page 648, Example 4.)*

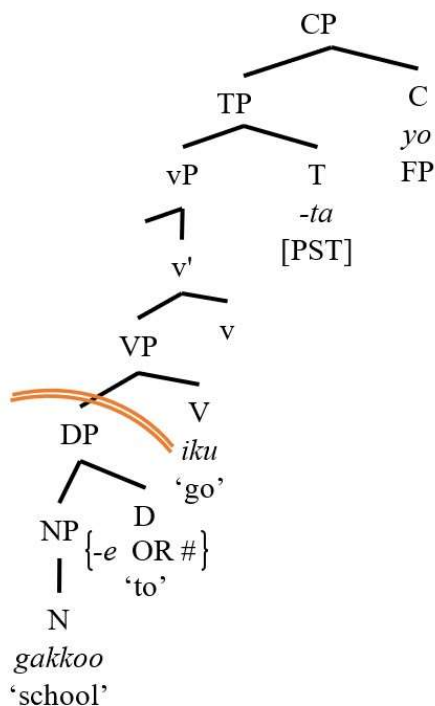


Figure 26. Structure of Example 22

Subjects present a possible problem for this analysis, as there would be no lexical head available to license the subject DP. However, if subjects in Japanese originate in the same location as in English (Spec vP), as suggested in Sakamoto 2016,<sup>16</sup> it seems plausible that once the V moves up to join with T, T becomes the “lexical” head which licenses the empty category in D of subject DPs. Example 23, and the corresponding structure in Figure 27 demonstrate how this might work for a subject DP. If we assume that vP is a phase in Japanese (following Sakamoto 2016), then there are two phase boundaries intervening between D and the V once V is in T. However, head-to-head movement is assumed to be maximally local, meaning that a head can only move to the next-higher head (Travis 1984). Therefore, in Figure 26, V must move first to v, and then the newly combined v and V complex moves to T to obtain Tense features. When V stops at v, this would

<sup>16</sup> This was not directly stated by Sakamoto 2016 but instead was implied by the examples he provided.

allow the subject DP to be properly head-licensed and the D to optionally contain an empty category (an omitted particle).

(23) *basu=ga/# kita*

*bus=NOM/# came*

‘The bus came. (The bus is here.)’

*(Adapted from Lee 2002b, page 698, Example 40c.)*

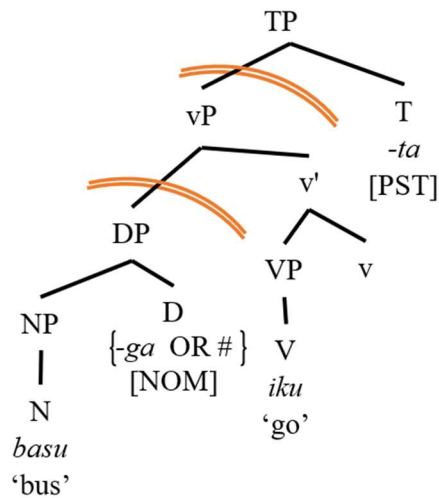


Figure 27. Structure of Example 23

A possible problem with this approach is the fact that other studies argue that Case particles in Japanese make the constituent “opaque” or invisible to ‘search’ (c.f. Miyagawa et al. 2018 and Saito 2016). If this is in fact true, then the Case particles would not be able to elide via the methods suggested here, as they would not be available to search by a higher head and therefore would not be properly licensed. However, it seems that such arguments were made under the assumption that Case particles are not their own heads but instead simply nominal suffixes which means they would

attach directly to N. As we are assuming the particles are the head D and at the phase edge of DP, then this problem should be avoided under the proposed analysis. However, further research into this should be conducted.

#### 4.3.2 Option 2: *Only What is Necessary is Merged*

A second option is that we could simply follow a basic idea of the Minimalist program that only what is necessary is merged (e.g. Chomsky 2000, 2001, and 2013). If we then posit that particles occur in D, then it follows that only NP would merge in if there is no particle in head D available to merge. If there is no D merging in, then no DP would be labeled and the result would be a bare NP (with adjuncts that might occur in the NP). This can be seen in (24) and the corresponding Figure 28 and (25) and corresponding Figure 29. However, if a particle is present, then the result would be a DP, as demonstrated in (26) and Figure 30.

(24) shibazuke=# tabe-tai!

shibazuke eat-want

‘I want to eat Shibazuke (pickles)!’

*(Adapted from Lee 2002b, page 666, Example 33c.)*

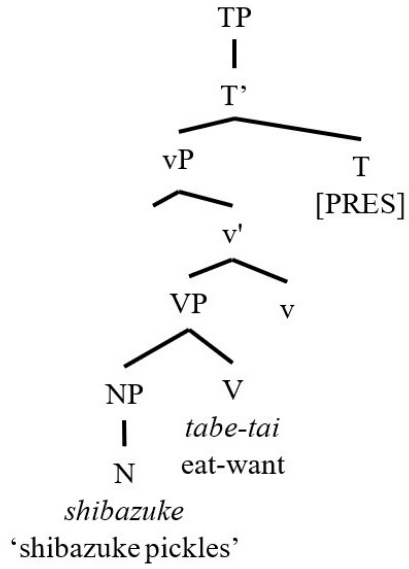


Figure 28. Structure of Example 24

(25) sono hon=# koko=# nai yo  
 that book here absent FP  
 ‘That book is not here.’

*(Adapted from Lee 2002a, page 652, Example 11a)*

(26) sono hon=ga koko=ni nai yo  
 that book=NOM here=LOC absent FP  
 ‘That book is not here.’

*(Adapted from Lee 2002a, page 652, Example 11a’)*

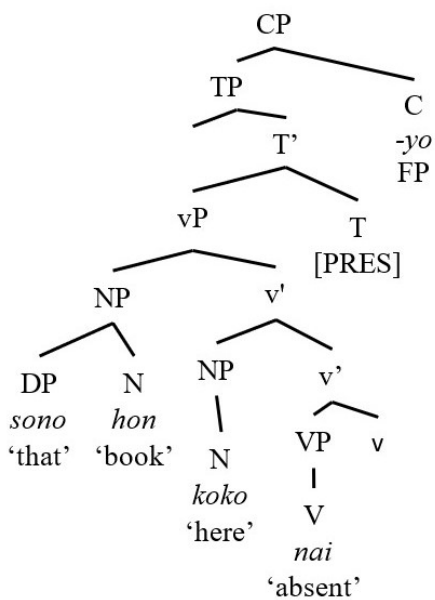


Figure 29. Structure of Example 25

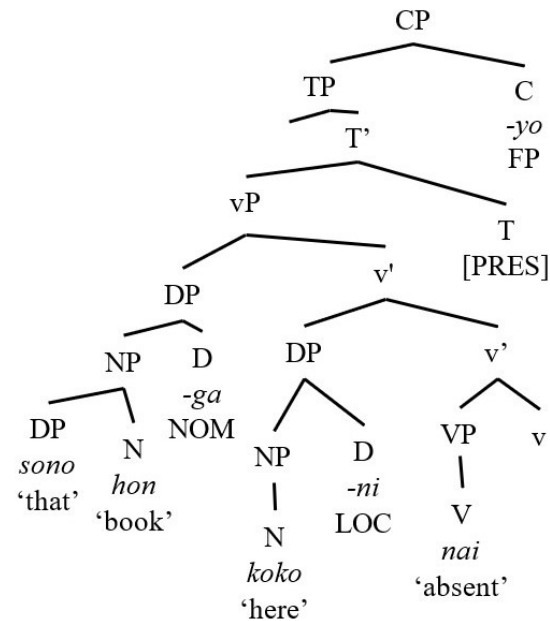


Figure 30. Structure of Example 26

Under this proposal, there would be no need to explain the phenomenon in terms of ECP-like analyses because there would simply be no empty category present; this is a nice model as we can rely solely on the Merge function and avoid adding any additional mechanisms.

This option could explain cases where two particles can be used together consecutively on the same nominal; the extra D (or K) would simply merge in after the first. Example 27 and Figure 31 demonstrate how such constructions could be treated under this option.

- (27) gakkoo=e=mo itta yo  
 school=to=also went FP  
 ‘(They) went to school too.’

(Adapted from Lee 2002a, Example 11d, page 652.)

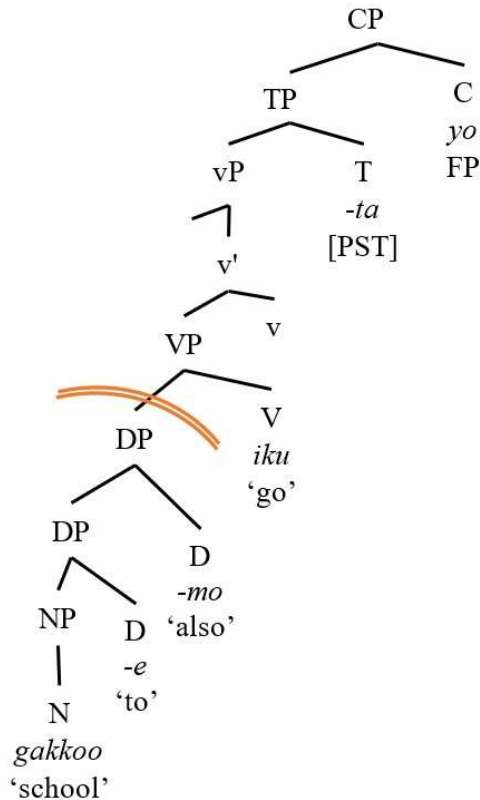


Figure 31. Structure of Example 27

However, whether or not we consider omitted particles to be “zero particles” or “particle ellipsis,” this option partially fails to capture both types. If we accept the “zero particle” hypothesis, we would expect there to be a DP structure present where the null particle occurs. If there were a zero particle, then we would expect to find an empty category in D, which was base generated and not based on a specific particle. However, under this option, we would find a bare NP, without a DP. Similarly, we lose the ability to reconstruct an elided particle (if we accept the “particle ellipsis” hypothesis); if there is no DP structure present, how can we determine which particle was elided? In both cases of particle omission, because there is no DP structure present in the derivation, it becomes difficult to determine that a particle was present at all.

What this option could capture are examples like (28b), where we see a bare nominal being uttered in answer to a question and (28c) where an NP occurs immediately before a copula with no particle between the NP and copula.

(28) a. Q: dare=ga kita?

Who=NOM came

‘Who showed up?’

b. A: tomodachi

friend

‘a friend’

c. gakusei desu

student COP

‘(I’m) a student.’<sup>17</sup>

This option is also reminiscent of the argument surrounding whether some languages lack DPs (and only have NPs), as explained in great detail by Köylü (2021). At a broad level, there is an ongoing argument related to a typological analysis presented by Bošković (2008) whereby languages exhibiting certain features were NP languages, and those without were DP languages. As explained by Köylü (2021), numerous researchers have pointed out problems with this typological treatment, and Köylü proposed that the best solution is that the determination must be made on a language-specific basis; a typological treatment is not ideal. This option is reminiscent

---

<sup>17</sup> The examples in (18) are my own.

of this argument, in that it proposes both a DP and NP option for Japanese – a language that is expected (according to the typology of Bošković 2008) to be an NP-only language.

Although this option has many merits, we need an analysis that easily manages particle omission as a whole (whether or not we consider these missing particles as zero particles or ellipsis it is clear a DP structure is necessary to capture a missing particle) and utterances containing bare nominals (such as copular constructions as in 18c).

#### *4.3.3 Option 3: Phases and Merge Explain Everything*

A possible solution to the shortcomings presented by Option 2 is to combine Option 1 (DP is a Phase, and an empty category in D can be properly head-licensed by a lexical head in the next Phase) and Option 2 (Merge only what is necessary). This would allow us to see a DP structure for particle omission (whether a case of particle ellipsis or the zero particle), and still account for cases with bare NPs. It would also allow us to explain cases of double particles, as described in 4.3.2. I therefore propose that the best solution, based on the limited scope of this thesis, is to explain the phenomenon in terms of both Merge and Phases. For the sake of simplicity, I will assume that particles occur in D (although the GEN particle *-no* and the topic marker *-wa* present complications to this and will be discussed briefly in Section 4.3.6).

As previously mentioned (Section 4.3.2), “double” particles can also be explained by the concept that only what is necessary is Merged. An interesting point made by Guekguezian (2021) (and previously by Dikken 2007) is that of Phase Extension, whereby a phase expands as more elements are Merged and the structure expands. If this is true, then this could help explain how the second particle in a “double” particle construction (typically a discourse marker, *-wa* or *-mo*), has scope over the whole DP that precedes them (see examples 29-33). Under the Phase Extension

theory, we could posit that the second DP is still in the same Phase as the first, which could help capture this phenomenon as well.

A summary of my proposal is therefore as follows:

- a) Nominal particles reside in D, head of DP
- b) The nominal particles in D can be omitted (whether an omitted particle is treated as a zero particle or particle ellipsis);
- c) Only what is necessary is merged (when no D is needed, a bare NP results);
- d) DP is a Phase in Japanese;
- e) D is at the phase edge, and therefore visible to a higher lexical head;
- f) An empty category (here, an omitted particle) in D is properly head-licensed by a higher lexical head (v+V or T+v+V in the case of subjects and V in the case of objects);
- g) Finally, phases expand as more structure is added (Phase Extension) (when a second D is merged, the resulting new DP is still a phase).

A final outstanding question presented in Section 4.1 is that of whether so-called Case and Postpositional particles can be treated in the same way in the syntax, even though they are different semantically. It is clear from the data examined in Chapter 3 that Case particles and Postpositional particles do not completely pattern in the same way (for instance, *-ga* and *-o* tend to be missing more often than other particles). In addition, according to Lee 2002a, and as demonstrated by Examples 29-30, Postpositional particles can either combine with or be replaced by the discourse markers *-wa* and *-mo*.

(29) a. sono hon koko=**ni=wa** nai yo.

that book here=**LOC=TOP** absent FP

b. sono hon koko=**wa** nai yo.

that book here=**TOP** absent FP

‘As for that book, it is not here.’

(30) a. gakkou=**e=mo** itta yo.

School=**to=also** went fp

b. gakkou=**mo** itta yo.

school=**also** went FP

‘(He) also went to school.’

*(Examples 29-30 are from Lee 2002a, Example 11, page 652.)*

However, according to Lee 2002a, *-ga* and *-o*, when used with discourse particles (such as *-wa*, *-mo*, or *-shikia*), must be replaced by the discourse particles as demonstrated by Examples 31-33.

(31) a. Tarou=**wa** kita yo.

Tarou=**NOM** came FP

‘As for Tarou, he has come.’

b. \*Tarou=**ga=wa** kita yo.

Tarou=**NOM=TOP** came FP

Intended: ‘As for Tarou, he has come.’

(32) a. Ken=**mo** kita yo.

Ken=**also** came FP

‘Ken has come too.’

b. \*Ken=**ga=mo** kita yo.

Ken=NOM=**also** came FP

Intended: ‘Ken has come too.’

(33) a. kono eiga=**shika** mitenai yo

this movie=**only** see.NEG FP

‘I have seen only this film.’

b. \*kono eiga=**o=shika** mitenai yo

this movie=ACC=**only** see.NEG FP

Intended: ‘I have seen only this film.’

*(Examples 31-33 adapted from Lee 2002a, page 651, example 9.)*

#### 4.3.4 Extending the Analysis to the COJADS Dialect Data

As demonstrated in Table 1 and discussed in Section 2.4, the phonological realization of the particles does not vary greatly from the phonological form of Standard Japanese. The Tokyo and Osaka dialect data suggest that the [g] in NOM *-ga* is devoiced resulting in [ka]. In addition, phonological reduction is observed for some particles in the Tokyo, Osaka, and Kitakyushu data (See Section 2.4 for details). However, these phonological variations do not appear to present much of a complication for the proposed analysis (see Section 4.4.3 for discussion of a possible complication relating to phonological reduction). I will therefore posit that the analysis works

identically in Standard Japanese and each of the dialects examined in Chapter 3, although further research is necessary to confirm this.

#### 4.4 Complications to the Proposal

Due to the limited scope of this thesis project, there are of course many reasons this proposal may prove to be wrong, and further research is certainly necessary. This section presents some possible complications to the proposal, and suggested solutions to those.

##### 4.4.1 The Genitive Particle and Demonstratives

The genitive particle *-no* presents a complication to the proposal outlined in Section 4.3.3; this particle is embedded inside of a larger nominal constituent, and typically acts as a possessive marker or as a link between two nouns. Sakamoto (2016) suggested that a possessive phrase occupies Spec N, as in Figure 32.<sup>18</sup>

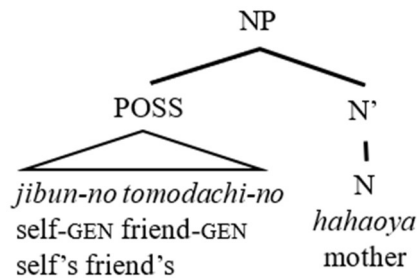


Figure 32. Structure of NPs with Possessive Phrases per Sakamoto 2016

Example 34 demonstrates that AP and possessive phrases must occur in different locations in the structure. As mentioned briefly in a footnote above, an AP is likely just an adjunct. In (34), we can see that the AP *kanashii* ‘sad’ must occur between the possessive phrase and N if the target

<sup>18</sup> Figure 32 was adapted from Example (32) on page 259 of Sakamoto 2016.

modified noun is *tomodachi* ‘friend’; it cannot occur before the possessive phrase. It would be possible to place an AP before the possessive phrase but only when the intended reading is that the head of the possessive phrase is modified (e.g. *kuroi neko=no sakana* ‘the black cat’s fish’).

(34) a. *jibun=no kanashii tomodachi=ga*

self=GEN sad friend=NOM

‘self’s sad friend’

b. \**kanashii jibun=no tomodachi=ga*

sad self=GEN friend=NOM

Intended: ‘self’s sad friend’<sup>19</sup>

If we adapt the idea that Possessive phrases occur in Spec N to the proposal presented in 4.3.3, the structure in Figure 33 can be proposed for (34a). Note that the AP is inserted into the sister of N position here for simplicity; however, it is an adjunct and would likely not occur in the sister of N slot.

---

<sup>19</sup> Examples in (34) are based on Sakamoto 2016, page 258, Example 30b; I added the adjective *kanashii* ‘sad’ and used a different predicate to shorten the sentence. The judgement in 25b regarding placement of the adjective is my own.

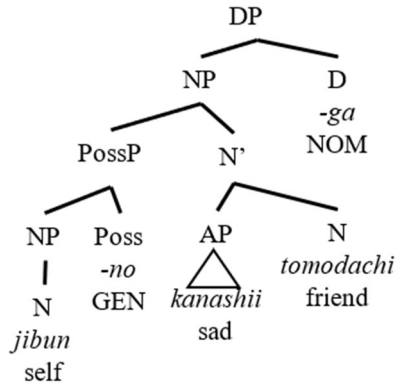


Figure 33. Structure of Example 25a

Placing a PossP in Spec N is very straightforward; but, based on this simple structure, it looks like it would also be possible to place these possessive phrases into a DP. However, the GEN particle cannot be replaced by any other particle. Therefore, it does not seem appropriate to state that it occurs in the same head as the other particles. By placing all particles including GEN *-no* into a DP, it also becomes difficult to visualize the intuition that particles (other than GEN *-no*) have scope over their whole constituent, and phases could further be complicated if multiple DPs are found embedded in a “main” DP. Figure 34 shows what such a complicated structure might look like for Example 35. (Note the brackets indicating phrase divisions are Sakamoto’s, but they match my own intuitions.) Figure 35 shows what this same phrase would look like using a combination of the proposal made in 4.3.3 and Sakamoto’s regarding the location of possessive phrases.

- (35) [[jibun=no tomodachi]=no hahaoya]=ni sensei=ga aitagatteiru to itta.  
 Self=GEN friend= GEN mother=DAT teacher=NOM wants.to.see C said  
 ‘Hanako told the mother of self’s friend that the teacher wanted to see (her friend)’  
 (Sakamoto 2016, page 258, Example 30b.)

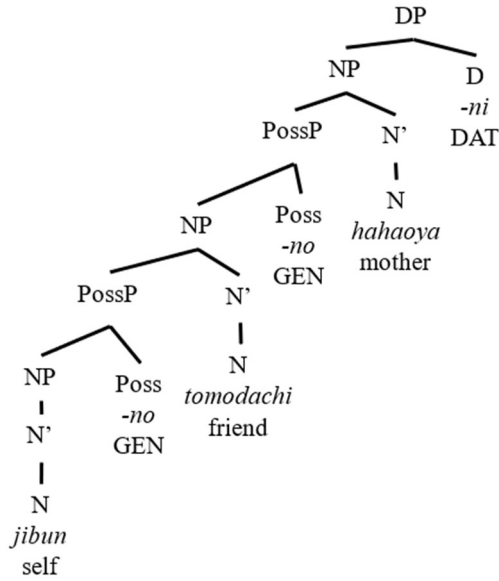


Figure 34. Structure of possessive phrase in (35), PossP and DP

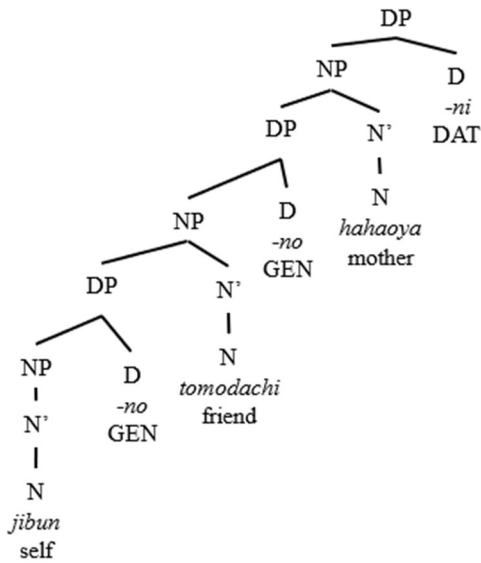


Figure 35. Structure of possessive phrase in (35), DP-only

As we can see in Figure 34, only one D occurs for the whole DP; this D is able to take scope over the whole nominal phrase, including any possessive phrases. This matches the brackets provided by Sakamoto (2016) on the original example (as seen in Example 35). Additionally, it is

structurally clear that it is difficult to replace the GEN marker located in Poss with any other particle, as all other particles (that have been examined in this thesis) would occur in D.

A third option would be to treat the possessive phrases as a DP which occurs not as a complement to N, but in Spec DP (just like English possessives). For example 35, this might look like the structure in Figure 36.

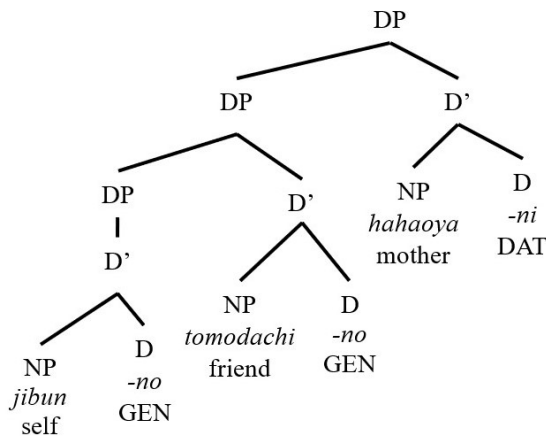


Figure 36. Structure of Example 35, PossP is a DP in Spec DP

This option may be the most concise, as it allows us to posit that possessive phrases in Japanese work similarly to English possessive phrases and ultimately it is easy to explain how the particle *-no* morphologically (and perhaps intuitively to native speakers) resembles the other particles. By positing that it is also D, we are maintaining that generalization. In addition, we can still see that the possessive phrase is more closely connected to the main NP than adjuncts are and indeed also that the possessive phrases take scope over NP and anything contained within NP as an adjunct.

A final point that must be made is that demonstratives ending in *-no* (such as *sono* ‘that’, *kono* ‘this’) likely also occur in Spec DP inside a DP. As discussed in Section 2.5.1, the demonstratives in Japanese could be considered noun phrases consisting of an NP and genitive particle (cf. Shibatani 1990). Because the demonstratives modify a noun, and can occur within a string of other possessive phrases (as in Example 13, copied here as 36), it seems plausible that they would also occupy a Spec DP slot and occur in a DP. The other demonstratives (such as *kore* ‘this’, *sore* ‘that’) likely also occur in a DP and may also occur in Spec DP.

(36) *ookina so=no issatsu=no Tarou=no hon*  
 big **that=GEN** one.CL=GEN Tarou=GEN book  
 ‘That big book of Tarou’s’

*(Adapted from Shibatani 1990, page 258)*

Example (37) and Figure 37 demonstrate how all the nominal elements (demonstratives, possessive phrases, adjective phrases, D and N) could fit into the structure based on my proposal and with the assumption that possessive phrases occur in Spec DP.

(37) *[[sono kodomo]=no [kawaii inu]=wa asondeiru.*  
 that child=GEN cute dog=TOP playing  
 ‘That child’s cute dog is playing.’<sup>20</sup>

---

<sup>20</sup> Example (37) is my own.

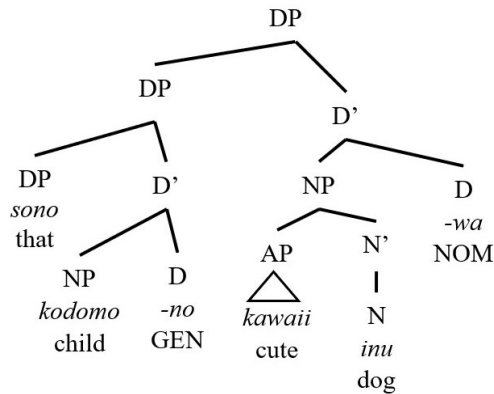


Figure 37. Structure of the subject constituent of Example 37<sup>21</sup>

A final indicator that the GEN particle should be treated separately from the other particles in the syntax (at least to some extent, even by simply positing that it can only occur in a DP located in Spec DP) is that the results presented in Chapter 3 indicated that *-no* was only omitted about 8% of the time in the data. This was the next-to-lowest rate of omission, second only to *-de* which may also have a different structure in some situations (for example when it acts as an instrumental as opposed to a LOC). (The particle *-de* should also be examined in greater detail in this regard, but this is beyond the scope of this thesis and will be left for further inquiry.)

However, we do need to explain how *-no* can be omitted at all. Under the PossP option, the GEN particle is in a location that is quite embedded within the DP phase, which should make it difficult (perhaps impossible) to omit. We can more easily explain the ability of this particle to elide by proposing that the possessive phrase is actually a DP in Spec DP; under this proposal, the GEN particle would occur in Spec DP which is at the edge of the DP as a whole (and could therefore be omitted as previously discussed). However, it would also be the head of its own

<sup>21</sup> Note that in Figure 37, the AP is an adjunct, as discussed and is placed in Spec N primarily to illustrate that the AP must occur within the NP (not DP); however, no claims are made as to the best treatment of such adjuncts.

embedded DP phase, which could present complications – and which might explain the apparent difficulty of omitting the genitive particle. A possible way to investigate this problem further would be to look at how frequently other particles elide when attached to a nominal phrase that includes a possessor, but this will be left for future research.

#### *4.4.2 The Discourse Particles*

One possible problem for my proposal is how the discourse markers, especially *-mo* and *-wa*, should be handled. As briefly discussed in Section 4.3.3, these particles can co-occur with most particles, but must replace *-ga* and *-o*. As discussed in 4.3.2, we could propose that the second particle simply merges in following the merge of the first. However, this does not consider the fact that nominal phrases typically are fronted when topicalized (marked with *-wa*). This does not happen with *-mo*, so perhaps *-mo* merges after the first DP as proposed in 4.3.2. In Japanese, the typical treatment is that the subject remains in Spec vP, and moves to Spec TP (or Spec CP) when topicalized. For *-wa*, perhaps the nominal argument is attracted by the TOP feature (which is likely on Spec T, as suggested by the diagrams presented by Sakamoto 2016 which posit the Subject raising to Spec T when topicalized). The TOP *-wa* would then merge with the DP, and if the original D is incompatible with *-wa* (*-ga* or *-o*), it would be elided; if the original D were compatible then a new DP level would merge to include the TOP *-wa*. However, this is a very tentative suggestion and further investigation into this is certainly needed.

#### *4.4.3 The Suffix Possibility*

For the purposes of this thesis, I assumed that particles are clitics and not suffixes, as suggested by Nakamura (2018). In addition to Nakamura's suggestion, some phonological reduction was observed in the data (discussed in Section 2.4 and 4.3.4). Based on the fact that phonological interactions are observed between word boundaries in other languages, it seems

plausible that even if these particles operate as clitics more than suffixes, we could still find phonological reduction.

However, as previously mentioned, it may prove necessary to treat them as suffixes if arguments made by Miyagawa et al. (2018) and Saito (2016) regarding argument ellipsis hold true; although their focus is on argument ellipsis in Japanese, the proposal that Case suffixes make an NP opaque could prevent Phases from being used as a viable explanation of particle omission. Further study is needed in light of other syntactic phenomena in Japanese, and if it is determined that the suffix option is the best treatment for these particles, then another analysis of particle omission would be necessary.

#### **4.5 Revised Proposal & Recap of Outstanding Questions**

Below is a revised version of the proposal presented in 4.3.3, incorporating the ideas discussed in Section 4.4.1 and 4.4.2. Although there are more questions to be addressed regarding the discourse markers, I tentatively include a point regarding these in this revised proposal, in line with the discussion presented in 4.4.2.

- a) Nominal particles (*-wa*, *-ga*, *-o*, *-ni*, *-de*, *-e*, *-no*) reside in D, head of DP;
- b) The genitive particle *-no* can only occur in a DP that is in Spec DP position;
- c) The nominal particles in D can be overt or empty (unpronounced);
- d) Only what is necessary is merged (when no D is needed such as in a copular construction, a bare NP results);
- e) DP is a phase in Japanese;
- f) D is at the phase edge, and therefore visible to the next phase and a higher lexical head;

- g) An empty category (omitted particle) in D is properly head-licensed by a higher lexical head (v+V or T+v+V in the case of subjects and V in the case of objects);
- h) Phases expand as more structure is added (Phase Extension) (when a second D is merged, the resulting new DP is still a Phase);
- i) A topicalized DP is likely moved up to a Spec T or Spec C position in order to obtain the topic marker *-wa*, with *-wa* still filling D.

Although this now incorporates one of the key complications the original proposal presented, there are still several outstanding questions that should be addressed via future research. In particular, would it be better (in light of other syntactic phenomena) to treat the particles as true suffixes, and if so, what is the best way to accommodate that in the syntax in light of particle omission? How should the discourse markers be handled? Does a TOP feature in Spec T or Spec C draw the DP up, as discussed in 4.4.2? I will leave these questions, and the many others unanswered here, for future research.

## **CHAPTER 5 - CONCLUSION**

Japanese nominal particle omission (which encompasses both the “zero particle” and “particle ellipsis” hypotheses) presents a puzzle for generative linguistics. This thesis set out to investigate this phenomenon by examining data from four Japanese dialects (Hokkaido, Tokyo, Osaka, and Kitakyushu) using a two-part analysis (quantitative and syntactic). The two main research questions were: Does evidence from the COJADS data uphold prior conclusions that particle omission varies by dialect, extra-linguistic and linguistic factors? How can we explain the phenomenon, and any variation, via a traditional generative analysis in the Minimalist framework (Chomsky 1995, 2000, 2001, 2013, etc.)?

In order to answer these questions, I first performed a quantitative analysis of natural conversation data of four Japanese dialects which was obtained from the companion metadata website of the Corpus of Japanese Dialects (COJADS) (NINJAL 2021). The quantitative analysis answered the first main research question with the goal of examining variation in particle omission with respect to dialect, particle, speaker age, speaker gender, presence of nominal honorific prefixes, and use of sentence-final particles. My hypotheses were specifically that: 1) the frequency of particle omission varies by dialect and some particles would exhibit more frequent particle omission than others; 2) the social factors of speaker age and gender would affect the frequency of particle omission and may interact with any dialectal variation; 3) the linguistic factors of sentence-final particle use and honorific prefix use would not impact frequency of nominal particle omission.

After presenting detailed graphs and tables to visualize these possible correlations, I conducted a mixed-ANOVA analysis in R in order to check for statistical significance of the effects of each factor on particle omission. The results of this test, combined with what was observed in the charts and tables, upheld the first and third hypotheses, but only part of the second hypothesis was upheld. In particular, speaker gender did not correlate with frequency of particle omission while speaker age did. Dialectal variation and variation by specific particle were observed. The presence of sentence-final particles and honorific prefixes did not correlate with particle omission.

Following the quantitative analysis, I performed a syntactic analysis in the Minimalist framework (Chomsky 1995, 2000, 2001, 2013, etc.). This analysis answered the second main research question, and more specifically addressed the following questions: 1) What is the best way to treat nominal particle omission syntactically, in the Minimalist framework? 2) Should Case particles and Postpositional particles be treated separately in the syntax?

In order to answer these questions, I explored several options for treating the phenomenon syntactically, and arrived at a tentative syntactic analysis couched in the Minimalist framework. The proposed analysis is essentially that first, nominal particles (which I take to be clitics following Nakamura 2018) occur in D, head of DP. Possessive phrases are DPs located in Spec DP and headed by the genitive particle *-no* which is head D. Demonstratives similarly occur in DPs in Spec DP. A particle in D can be omitted (whether “elided” or a “zero particle”, it constitutes an Empty Category) as D is at the phase edge of DP and therefore visible to the next Phase and higher lexical head. If a bare nominal occurs, then there is no need for a DP to be merged, as there is simply no D to be merged. Although this analysis appears to work for the examples I examined, further research is needed to verify that it works in all situations and in particular in light of other syntactic phenomena in Japanese.

Additionally, I presented some topics that should be investigated further; in particular, the topic marker *-wa* and discourse marker *-mo*, as well as the idea that the particles could be suffixes rather than clitics. Although further study is needed, I did achieve my goal of presenting a preliminary look at how we might address particle omission from a more modern standpoint (Minimalism) while attempting to take into consideration the results of a quantitative analysis.

Although this thesis arrived at some interesting results, it is by necessity still quite small in scope, and could therefore be improved in many ways. Although I have briefly mentioned in some sections ways to further improve the study, some additional methods are briefly discussed here. First, it was difficult to obtain negative evidence; I was only able to do this by consulting the literature, and by relying on my own knowledge of Japanese (which, as a non-native speaker, is imperfect). However, the ideal situation would be one in which at least one native speaker of each dialect was available to consult and test grammaticality judgements in order to confirm theories about the syntactic nature of this phenomenon. Second, after completing the syntactic analysis I understand that I should not have lumped the genitive marker *-no* with the other particles in the overall summary charts in Chapter 3; in future, this particle should likely be split out from the rest due to its slightly different syntactic structure. Finally, I was only able to annotate the data by particle name, and not by the case/use in each line, due to the difficulty of identifying grammatical use in all cases. Here again, a native speaker consultant (per dialect) would be helpful. Additionally, it should be noted that due to the large scale of the data and method of hand-annotating it, there could be errors in my annotations (despite careful review of the completed spreadsheet).

Despite, or indeed because of, these shortcomings, this thesis was intended to be of a more exploratory nature; I am hesitant to make any hard claims regarding my conclusions and suggested syntactic treatment of the phenomenon. However, it is my hope that this thesis encourages further

investigation into the field of Japanese dialectal variation especially as it relates to syntactic and morphological variation. I also hope that this thesis will add to a growing body of research in syntax which is informed by quantitative data; syntax does not operate in a vacuum, and many factors should be considered as demonstrated here. Future studies on this phenomenon or building on this thesis should attempt to arrive at a more complete description of particle omission, taking into account many more factors than were examined here; for example, some factors that may interact with particle omission include semantic and contextual constraints, composition of the audience and interlocutors, predicate type, honorific and formality levels of the predicate or sentence as a whole, and mood/aspect. The discourse particles, *-wa* and *-mo* should also be closely and separately examined, as should the distance between an omitted particle and the verb. Due to the exploratory nature of this thesis, it was rather broad; future studies might therefore take a narrower focus, in order to focus on a small subset of the factors examined or suggested here. Two final suggestions for further study are that more data contained in the COJADS corpus could be examined; more data would allow a better picture of the phenomenon. A final suggestion is that more data, with younger speakers, could be collected in a similar manner as that collected for the COJADS corpus. This would enable an investigation into generational change, and topics such as leveling towards the Standard could be examined.

Despite the many ways this research project could be improved or extended, I arrived at many interesting conclusions, and I came to better understand many aspects of the Japanese language as it relates to syntax, morphology, and sociolinguistics. Although I was able to at least tentatively answer all of the research questions I set out to answer, there is still much to learn, and many questions remain to be answered. It is my hope that this thesis will provide a basis for future

research into Japanese particle omission and Japanese dialectal variation, and that future research will continue to explore this phenomenon.

## REFERENCES

- Ahlström, Kim, Miwa Ahlström & Andrew Plummer. 2022. Jisho. <https://jisho.org/> (March, 2022).
- Bates, Douglas, Martin Maechler, Ben Bolker & Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1). 1-48.  
doi:10.18637/jss.v067.i01.
- Bilous, Frances R. & Robert M. Krauss. 1988. Dominance and accommodation in the conversational behaviours of same- and mixed-gender dyads. *Language & Communication* 8(3). 183-94. doi: 10.1016/0271-5309(88)90016-X.
- Bošković, Željko. 2008. What will you have, DP or NP? *Proceedings of the North East Linguistics Society (NELS)* 37(1). 101-114.
- Celik, K. & N. Kibe. 2019. Raising language diversity awareness in Japan through web-based open access application. International Conference on Language Documentation & Conservation (ICLDC 6). University of Hawai'i at Manoa. Presentation Slides.  
<http://kikigengo.ninjal.ac.jp/en/index.html> (March, 2022).
- Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.
- Chomsky, Noam. 1995. A Minimalist Program for Linguistic Theory. In *The Minimalist program*. Cambridge, Massachusetts: The MIT Press.
- Chomsky, Noam. 2000. Minimalist enquiries. In Roger Martin, David Michaels & Juan Uriagareka (eds.), *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*, 89-156. Cambridge, Massachusetts: MIT Press.

- Chomsky, Noam. 2001. Derivation by Phase. In Michael Kenstowicz (ed.), *Ken Hale. A Life in Language*, 1-52. Cambridge, Massachusetts: MIT Press.
- Chomsky, Noam. 2013. Problems of projection. *Lingua* 130. 33-49. doi: 10.1016/j.lingua.2012.12.003.
- Collins, Chris & Edward Stabler. 2016. A Formalization of Minimalist Syntax. *Syntax* 19(1). 43-78. doi: 10.1111/synt.12117.
- Coupland, Nikolas. 1980. Style-Shifting in a Cardiff Work-Setting. *Language in Society* 9(1). 1-12. <http://www.jstor.org/stable/4167108>.
- Dib, Firas. 2021. Regular Expressions 101. <https://regex101.com/> (April 2021).
- Didi-Ogren, Holly. 2020. Sociocultural linguistic approaches to code switching in Japanese women's talk in interaction: Region, gender, and language. *Multilingua* 36(6). 685-720. doi: 10.1515/multi-2018-0096.
- Dikken, Marcel den. 2007. Phase Extension Contours of a theory of the role of head movement in phrasal extraction. *Theoretical Linguistics* 33(1). 1-41. doi:10.1515/TL.2007.001.
- Eckert, Penelope & Sally McConnell-Ginet. 1992. Think Practically and Look Locally: Language and Gender as Community-Based Practice. *Annual Reviews of Anthropology* 21. 461-90.
- Endo, Yoshio. 2007. *Locality and Information Structure: A cartographic approach to Japanese*. (Linguistik Aktuell/Linguistics Today). John Benjamins Publishing Company. doi: 10.1075/la.116.
- Fry, John. 2003. *Ellipsis and wa-marking in Japanese conversation*. New York, NY: Routledge.
- Fukuda, Minoru. 1993. Head Government and Case Marker Drop in Japanese. *Linguistic Inquiry* 24(1). 168-72. Cambridge, MA: The MIT Press. <http://www.jstor.org/stable/4178806>.

- Guekguezian, Peter Ara. 2021. Aspectual phase heads in Muskogee verbs. *Natural Language & Linguistic Theory* 39(4). 1129-1172. doi: 10.1007/s11049-020-09495-7.
- Haibuihoang. 2015. Japanese Translator. RomajiDesu: Japanese Dictionary & Translator. Ver. 4.1. <http://www.romajidesu.com/translator> (April 2021).
- Hasegawa, Yoko. 2018a. *The Cambridge handbook of Japanese linguistics*. New York, NY: Cambridge University Press.
- Hasegawa, Yoko. 2018b. Introduction. In Yoko Hasegawa (ed.), *The Cambridge handbook of Japanese linguistics*, 3-14. New York, NY: Cambridge University Press.
- Hoji, Hajime, Satoshi Kinsui, Yukinori Takubo & Ayumi Ueyama. 2003. The demonstratives in modern Japanese. In Andrew Simpson & Audrey-Li Yen-hui (eds.), *Functional Structure(s), Form and Interpretation: Perspectives from East Asian Languages*, 97-128. London: RoutledgeCurzon Taylor & Francis Group.  
<https://doi.org/10.4324/9780203987667>
- Hornstein, Norbert. 2009. *A Theory of Syntax: Minimal Operations and Universal Grammar*. Cambridge: Cambridge University Press.
- Igarashi, Yosuke. 2007. Typology of Prosodic Phrasing in Japanese Dialects. *Workshop on Intonational Phonology: Understudied or Fieldwork Languages, ICPHS 2007 Satellite Meeting*. Saarland University, Saarbrücken: Germany.  
<https://linguistics.ucla.edu/people/jun/Workshop2007ICPhS/Papers/YosukeIgarashi.pdf>  
(March 2022).
- Kibe, Nobuko, Kumiko Sato, Taro Nakanishi & Kohei Nakazawa. 2020. Corpus-based study of Japanese dialects: Regional differences in accusative case marking system. In Asahi Yoshiyuki (ed.), *Proceedings of Methods XVI: Papers from the Sixteenth International*

- Conference on Methods in Dialectology 2017*. (Bamberger Beiträge zur Englischen Sprachwissenschaft [Bamberg Studies in English Linguistics] 59). Berlin: Peter Lang GmbH. doi: 10.3726/b17102.  
[https://researchmap.jp/KIBE\\_Nobuko/published\\_papers/27685987](https://researchmap.jp/KIBE_Nobuko/published_papers/27685987) (27 January 2022).
- King, Sara, Yi Ren, Kaori Idemaru & Cindi Sturtzsreetharan. 2021. Sounding like a father: The influence of regional dialect on perceptions of masculinity and fatherhood. *Language in Society*. 1-24. doi:10.1017/S0047404520000925.
- Kishimoto, Hideki. 2017. Case Marking. In Shibatani Masayoshi, Miyagawa Shigeru, Noda Hisashi (eds.), *Handbook of Japanese Syntax*. De Gruyter Mouton.  
doi:10.1515/9781614516613.
- Koiso, Hanae, Masayuki Asahara, Salvatore Carlino, Ken'ya Nishikawa, Kazuki Aoyama, Yuichi Ishimoto, Aya Wakasa, Michiko Watanabe, Yoshimi Yoshikawa, Nobuko Kibe & Kikuo Maeka. 2019. Speech corpora in NINJAL, Japan demonstration of corpus concordance systems: Chunagon and Kotonoha. *Proceedings of LPSS [Linguistic Patterns in Spontaneous Speech] 2019*. 8-12. Taipei: Institute of Linguistics, Academia Sinica. doi: 10.15084/00003051.
- Köylü, Yılmaz. 2021. An overview of the NP versus DP debate. *Language and Linguistics Compass* 15(3). e12406. 10.1111/lnc3.12406.
- Kuznetsova, Alexandra, Per B Brockhoff & Rune H. B. Christensen. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82.1-26. doi: 10.18637/jss.v082.i13.
- Lee-Schoenfeld, Vera. 2008. Binding, Phases, and Locality. *Syntax* 11(3). 281-98. doi: 10.1111/j.1467-9612.2008.00118.x.

- Lee, Duck-Young. 2002a. The function of the zero particle with special reference to spoken Japanese. *Journal of Pragmatics* 34. 645-82. doi: 10.1016/S0378-2166(02)00042-5.
- Lee, Kiri. 2002b. Nominative case-marker deletion in spoken Japanese: an analysis from the perspective of information structure. *Journal of Pragmatics* 34(6). 683-709. doi: 10.1016/S0378-2166(02)00048-6.
- Levon, Erez. 2014. Categories, stereotypes, and the linguistic perception of sexuality. *Language in Society* 43. 539-66. doi: 10.1017/S0047404514000554.
- McCloskey, James. 2000. Quantifier Float and Wh-Movement in an Irish English. *Linguistic Inquiry* 31(1). 57-84.
- Miyagawa, Shigeru, Danfeng Wu & Masatoshi Koizumi. 2018. A Revised Labeling Approach to Long-Distance Agreement. Paper presented to the MAPLL × TCP × TL × TaLK (MT3), 2018. <http://lingphil.mit.edu/papers/dfwu/MT3%20Proceedings.pdf> (February 2022).
- Nakagawa, Natsuko, Masahiro Yamada, Nobuko Kibe & Yukinori Takubo. 2019. Archiving System of Endangered Languages in Japan: A Preliminary Report. *Proceedings of the Language Technologies for All (LT4All)*. 371-374. UNESCO Headquarters, Paris: European Language Resources Association (ELRA). <https://lt4all.elra.info/proceedings/lt4all2019/pdf/2019.lt4all-1.93.pdf> (January 2022).
- Nakamura, Wataru. 2018. Case. In Yoko Hasegawa (ed.), *The Cambridge handbook of Japanese linguistics*, 250-75. New York, NY: Cambridge University Press.
- Nariyama, Shigeko. 2003. *Ellipsis and reference tracking in Japanese*. (Studies in Language Companion Series 66). Philadelphia, PA: John Benjamins Publishing Company.
- Negre, Xavier. 2021. Multilingual keyboard: Conversion Katakana <> Romaji. Lexilogos. [https://www.lexilogos.com/keyboard/katakana\\_conversion.htm](https://www.lexilogos.com/keyboard/katakana_conversion.htm) (April 2021).

- Nishiguchi, Sumiyo. 2009. Quantifiers in Japanese. In Peter Bosch, David Gabelaia, Jérôme Lang, *Logic, Language, and Computation (TbiLLC 2007)*, Berlin: Springer.  
[https://doi.org/10.1007/978-3-642-00665-4\\_13](https://doi.org/10.1007/978-3-642-00665-4_13).
- The National Institute for Japanese Language and Linguistics/ 国立国語研究所 (NINJAL).  
2021. 日本語諸方言コーパス(Corpus of Japanese Dialects/COJADS). Version 2021.01.  
<https://www2.ninjal.ac.jp/cojads/index.html?targ=data> (Files downloaded March 2021; site last accessed March 2022).
- Okamoto, Shigeko. 2008. The Use of “Regional” and “Standard” Japanese in Conversations: A Case Study from Osaka. In Junko Mori & Amy Snyder Ohta, *Japanese Applied Linguistics: Discourse and Social Perspectives*. 132-59. United Kingdom: Continuum.
- Oku, Satoshi. 1998. LF Copy Analysis of Japanese Null Arguments. *Proceedings of Chicago Linguistic Society CLS 34*. 299-314. Chicago: Chicago Linguistic Society.  
<http://hdl.handle.net/2115/44867> (16 February 2022).
- Onishi, T. 2017. Dialects of Japanese. In Charles Boberg, John Nerbonne, Dominic Watt, *The Handbook of Dialectology*. (Blackwell Handbooks in Linguistics). Wiley. doi:  
10.1002/9781118827628.ch35.
- Podesva, R. J. 2007. Phonation type as a stylistic variable: The use of falsetto in constructing a persona. *Journal of Sociolinguistics* 11. 478-504. doi: 10.1111/j.1467-9841.2007.00334.x.
- RStudio Team. 2021. RStudio: Integrated Development Environment for R. Version 1.4.1103. Boston, MA: RStudio, PBC. <http://www.rstudio.com/> (February 2022).
- Saito, Mamoru. 2016. (A) Case for labeling: labeling in languages without  $\phi$ -feature agreement. *The Linguistic Review* 33(1). 129-175. doi: 10.1515/tlr-2015-0017.

- Saito, Mamoru. 2017. Ellipsis. In Shibatani Masayoshi, Miyagawa Shigeru, Noda Hisashi, *Handbook of Japanese Syntax*. De Gruyter Mouton. doi: 10.1515/9781614516613.
- Sakamoto, Yuta. 2016. Phases and argument ellipsis in Japanese. *Journal of East Asian Linguistics* 25. 243-274. doi: 10.1007/s10831-016-9145-6.
- Sharma, Devyani. 2018. Style dominance: Attention, audience, and the 'real me'. *Language in Society* 47(1). 1-31. Cambridge: Cambridge University Press. doi: 10.1017/S0047404517000835.
- Shibamoto, Janet S. 1985. *Japanese women's language*. Orlando, FL: Academic Press.
- Shibatani, Masayoshi. 1990. *The Languages of Japan*. (Cambridge Language Surveys). Cambridge: Cambridge University Press.
- Shimojo, Mitsuaki. 2005. *Argument encoding in Japanese conversation* New York: Palgrave Macmillan.
- Shimojo, Mitsuaki. 2006. Properties of particle "omission" revisited. *Toronto Working Papers in Linguistics* 26. 123-140. <https://twpl.library.utoronto.ca/index.php/twpl/article/view/6178>
- Sum, Enirac. 2008. Map of Japanese Dialects (English version). [https://en.wikipedia.org/wiki/Japanese\\_dialects#/media/File:Japanese\\_dialects-en.png](https://en.wikipedia.org/wiki/Japanese_dialects#/media/File:Japanese_dialects-en.png) (April 2021).
- Svenonius, Peter. 2004. On the Edge. In David Adger, Cécile de Cat & George Tsoulas, *Peripheries: Syntactic Edges and their Effects*. (Studies in Natural Language and Linguistic Theory 59). Springer, Dordrecht. doi: 10.1007/1-4020-1910-6\_11.
- Takano, Shoji. 1998. A quantitative study of gender differences in the ellipsis of the Japanese postpositional particles -wa and -ga: Gender composition as a constraint on variability. *Language Variation and Change* 10(3). 289-323. 10.1017/S0954394500001344.

- R Core Team. 2020. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Travis, Lisa. 1984. *Parameters and effects of word order variation*. Cambridge, MA: Massachusetts Institute of Technology Doctoral Dissertation.
- Trudgill, Peter. 1972. Sex, covert prestige and linguistic change in the urban British English of Norwich. *Language in Society* 1. 179-195. doi: 10.1017/S0047404500000488.
- Tsujimura, Natsuko. 2014. *An Introduction to Japanese Linguistics*. 3<sup>rd</sup> Edition. United Kingdom: Wiley Blackwell.
- West, Candace & Don H. Zimmerman. 1987. Doing Gender. *Gender and Society* 1(2). 125-151. Sage Publications, Inc. <http://www.jstor.org/stable/189945>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Golemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo & Hiroaki Yutani. 2019. Welcome to the tidyverse. *Journal of Open Source Software* 4(43). 1686. doi: 10.21105/joss.01686.

## APPENDICES

### Appendix A. Abbreviations and Transcription System

#### List of Abbreviations

A	Adjective	N	Noun
ACC	Accusative	NOM	Nominative
AP	Adjective Phrase	NP	Noun Phrase
CONJ	Conjunction	P	Preposition/Postposition
D	Determiner	PP	Prepositional/Postpositional Phrase
DAT	Dative	PRES	Present
DP	Determiner Phrase	Q	Question
ECP	Empty Category Principle	T	Tense
FP	Focus Particle	TOP	Topic
GEN	Genitive	TP	Tense Phrase
I	Inflection	V	Verb
IP	Inflection Phrase	VP	Verb Phrase

#### Transcription System Used

The Hepburn orthography (current *romaji* system) is used in this thesis, as opposed to the older *kunrei* system. The Hepburn style is more consistent with newer publications (such as *Nariyama 2003*), and also with how native speakers actually pronounce sounds (especially consonants). For example, I would transcribe *zibun* as *jibun*. In the data itself (discussed in Chapter 3), I had to use different conventions to ensure readability in R; for example, long *o* was transcribed as *oo* rather than the Hepburn standard *ou*.

## Appendix B. Sample Datasets

### Example of Original COJADS Dataset (Hokkaido)

This is a sample of the raw data downloaded from the COJADS (NINJAL 2021) companion website, <https://www2.ninjal.ac.jp/cojads/index.html?targ=data>, which was used in the analysis presented in Chapter 3. This is a small sample of the Hokkaido file, 01\_b\_099.csv, which was downloaded in March 2021. The data version at the time of download was Ver.2021.01.

県	地点	file番号	地点ID	ID	話者	方言テキスト	標準語テキスト	データ名	収録年月日	収録場所	編集担当者	話者生年	話者年齢
北海道	中川郡豊頃町	1b	1A			トシクレート ユーノウ マー	歳末 というのは (F: まあ)	ふるさとことば集成	1978年10月20日	豊頃町牛首別 二宮生活館	佐藤亮一、江川清、田原広史、井上文字	1907年	71歳
北海道	中川郡豊頃町	1b	2A			ライネン/ ジュンピダト オモーンダナ。	来年の 準備だと 思うんだね。	ふるさとことば集成	1978年10月20日	豊頃町牛首別 二宮生活館	原広史、井上文字、佐藤亮一、江川清、田	1907年	71歳
北海道	中川郡豊頃町	1b	3C			ハイ。	はい。	ふるさとことば集成	1978年10月20日	豊頃町牛首別 二宮生活館	原広史、井上文字、佐藤亮一、江川清、田	1912年	66歳
北海道	中川郡豊頃町	1b	4A			ンー	うん	ふるさとことば集成	1978年10月20日	豊頃町牛首別 二宮生活館	佐藤亮一、江川清、田原広史、井上文字	1907年	71歳
北海道	中川郡豊頃町	1b	5A			デ ドンナ テードニ マー	それで どのな 程度に (F: まあ) (2 PL: あなたたち) [は] (1 SG: 私) の 考え [で] は (F: まあ)	ふるさとことば集成	1978年10月20日	豊頃町牛首別 二宮生活館	原広史、井上文字、佐藤亮一、江川清、田	1907年	71歳
北海道	中川郡豊頃町	1b	6A			アンタカ 'タ オレノ カンカ 'エウ マー		ふるさとことば集成	1978年10月20日	豊頃町牛首別 二宮生活館	原広史、井上文字、佐藤亮一、江川清、田	1907年	71歳
北海道	中川郡豊頃町	1b	7A			アー イマ ユー トリ マー	(F: まあ) 今 言う とおり (F: まあ)	ふるさとことば集成	1978年10月20日	豊頃町牛首別 二宮生活館	原広史、井上文字、佐藤亮一、江川清、田	1907年	71歳
北海道	中川郡豊頃町	1b	8A			ユキ フル クニナ モンダガラー	雪 [が] 降る 国だ ものだから	ふるさとことば集成	1978年10月20日	豊頃町牛首別 二宮生活館	原広史、井上文字、佐藤亮一、江川清、田	1907年	71歳
北海道	中川郡豊頃町	1b	9A			ア ニワノ フユカ 'コイガラ	(F: まあ) 庭の 冬園いから 始めて だんだんと 正月の 元旦には 休むんだと い	ふるさとことば集成	1978年10月20日	豊頃町牛首別 二宮生活館	原広史、井上文字、佐藤亮一、江川清、田	1907年	71歳
北海道	中川郡豊頃町	1b	10A			ムンダト ユー ゴトデ イッシューケンメー	う ことで 一生懸命 [に]	ふるさとことば集成	1978年10月20日	豊頃町牛首別 二宮生活館	原広史、井上文字、佐藤亮一、江川清、田	1907年	71歳
北海道	中川郡豊頃町	1b	11A			ユギー フル フランウ ベツニ シテモ	雪が 降る 降らないは 別に しても おそらく 降る 年も あれば 降らない 年も あった	ふるさとことば集成	1978年10月20日	豊頃町牛首別 二宮生活館	原広史、井上文字、佐藤亮一、江川清、田	1907年	71歳
北海道	中川郡豊頃町	1b	12A			ツタワ。	よね。	ふるさとことば集成	1978年10月20日	豊頃町牛首別 二宮生活館	佐藤亮一、江川清、田	1907年	71歳

### Sample of Transcribed Dataset (Hokkaido)

This is a sample of the COJADS data contained in the above screenshot, transcribed into Roman characters and annotated for particle omission and other variables as described in Chapter 3.

Prefecture	Speak er	Line ID	Dialect_Text	SJ_Translation	Nominal_Particle_Possible	Nominal_Particle_Omitted	WA	GA	O	NI	DE	E	NO	Honorific_Prefix_Present	Sentence_Final_Partic le_Present	Birth_Year	Age	Sex	Date
Hokkaido	H_A	H-1	toshikuree to yuu no wa maa	saimatsu to iu no wa (F: maa)	NO	NA	NA	NA	NA	NA	NA	NA	NA	NH	NP	1907	71	male	Hon Lan Coll
Hokkaido	H_A	H-2	rainen no jumbida to omoonda na.	rainen no jumbi da to omoonda ne .	YES	NO	NA	NA	NA	NA	NA	NA	PRESENT	NH	YES	1907	71	male	Hon Lan Coll
Hokkaido	H_C	H-3	hai.	hai .	NO	NA	NA	NA	NA	NA	NA	NA	NA	NH	NP	1912	66	female	Hon Lan Coll
Hokkaido	H_A	H-4	n-	un	NO	NA	NA	NA	NA	NA	NA	NA	NA	NH	NP	1907	71	male	Hon Lan Coll
Hokkaido	H_A	H-5	de donna teedo ni maa	sorede donna teido ni (F: maa)	NO	NO	NA	NA	NA	NA	NA	NA	NA	NH	NO	1907	71	male	Hon Lan Coll
Hokkaido	H_A	H-6	antaka-ta [wa] ore no kanka-e [de] wa maa	[2 PL: anatatachi] [wa] [1 SG: watashi] no kangae [de] wa (F: maa)	YES	YES	OMITTED	NA	NA	NA	OMITTED	NA	PRESENT	NH	NO	1907	71	male	Hon Lan Coll
Hokkaido	H_A	H-7	aa ima yuu tori maa	(F: aa) ima iu toori (F: maa)	NO	NA	NA	NA	NA	NA	NA	NA	NA	NH	NO	1907	71	male	Hon Lan Coll
Hokkaido	H_A	H-8	yuki [ga] furu kunina mondagaraa	yuki [ga] furu kunida mondakara	YES	YES	NA	OMITTED	NA	NA	NA	NA	NA	NH	NO	1907	71	male	Hon Lan Coll
Hokkaido	H_A	H-9	a riwa no fuyuka-koigara	(F: aa) riwa no fuyu kakoi kara	YES	NO	NA	NA	NA	NA	NA	NA	PRESENT	NH	NO	1907	71	male	Hon Lan Coll
Hokkaido	H_A	H-10	dete dandan to shyooka-tsu no gantan ni wa yasumunda to yuu goto de issyookernee [n]	hajimele dandan to shyogatsu no gantan ni wa yasumu nda to iu kotode issyookernee [n]	YES	YES	PRESENT	NA	NA	PRESENT	NA	NA	PRESENT	NO	NO	1907	71	male	Hon Lan Coll
Hokkaido	H_A	H-11	yugii [ga] furu furan wa betsu ni shitemoo	yuki [ga] furu furanai wa betsu ni shite mo	YES	YES	PRESENT	OMITTED	NA	NA	NA	NA	NA	NH	NP	1907	71	male	Hon Lan Coll
Hokkaido	H_A	H-12	osoraku furu toshimo areba furan toshi mo attawana.	osoraku furu toshi mo areba furanai toshi mo attayone .	YES	NO	NA	NA	NA	NA	NA	NA	NA	NH	YES	1907	71	male	Hon Lan Coll

Sample of Long-form Dataset (Hokkaido) after R Processing:

This is a sample of the COJADS data after it was processed in R and turned into a “long-form” dataset, as described in Chapter 3.

Pronounced_ or_Not	Particle	Dialect_Text	SJ_Translation	Nominal_Particle_Ossible	Nominal_Particle_Omitted	Honorific_Prefix_Present	Sentence_Final_Particle_Present	Line_ID	Prefecture	Speaker	Birth_Year	Age	Sex	Data_Name	Recording_Place	Collection_Position	Topic
OMITTED	WA	antaka-ta [wa] ore no kanka-e [de] wa maa	(2 PL: anatachi ) [wa] (1 SG: watashi ) no kangae [de] wa (F: maa)	YES	YES	NH	NO	H-6	Hokkaido	H_A	1907	71	male	Hometown Language Collection	Toyokoro Town Ushshubetsu Niromiya Sekatsukan	Toyokoro Town, Nakagawa District	Annual ev of life in ok
PRESENT	WA	dete dandan to shyooka-tsu no gantan ni wa yasumunda to yuu goto de isshyokenmee [ni]	hajimete dandan to shoogatsu no gantan ni wa yasumi nda to iu kotode isshyokenmee [ni]	YES	YES	NO	NO	H-10	Hokkaido	H_A	1907	71	male	Hometown Language Collection	Toyokoro Town Ushshubetsu Niromiya Sekatsukan	Toyokoro Town, Nakagawa District	Annual ev of life in ok
PRESENT	WA	yugi [ga] furu furan wa betsu ni shitemoo	yuki [ga] furu furanai wa betsu ni shite mo	YES	YES	NH	NP	H-11	Hokkaido	H_A	1907	71	male	Hometown Language Collection	Toyokoro Town Ushshubetsu Niromiya Sekatsukan	Toyokoro Town, Nakagawa District	Annual ev of life in ok
OMITTED	WA	mukashi wa maa muronantee yuu goto ka-attero ima no yona karabi shita mono de [wa] nagatta to moonda .	mukashi wa (F: maa) shitsu nante iu koto ga atte mo ima no yona karabi shita mono de [wa] nakatta to omoonda .	YES	YES	NH	NO	H-31	Hokkaido	H_A	1907	71	male	Hometown Language Collection	Toyokoro Town Ushshubetsu Niromiya Sekatsukan	Toyokoro Town, Nakagawa District	Annual ev of life in ok
OMITTED	WA	maa uchini [wa] chisana murode	(F: maa) uchi ni [wa] chisana muro (D: de)	YES	YES	NH	NO	H-32	Hokkaido	H_A	1907	71	male	Hometown Language Collection	Toyokoro Town Ushshubetsu Niromiya Sekatsukan	Toyokoro Town, Nakagawa District	Annual ev of life in ok
PRESENT	WA	konkuri de naikara tada ana [o] hokkara mizu wa wa wakushi.	konkurito de naikara tada ana [o] < honokara = hotta daikedakara > mizu wa (D: wa) wakushi .	YES	YES	NH	NO	H-42	Hokkaido	H_C	1912	66	female	Hometown Language Collection	Toyokoro Town Ushshubetsu Niromiya Sekatsukan	Toyokoro Town, Nakagawa District	Annual ev of life in ok
OMITTED	WA	imawa sonna goto [wa] naigedona.	ima wa sonna koto [wa] naikedone .	YES	YES	NH	YES	H-51	Hokkaido	H_C	1912	66	female	Hometown Language Collection	Toyokoro Town Ushshubetsu Niromiya Sekatsukan	Toyokoro Town, Nakagawa District	Annual ev of life in ok
PRESENT	WA	oshyooka-tsu wa tanoshimi de maa wareware wa moo.	shoogatsu wa tanoshimi de (F: maa) (1 PL: watashitachi) wa moo .	YES	NO	YES	NO	H-53	Hokkaido	H_A	1907	71	male	Hometown Language Collection	Toyokoro Town Ushshubetsu Niromiya Sekatsukan	Toyokoro Town, Nakagawa District	Annual ev of life in ok
PRESENT	WA	maa ima motte ore wa gantanmairi [o] suru waa.	(F: maa) ima motte (1 SG: watashi) wa gantan mairi [o] suruyo .	YES	YES	NH	YES	H-55	Hokkaido	H_A	1907	71	male	Hometown Language Collection	Toyokoro Town Ushshubetsu Niromiya Sekatsukan	Toyokoro Town, Nakagawa District	Annual ev of life in ok
PRESENT	WA	X1santo ore wa.	(R: X 1) san to (1 SG: watashi) wa .	YES	NO	NH	NO	H-56	Hokkaido	H_A	1907	71	male	Hometown Language Collection	Toyokoro Town Ushshubetsu Niromiya Sekatsukan	Toyokoro Town, Nakagawa District	Annual ev of life in ok
PRESENT	WA	wachira mo gantanmairi wa.	(1 PL: watashitachi) mo gantan mairi wa .	YES	NO	NH	NO	H-58	Hokkaido	H_C	1912	66	female	Hometown Language Collection	Toyokoro Town Ushshubetsu Niromiya Sekatsukan	Toyokoro Town, Nakagawa District	Annual ev of life in ok
PRESENT	WA	kyonen wa j(tsu)chan nagunattagara.	kyonen wa < oisan = otto > [ga] nakunattakara .	YES	YES	NH	NO	H-62	Hokkaido	H_C	1912	66	female	Hometown Language Collection	Toyokoro Town Ushshubetsu Niromiya Sekatsukan	Toyokoro Town, Nakagawa District	Annual ev of life in ok
PRESENT	WA	oshyooka-tsu wa mada tanoshimida wa na.	shoogatsu wa mata tanoshimidayone .	YES	NO	YES	YES	H-65	Hokkaido	H_A	1907	71	male	Hometown Language Collection	Toyokoro Town Ushshubetsu Niromiya Sekatsukan	Toyokoro Town, Nakagawa District	Annual ev of life in ok

## Appendix C. R Code

This appendix presents the R code used to format and analyze data from the dataset which was derived from four individual files downloaded from the NINJAL COJADS website (see 2.2.2 and 3.2 for details). After the data was converted to Roman characters (*romaji*), the following code was performed to clean up and analyze the data. Code used to generate the plots and tables presented in Chapter 3 was omitted due to space constraints.

```
#Load the FourDialects_byParticle_ExtraFormatting_Thesis.csv file into R as a dataframe:
FourDialects_Thesis.df <- read.csv(file="E:/DATA Rachel/Documents/MA UGA/THESIS-Spring 2022/R
PROJECT_THESIS/FourDialects_byParticle_ExtraFormatting_Thesis.csv", stringsAsFactors = T)
View(FourDialects_Thesis.df)

#Use the TidyVerse package to turn the data into long format, with 1 row per particle:
library(tidyverse)
LongFourDialects_Thesis = FourDialects_Thesis.df %>%
  gather("WA", "GA", "O", "NI", "DE", "E", "NO", key = Particle, value = Pronounced_or_Not)
View(LongFourDialects_Thesis)

#Use the na.omit() function to omit all NA values:
LongFourDialects_Thesis.NAomit <- na.omit(LongFourDialects_Thesis)
View(LongFourDialects_Thesis.NAomit)

#Reorder the columns to be in a logical order, with "Particle" and "Pronounced_or_Not" first:
LongFourDialects_Thesis.NAomitReordered.df <- LongFourDialects_Thesis.NAomit[, c(21, 20, 4, 5,
6, 7, 8, 9, 3, 1, 2, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19)]
View(LongFourDialects_Thesis.NAomitReordered.df)

#Export the re-ordered dataframe:
write.csv(LongFourDialects_Thesis.NAomitReordered.df, "E:/DATA Rachel/Documents/MA UGA/THESIS-
Spring 2022/R PROJECT_THESIS/LongFourDialects_Thesis.NAomitReordered.csv", row.names = FALSE)

#####

#Import the separate csv file that contains only columns for GA, O, and NI (this file was used
to create plots showing only GA, O, and NI).
```

```

FourDialects_GAONI_Only_Thesis.df <- read.csv(file="E:/DATA Rachel/Documents/MA
UGA/THESIS-Spring 2022/R
PROJECT_THESIS/FourDialects_byParticle_ExtraFormatting_Thesis_GAONI_ONLY.csv",
stringsAsFactors = T)

View(FourDialects_GAONI_Only_Thesis.df)

#Use the TidyVerse package to turn the data into long format, with 1 row per particle:
library(tidyverse)

LONGFourDialects_GAONI_Only_Thesis = FourDialects_GAONI_Only_Thesis.df %>%
  gather("GA", "O", "NI", key = Particle, value = Pronounced_or_Not)

View(LONGFourDialects_GAONI_Only_Thesis)

#Use the na.omit() function to omit all NA values:

LONGFourDialects_GAONI_Only_Thesis.NAomit <-
na.omit(LONGFourDialects_GAONI_Only_Thesis)

View(LONGFourDialects_GAONI_Only_Thesis.NAomit)

#Reorder the columns to be in a logical order, with "Particle" and "Pronounced_or_Not"
columns first:

LONGFourDialects_GAONI_Only_Thesis.NAomitReordered.df <-
LONGFourDialects_GAONI_Only_Thesis.NAomit[, c(21, 20, 4, 5, 6, 7, 8, 9, 3, 1, 2, 10,
11, 12, 13, 14, 15, 16, 17, 18, 19)]

View(LONGFourDialects_GAONI_Only_Thesis.NAomitReordered.df)

#Export this newly re-ordered dataframe to save a backup:

write.csv(LONGFourDialects_GAONI_Only_Thesis.NAomitReordered.df, "E:/DATA
Rachel/Documents/MA UGA/THESIS-Spring 2022/R
PROJECT_THESIS/LONGFourDialects_GAONI_Only_Thesis.NAomitReordered.csv", row.names =
FALSE)

#####

#ANOVA Analysis

#Run a mixed ANOVA analysis, because the assumption of independence is violated in this data
(there are many utterances from the same speakers).

#Fit a mixed-methods model that predicts omitted vs. overt particle use using dialect as a
fixed effect, speaker as a random effect, and the other factors (age, gender, sentence-final
particle use, and honorific prefix use) as additional fixed effects.

# Check the assumptions for a mixed-effects model:

# (1)The observations in the samples are independent from one another. <- No, in this case
we have data from conversations, in which one speaker usually says many utterances.

# Because this first assumption is violated, we do not need to test the rest; this confirms
that we need to use a mixed ANOVA analysis (or repeated-measures).

# Next, choose the best model of Zero vs. Overt particles (Pronounced_or_Not). Use the final
formatted spreadsheet for the analysis (LongFourDialects_Thesis.NAomitReordered.df).

```

```

#Load the appropriate package:

library(nlme)

#Convert the output of the Particle ellipsis column (Pronounced_or_Not) into numbers so
the model will recognize them.

#Because we have binary output (Omitted vs. Present), we can convert them to binary
numerals (0 and 1).

#First confirm this column is in fact character class:

class(LongFourDialects_Thesis.NAomitReordered.df$Pronounced_or_Not)

#It is, so now convert all the Omitted and Present tokens to 0 and 1.

#Create a separate dataframe to work with:

LongFourDialects_Thesis.NAomitReordered_Binary.df <-
LongFourDialects_Thesis.NAomitReordered.df

View(LongFourDialects_Thesis.NAomitReordered_Binary.df)

#Then convert OMITTED to "0" and PRESENT to "1":
LongFourDialects_Thesis.NAomitReordered_Binary.df[LongFourDialects_Thesis.NAomitReordere
d_Binary.df == "OMITTED"] <- "0"

View(LongFourDialects_Thesis.NAomitReordered_Binary.df)
LongFourDialects_Thesis.NAomitReordered_Binary.df[LongFourDialects_Thesis.NAomitReordere
d_Binary.df == "PRESENT"] <- "1"

View(LongFourDialects_Thesis.NAomitReordered_Binary.df)

#Check the class:

class(LongFourDialects_Thesis.NAomitReordered_Binary.df$Pronounced_or_Not)

#Convert it to a numeral class:

LongFourDialects_Thesis.NAomitReordered_Binary.df$Pronounced_or_Not <-
as.numeric(as.character(LongFourDialects_Thesis.NAomitReordered_Binary.df$Pronounced_or_
Not))

View(LongFourDialects_Thesis.NAomitReordered_Binary.df)

class(LongFourDialects_Thesis.NAomitReordered_Binary.df$Pronounced_or_Not)

#Now the response variable is in a format that should work with the lme() function.

#Save a copy of this version of this new dataframe:

write.csv(LongFourDialects_Thesis.NAomitReordered_Binary.df, "E:/DATA
Rachel/Documents/MA UGA/THESIS-Spring 2022/R
PROJECT_THESIS/LongFourDialects_Thesis.NAomitReordered_Binary.csv", row.names = TRUE)

#Next we need to ensure all speaker birth years are numeric in order to use scale(). First
confirm the class of this column:

class(LongFourDialects_Thesis.NAomitReordered_Binary.df$Birth_Year)

#It is a factor class, so now let's convert it to a numeral class.

#Create a separate dataframe to work with:

```

```

LongFourDialects_Thesis.NAomitReordered_Binary_BirthYear.df <-
LongFourDialects_Thesis.NAomitReordered_Binary.df

View(LongFourDialects_Thesis.NAomitReordered_Binary_BirthYear.df)

#Convert the column to a numeral class:

LongFourDialects_Thesis.NAomitReordered_Binary_BirthYear.df$Birth_Year <-
as.numeric(as.factor(LongFourDialects_Thesis.NAomitReordered_Binary_BirthYear.df$Birth_Year))

View(LongFourDialects_Thesis.NAomitReordered_Binary_BirthYear.df)

class(LongFourDialects_Thesis.NAomitReordered_Binary_BirthYear.df$Birth_Year)

class(LongFourDialects_Thesis.NAomitReordered_Binary_BirthYear.df$Pronounced_or_Not)

#After comparing (spot-checking visually) the new dataframe and the original
(LongFourDialects_Thesis.NAomitReordered.df), it looks like the

#conversion was successful; each birth year is now corresponding to a number 1-12 (with
12 being the "unknown" age speakers)

#Check the class:

class(LongFourDialects_Thesis.NAomitReordered_Binary_BirthYear.df$Birth_Year)

#Now both the response variable and birth year variable are in a format that should work with
the lme() function.

#Save a copy of the final version of this new dataframe:

write.csv(LongFourDialects_Thesis.NAomitReordered_Binary_BirthYear.df, "E:/DATA
Rachel/Documents/MA UGA/THESIS-Spring 2022/R
PROJECT_THESIS/LongFourDialects_Thesis.NAomitReordered_Binary_BirthYear.csv", row.names
= TRUE)

#Now make a baseline model with no predictors, with just the response variable
(Pronounced_or_Not), and random effect (Speaker), included.

#Load the lme4 package & lmerTest package:

library(lme4)

library(lmerTest)

#Use the ML (Maximum Likelihood) method of fitting the model, because we are comparing
different models with different numbers of parameters.

ZeroPrt_LMER_0 <- lmer(Pronounced_or_Not ~ 1 + (1 | Speaker), data =
LongFourDialects_Thesis.NAomitReordered_Binary_BirthYear.df, REML = FALSE)

#View the model:

summary(ZeroPrt_LMER_0)

#Next start adding the other predictors:

#Add particle:

```

```

ZeroPrt_LMER_1 <- lmer(Pronounced_or_Not ~ Particle + (1 | Speaker), data =
LongFourDialects_Thesis.NAomitReordered_Binary_BirthYear.df, REML = FALSE)
summary(ZeroPrt_LMER_1)

#Add additional predictor variables using update():

#Add prefecture (dialect):
ZeroPrt_LMER_2 <- update(ZeroPrt_LMER_1, .~. + Prefecture)
summary(ZeroPrt_LMER_2)

#Add speaker age:
ZeroPrt_LMER_3 <- update(ZeroPrt_LMER_2, .~. + scale(Birth_Year))
summary(ZeroPrt_LMER_3)

#Add speaker gender:
ZeroPrt_LMER_4 <- update(ZeroPrt_LMER_3, .~. + Sex)
summary(ZeroPrt_LMER_4)

#Add sentence-final particle:
ZeroPrt_LMER_5 <- update(ZeroPrt_LMER_4, .~. + Sentence_Final_Particle_Present)
summary(ZeroPrt_LMER_5)

#Add honorific prefix use:
ZeroPrt_LMER_6 <- update(ZeroPrt_LMER_5, .~. + Honorific_Prefix_Present)
summary(ZeroPrt_LMER_6)

#Next, use ANOVA to determine which models are most useful:
anova(ZeroPrt_LMER_0, ZeroPrt_LMER_1, ZeroPrt_LMER_2, ZeroPrt_LMER_3, ZeroPrt_LMER_4,
ZeroPrt_LMER_5, ZeroPrt_LMER_6)

#Three of the models look promising; LMER_1, LMER_2, and LMER_3

#Because models 1-3 have relatively low AIC, BIC, logLik and p-values, and because
the visualizations generated above indicated that prefecture, age and particle all
correlate with omission, the final model will include only these three factors
(excluding gender, honorific prefix use and sentence-final particle use)

#The final model is therefore LMER 3:
ZeroPrt_LMER_Final <- lmer(Pronounced_or_Not ~ Particle + Prefecture + Age + (1 |
Speaker), data = LongFourDialects_Thesis.NAomitReordered_Binary.df, REML = FALSE)
summary(ZeroPrt_LMER_Final)

```