ABSTRACT

Title of Dissertation:	VARIATION IN INTERLANGUAGE: EVIDENCE FROM INTERNAL AND EXTERNAL PATTERNING OF MORPHOSYNTACTIC VARIABILITY IN THE SPEECH OF SECOND LANGUAGE LEARNERS
	Qi Zheng, Doctor of Philosophy, 2022
Dissertation directed by:	Professor Nan Jiang Department of Second Language Acquisition

Language is inherently variable, and learner language is particularly variable. The variationist paradigm considers learner language a heterogeneously variable yet inherently rule-governed system. Specifically, learners' alternation between native-like and nonnative-like variants of a variable or invariable target native speaker (NS) form constitutes learner language variation. Variation is also viewed as an indication of a transitional phase towards acquisition (e.g., Regan, 2013; Tagliamonte, 2011). With a particular concentration on second language (L2) morphosyntactic variation, this dissertation explored inter-learner variation and intra-learner variability together with interlanguage development by analyzing Japanese L2 learners' oral performances in English oral proficiency interviews. The research observed and studied the variation pattern in the interview data and identified the linguistic, paralinguistic, and nonlinguistic factors and factor groups which may give rise to Japanese L2 learners' repeated exercise of their interlanguage grammar for four morphosyntactic features: preposition/particle,

article, object pronoun-dropping, and modal auxiliary verb. The data were analyzed by using classification trees, random forests, and mixed-effects variable rule methods which together identified a hierarchy of variable importance among potential factors and factor groups and the influential factor levels within each significant factor group. With modern mixed models, the dissertation concluded that the observed morphosyntactic variation is subject to inter-lingual and intra-learner factors. Additionally, learners may also have individualized baselines and grammar. More importantly, the findings of the current research have provided important theoretical and empirical justification on whether and how individual patterns mirror the interlanguage patterns and hence an inter-lingual development understanding of L2 morphosyntactic competence.

VARIATION IN INTERLANGUAGE:
EVIDENCE FROM INTERNAL AND EXTERNAL PATTERNING OF MORPHOSYNTACTIC VARIABILITY
IN THE SPEECH OF SECOND LANGUAGE LEARNERS

by

Qi Zheng

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2022

Advisory Committee:
Professor Nan Jiang, Chair
Professor Kira Gor
Dr. Martyn Clark
Dr. Megan C. Masters
Associate Professor Donald J. Bolger, Dean's Representative

Dedication

To Professor Steven J. Ross

Who inspired this research but was not able to participate in its completion.

Acknowledgements

This dissertation would not have been possible without many people who supported me during my time at the University of Maryland. From the bottom of my heart, the most sincere thank you to my advisor and dissertation committee chair, Professor Nan Jiang, for his invaluable assistance, thoughtful guidance, and full support with this dissertation. And I am more than grateful to Professor Kira Gor, Dr. Martyn Clark, Dr. Megan Masters, and Associate Professor Donald Bolger for serving on the doctoral committee and providing detailed feedback to improve this work.

Many thanks to our SLA family. It's an honor to learn from Professor Michael Long, Professor Robert DeKeyser, and Adjunct Professor Catherine Doughty. I greatly appreciate my QP1 advisor Professor Kira Gor and my QP2 advisor Professor Steven Ross for their supervision and encouragement.

My special thanks also goes to Professor Gregory Hancock, Associate Professor Yang Liu, Professor Jeffrey Harring, Professor Hong Jiao, Professor Laura Stapleton, and Associate Professor Tracy Sweet at the Measurement, Evaluation, and Statistics Program for their excellent courses and guidance with statistical analysis.

Thank you, Brandie Shackelford, Jannitta Graham, and Jody S. Heckman-Bose, for your kind assistance.

Most importantly, my heartfelt thanks to my family, my cohort, and numerous friends who offered endless patience, support, and love.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

SLA	Second Language Acquisition
L1	First Language
L2	Second Language
NS	Native Speaker
NYC	New York City
Varbrul	Variable Rule Analysis
RQ	Research Question
TOEIC	The Test of English for International Communication
LPI	Language Proficiency Interview
NO	Natural Order
COCA	Corpus of Contemporary American English
IR	Index of Reproducibility
AIC	Akaike Information Criterion
OOB	Out-of-bag Classification Accuracy

Chapter 1: Introduction

Variation in learner language can be observed everywhere in everyday learner speech. A simple definition of variation is "two or more ways of saying the same thing (Labov, 1972, p.8)". For example, an English learner may say "*I want dog*" which is ungrammatical in English, while the same learner another time may also say "*I want a dog*" which is grammatical. Another example in contemporary spoken French is L2 learners' and French native speakers' variable use of the French subject pronoun *nous*/*on* (Rehner, Mougeon, & Nadasdi, 2003; p.128):

(1) *Ma soeur et moi nous allons à la même école*

(2) *Ma soeur et moi on va à la même école*

Both (1) and (2) mean 'my sister and I, *we* go to the same school', with *on* as an informal variant and *nous* the formal variant of the spoken French subject pronoun 'we'.

A variation-oriented linguist, i.e., a variationist, views the occurrences of standard and non-standard features within individuals and groups as variable yet systematic and developmental properties of the speech community individuals belong to (Labov, 1964; 1972). In fact, although heterogeneity was claimed to be a fundamental property and an integral part of the language in any speech community, many believed the occurrence of variation is heterogeneous but not random (Labov, 1969, Weinreich, Labov, & Herzog, 1968).

Indeed, the key to understanding variation in the use of a language is to find whether there is inherent systematicity in the observed chaos (Tagliamonte, 2011).

Traditional variation studies dealing with sociolinguistic observations of native speech linked historical, social, and cultural phenomena with the system of language variation and change. The way second language acquisition (SLA) applies the variationists approach took a similar trajectory by shifting the native paradigm and variation change over time into learner language variation and L2 development in the long run.

The SLA variationist deals with the systematicity of linguistic variables in the learner language. Linguistic variables in the field of SLA focus on native-like and non-native-like linguistic features as variants that convey the same target language speech in one or more standard and non-standard native forms. Learners' alternation between nonnative-like and native-like variants of an invariant second language (L2) form as well as learners' approximation to native-like use of target native speaker (NS) variation patterns are viewed as evidence that the learners are en route to fully acquire the target L2 form (e.g., Mougeon, Rehner, & Nadasdi 2004; Regan, 2013; Tagliamonte, 2011). Thus, whether variability in learner language is inherently consistent, how such variability posits relative to learners' native and target language, and to what extent could inter-lingual and individual grammar explain the observed variation is crucial for researchers in the field of SLA (Regan, 2013).

Against this background, this dissertation is designed to concentrate on learners' morphosyntactic competence in L2 speech with the goal to explore the systematicity of learner language. The purpose of the current research is to observe, study, and interpret L2 morphosyntactic variation patterns in learners' actual spontaneous speech to justify relationships between individual and inter-lingual and

intra-learner patterns. By exploring the native-like variant (accurate use) and nonnative-like variant (inaccurate use, or, ungrammatical use) on their probabilities and distributions over various linguistic, paralinguistic, and non-linguistic conditions, the current research hopes to extend the field understanding of the structure, processing, and development of L2 competence in morphosyntax.

The dissertation is organized as follows: Chapter 1 gives a brief introduction of the study. Chapter 2 synthesizes an extensive review of theoretical and empirical discussions on the role of language variation and change play in the field of SLA, the history of variation study in SLA, different types of variation studies, their contributions to our understanding of SLA, as well as methodological and analytical strategy developments within the variationist paradigm. A set of research questions are proposed at the end of Chapter 2 to set the primary goals for the current study. Chapter 3 then considers the methodological steps with detailed illustrations on data transcription, feature selection, and factor group identification, followed by Chapter 4 which demonstrates how quantitative analyses describe and explain the variation patterns observed in the L2 speech data. Chapter 5 presents results for four sets of analyses using conditional inference tree, conditional inference forest, and mixed-effects variable rule analysis regarding the transcribed variation data. Implicational scale analysis was also conducted to compare the data cross-sectionally. Lastly, Chapter 6 closes the dissertation with discussions and interpretations of the patterns observed and analyzed in Chapter 5. Important implications and questions remained for further research could be found in Chapter 7.

Chapter 2: Background

This chapter reviews studies in the field of variationist SLA, with a particular focus on the historical development of variation study in the field of SLA. An extended discussion on how variation studies have shed light on the theoretical and empirical understandings of critical SLA questions is also presented. Research questions for the current study are presented at the end of this chapter.

2.1 Language Variation and Change

Researchers may notice in a speech community, speakers may alternate between more than one form, formal and informal, standard and non-standard, to convey the same meaning. Before variation studies, the choice between forms was believed to be arbitrary. The variationist paradigm was first proposed by Labov (1963, 1969) in which the researcher discussed language being inherently variable yet systematic in a speech community and the possibility of quantitatively describing and evaluating the observed variation.

Variation studies since the 1970s have explored the systematic nature of variation in the use of language, both native and nonnative, by analyzing social and linguistic factors that were believed to account for the observed variation, e.g., gender, age, socio-economic status, linguistic context, and phonological, lexical, morphosyntactic, pragmatic properties, etc., (see a summary in Tagliamonte, 2011). The primary goal for variationists is to identify and describe the variation that occured, quantitatively measure the variable feature of the linguistic system, and interpret whether and why the occurrence of language variation is conditioned by

internal and external factors (e.g., Adamson, 2009; Chambers, 2009; Regan, 2004; 2013; Tagliamonte, 2011; Young, 1991).

The goal of studies on variation in language use includes but is not limited to analyzing the linguistic variation features whose recurring patterns potentially constitute the systematicity of a language system. More importantly, they focus on the changing nature that comes along with language variation in the speech community (Labov, 1994; 2001). As variationists investigate the distribution of variation longitudinally, many observed the rise of new variants, the cessation of existing variants, and the maintenance of an existing variant that occurs more or less frequently-all are important indications of language change and development over time. A particular benefit variation studies have on SLA is the extensive documentation and examination of learner language change over time. Variationists in the field of SLA are especially interested in both how the observed variation accounts for the development of L2 competence, as well as how the changing linguistic and nonlinguistic contexts which co-occur with the observed variation implicate the L2 development trajectories (e.g., Hansen, 2006; Young, 1991).

To investigate language variation and change, a typical procedure would be followed: (1) observe recurring variation patterns in language use; (2) select linguistic variables of interest; (3) determine potential factors that co-vary or co-occur with the selected linguistic feature; (4) form hypothesis and evaluate the potential factors with their effect on the variation using quantitative methods; (5) explain language variation and change with the quantitative results (Tagliamonte, 2011).

2.2 The Linguistic Variable

All variation studies begin with observations of recurring patterns in which speakers choose from two or more ways to express the same lexical, phonological, morphosyntactic, or discourse-pragmatic features in a given speech community. Speakers' choice among the available language variation is termed as *linguistic variables* in the field. In fact, choosing from two or more ways of saying the same thing is only a basic description of the linguistic variable, and researchers have emphasized several other properties a linguistic variable should have (Tagliamonte, 2011): Firstly, the linguistic variable must contain two or more alternative variants. The variants must belong to a linguistically definable category, e.g., a phonological segment, a lexical item, a morphosyntactic feature, a pragmatic expression, etc. Additionally, the variants of the same linguistic variable are not simply semantically more or less synonymous to each other, rather, scholars have restricted the variants to be interchangeably used "within the same grammatical system which has the same referential value (meaning) in running discourse (p.4, Tagliamonte, 2011; adapted from Sankoff, 1988)". Researchers also believed that the occurrence of variants is highly systematic and is correlated with patterns of linguistic, social, and cultural aspects of the speech community (Tagliamonte, 2011).

Specifically, linguistic variables in second acquisition studies could also apply the variationist paradigm with an adapted framework. L2 Learners' choice among the native-like variants and non-native-like variants of an L2 native form match the definition of a linguistic variable: both native-like and non-native like variants of the L2 linguistic variable state the same meaning, and they are perfect alternatives in every possible context under every grammatical scheme in L2 leaners' discourse.

Additionally, learners' choice of non-native forms over native forms, as much previous research which will be reviewed later in this chapter has reported, co-varies with a great number of linguistic and non-linguistic phenomena (e.g., Adamson & Regan, 1991; Drummond, 2012; Hoffman & Walker, 2010; Mougeon, Nadasdi, & Rehner, 2002; Mougeon, Nadasdi, & Rehner, 2010; Nance, McLeod, O'Rourke, & Dunmore, 2016; Noels, 2014; Regan & Nestor, 2010; Regan, 2013; Sax, 2003; Young, 1988; 1989; 1991; 1996; Young & Bayley, 1996).

2.3 Variation and Interlanguage

The variationist paradigm and the SLA research have shared a joint interest in particular speech properties: the former in standard and non-standard variants in a speech community, and the latter in native-like and nonnative-like variants in interlanguage. In short, variationists in SLA investigate language learners' variation pattern in their interlanguage.

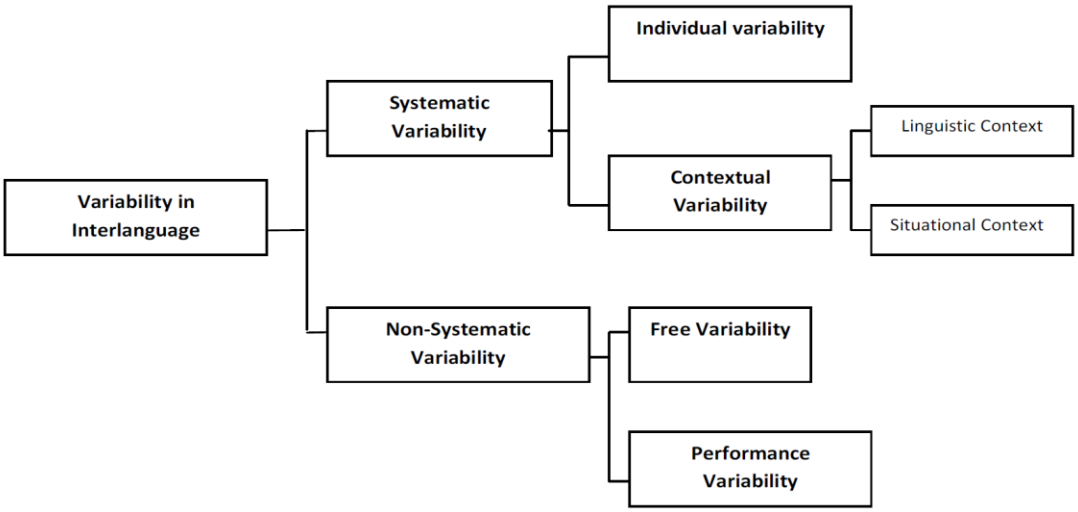
Selinker's Interlanguage theory (1972) was proposed as a theoretical hypothesis in describing and explaining learners' performance in acquiring a non-native language. This learner-focused perspective of language acquisition considers learner language as an independent linguistic system (Selinker, 1972, 1974; Tarone, 1988, 1990, 1994; 2012). The acquisition trajectory of a target language is therefore viewed as a transitional progression of interlanguage from learners' native language to the target language (e.g., Corder, 1981; Fauziati, 2011; Richards & Schmidt, 2013; Selinker, 1972).

The interlanguage system is dynamically changing before stabilized: during L2 acquisition, learners gradually establish and evaluate their hypothesized rules

against the target language system, and as learners attempt to approach the target language by adding and removing hypothesized rules, their performance could be highly variable.

Figure 2.1

Variability in Interlanguage



Note. Variability in Interlanguage introduces the composition of types of variability in interlanguage. From "A Review Study of Interlanguage Theory," by M. H. Al-Khresheh, 2015, *International Journal of Applied Linguistics and English Literature, 4*(3), p.127. Originally adapted from "The Definition and Measurement of L2 Explicit Knowledge," by R. Ellis, 2004, *Language learning, 54*(2).

Variability in interlanguage emerged as a theoretical and empirical research of interest in the field of SLA since the 1980s (e.g., Adamson & Kovac, 1981; Adamson & Regan, 1991; Bayley & Langman, 2004; Beebe, 1980; Othman, 2003; Tarone, 1988; Wolfram, 1985; Young, 1988; 1989; 1991; 1996; Young & Bayley, 1996). Many have demonstrated that although learners' performance is highly variable, the

interlanguage system is not random or chaotic but a systematic collection of linguistic patterns (e.g., Bayley, 2004; Foulkes & Docherty, 2005; Regan, 2004; 2013; Tarone, 2001). Ellis (2004) suggested that observed variability in interlanguage could be further classified into different sub-categorizations of variability (see Figure 2.1). The systematic variability, for example, is composed of individual variability and contextual variability which further splits into linguistic (e.g., phonological, syntactical, and semantic elements) and situational context (e.g., pragmatic and discursive). The non-systematic variability, on the other hand, is composed of free variability and performance variability.

2.4 Variation and SLA

Since the 1970s, variation-based perspectives have contributed significantly to our understandings of SLA with theoretical discussions supported by empirical evidence on a number of aspects that other SLA approaches often neglect and/or are not able to examine (Ellis, 1994; Regan, 2013). Mainstream SLA approaches, as Ellis has demonstrated (1994), often ignore variation as an indicator of competence, but rather, consider it as a part of the performance. The variationist paradigm was found to best describe and quantitatively analyze the learner speech which, as many have stated, is highly variable in various aspects.

The variation studies in SLA were marked by a few stages. As early as the 1970s, early connections between variationist sociolinguistics and SLA were established when researchers start to apply the variationist paradigm to SLA (e.g., Dickerson,1974; 1975; Tarone, 1979, 1982; 1983). Later in the 1980s, a number of pioneering studies used the variation methods in SLA research in support of the idea

that L2 variation, like L1 variation explored in the field of variationist sociolinguistics, was also inherently systematic (e.g., Adamson & Kovac, 1981; Beebe, 1980; Beebe & Zuengler, 1983; Meisel, 1983; Tarone, 1988; Tarone & Parrish, 1988; Wolfram, 1985). These early variationist SLA studies have been criticized on their one-sidedness, as many attributes the observed variation to a single factor, e.g., linguistic environment, proficiency, etc. To address this multicausality issue, the 1990s studies conducted a series of variable rule approach-based analyses and implemented detailed variation frameworks with discussions on interlanguage hypothesis and L2 development (e.g., Adamson & Regan, 1991; Bayley & Preston, 1996; Young, 1988; 1989; 1991; 1996; Young & Bayley, 1996). It is worth mentioning that systematic implementation and justification of quantitative modeling with regard to multicausality in variation studies also emerged in the 1990s and the early 2000s (Bayley, 2002; Young & Bayley, 1996). By the late 1990s, most variation studies in SLA were concentrated on language learners' variable use of invariant target native language forms, while starting from the 2000s, a new strand of variation research starts to recognize and address learners' acquisition of target NS variation patterns: Like what has been examined in variationist sociolinguistics that native speakers may choose from more than one variant of a native variation, language learners, too, can choose from more than one variant of a target language form. Studies investigating language learners' acquisition of target NS variation patterns under classroom, immersion, naturalistic, etc., settings have emerged around this period (e.g., Howard, Lemée, & Regan, 2007; Mougeon, Nadasdi, & Rehner, 2002, 2010; Mougeon & Rehner, 2001; Nadasdi, Mougeon, & Rehner, 2003; Sax, 2003).

The recent development of variationist SLA sees the demand for a nuanced ethnographic description of the L2 learners. This emphasizes the qualitative exploration of age, gender, identity, motivation, etc., as social constructs to obtain a more comprehensive understanding of how variation in L2 speech interacts with learners' multilayered experience (e.g., Drummond, 2012; Hoffman & Walker, 2010; Meyerhoff, & Schleef, 2012; Nance, McLeod, O'Rourke, & Dunmore, 2016; Noels, 2014; Regan & Nestor, 2010; Schleef, Meyerhof, & Clark, 2011).

2.5 Vertical Variation and Horizontal Variation in SLA

While in variationist sociolinguistics only social characteristics of native speakers in the speech community were considered, variationists in SLA have proposed for an adaptation of the sociolinguistic paradigm into *Vertical variation* and *Horizontal variation* (some also called them *Type 1 variation* and *Type 2 variation*, see Rehner, 2003). This differentiation could be understood as two axes, vertical and horizontal. *Vertical variation* represents the developmental dimension in L2 competence, and it primarily evaluates learners' categorical knowledge of alternating between native (grammatical) and nonnative (ungrammatical) forms. *Horizontal variation* recognizes learners' capacity to produce socially appropriate speech in the given speech community, hence the dimension of progression in sociolinguistic competence (Mougeon & Dewaele, 2004; Regan, 2013).

Vertical variation in learner language was visited by variationists in SLA from the early 1980s till the present day. It has offered an alternative perspective to traditional non-variationist works in a number of domains and has explored various linguistic and nonlinguistic factors that conditioned the occurrence of L2 variation.

Meanwhile, theoretical discussions in SLA were also evaluated by the findings and implications of these variation studies. Tense in L2 use, for example, has been investigated by variationists on a variety of ranges (e.g., tense marking, perfective aspect marking, irregular past tense marking, and aspectuo-temporal marking) with a number of L2s such as Chinese, Spanish, English, and French (e.g., Adamson, Fonseca-Greber, Kataoka, Scardino & Takano, 1996**;** Bayley, 1994; Berlin & Adamson, 2009; Howard, 2002; 2004; 2005; Jia & Bayley, 2008; Wolfram, 1985). For instance, Berlin and Adamson (2009) in their study investigated the acquisition of English past tense marking by a group of children whose native language is Chinese. The use of multivariate modeling identified the best combination of factors on learners' variation in irregular past tense marking such as the telicity of the verb and the rate of accuracy of the background clause. The findings were considered empirical evidence in supporting the overarching SLA discussions on language transfer (Bayley, 1994), prototype and schema theory (Anderson, Reynolds, Schallert, & Goetz, 1977; Adamson, 2009), the role of clause type in tense marking (Young, 1991), and the notion of salience in L2 acquisition (Wolfram, 1985).

The primary goal of a typical vertical variation study is to examine the relationships between the observed variation in interlanguage and the linguistic and nonlinguistic contexts that condition the variation in support of the hypothesis that interlanguage is a rule-governed system (e.g., Adamson et al., 1996; Bayley, 1994; Berlin & Adamson, 2009; Ellis, 1987; Jia & Bayley, 2008; Langman & Bayley, 2002; Selinker & Douglas, 1985; Tajika, 1999; Young & Bayley, 1996; Young, 1989, 1991). Young (1989, 1991) on variation in L2-English -s plural marking, for example,

tested the hypothesized effect of four categories of factors: the speech patterns of the interlocutor, the proficiency level of the speaker, the linguistic (e.g., phonological, syntactic and semantic) environment of the -s inflection, and its functionality in delivering necessary information. The study identified multiple factors that influenced the choice of pluralization rules differently at different stages of acquisition: lower proficiency learner is most strongly promoted by phonotactic factors while higher proficiency learner is least and insignificantly affected by the same factor. Young therefore concluded that there was a certain degree of systematicity in learner language, and that learners' mental rule of -s plural marking is less complicated than many non-variationist SLA researchers had implicated. Langmam and Bayley (2002) in analyzing native speakers of an analytic language learning verbal morphology of an agglutinative language found that perceptual salience, morphophonological regularity (vowel harmony), as well as the definiteness of the object on the verb have all significantly promoted the acquisition of complex agglutinative morphological system in L2. Proficiency and verb frequency have an interactive effect, that learners at an early stage of acquisition weight heavily on frequency in acquiring the L2 morphology. These authors also concluded that variation in learner speech is systematic based on their findings. Learners' performance on phonological variables has also been documented, such as Hansen (2001) who analyzed L2 acquisition of English consonantal syllable codas, and Hurtado and Estrada (2010) on L2 Spanish vibrant acquisition. Hansen (2001) demonstrated that linguistic constraints such as length of coda, markedness, sonority, natural phonological processes, and L1 transfer were found to account for the observed consonantal coda variation. Hurtado and

Estrada (2010) detected influential linguistic factors including vibration and vibration position, and articulatory effort of the preceding segment. The authors also found that the level of class taken abroad, instead of study abroad experience (yes, no), significantly co-occurs with the L2 acquisition of Spanish vibrants. Taken together, these phonological variation studies have also come up with the conclusion that variation observed in different aspects L2 use is reflective of the L2 acquisition process, and more research is needed for further evaluation.

Vertical variation studies also paid close attention to the theoretical discussion on the relationship between individual and group variation patterns. While the systematic similarity between individual and group variation patterns has been demonstrated within the variationist sociolinguistics paradigm (see a detailed summary in Tagliamonte, 2011), the same issue in the field of SLA is far more complicated. A typical difference between native language variation and interlanguage variation is the rapid changing of learner variation as their L2 competence develops. As a result, factors operate on a linguistic variable at one stage of acquisition may not be influential for the same variable at another stage (Bayley, 1996; Bayley & Langman, 2004; Blondeu & Nagy, 2003). L1 transfer, on the other hand, may also affect the influence of contextual constraints on L2 variation patterns (e.g., Nagy, Blondeu, & Auger, 2003; Norton, 2000; Regan, 1996). Bayley and Langman (2004) looked for evidence in relation to individual and group variation in SLA and investigated L1-Chinese learners' acquisition of L2-English and L2-Hungarian verbal morphology. The authors demonstrated that individuals do mirror group patterns on a number of domains such as aspect, frequency, and perceptual

saliency, however, they also pointed out that their grouping of L2 learners shared an overwhelming similarity: the Hungarian L2 learners were all a part of the local community of the same generation, with similar causes of immigration, and importantly, engaged in very similar occupations. SLA researchers have therefore called for more research to explore L2 speech communities with more diversity introduced by individual learners (Regan, 2013).

Horizontal variation concentrates on aspects that are not only restricted to L2 proficiency, but also the context of acquisition, interactions with native-speaking interlocutors, socioeconomic status, gender, etc. (e.g., Adamson & Regan, 1991; Hansen Edwards, 2011; Howard, Lemée, & Regan, 2007; Mougeon, Nadasdi, & Regan, 1995; 1996; 1997; 1998; Mougeon, Nadasdi, & Rehner, 2002; Mougeon & Rehner, 2001; Sax, 2003; Nadasdi, Mougeon, & Rehner, 2003; Rehner, 2010). This line of research is building upon the theoretical understanding that the knowledge of linguistic variation and variability is also an integral part of native speaker competence, and hence an important L2 competence to acquire for some language learners whose goal is to become native-like in L2 performance (Chambers, 2002; Guy, 2005). Specifically, studies exploring L2 horizontal variation are interested in how L2 variation patterns resemble the target native variation patterns, whether the linguistic and nonlinguistic constraints on the variation pattern is shared and not shared by the L2 and L1 speakers, as well as the acquisition process of the target language variation patterns by the L2 learners. One of the earliest studies of L2 horizontal variation, for example, investigated L1-Vietnamese and L1-Cambodian immigrants' acquisition of the English nominal and verbal *-ing* form (Adamson &

Regan, 1991). Though spelled in a similar fashion, a historical perspective views the nominals and verbals differently as the former was developed from *-ind* (pronounced as [ɪnd]) until the end of the 14th century. The verbal *-ing* form (pronounced as [ɪŋ]), on the other hand, possessed features of verbs such as direct object-taking, adverbial modifiers, and aspect (Houston, 1985). As Adamson and Regan stated, the variability of the use of the non-prestige and the prestige form [ɪn]/[ɪŋ] still exists, and they found native speakers and language learners exhibit different patterns of linguistic conditionings on the variability in the use of [ɪn]/[ɪŋ] variation. Additionally, the study also reported a gender effect on learners' desire to approximate male native speaker patterns.

Context of acquisition has shown a dominating influence on learner's acquisition of native variation patterns, and learners who have sustained context with native speakers (e.g., study abroad and immersion programs) tend to approximate native speaker variation patterns with similar factor conditionings on such patterns (e.g., Byram & Feng, 2006; Dewaele & Regan, 2001; Lemée, 2002; Mougeon & Rehner, 2001; Regan, Howard, & Lemée, 2009). Dewaele (1992), Regan (1996), and their later works (2001) have shown that the lack of interaction with native speakers and limited time spent in the native speech community (e.g., classroom learning) has an important negative effect on language learners' acquisition of target language norms and variation. Rehner and colleagues (2003) summarized a series of immersion learner studies on L2-French variation patterns categorized by different degrees of formality (e.g., Dewaele & Regan; 2000; Mougeon & Rehner, 2001; Rehner, 1998). Together with their own research findings, these authors implicated learners' less

frequent use of *on* in the *on/nous* variation than that of L1 speech patterns is associated with teachers' classroom speech and textbook materials in which explicit use of *nous* and infrequent appearance of the informal *on* were observed. On the other hand, Lemée (2002) investigated the variability in the speech of L1-English learners of L2-French in their use of informal and formal pronouns *on/nous* 'we'. The author observed the use of both formal and informal variants among learners regardless of their level of education and their interaction with native speakers, however, the native-like preference of the informal variant *on* in L2 speech was found to be systematically associated with learners' study-aboard experience. Lemée therefore concluded that learners seek to more authentic use of the target language while spending time in the native speech community, an indication of improvement in their sociolinguistic competence. Dewaele and Regan (2002) based on their findings of omission of French *ne* 'not' (the informal variant) also argued that the native-like use of informal variants is not a direct product of the intensity and amount of formal instruction. Instead, the authors have suggested that living in the native speech community may be a prerequisite for a native-like use of the informal variant. Together, these findings have important implications on language learning (e.g., classroom instruction versus study abroad), language teaching (e.g., native and nonnative teachers), as well as policy-making in language education (e.g., immersion programs). Other potential factors that may constitute a derivative discussion to the context of acquisition also include naturalistic learning (e.g., Bayley, 1996; Major, 2004), gender (Adamson & Regan, 1991; Major, 2004; Mougeon, Nadasdi & Rehner, 2002; Sax 2003; Uritescu, Mougeon, Rehner, & Nadasdi, 2004), and length of

residence in the target language-speaking country (see a discussion in Regan. 2013). As Hansen Edwards (2008) has concluded, language learners "may be active agents in targeting which variants to use and acquire and may use the variants purposefully to mark gender, social, and ethnic identity (p.260)."

Like vertical variation studies, variationists exploring horizontal variation are also interested in the role variation played in the individual and in the group. Regan (2004) followed a group of five L1-Irish L2-French learners on their acquisition of the informal *ne* 'not' deletion in French, a sociolinguistically sensitive variation in spoken French, over a three-year period in the classroom setting (Year 1), in a Francophone country (Year 2), and back in the classroom setting (Year 3). The author found similarities in patterns of *ne* deletion among four out of five learners across the longitudinal stages, from a significant increase in *ne* deletion in Year 2, and overall maintenance of the probability of *ne* deletion in Year 3. Regan (2004) hence concluded that learners' L2 *ne* deletion, although inherently variable on individual-basis, is generally reflective of group variation patterns as they develop towards the target native speech norms. Nevertheless, Regan did mention that the implications of her findings should remain tentative as only a limited number of L2 speakers were examined.

2.6 L2 Developing Sequences and L2 Acquisition Orders

The relationship between individual and group patterns, as reviewed in the previous section, has been one of the theoretical and empirical research of interest for variationists in SLA. A fair number of the L2 variation studies have demonstrated that variability in L2 performances was inherently systematic and that individuals mirror

the group variation patterns (e.g., Bayley, 2004; Regan, 2004; 2013)-note that this still remains on the table for non-variationist SLA researchers (see a discussion in Herschensohn & Young-Scholten, 2013; Lardiere, 2013). The exploration of individual and group variation patterns has also led to an extended discussion on L2 developmental sequences and L2 acquisition orders which is also under debate at this point.

The issue of L2 acquisition order has been explored by SLA researchers in various domains. A rich number of L2 morpheme order studies in the 1970s and 1980s, for example, have demonstrated a predetermined order of L2 morpheme acquisition (e.g., Dulay & Burt, 1975; Krashen, 1982; Larsen-Freeman, 1975; 1976; Pica, 1983) although later some disagreed after considering learners' L1 backgrounds (Khor, 2013; Luk & Shirai, 2009; Murakami, 2014; Murakami & Alexopoulou, 2016; Weitze, McChee, Graham, Dewey, & Eggett, 2011; Zobl, 1982) and other cognitive and linguistic factors such as psychological salience, frequency, syntactic category, and so on (Goldsneider & DeKeyser, 2001; Kwon, 2005; McFerren, 2015). Studies in L2 formulaic sequences and L2 constructions have also suggested an underlying systematicity toward L2 acquisition (e.g., Eskildsen, 2012; Schmitt, 2004; Yuldashev, Thorne, & Fernandez, 2013), but researchers nevertheless argued that the acquisition might also be lexical-, item-, and language experience-dependent (Ellis, 2002; Tomasello, 2005) as limited research at present is not enough to defend such implication.

In fact, a number of more recent research have proposed for an individually owned developmental view of L2 acquisition (e.g., de Bot, Lowie, & Verspoor, 2007;

Lowie & Verspoor, 2011; 2015). Research in support of this process-based view of acquisition has criticized traditional arguments on predicative acquisition orders treating learning as a product but rather a process. Consequently, traditional measures equal the ordered group means as the developing sequence would face a major challenge when the order fails to reflect the actual process of any individuals.

Indeed, the failure to recognize inter-learner variation and intra-learner variability during the language learning process would threaten the validity in generalizing the individual-to-group developmental patterns. For example, in observing a 13-year-old L1-Spanish L2-English learners' morpheme acquisition in an English immersion program, Rosansky (1976) found highly variable monthly ranks of morpheme development within the same learner and an overall disagreement of order ranks by month to the overall development trajectory. Cancino and colleagues (1978) reported on the development of four negative constructions by a 13-year-old English L2 learner concluded a generally free variation during the early stages of acquisition and a gradual pattern toward a more organized variation system. Similar conclusions were also drawn from L2 phonology studies: In an analysis of weekly recordings of L1-English L2-Dutch learners' oral production of native-like Dutch vowels during their one-year stay in the Netherlands, highly variable individual productions were observed which implicate an unsystematic vowel developmental process in L2 acquisition (Lowie, 2013). Also, Derwing and colleagues (2006, 2013) compared L2 English learners' development trajectories on L2 accentedness and L2 comprehensibility from different L1 backgrounds (Slavic and Mandarin). The later study (Derwing & Munro, 2013) shows that not a single individual trajectory

resembled the mean pattern which was supposed to indicate an overall L2 development.

Longitudinal studies as well as corpus-based reanalyzes of L2 acquisition over time have also documented dynamic inter-learner variation and intra-learner variability in L2 data (e.g., Brown, 1973; Eskildsen, 2012; 2015; Hakuta, 1975; 1976; Yuldashev, Thorne, & Fernandez, 2013). For instance, in an L2 morpheme development study, Hakuta (1975; 1976) reported a 60-week longitudinal observation of an L1-Japanese L2-English learner *Uguisu* from age 5;4 until age 6;5. The highly fluctuating acquisition curves Hakuta demonstrated for English definite and indefinite article, auxiliary *be*, how-embedding, and others showed the highly variable patterns for L2 grammatical morpheme development. More recent studies also unveil the inter-learner variation and intra-learner variability in acquiring L2 constructions: Eskildsen (2012) on L2 negation construction use in classroom interactions showed a longitudinally developing and dynamic trajectory of acquisition. Yuldashev and colleagues (2013) in analyzing multi-word unit use over time also found a varied repertoire of Spanish *es/ques* construction in learners' out-of-class text messaging. Although, unlike studies criticizing the use of group mean statistics, most of these studies supported an underlying systematic L2 development based on accuracy orders, implicating an overall similarity between individuals and group patterns.

Building upon the observation of variable inter-learner and most importantly intra-learner patterns summarized above, Lowie & Verspoor (2015) has argued for an over-generalization based on group mean statistics to indicate the L2 developmental trends in these studies. The authors extended a theoretical discussion by van Geert

(2014) in which L2 development patterns were considered multi-dimensional: potential influence by multiple variables and interactions between these variables, individual trajectories that develop over time, and a joint effect of these two dimensions. As a result, valid conclusions could not be guaranteed when traditional regression models of language processing attempt to aggregate group means regardless of the multi-level and multi-dimensional nature of their observations which violates the assumption of homogeneity of regression slopes.

Modern mixed-model designs have offered an alternative examination of L2 variable data while taking its multi-dimensional nature into consideration. The design used the maximum likelihood method in estimating probabilities and is therefore free from assumptions that traditional methods require. By including individuals as random effects together with other fixed effects, learner variation within the group and their joint effect with other variables would be estimated to allow for a valid generalization. However, concerns arise as some researchers argue that the mixed-model designs are ineffectual to capture the intra-individual variability (e.g., Molenaar, Huizenga, & Nesselroade, 2003; Molenaar, 2008). As a result, it is equally important, as some of the discussions implied, to also include a time dimension to effectively account for the intra-learner variability in discussing L2 acquisition orders (Lowie & Verspoor, 2015; van Geert, 2014).

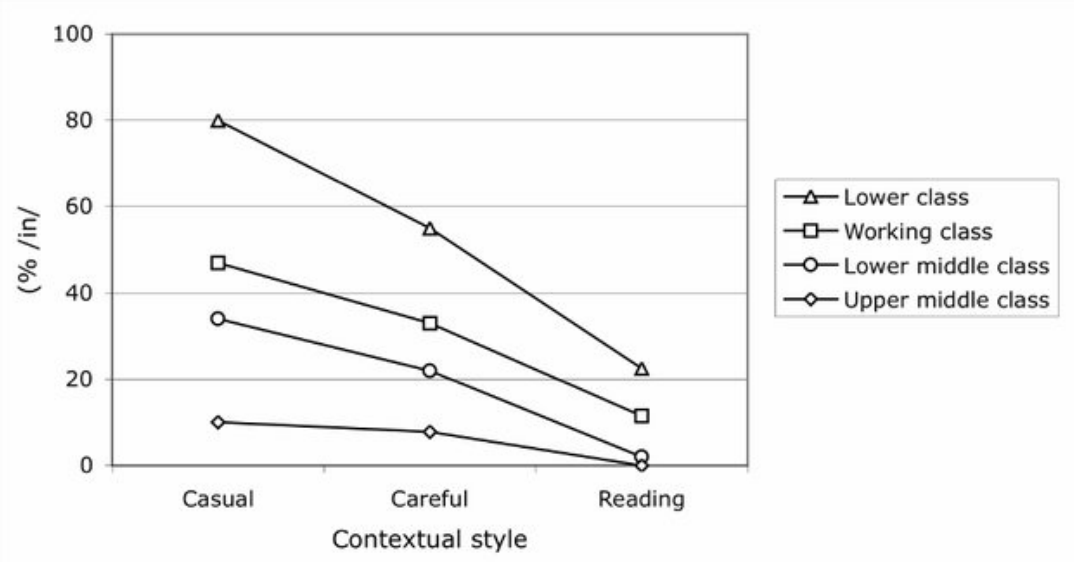
2.7 Methodological Development

The choice of analytical strategies in measuring linguistic variation has been long debated in the application and selection of the best statistical models (e.g., Bickerton, 1973; Cheshire, 2005; Downes, 1984; Rickford, 1975; Tagliamonte & Baayen, 2012).

Quantitative approaches to demonstrating and measuring linguistic variation have also followed a developmental trajectory: (1) Style stratification diagram; (2) Implicational scale; (3) The variable rule program; (4) Mixed-effects probabilistic models and new tools such as conditional inference trees (classification trees) and conditional inference forests (random forests).

Figure 2.2

Stylistic Stratification Diagram



Note. The social class and contextual style stratification diagram introduces proportions of target-like variants (the vertical axis, %) by contextual style (the horizontal axis) and social classes (lower, working, lower-middle, and upper) for the Labov NYC study (1966). From "Listeners' Sensitivity to the Frequency of Sociolinguistic Variables," by W. Labov, S. Ash, M. Baranowski, N. Nagy, M. Ravindranath, & T. Weldon, 2006, *University of Pennsylvania Working Papers in Linguistics, 12*(2), p.19.

Style stratification diagram with percentages of target-like variants in the

observed variation as the outcome of interest was first implemented in the field of variationist sociolinguistics by Labov (1964; 1966) in exploring the phonology of speakers in New York City (NYC). Figure 2.2 presents an adapted style diagram of the Labov (1966) NYC study which involves degrees of attention to speech (casual, careful, and reading) and social class (lower, working, lower-middle, and upper) as independent variables. Conclusions were carried out by describing slopes of each line (see Figure 2.2), but such slope tendency-based inference was criticized by later researchers on the lack of statistical significance (Beebe, 1987; Sato, 1984; Young, 1986). Later in the 1970s and the 1980s, the Labovian approach was adapted to also explore issues in SLA by presenting L2 learners' proportion of native-like variants as indications of variable accuracy, and these SLA research chose to compare mean statistics and used t-tests in support of statistical inferences (e.g., Beebe, 1987; Dickerson, 1974; 1975; Dickerson & Dickerson, 1977; Sato, 1984; Tarone, 1983; Young, 1986). Nevertheless, the adapted Labovian approach was still under debate on a number of aspects: the difficulty to apply interval scales in describing morphosyntactic variables, a threat to validity when placing the categorically identified speech style in a continuum, and more importantly, little indication on the systematicity of interlanguage variation.

Variationist SLA in search of an adequate measure of interlanguage variation notice the use of implicational analysis in sociology. SLA researchers saw the advantage of the implicational approach which presents the variable of interest and groupings of individuals in a way that the presence of one property could be implied by the presence of another property in a predicative order-well suited to describe the

systematicity and development of interlanguage. The approach was introduced to SLA in the late 1970s (e.g., Andersen, 1978; Gatbonton, 1978; Hyltenstam, 1977; Meisel, 1983). Gatbonton's gradual diffusion model of acquisition was a classic theoretical realization of implicational analysis in SLA research (see Figure 2.3). The acquisition phase marks a start from nonnative-like variants (0) to the replacement phase in which alternation between native-like and nonnative-like variants (v) emerge before the native-like variants gradually replace the nonnative-like variants. At the end of the replacement phase, the cessation of nonnative-like variants indicates the complete acquisition.

Figure 2.3

Gatbonton's Gradual Diffusion Framework

Stages	$Environment_1$	$Environment_2$	$Environment_3$
	Acquisition Phase		
a	0	0	0
b	(V)	0	0
c	(V)	(V)	0
d	(V)	(V)	(V)
	Replacement Phase		
e	1	(V)	(V)
f	1	1	(V)
g	1	1	1

Note. Gatbonton's gradual diffusion model of acquisition. (V) = observed variation; 0 = before acquisition; 1 = successful acquisition. From “Variation in Interlanguage Morphology:(s) Plural-marking in the Speech of Chinese Learners of English,” by R.F. Young, 1989, *Doctoral dissertation, University of Pennsylvania*, p.37.

Table 2.1

Sample Implicational Scale of L2 Development

		The L2 Learners							
		L2 Stages							
Morphological Features		A	B	C	D	E	F	G	…
Hypothesized Acquisitional Order ↓	Feature 1	-	-	-	-	+	-	-	…
	Feature 2	-	-	-	+	+	+	/	…
	Feature 3	-	-	+	+	+	-	+	…
	Feature 4	-	+	+	+	+	-	+	…
Plusses		0	1	2	3	4	1	2+	…

Note. (+): acquired; (-): not acquired; (/): not emerge.

The table illustrates a hypothesized implicational scale of morphological development. Adapted from "Implicational Scales," by J.R. Rickford, 2002, *The Handbook of Language Variation and Change*, p.156. and "Portraying the English Morphological Development of Indonesian-English Interlanguage Learners," by R. Suherman, L. M. Indrayani, and E. Krisnawati, 2020, *Indonesian Journal of English Language Teaching and Applied Linguistics, 5*(1), p.91.

In practice, early literature using implication analysis has favored the order and re-order of trinary scales (+, -, /) to indicate a hypothesized developing sequence of L2 competence. Table 2.1 depicts one of the implicational scales following the early day fashion in describing the acquisition of morphological features by re-ordering the learners based on their performances (i.e., plusses). If all learners followed the hypothesized step-like line pattern (e.g., Learner A-E), then it would be considered a strong piece of evidence in support of the L2 acquisitional order from Feature 1 to Feature 4 and furthermore the processability hierarchy in SLA (Glahn et al., 2001; Pienemann, 1998). The observation of Learner F, for example, are major concern to implicational research in SLA-A fundamental criticism regarding the

implicational approach lies in the strict regularity it implies, while in language learning, data points not conforming to the implicational framework may be easily observed (Huebner, 1985). Researchers then would have to seek further explanation to understand the irregularity Learner F has introduced to the implicational pattern. The observation of Learner G with an empty cell, hypothetically, on the other hand, might not violate the implicational pattern but sits in between Learner C and Learner D if the implicational pattern implied by Learner A-E is well supported. Researchers in the later development of implicational analysis have proposed more accurate and generalizable forms of scales (see a detailed discussion in Rickford, 2002; see also Dunn-Rankin, Knezek, Wallace, & Zhang, 2014; Suherman, Indrayani, & Krisnawati, 2020; Trofimovich, Gatbonton, & Segalowitz, 2007). See more discussion on implicational scales in Chapter 4.

Figure 2.4

Rule Structures: Optional Rule (Up) and Variable Rule (Down)

$$\mathbf{X} \rightarrow (\mathbf{Y}) \,/\, \left\{\begin{matrix}[\text{fea A}]\\ [\text{fea B}]\\ \vdots\end{matrix}\right\} \left\{\begin{matrix}\overline{\qquad}\\ [\text{fea I}]\\ [\text{fea J}]\\ \vdots\end{matrix}\right\} \left\{\begin{matrix}[\text{fea P}]\\ [\text{fea Q}]\\ \vdots\end{matrix}\right\} [\text{fea Z}]$$

$$\mathbf{X} \longrightarrow \langle \mathbf{Y} \rangle \,/\, \left\langle\begin{matrix}[\text{fea A}]\\ [\text{fea B}]\\ \vdots\end{matrix}\right\rangle \left\langle\begin{matrix}\overline{\qquad}\\ [\text{fea I}]\\ [\text{fea J}]\\ \vdots\end{matrix}\right\rangle \left\langle\begin{matrix}[\text{fea P}]\\ [\text{fea Q}]\\ \vdots\end{matrix}\right\rangle [\text{fea Z}]$$

Note. Rule structures present the original optional rule structure applied to generative grammar (up) and the adaptation of the structure into variable rule notation (down). From “Variable Rules: Performance as a Statistical Reflection of Competence,” by H. J. Cedergren, & D. Sankoff, 1974, *Language*, p.340.

Variable rule analysis (Varbrul), a heuristic tool designed to deal with natural speech data in the 1980s, is particularly effective in exploring the multicausality nature of variation data. In fact, most of the empirical variation studies reviewed in this chapter have used one or another version of the variable rule method (e.g., *Varbrul*: Cedergren & Sankoff, 1974; *Goldvarb* and *Goldvarb 2.0*: Rand & Sankoff, 1990; *Goldvarb X*: Sankoff, Tagliamonte, & Smith, 2005; *Rbrul*, Johnson, 2009; *Shiny Rbrul*: Johnson, 2017). The variable rule program models variation patterns in spoken data and evaluates the relative influence of multiple factors and factor groups on the choice of variants (i.e., the operation of the *rule*). The term variable rule was originally adapted from the rewrite of optional rules applied to generative syntax (Cedergren & Sankoff, 1974; Sankoff & Labov, 1979). Figure 2.4 presents the structural description of the variable rule approach that Cedergren and Sankoff (1974) proposed to adopt in which the operation of the rule is conditioned by a number of environmental features and feature bundles. Each pair of angled brackets (<>) represents an environmental feature bundle that is composed of a mutually exclusive and exhaustive list of features: only one feature in the list occurs at a time to constitute the environment X. A minimally required environmental feature is presented in a pair of square brackets ([fea Z]) and must occur to guarantee the operation of the rule. The current field term *factor group* is one of the adaptations of feature bundles in the original variable rule structure. The list of features within each feature bundle are now called *factors* (or *factor levels, factor nodes, factor categories*) within a *factor group*. Taking the Adamson & Regan (1991) study in which a gender effect on learners' [ɪn]/[ɪŋ] (-in/-ing) alternation was reported as an

example, the *factor group* gender is composed of two *factor levels*, male and female. Their variable rule analysis showed that male (p = 0.77) uses [ɪn] more frequently than females (p = 0.24), hence the factor "male" is considered a more important factor than "female" in the use of [ɪn] (-in).

Daniel Johnson designed the most recent version of the variable rule program *Rbrul* (2009) and *Shiny Rbrul* (2017) in which he made a number of improvements on the functionality of the previous versions (e.g., *GoldVarb* and *GoldVarb X*): Rbrul runs in an opensource platform R which is freely accessible to anyone, and it is more flexible in data format and data presentation. For example, GoldVarb would crash if any factor level in any factor group has invariant data (i.e., either 0% or 100% probability), and Rbrul is more flexible in taking extreme invariance (Johnson, 2010). Moreover, Rbrul can adjust the level of significance according to predictor numbers (the Bonferroni correction) and can group factors together to avoid situations when an invariant outcome is observed (Johnson, 2009). Most importantly, variable rule program prior to Rbrul tends to ignore the necessity to have a factor group for speaker and assumes that speech data are independent while this is not true-each of the elicited token in the observed variation are nested within the speakers who produced them. As a result, these variable rule programs tend to overestimate the significance of factor groups (Johnson, 2009; Sankoff, 2004; Tagliamonte, 2006; Young & Bayley, 1996). Rbrul under the modern mixed-effects modeling framework overcomes these shortcomings by treating factor groups as fixed effects and speakers as random effects. Its superior performance in avoiding Type 1 error than *GoldVarb* was tested with both real and simulated data (see Johnson, 2009).

Other than mixed-effects probabilistic models, an expanded tool kit including conditional inference trees (*classification trees*; Hothorn, Hornik, & Zeileis, 2006) and conditional inference forest (*random forest*; Hothorn, Bühlmann, Dudoit, Molinaro, & Van Der Laan, 2006; Strobl, Boulesteix, Zeileis, & Hothorn, 2007; Strobl, Boulesteix, Kneib, Augustin, & Zeileis, 2008) have come into language variation and change studies to evaluate the relative importance of internal and external effects when an overwhelming pool of factor groups in exploratory designs were involved (Tagliamonte & Bayaan, 2012; see also a detailed demonstration in Chapter 4 and Chapter 5).

2.8 The Current Study

In spite of variation studies in SLA on a number of domains reviewed above, researchers have called for more fine-tuned studies with an expanded tool kit to provide cross-linguistic, cross-sectional, and longitudinal evidence in exploring the individual-based and group-based learner language variation. To better account for inter-learner variation and intra-learner variability, and to further examine the relationship between individual and group patterns, the current research conducted a cross-sectional variation study with an expanded modern statistical tool kit to investigate L2 vertical variation on multiple morphosyntactic variables with a diverse group of L2 learners.

Morphosyntactic features were considered more dynamically reflective of inter-lingual patterns among other linguistic variables-there are rich numbers of morpheme order studies that has argued for a pre-determined or individually owned L2 developmental trajectory (e.g., Dulay & Burt, 1975; Krashen, 1982; Larsen-

Freeman, 1975; 1976; Luk & Shirai, 2009; Murakami, 2014; Murakami & Alexopoulou, 2016; Pica, 1983). On the other hand, as variation studies attempt to use mixed-effects modeling to account for the multi-dimensional nature of L2 development, a major methodological difficulty is to include the time dimension in addition to the already complex variation with a relatively large number of individuals. The current study seeks to provide a cross-sectional perspective complement to the time dimension by examining multiple observations of multiple L2 morphosyntactic elements-different elements are less likely to be developed and acquired at the same time. Moreover, the design of multiple observations from multiple learners with crossed random effects (i.e., adding not only individual-level but also task-level and item-level random effects in the mixed-effects modeling) would neutralize the underlying intra-learner variability. Consequently, unlike studies focusing on one linguistic variable whose conclusions on acquisitional orders may be questioned of external validity, the current research on multiple linguistic variables offered a more generalizable understanding of L2 competence development cross-sectionally. And unlike studies centering on one or a few learners longitudinally to extensively account for intra-learner variability while little can tell from inter-learner variation, the current research performed a multi-dimensional evaluation on both inter-learner variation and cross-sectional intra-learner variability.

Additionally, though statistical support has been found for the idea that certain linguistic and nonlinguistic (e.g., gender, social-economic status, policy) aspects are shared by the group (e.g., Goldschneider & DeKeyser, 2001; Regan, 2013; Tagliamonte & Bayley, 2012; Young 1989, 1996), empirical evidence is still in great

need to further validate such hypothesis in variationist SLA. Theoretically, as the interlanguage theory has been criticized of its limited explanatory power and vague prediction of L2 competence and incompetence (Al-Khresheh, 2015; Fauziati, 2011; Henderson, 1985; Richard, 1971), there is a clear need in the field to investigate whether certain grammar is subject to interlanguage factors and whether individuals have their own grammar independent of the shared interlanguage patterns. The current research is designed to examine what aspects of interlanguage grammar are shared or not shared by the L2 speech community, and whether the aspects found are generalizable across different linguistic variables.

Moreover, although a fair number of variation research in SLA has investigated a number of linguistic and nonlinguistic factors that promotes the occurrences of the observed variation (e,g,, Adamson et al., 1996; Berlin & Adamson, 2009; Langman & Bayley 2002; Major, 2004; Regan, 2013; Young, 1989, 1991; Young & Bayley, 1996), no variation study, to the best of our knowledge, has systematically investigated the typical paralinguistic properties in the spoken data, i.e., the nonlinguistic and non-verbal communicative features such as hesitation and speech modification. Paralinguistic features are especially important to consider in SLA research, since unlike native speech which may have limited numbers of modifications, spontaneously produced L2 speech often carries a rich number of repairs, hesitations, pauses, fillers, and repeated segments. These features were rather analyzed piecemeal with phonological variables (Hansen, 2001) instead of being systematically considered together with linguistic and nonlinguistic factors on their co-occurrences with the observed variation. Hence, the current study set out to test

the potential factors and factor hierarchies of a group of paralinguistic features on L2 learners' variation in morphosyntactic use.

For these reasons, the aim of this study is to describe and explain the highly variable learner language by investigating whether and how linguistic, paralinguistic, and nonlinguistic aspects co-vary with the acquisition of multiple L2 morphosyntactic features. The study also seeks to determine interlanguage patterns that are shared by or exclusive to each L2 learners with regard to the following research questions (RQ):

- RQ1: How do modern statistical and analytical tools such as conditional inference tree and conditional inference forest perform in identifying important factors and factor groups for the corresponding variable rule analyses?
- RQ2: What are the relative contributions of the linguistic, paralinguistic, and nonlinguistic factors and factor groups on Japanese L2 learners' alternation between native-like and nonnative-like use of English morphosyntactic features, i.e., what factors and factor groups are significantly associated with the linguistic variable examined?
- RQ3: What is the hierarchy of factor levels within the linguistic, paralinguistic, and/or nonlinguistic factor groups which systematically operates on Japanese L2 learners' competence of English morphosyntactic use?
- RQ4: Do individual patterns for L2 morphosyntactic use conform to inter-lingual (group) patterns, i.e., do individual patterns indicate an inherently consistent inter-lingual variation system?

- RQ5: To what extent does individuals' competence in the use of one morphosyntactic feature resemble the other? Is there a systematic implicational pattern across different morphosyntactic features that indicate a fixed L2 developmental order?

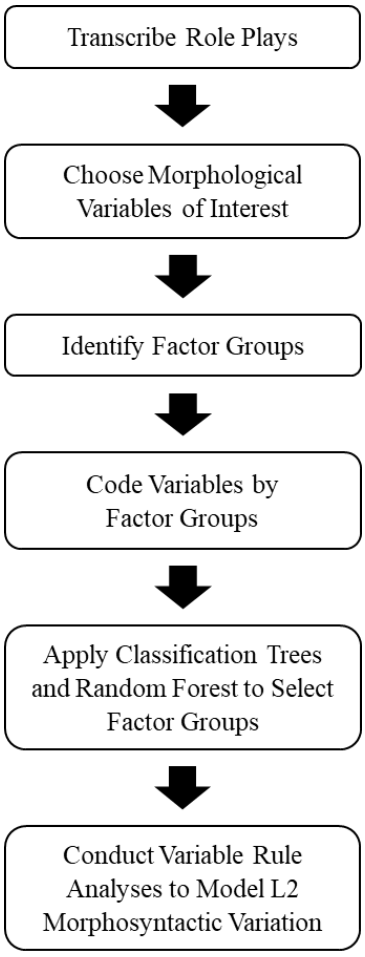
Importantly, for RQ2 and RQ3, detailed linguistic, paralinguistic, and nonlinguistic factors (or factor groups) and factor levels within each factor (or factor group) are illustrated in detail in Chapter 3 after the morphosyntactic features of interest for the current research is introduced.

Chapter 3: Data and Methodology

This chapter aims to describe the data for the current research and introduce the methodological steps for demonstrating the systematicity of morphosyntactic variation in L2 speech. Figure 3.1 displays an illustration of the methodological and analytical steps for modeling morphosyntactic variation in L2 speakers, and the current chapter focuses on the first three of the steps.

Figure 3.1

Illustration of Steps for Modeling Morphosyntactic Variation in L2 Speakers



Three standard terms that will be frequently used in this and the following chapters are *variable, factor group,* and *factor level*. In variationist linguistics, a variable is the short form of the term linguistic variable, "the feature that varies and that is under investigation (p.9, Tagliamonte, 2011)". In the current research, variables are morphosyntactic features observed and extracted from the L2 speech for further examination. Factor groups are "aspects of … the linguistic context that influence the variable phenomena (p.9, Tagliamonte, 2011)", therefore, the independent variables in statistical terms. Factors levels are levels within each factor group that may differ in their influence on the occurrence of the dependent variable of interest.

3.1 Data

The data for the current research contains 46 role play audio recordings from the TOEIC-LPI test (The Test of English for International Communication-Language Proficiency Interview), an optional language proficiency interview speaking test used before the internet version of TOEIC was launched. The language of the speaking test was English, the interviewers were certified native speakers of English, and the 46 candidates were native speakers of Japanese learning English as a second language. It is worth mentioning that the variationist sociolinguistics paradigm requires spontaneous and unmonitored speech in order to reliably describe and evaluate the individual and group variation patterns (Regan, 2013; Tagliamonte, 2011). Unmonitored L2 speech, however, is difficult to obtain and to analyze systematically. The current TOEIC-LPI data is not unmonitored but still spontaneous L2 speech: learners were emotionally engaged in playing the role and would react spontaneously

to unprepared situations provided by the interviewer in real time.

3.1.1 The TOEIC® Test

The Test of English for International Communication (the TOEIC), first introduced by the Educational Testing Service (ETS) and administered by the Institute for International Business Communication (IIBC), was designed to evaluate English language skills since the late 1970s. The original idea, proposed by a man named Yasuo Kitaoka, was to measure achievement in using English within a business environment. The first TOEIC test was created in 1977. In December 1979, 2,710 test-takers took the very first TOEIC test in Sapporo, Nagoya, Tokyo, Osaka, and Fukuoka in Japan (Chapman, 2003, 2004; McCrostie, 2010; Woodford, 1982).

The TOEIC test as well as the TOEIC Bridge test (i.e., a shortened version of the TOEIC test) have been widely used across countries, regions, populations, and industries since the 1980s. In Japan, a large number of companies, both internationally and domestically based, require TOEIC results from current and future employees (Templer, 2004).

The data for the current research was collected before the internet version of TOEIC was launched, and the multiple-choice based TOEIC tests at that time is composed of the two sections: listening comprehension and reading comprehension. Each section takes approximately 45 minutes and 75 minutes. The scores range from 4 to 459 for each section, and the total scores range from 10 to 990.

3.1.2 The TOEIC® Test and the TOEIC-IPL

Although only listening and reading skills were directly measured, efforts had been

made to claim that the TOEIC test could also be considered an indirect measure of English communicative abilities (e.g., Cunningham, 2002; Miyata, 2004; Woodford, 1982). On the other hand, researchers argue that test-takers might misinterpret the purpose of TOEIC, as no spoken section but only listening and reading sections were included (e.g., Childs, 1995).

Against this background, the TOEIC-LPI (Language Proficiency Interview) was first implemented in June 1980 as an optional speaking test component for candidates with a TOEIC score of 730 or higher (MacGregor, 2000; Miyata, 2004). The TOEIC-LPI normally took 20 to 30 minutes, during which candidates would be evaluated on their overall spoken English proficiency (MacGregor, 2000). The speaking test was conducted by TOEIC-IPL certified native speakers, and the interviews were recorded and rated independently by two TOEIC-IPL raters. The TOEIC-LPI test was composed of three phases: *warm-up, level checks and probes*, and *wind-down* (MacGregor, 2000; Quinn, 2000).

Warm-Up. The purpose of the *warm-up* phase is to familiarize the candidate with the target language (i.e., English) as well as the task mode. During this phase, the interviewer will have a preliminary judgment of the candidate's spoken proficiency level. Simple conversations such as greetings and self-introductions often happen at this phase.

Level Checks and Probes. The purpose of the *level checks and probes* phase is to find the highest sustainable level of the candidate's proficiency: First, the interviewer identifies a sustainable level based on the initial oral performance of the candidate; Second, the interviewer "probes" to a higher level than the identified level

to see if the candidate could sustainably exhibit the higher level; Third, the interviewer continues with an even higher level, repeats the second step until the candidate fails to sustainably demonstrate the language properly. Fourth, the interviewer goes back to the highest level before the last unsuccessful "probe" and finishes the interview.

Wind-Down. Lastly, the TOEIC-IPL closes with the *wind-down* phase, which aims to restore the candidates' confidence after probing their proficiency limit and provide the candidates with a feeling of satisfaction when they leave the interview. Sometimes the *wind-down* phase could also serve as a final check, and therefore the wind-down phase may be maintained at the previously identified level as well.

As the essential part of the oral proficiency test, the level checks and probes phase are often composed of one or more role plays, which will be detailed in the following section.

3.1.3 Role Playing and Proficiency Levels

Role plays in proficiency interviews are primarily used to simulate realistic situations with the purpose of identifying test takers' level of proficiency. Role playing is a useful technique to assess proficiency as it involves contextualized language use and unexpected real-life communications.

The TOEIC-LPI is composed of 6 levels with a rating scale from 0 to 5, and the interviewers use different role plays for candidates at different levels (Educational Testing Service, 1999; MacGregor, 2000; Quinn, 2000). Level 0 candidates are almost incapable of forming a complete speech. Level 1 candidates can ask and answer questions with short and simple conversations. Common Level 1 role plays

are ordinary real-life situations, for example, making a dinner reservation, asking to have a camera repaired, asking a dry cleaner to clean a suit jacket, etc. These role plays normally involve direct exchanges of names and dates, and importantly, without involvement of any complication-the reservation is always available, the repair is always possible, the dry-cleaning service is always manageable, etc. Level 2 candidates may be asked to perform a current, past, or future narration (e.g., describe your bedroom in detail). Role play situations at this level are still commonly occurring situations, but with complications-the reservation is not available for some reason, the camera could not be repaired, the dry-cleaning service could not be finished on time, etc. An important skill-to-test for a level 2 task is candidates' ability to form causal conversations-candidates need to understand the cause of the complication (e.g., reservations are not available for a busy Saturday night) and fully participate in negotiating for an alternative solution (e.g., reschedule another reservation). Level 3 speech involves non-routine topics outside common daily experiences. Role plays for a Level 3 candidate include situations where something unforeseeable has happened (e.g., the teacher wants to know why the homework has been missing for several weeks). Level 4 candidates can accurately perform persuasion and negotiation in the role play, during which the candidate may be asked to play a role of an employee and persuade his or her boss to remove the boss' son from an undergoing project the candidate oversees. The candidate is also expected to adjust his/her language to fit the target audience (e.g., do a formal speech in the city hall, have a conversation with children at the kindergarten, etc.). Level 5 candidates perform equally well as native speakers, and as the ETS manual of TOEIC-IPL

particularly emphasized, the candidates' performance should be comparable to an *educated* native speaker (Educational Testing Service, 1999).

A total of 50 role plays involving 18 different scenarios were used in the current research (see a full list of the scenarios in Appendix S3.1). Excerpt 3.1 is a sample transcription of one of the role plays for the current research. It features a common occurring daily activity that Level 1 and Level 2 candidates ("C") would encounter. The interviewer ("I") first introduced a verbal description of a scenario (see Line 1-5, 7-9), oftentimes the candidate is also allowed to read a written description of the scenario in his/her native language (i.e., Japanese). During this role playing, the interviewer added a complication (see Line 19-26) and expect the candidate to negotiate. Unlike level 2 candidate, level 1 candidate failed to detect and/or solve the complication (see Line 34-38), and he/she was likely to be rated as level 1.

Excerpt 3.1

Visiting a Dry Cleaner

01	I:	we are going to finishing out the interview by doing a role play
02		(1.2)
03		we are gonna act out the situation
04		(1.7)
05		a::nd let's see let me take you to London
06	C:	(0.4) far ·hhh
07	I:	you are in London on business and noticed that your suit jacket is dirty
08		(1.1) ·hhh I'm going to play the role of a dry cleaner
09		ask me to clean your jacket (.) in Japanese if you prefer
10		(0.7) hi how can I help you today
11	C:	uhm I got a s- (.) a spot here and I I spray my orange juice here ·hhh
12		so:: can I take it off?

I: ORANGE juice
C: yes
I: oww::
C: and (.) this is uhm (.) very expensive an Italian wool ·hhh so
(0.7) and I I really need (1.0) in two hours
(1.4) can you do it quickly for me?
I: well um
(2.0)
well let's see that kind of spot on that type of cloth ·hhh takes special
attention (.) and usually we have to send that out (.) to:: the factory
(0.8) and so if I were to do that it's going to be VERY hard to ready
in two hours (1.1) uhm::
(1.2) of course if you're willing to stay- t- to pay for an extra fee
(0.9) I can (1.0) probably do something (.) extra
C: (0.9) okay
I: so uh normally it cost- to clean that kind of jacket (.) would be about
eight pounds and um (0.7) with that if you want extra service
I'm gonna double the charge it's gonna be sixteen pounds to clean up
C: so it IS possible for you to [prepare two hours.
I: [it's possible-
it's possible (.) but it will cost too.
C: (1.2) okay uhm
(3.0)
the::
(1.5) okay I'm- (0.7) can- can- can you let me know any- anyway I can-
I can replace (.) I can buy new new jacket?
I: well I mean I CAN do this in in (.) two hours (.)
but you know you have to pay sixteen pounds
(0.8) and if you buy a new jacket
new jacket will cost you probably a hundred and sixty pounds
C: (0.9) okay I'm sorry I have no idea about the pounds so
pounds okay okay uh ·hhh okay (1.0)
I: okay
C: okay
I: alright [so
C: [so I'll be back in in (.) two (.) hours
I: okay that'll be sixteen pounds please
C: thank you
I: okay thank you very much and your change and here’s your ticket
and we'll see you (.) very soon

53 C: okay (.) thank you
54 I: welcome

3.1.4 Data Transcription

All 50 role plays were recorded and fully anonymized before data transcription: the warm-up conversations involving names, occupations, and other identifying information were removed. The research was initially reviewed by the IRB (Internal Review Board) and was then exempted from further review according to IRB decisions.

All audio-recorded data were transcribed manually for several reasons. First, automatic speech-to-text transcription software was relatively not flexible in transcribing the current L2 speech. As a large proportion of the current data set are TOEIC-LPI Level 1 and Level 2 speech with limited proficiency and a relatively heavy accent, the quality of automatic transcriptions fails to meet the required precision for the current research. Furthermore, fluency features such as pauses, hesitations, fillers, repairs, etc., are important factor groups of interest in the analyses. Manual transcription is necessary because automatic transcription is unavailable in recognizing and forming detailed speech fluency markings.

A reduced discourse transcription convention (adapted from Du Bois, 1991; Du Bois, Cumming, Schuetze-Coburn, & Paolino, 1992) was followed to retain linguistic, paralinguistic, and nonlinguistic information in written form for further analysis (see Excerpt 3.1 as an example). A hyphen (-) indicates a break in the intonation unit and usually an indication of repair in the speech (see Line 37 in Excerpt 3.1 as an example). Square brackets ([word]) indicate speech overlap

between the candidate and the interviewer. The overlapped speech is usually aligned vertically together (see Line 31, Line 32 in Excerpt 3.1 as an example). The left bracket ([word) or the right bracket (word]) marks the beginning and/or the ending of overlap, respectively. A double colon (::) is used to indicate an unusual lengthened segment (see Line 5 in Excerpt 3.1 as an example). The double colon within or after a non-filler word (e.g., um, uh, huh, etc. are filler words) is considered a hesitation in the speech in the current research. A dot within single parentheses ((.)) indicates an untimed pause, usually an audible inhalation which is shorter than 300ms. Timed pauses, usually 300 milliseconds or longer, were transcribed as numbers rounded to the nearest tenth of a second within parentheses, such as (0.7), (3.5), (4.1), etc.

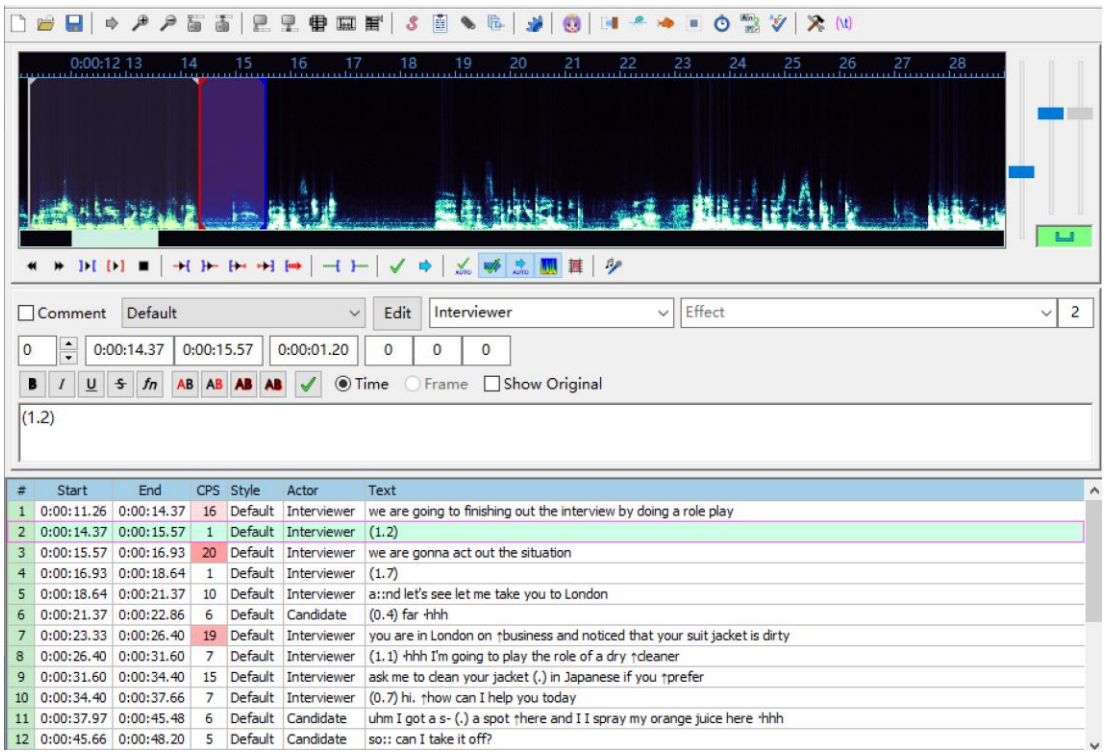
Additionally, punctuations (e.g., commas, period, slash, etc.) were not transcribed since they are non-spoken elements. The only exception is the use of question marks which appears in transcription where a rising tone is detected to indicate a question speech in the superficially declarative utterance (see Excerpt 3.1, Line 38: *I can buy new new jacket?*)-question speech was coded as an important factor group in the research. Written forms such as apostrophe expressions and hyphen words were transcribed and counted as separate words, this is necessarily conducted due to the coding difficulty for spoken errors where part of the apostrophe expression (e.g., *I I I'd like to go to lunch with you*) or hyphen words (e.g., *my mo-mother:: (.) in law is waiting for me*) were involved. L2 speech accent was not transcribed since the current research focuses on morphosyntactic variation.

An open-source subtitle editor software named Aegisub (version 3.2.2; Monteiro & Hansen, 2013) was employed in the manual transcription process. Figure

3.2 presents a sample Aegisub interface for the current analysis in transcribing a piece of audio recording. The upper right corner of the interface features three zooming sliders which allow audio quality and noise level adjustment for transcription. The vertical zooming sliders, especially, were used to adjust background noises in the recordings. Timestamps marking utterance start and end time were displayed with utterance duration and pause duration (e.g., Line 2, "(1.2)") being accurately measured and automatically generated. The software-specific script file was then converted to plain text for further coding.

Figure 3.2

Sample Aegisub Interface



#	Start	End	CPS	Style	Actor	Text
1	0:00:11.26	0:00:14.37	16	Default	Interviewer	we are going to finishing out the interview by doing a role play
2	0:00:14.37	0:00:15.57	1	Default	Interviewer	(1.2)
3	0:00:15.57	0:00:16.93	20	Default	Interviewer	we are gonna act out the situation
4	0:00:16.93	0:00:18.64	1	Default	Interviewer	(1.7)
5	0:00:18.64	0:00:21.37	10	Default	Interviewer	a::nd let's see let me take you to London
6	0:00:21.37	0:00:22.86	6	Default	Candidate	(0.4) far ·hhh
7	0:00:23.33	0:00:26.40	19	Default	Interviewer	you are in London on ↑business and noticed that your suit jacket is dirty
8	0:00:26.40	0:00:31.60	7	Default	Interviewer	(1.1) ·hhh I'm going to play the role of a dry ↑cleaner
9	0:00:31.60	0:00:34.40	15	Default	Interviewer	ask me to clean your jacket (.) in Japanese if you ↑prefer
10	0:00:34.40	0:00:37.66	7	Default	Interviewer	(0.7) hi. ↑how can I help you today
11	0:00:37.97	0:00:45.48	6	Default	Candidate	uhm I got a s- (.) a spot ↑here and I I spray my orange juice here ·hhh
12	0:00:45.66	0:00:48.20	5	Default	Candidate	so:: can I take it off?

Importantly, both interviewer speech and candidate speech were transcribed

(see Excerpt 3.1 and Figure 3.2 as examples). Although interviewer speech is not evaluated in the TOEIC-LPI rating, and conventional practices in speech analysis normally remove rater speech and analyze the candidates' speech solely, the interviewer speech is considered as a secondary source of speech information indicators where potential contextual clues exist as the candidate listens and responds.

3.2 The Linguistic Variable: Selecting Morphosyntactic Features

Starting with the observational data, the research set out to look for patterned morphosyntactic features in the L2 speech transcription. Several steps and criteria were followed for selecting the appropriate morphosyntactic features for statistical analysis (Tagliamonte, 2011):

- Step1: Find a morphosyntactic feature that occurs in both accurate and inaccurate forms. The forms must recur repeatedly, within and across individuals.
- Step2: Formulate hypotheses in which the occurrence of accurate and inaccurate forms of the variable could be tested.
- Step3: Coding the transcribed data to verify the hypothesis. It’s important to confront the data with the hypothesis at an early stage.
- Step 4: Decide the number of tokens of the variable of interest within a reasonable amount of data (i.e., manually transcribed role plays). If not enough tokens were found within a reasonable amount of data to warranty an analysis, researchers could choose from the following two options:
    - Step 4.1: If the amount short is within a practical range, transcribe more role plays;

    - Step 4.2: Find another morphosyntactic feature;
- Step5: Start with another morphosyntactic feature and repeat the previous steps.

The current study set a token number per feature at the minimum of 200, with initially 45 role plays transcribed, and an addition of 5 to meet the token number requirement. Accurate and inaccurate tokens were then extracted from the transcribed audio-recorded data by two native speakers of American English on a native speech standard.

The current study, designed to understand how learner language is situated in between the target L2 and their native language, focuses on several morphosyntactic features. Firstly, English preposition/particle usage, one of the most frequently observed morphological confusion for L2 learners, requires looking into the typological differences between learners' L1 and L2. The second, article use, a perennial morphological tangle that even advanced L2 learners of English fail to handle perfectly, implicates the effect of a missing grammatical morpheme in learners' L1 on their acquisition of an L2. The third, pronoun-dropping, offers a cross-linguistic insight in L2 morphosyntactic acquisition, as the native language of the L2 speakers in the study, Japanese, allows extensive pronoun dropping while English (L2) does not. Lastly, modal auxiliary verb usage in the L2 speech provides both a morphosyntactic and a pragmatic perspective on language variation studies. In the following section, each of the morphosyntactic features will be detailed.

3.2.1 Morphosyntactic Feature—Preposition/Particle

English preposition was found to be one of the most frequently observed grammar

and usage errors in both L2 written and oral productions (e.g., Bitchener, Young, & Cameron, 2005; Izumi, Uchimoto, Saiga, Supnithi, & Isahara, 2003; Leacock, Chodorow, Gamon, & Tetreault, 2010; Ting, Mahadhir, & Chang, 2010). Besides the idiosyncratic use which confuses learners (Tetreault & Chodorow, 2008), errors in L2 English preposition usage were also considered as interlingual errors (Chelli, 2013; James, 2013). The inaccurate forms appear to be associated with language transfer from the learners' native language-direct erroneous translation from learners' L1 was suggested to account for the inaccurate use of English preposition.

Japanese L2 learners whose native language is a verb-framed language may find it particularly difficult to conceptualize the extraneous prepositions in prepositional phrases in addition to the main verb, and similarly, the particles in the phrasal verb (e.g., Cadierno, 2008; Slobin, 1997). This difficulty stems from the typological differences between the learners' L1 and L2: English, as a satellite-framed language, expresses the path of a motion event in non-verb elements, e.g., by an adverbial particle or a preposition after the main verb. On the other hand, Japanese, a verb-framed language, expresses the information about the motion path in the main verb (Matsumoto, 1996; Miyajima, 1984; Talmy, 1985; 1991). For example, the phrasal verb *wake up* in Japanese is *okiru*, and the verb phrase *look for* in Japanese is *sagasu*. As a result, Japanese learners of English are less aware of the necessity of the non-verb elements, usually the particle or the preposition, and would not fully understand the choice of one particle/preposition over the other (Neagu, 2007; Yasuda, 2010).

Example 3.1 Prepositional Phrases

(a) because I think I just forget my wallet in the office
(b) would you mind to uh lend me some money for lunch
(c) um (.) I think the cheapest start from eight dollars and then ah nonono- six dollars
(d) so I will met uh (.) yokohama and I need to go back to yokohama
(e) I want to live (0.6) umm:: to area that (2.2) I can go to university by:: within one hour
(f) and:: she want to accomp accompany with me

Against this background, linguistic variable Preposition/particle in this research involves the accurate and inaccurate forms of prepositions in prepositional phrases, as in Example 3.1, and the accurate and inaccurate forms of particles in phrasal verbs, as in Example 3.2. A total of 671 tokens (83 inaccurate and 588 accurate variants) were extracted among the 50 role plays in the current research.

In Example 3.1, (a), (b), and (c) present the accurate use of prepositions in, for, and from in each of the prepositional phrases. Token (d) displays an inaccurate omission of the preposition at in the prepositional phrase meet (her) [at] Yokohama. Token (e) shows another inaccurate use of preposition to, and the accurate use should be live [in] (an) area. Note that for (d) and (e), not only did the L2 learner inaccurately used the proposition, omission of the object pronoun her and the indefinite article an were also observed. Token (f) is an addition of the preposition with where it is not needed.

Example 3.2 Phrasal Verbs

(a) ohhh I like going out and working in the worldwide so (.) I prefer en jacket
(b) uhm please see this (0.6) place uhm (.) uhm the dirt the very dirty place uhm please clean up (.) it
(c) but I (1.2) but I but I don't remember I bumped in someone ep so (.) could be wine but I just jacket this is bit uh classic you can't find out
(d) oh I think I can pop in here

Example 3.2 above presents the accurate use of an adverbial particle *out* in a phrasal verb *going out* (a), and three inaccurate uses of particles (b) and (c) which reflect two different error types: misordering and misinformation. Specifically, token (b) marks the misordering of the particle *up* before the pronoun *it* where *up* should be placed after *it* to form a phrasal verb *clean it up*. Token (c) is a misinformation of the particle *in* with the accurate phrasal verb form being *bumped into*.

3.2.2 Morphosyntactic Feature—Article

Variation in article usage has been well documented in both empirical and theoretical studies in second language acquisition (e.g., Abe, 2003; Freeman, 1975; Izumi et al., 2003; Izumi & Isahara, 2004; Han, Chodorow, & Leacock, 2006; Hakuta, 1976; Nomura, 2012; Tarone, 1985). Research in L2 error analysis and related fields has reported that article errors were among the most frequently observed morphosyntactic and usage errors in L2 learners' oral and written productions with a range of frequency from 10% to more than 30% (e.g., Chodorow, Gamon, & Tetreault, 2010; Izumi et al., 2003; Izumi & Isahara, 2004; Ting, Mahadhir, & Chang, 2010).

English article is particularly difficult for learners whose native language has no article system such as Japanese, Korean, and Chinese (e.g., DeKeyser, 2005;

Freeman, 1975; Han, Chodorow, & Leacock, 2006; Lightfoot, 1998; Nomura, 2012; Thomas, 1989; Odlin, 1989). Previous studies have suggested that the lack of noun countability and plural marking in L1 Japanese may confuse Japanese L2 learners in distinguishing between mass and count nouns (Butler, 2002; Hiki, 1991). In addition to the novelty and abstractness of applying articles to where it is necessary, the difficulty for Japanese learners to master article usage also lies in the ambiguous use between definite and indefinite article (Hakuta, 1976).

Particularly, researchers have indicated that the difficulty for Japanese learners of English in article acquisition may also stem from the acquisitional order which was later than predicted by the natural order (e.g., Izumi & Isahara, 2004; Tono & Aoki, 2000). Luk and Shirai (2009) summarized and reanalyzed previous studies reporting acquisitional order on plural (-s), possessive (’s), and articles of learners from various L1 backgrounds (Chinese, Japanese, Korean, and Spanish) and compared their L2 acquisition orders against natural order (NO; Krashen, 1977). The authors observed a later-than-NO acquisition on articles for Japanese L2 learners and an acquisition order consistent with NO on articles for Spanish L2 learners. As a result, the authors concluded the consistency of acquisition order gives L1-Spanish L2-English learners advantages in learning articles over those whose L1 is Japanese.

For the current study, the total tokens detected for variable article for the current analysis are 376, which involves the variation in the usage of the definite article *the*, the indefinite article *a* and *an*, and in some cases, the zero article (error of addition). A total of 164 inaccurate uses were detected, marking the highest percentage of inaccuracy (43.62%) among all morphosyntactic features analyzed in

the current study. Example 3.3 and Example 3.4 illustrate the inherent variability of definite and indefinite article variation in L2 speech, as the accurate and inaccurate uses are found in the speech of the same candidate with the same interviewer in the same role play activity (Tagliamonte, 2011).

Example 3.3 Definite article

(a) y- yes but sometimes he doesn't so hey you know inaffectation or keep the attentions to the:: same subject
(b) maybe (1.0) she likes so many things and (1.4) favor she find very interesting that she (.) very you know (1.5) throw away the homeworks
(c) en (.) that's a problem
(d) well of course you know we ask her you know the:: every morning that if-uh she has finish the homework

In Example 3.3, token (a) and (b) are accurate uses of the definite article *the*. Token (c) is a misinformation error in which the definite article *the* should be used according to prior context. The token (d) shows another type of error: no article is needed before the noun phrase every morning. Example 3.4 below presents accurate (a) and inaccurate (b, c, and d) uses of indefinite articles. In (b), the L2 learner made a misinformation error where the indefinite article *an* were mistakenly replaced by the definite article *the*. Token (c) marks the inaccurate addition of the indefinite article *a* as spaghetti in English is an uncountable (mass) noun. Token (d) displays a case of article omission: the indefinite article *a* was erroneously dropped in the phrase *in a rush*.

Example 3.4 Indefinite article

(a) how- how long does it take to a time a day?
(b) when-I I just (0.6) Well I was here for the business trip
(c) It's a spaghetti
(d) I need it ssoon as possible but uh w- (0.5) so if I (1.0) uh: were in rush

3.2.3 Morphosyntactic Feature—Object Pro-Drop

One of the biggest differences between Japanese and English is that Japanese is a radical pro-drop language that allows empty subjects and empty objects. For example:

(1) Deki-ta.
did-Past
I did it.
(2) Watashi-ga deki-ta.
I-Top did-Past
I did it.

In Japanese, both (1) and (2) are perfectly grammatical only that (2) emphasizes the topic *I* therefore only null object occurs. In contrast, both subject and object are necessary in English to grammatically convey the complete meaning in (1) and (2). It is argued that languages with a rich (subject-verb) agreement system (e.g., Italian, Spanish, Arabic, and Portuguese) permit empty subjects as the missing argument could be determined by the agreement system. Therefore English, as a language that lacks a rich subject-verb agreement system, does not allow pro-drop generally. On the other hand, Languages such as Japanese, Chinese, and Korean, although having no subject-verb agreement system, are known as radical pro-drop languages, or more frequently, discourse pro-drop languages that allow null subjects and null objects (e.g., Huang, 1984; Nakamura, 1991). Researchers have proposed that null objects in

radical pro-drop languages are licensed at the syntax level (Park, 2004; see also the agglutinative pronominal paradigm; Koeneman & Zeijlstra, 2019; Neeleman & Szendrői, 2005; 2007; 2008) as well as the pragmatic level by a preceding noun phrase or a topic in the discourse (e.g., Cote, 1996; Huang, 1984; 1989; Ohara, 2007; O'Grady, Yamashita, & Cho, 2008; Tomioka, 2003; see also White, 2016 for a review).

The pro-drop phenomenon is also one of the earliest parameters investigated in second language acquisition (e.g., Camacho, 2013; Hilles, 1989; Lakshmanan, 1991; Liceras, 1988; Park, 2004; Phinney, 1987; White, 1985; 2016; Yuan 1997; Zobl, 1990). White (2016) in a review of pro-drop research in second language acquisition from the past 30 years has concluded that investigating the pro-drop phenomenon has provided important insights in understanding the development of interlanguage grammar. That is, with the default L1 parameter setting (i.e., null subject and null object), how L2 learners unlearn null subject and null object and to what extent can these L1 parameter settings be reset in L2 has provided insights into second language acquisition from the perspective of L1-L2 language transfer.

While a rich number of L2 pro-drop research invested in the unlearning of null subject in L2 as evidence in support of successful L2 parameter resetting (e.g., Camacho, 2013; Hilles, 1989; Lakshmanan, 1991; Liceras, 1988; Phinney, 1987; White, 1985; see White, 2016 for an overview), a limited number of them paid attention to the null object phenomenon (Park, 2004; Yuan, 1997; Zobl, 1990). Yuan (1997) was one of the early studies that investigated both the null subject and the null object phenomenon. With L2 English learners whose L1 was Chinese, Yuan

discovered an asymmetric pattern between ungrammatical null subject acceptability and ungrammatical null object acceptability: L2 learners were able to reject ungrammatical null subject sentences in English while they experienced great difficulty in detecting the problematic null object sentences. Yuan, therefore, concluded while the unlearning of null subjects in L1 (Chinese) is successfully reset by the lack of rich subject-verb agreement in L2 (English), the learners failed to unlearn the object drop, indicating the fundamental mechanism between subject drop and object drop is different. Ungrammatical null object in L2 is, therefore, a product of L1 transfer.

The current research employed the object pro-drop variation in the observed L2 speech data and attempted to investigate whether the occurrence of incorrect object pro-drop in L2 English is a product of L1 transfer as Japanese (L1) is a radical pro-drop language. The overall token numbers for this feature are 247, with 43 inaccurate and 204 accurate occurrences. Again, the best examples that show inherent variability are found within the speech of the same L2 speaker, as in Example 3.5.

As demonstrated, token (a) and (b) in Example 3.5 accurately used the personal and impersonal pronoun to form a grammatical sentence. Token (c) and (d), on the other hand, inaccurately dropped the object pronoun *it* and *you*, respectively.

Example 3.5 Pronoun Dropping

(a) em so uhh where is <u>it</u>
(b) some money from <u>you</u>
(c) can I <u>borrow</u>
(d) uhh (0.9) okay I will certainly <u>repay</u> (0.4) the end of (.) month so

3.2.4 Morphosyntactic Feature—Modal Auxiliary Verbs

Modal auxiliary verbs are used to express speakers' stance on possible situations. Previous studies have reported that accurate English modal usage is difficult for L2 learners with various L1 backgrounds (e.g., Aijimer, 2002; Cook 1978; DeCarrico, 1986; Hinkel, 1995; Hu & Li, 2015; Kasper, 1979; McEnery & Kifle, 2002; Nakayama, 2021). The difficulty, as some researchers have suggested, may stem from learners' reliance on one-to-one form-meaning association while English modal usage is ambiguous (e.g., Hu & Li, 2015; McEnery & Kifle, 2002; Nakayama, 2021). For example, the modal verb can could be used to express possibility, permission, as well as ability, and one can choose from can, could, may, might to convey different degrees of possibility. On the other hand, some also indicated that learners may not be aware of the pragmatic property and understand the contextual implications English modality carries (e.g., Altman, 1990; Hinkel, 1995; Kasper, 1979; Robberecht & Van Peteghem, 1982) as modality in some languages is obligatory (e.g., Korean; Choi, 1995; McDouall, 2012) or is strictly restricted by morphosyntactic environment (Japanese; Hirakawa, & Suzuki, 2010; Moriya & Horie, 2009; Takayama, 2002).

L2 spoken data was believed to contain a rich modal usage-it involves spontaneous processing of and responding to real-time information and hence an increased level of uncertainty as compared to written data (Chafe & Danielwicz, 1987). For Japanese-L1 English-L2 speakers, research indicated the added politeness addressed by the modality speech in English (e.g., Boncea, 2013; Kennedy, 2002; Palmer, 1990; 2001) may be pragmatically favored by Japanese learners as their L1 accommodates different levels of politeness which reflects the Japanese social and

cultural manners (e.g., Nakayama, 2021; Nozawa, 2010).

Most research reporting on L2 learners' modality use were conducted with written data, and they focus on perspectives such as overuses and underuses of modality as an indicator of interlanguage development (e.g., Aijimer, 2002; Back & Lee, 2017; Basham & Kwachka, 1989; Hinkel, 1995; 2009; Kwachka & Bascham, 1990; McDouall, 2012; Nakayama, 2021; Pemberton, 2020; Seog & Choi, 2018; Stapleton, 2007; Shirato & Konakahara, 2011). On the other hand, a number of L2 error analyses also picked up modality error in learners' production primarily in written situations, involving modal verb tense errors, inflected main verbs, or missing of the main verbs after the modal (e.g., Dutra, 1998; McDouall, 2012; Nakayama, 2021; Sembiring, Sibarani, & Mangaraja, 2021).

Interestingly, the current spoken data observed a missing pattern which has been less frequently documented in previous error analyses-the missing of the modal verb itself (Leli, 2017; Pertiwi & Adityarini, 2019). In fact, missing modal verb seems to be the primary source of modal errors in the current data: Thirty-four out of 209 tokens were found to demonstrate modal usage errors in the L2 speech, among which only 2 were misuses of other modal verbs. Among all 209 tokens involving modal verb use, only 3 involve the misuse of inflectional forms of the main verbs, 2 involve the omission of the modal verb and the main verb, and 3 involve the omission of the main verb only-note that the missing main verb are all "be"s. This error distribution observed is very different from previous error analysis in which misinformation (i.e., incorrect modal verbs or forms) was found to be the most frequently observed modal error while omission is the least or the second least

observed error (e.g., Pertiwi & Adityarini, 2019; Sembiring, Sibarani, & Mangaraja, 2021). The exploration on how and potentially why missing modal verb has become a dominant source of error in the current L2 speech data may provide important evidence and implications in understanding L2 modality acquisition. Example 3.6 below presents the modal auxiliary variable in the current data, and all were from the same candidate to demonstrate inherent variability. Specifically, token (a), (b), and (c) in Example 3.6 present the accurate use of the modal verb *would* and *can*, while token (d) and (e) display the inaccurate omission of the modal verb *would* and *will*.

Example 3.6 Modal Verb Omission

(a) would you mind to uh, lend me some money for lunch
(b) and we can go to lunch two days
(c) oh that's too ba::d so I cannot borrow some money huh
(d) I like to go to sushi bar too but tempura sounds good
(e) because I think I just forget my wallet in the office so I just go get ehn my money (.) and come back

3.3 The Linguistic Variable: Identifying Factor Groups

After detecting variants (i.e., the inaccurate and accurate forms) of the morphosyntactic features in the L2 speech, the essential question to ask is why the inaccurate or the accurate form occurs or why the accurate form does not occur when it could have occurred. To model the occurrences of the linguistic variable of interest, the research examined the potential factor groups that might explain the variation observed.

Factor groups are linguistic, paralinguistic, and non-linguistic aspects that are

hypothesized to account for the observed variation in the data. As mentioned in the beginning this chapter, factor group is the field standard term in variation studies, and in statistics, these terms are often referred to as independent variables or predictors (Tagliamonte, 2011).

Based on theoretical and empirical research in second language acquisition and related fields, three types of factor groups were examined in understanding the observed variation: linguistic factor groups, paralinguistic factor groups, and nonlinguistic factor groups. The linguistic factor groups are variable-dependent linguistic contexts that were believed to conditions the occurrence of the variable. The other two types of factor groups, paralinguistic, and nonlinguistic factor groups are shared across variables to provide an overall understanding of the variant occurrence across variables and individuals.

3.3.1 Linguistic Factor Groups

Like many variables in phonological, morph-syntactical, or semantic variation studies, L2 use in morphosyntactic feature preposition/particle, article, object pronoun, and modal auxiliary verb all have standard (accurate) and nonstandard (inaccurate) forms. The observed variation is likely to be conditioned by their linguistic properties and contexts which constitutes the linguistic factor groups of the current research.

3.3.1.1 Linguistic factor groups—Preposition/Particle

Variation in preposition/particle use in L2 speech may be conditioned by contextual features such as the preceding verb (Chodorow, Gamon, & Tetreault, 2010). Whether

the preceding verb is a transitive or an intransitive verb will directly determine the linguistic structure followed by the verb. In addition, previous studies have shown that prepositional phrases acting as adverbs and phrasal verbs are more frequently perceived as purely idiomatic or as formulaic sequences for L2 learners (Yasuda, 2010). It is therefore reasonable to hypothesize that frequency of exposure, i.e., the word frequency of the preceding verb and phrasal verb (also their root forms) may be a good indicator of their co-occurrences with the variation in preposition/particle in the current data set (e.g., Cooper, 1999).

Along with linguistic context, linguistic structures also play a role in estimating the observed preposition/particle variation. When using a satellite-framed L2 like English, it is reported that verb-framed language speakers like Japanese L2 learners prefer using a single verb structure and are less aware of the orientational and spatial meanings carried by the prepositions and/or particles in English (Cadierno, 2008; Neagu, 2007; Slobin, 1997; Yasuda, 2010). As a result, particles or prepositional phrases acting as adverbs can be expected to be a greater challenge and hence trigger more inaccurate form occurrences for Japanese L2 learners of English. Meanwhile, due to the great similarity between the English V – [P – NP] (i.e., verb + prepositional phrase) and the English V – Prt – N/NP (i.e., verb + particle + noun/noun phrase) constructions, additional confusions are expected when Japanese L2 learners encounter them in their L2 use (Cappelle, 2005). Therefore, construction (phrase type) is included as a factor group in accounting for preposition/particle variation with the prediction that English V – [P – NP] and V – Prt – N/NP constructions are more likely to co-occur with inaccurate preposition/particle uses.

Table 3.1 below presents the factors and factor groups for L2 preposition/particle variation.

Table 3.1

Factors and Factor Groups for Accurate/Inaccurate Preposition/Particle Use

Factor group	**Factor Levels (Coding)**
Preceding Verb	T=Transitive I=Intransitive
Verb Frequency	Word frequencies in the COCA spoken corpus
Verb Frequency (Root Form)	Word frequencies in the COCA spoken corpus
Phrase type	P=Particle I=Infinitive Verb+Prep=P1 (Prepositional phrases acting as adverbs, verbs immediately followed by prepositions; e.g., look for an apartment) Verb+Obj+Prep =P2 (Prepositional phrases acting as adverbs, with direct/indirect objects in between the verbs and the prepositions; e.g., borrow money from other people) Other Preps=P4 V=Verb only (addition error where the preposition/particle is not needed; e.g., he dropped out the wine on my jacket)

3.3.1.2 Linguistic factor groups—Article

For article acquisition, the problem accompanying the absence of an article system in L1 Japanese is that learners can hardly differentiate the use of definite, indefinite, and zero articles (Hakuta 1987; Parrish, 1987). Definiteness, therefore, is considered an important factor group on the use of article variation in L2 speech. With Japanese L2 learners, Nomura (2010, 2012) reported a significantly higher percentage of definite article usage than indefinite article usage in spoken and written learner corpora. In

contrast, the accuracy rate observed showed an opposite pattern: the indefinite article accuracy rate was significantly lower in the spoken corpus (29.17% versus 40.45% in written), compared to definite article uses (38.81% in spoken versus 43.60% in written). The hypothesized effect of definiteness, therefore, favors indefinite over definite in their co-occurrences with inaccurate English article use.

In addition, the difficulty of accurate article usage may be conditioned by a number of lexical, syntactical, and general knowledge-based factors (e.g., Barrett & Chen, 2011; Chodorow, Gamon, & Tetreault, 2010; Han, Chodorow, & Leacock, 2004; 2006; Lee Amuzie & Spinner, 2013; Young, 1996; Young & Barley, 1996). A frequency effect of the following noun after the article position was hypothesized to be tested. Furthermore, a great amount of research has supported that learners' article use is largely influenced by the syntactic type of the following noun or noun phrase (e.g., Bond & Ikehara, 1996; Chodorow, Gamon, & Tetreault, 2010; Han, Chodorow, & Leacock, 2004; 2006; Knight & Chander, 1994; Lee Amuzie & Spinner, 2013; Murata & Nagao, 1994; Young, 1996). Article a can only be used before the countable nouns in their singular form and are not allowed to be used when the following noun is in plural form. Proper nouns such as unique entities (e.g., the Sun), places, things, etc., (e.g., the San Francisco station) are modified by the definite article the (e.g., Leech & Svartvik, 2013). The current study hence hypothesized that nouns superficially singular are more likely to co-occur with inaccurate article uses, as complex rules are involved in article usage before a singular noun. Another crucial property of the following noun is whether a premodifier is attached, such as a superlative adjective (e.g., best) which prefers the use of the definite article the. The

noun phrase without a premodifier is hypothesized to account for more inaccurate uses of articles than the ones with premodifiers. Additionally, the factor group formulaic language includes the article usage in idiomatic phrases, formulaic sequences, and general knowledge. As idiomatic phrases and formulaic sequences are often taught and stored as a non-compositional lexical unit in L2 learners' lexicon (e.g., Gibbs, 1990; Nippold & Taylor, 1995; Nippold, 1998; Yasuda, 2010), the study considers it an important independent variable in predicting the outcome of interest.

Table 3.2

Factors and Factor Groups for Accurate/Inaccurate Article Use

Factor group	**Factor Levels (Coding)**
Definiteness	D=Definite I=Indefinite N=Zero article (in the case of addition)
Syntactic type (Head Noun)	S=Singular noun P=Plural noun R=Proper Noun Q=Quantifier X=Other
Word Frequency (Head Noun)	Word frequencies in the COCA spoken corpus
Premodifier	Numerical (0, 1, 2, 3)
Formulaic Language	F=Idiomatic or formulaic N=None

Importantly, English article use does not always follow the rules and is always accompanied by exceptions and interactions with other grammatical rules. It is therefore important to take all factor groups into consideration and see how they contributes to the use of English article found in the speech of Japanese L2 learners in this study. Table 3.2 list the factors and factor groups that would likely to co-occur

with accurate/inaccurate article uses variation.

3.3.1.3 Linguistic factor groups—Object pro-drop

As mentioned earlier in this chapter, several theories were proposed to account for the pro-drop phenomenon in languages without a subject-verb agreement system, with one emphasized on morphological pronominal properties (Neeleman & Szendrői, 2005; 2007; 2008) and the other at the discourse and pragmatic levels (e.g., Huang, 1984. Tomoka, 2003; White, 2016).

From a syntax-level perspective, Park (2004), Neeleman and colleagues (2005; 2007; 2008), and other researchers (Koeneman & Zeijlstra, 2019) have argued that languages having no pronominal variation (Chinese), or the pronominal inflection of which is independent of the stem (Japanese) allows pro-drop. In contrast, English is a fusional language which radical pro-dop is not permitted. The current research, therefore, included the pronoun and pronoun type factor groups in which demonstrative, impersonal, personal, indefinite, relative, and reflexive pronouns are coded in the data to test their co-occurrences with accurate and inaccurate object pro-dropping in English. Impersonal pronouns are hypothesized to be more related to the pro-drop phenomena, as observed in Yuan (1997)-inanimate null object sentences are classified as grammatical by even verb advanced L2 learners.

On the other hand, the pragmatic level approach on pro-drop is understood to be closely related to the recovery of the missing argument from the discourse (e.g., Huang, 1989; Park, 2004; Tomika, 2003; White, 2016; Yuan, 1997). Huang (1989) explored the distribution of null objects and proposed that for a radical pro-drop language like Japanese, missing arguments are licensed at the pragmatic level and

could possibly be implicitly retrieved from a topic in the discourse. White (2016) in an overview of pro-drop research in second language acquisition also indicated that the pro-drop phenomenon involves the considerations of discourse contexts. To further explore the effect of discourse context on L2 object pro-drop, a number of factor groups describing the prior context were included in the current research, including whether a topic was mentioned in the discourse, by the interviewers or by the candidates themselves, the form of the prior mentioning (i.e., in its pronoun form or in its full form), and the distance between the current pronoun, presence or absence, and the prior pro- and/or the full form. As a hypothesized effect of L1 transfer, the occurrence of a topic in the discourse was considered to account for more object pro-drop since the L1 default setting permits mull object when a topic was mentioned, and the closer the distance between the topic and the supposed object position, the more likely the null object is granted in L1. Furthermore, since the data is from a proficiency test, candidates may be more aware of their own speech than the interviewers' speech, therefore a topic in the candidate speech is favored for the occurrence of L2 pro-drop.

While null object in L1 Japanese seems to be either grounded by the syntactic or pragmatic properties, the licensing of null objects in L2 English is also taken into consideration, as it is more likely a lexically constrained process. Many studies have proposed that the selection of the object is a subcategorization by the preceding verb (e.g., Allerton, 1975; Cote,1996; Lehrer, 1970; Ohara, 2007; O'Grady, Yamashita, & Cho, 2008; Park, 2004). For example, some verbs require an object to convey a complete and grammatical meaning (i.e., transitive verbs) and others (i.e., intransitive

verbs) cannot have a direct object. Therefore, the occurrence of null object is only possible when the preceding verb allows it. In addition, added difficulty was assumed for L2 learners of English as some verbs could be transitive in one sentence and intransitive in another. The transitivity issue is also intertwined with preposition/particle usage, as intransitive verbs take prepositional phrases as adverbs (e.g., *I think the cheapest start* ***from eight dollars***), transitive phrasal verbs take direct objects (e.g., *I can take off* ***my jacket***)-these two constructions are superficially identical, and more-the same phrasal verb could also be intransitive (e.g., *The flight just* ***took off***). The complexity of understanding the correct use of null object underlying this cross-linguistic variation for L2 learners of Japanese, therefore, hypothesized preceding verb transitivity as an important factor group to be investigated in the current research. Since intransitive verbs introduce more complexity than the transitive verbs to the determination of null object permission, it was hypothesized that intransitive verbs favor the incorrect object drop in the current data. Table 3.3 below summarizes the factors and factor groups proposed to account for the occurrence of inaccurate object pro-drop in the current study:

Table 3.3

Factors and Factor Groups for Accurate/Inaccurate Object Pronoun Use

Factor group	**Factor Levels (Coding)**
Pronoun	The pronoun used, e.g., it, me, that, myself, etc.
Pronoun Types	D=Demonstrative I=Impersonal N=Indefinite P=Personal R=Relative

	F=Reflexive
Preceding Verb	T=Transitive I=Intransitive D=Do B=Be N=Others
Prior	P=Pro-form mentioned somewhere prior to the conversation token (distant) N=Pro forms NOT mentioned anywhere prior to the conversation token
Scenario	F=Full form mentioned somewhere prior to the conversation token (distant) N=Full forms NOT mentioned anywhere prior to the conversation token
Pro-Form	P=Pro-form mentioned in the prior context N=Pro-form NOT mentioned in the prior context
Full Form	F=Full form mentioned in the prior context N=Full form NOT mentioned in the prior context
Interviewer	I=Pro- or full form mentioned by the interviewer N=Pro- or full form NOT mentioned by the candidate
Candidate	C=Pro- or full form mentioned by the candidate N=Pro- or full form NOT mentioned by the candidate
Distance (Pro-Form)	Distance from the most recent pro-form to the pronoun variable
Distance (Full Form)	Distance from the most recent full form to the pronoun variable

3.3.1.4 Linguistic factor groups—Modal auxiliary verbs

For L2 use of modal auxiliary verbs, previous studies on L2 acquisition of English modals have suggested L2 learners intend to overuse and/or underuse certain types of modal auxiliaries (e.g., Aijimer, 2002; Dutra, 1998; Hinkel, 2009; McDouall, 2012; Nakayama, 2021). Specifically, Japanese L2 learners have been reported to have a higher-than-native reliance on the use of can and are unfamiliar with would, could, might, and will in their L2 productions (e.g., Nakayama, 2021;

Nozawa, 2014). This study, therefore, included modals as a factor group and hypothesized that the unfamiliarity with would, could, might, and will may account for occurrences of inaccurate modal verb usage in the current data.

On the other hand, the underlying reason for the underuses of would, could, might, and will might be fundamentally different, as would, could, and might involve speech politeness (Boncea, 2013; Kennedy, 2002; Nozawa, 2010; Palmer, 1990, 2001) while will and would may reflect the effect of salience as in conversations these are frequently abbreviated as ’d or ’ll whose physical and psychological form could be easily ignored, with articulations being of low salience to L2 learners. To further account for the potential effect of politeness and salience, two extended factor groups, politeness and salience were included in the analysis. Their hypothesized effects on inaccurate L2 productions are based on the observed underuses (e.g., Nakayama, 2021; Nozawa, 2014).

As reported, Japanese L2 learners also showed significant preferences of dynamic and deontic modalities over epistemic modalities as compared to that of the native speakers (e.g., Moriya & Horie, 2009; Nakayama, 2021; Pemberton, 2020). Related studies have associated such L2 modality preference with either learners’ over-reliance on one-to-one modal-meaning association (e.g., Hu & Li, 2015; McEnery & Kifle, 2002; Nakayama, 2021) or with the L1 developmental pathways and L2 acquisitional orders in which English dynamic and deontic modality was developed and acquired prior to epistemic modality (e.g., Bybee, Perkins, & Pagliuca, 1994; Gibbs, 1990; Horie, 1997; Moriya & Horie, 2009; Stephany, 1995; Sweeter, 1990). In fact, these two explanations are not exclusive to each other, and many argue

the epistemic modality is an extension of the root modality (i.e., the deontic and the dynamic modality). Consequently, the systematic ambiguity of more than one meaning associated with the same modality form in English maybe a potential result of historical development (Bybee, Perkins, & Pagliuca, 1994; Sweeter, 1990). The current research applied a 3-modality factor group which was originally developed by Coates (1983) and further extended by Palmer (1990, 2001), Biber and colleges (1999): (1) Epistemic modality: possibility (*can*, *could*, *may*, *might*); prediction (*will*, *would*, *shall*); certainty (*should*, *must*); (2) Deontic modality: permission (*can*, *could*, *may*, *might*); necessity (*should*); advice (*should*); obligation (*must*); (3) Dynamic modality: ability (*can*, *could*); volition (*will*, *would*, *shall*). Hypothesized effect of modality preferences on the acquisition of accurate modal verb usage favor dynamic and deontic modalities over epistemic modality based on the findings concluded earlier in this paragraph.

Moreover, factor groups describing morphosyntactic environment was also tested as it enables modal auxiliaries to occur in L1 Japanese modality speech (Moriya & Horie, 2009; Takayama, 2002). For example, in Japanese *beki* (obligation) and *te-mo-ii* (permission) cannot occur in adverbial clauses expressing time, while *ka-mo-shire-nai* (possibility) and *ni-chigai-nai* (certainty) cannot occur in yes-or-no questions, negations, if clauses, and adverbial clauses expressing time (see Table 3.4 for a full list of this linguistic factor). Interestingly, although a few comparative analyses have investigated the cross-linguistic differences of morphosyntactic restrictions on modality use between Japanese and English (e.g., Hirakawa, & Suzuki, 2010; Moriya & Horie, 2009), no study, to the best of our knowledge, has discussed

whether such restrictions have a potential effect on L2 modality variation. This study therefore attempts to investigate whether modality functions (e.g., obligation, permission, possibility, etc.) and morphosyntactic environments (see Table 3.4) either independently or jointly predict the modality usage variation observed in the current data. Hypothesized effects of these factor groups were based on the scope of restrictions (see Table 3.4; the x mark indicates a no-co-occurrence restriction), and the more restricted the morphosyntactic environment, the more likely an error might be observed.

Table 3.4

Restrictions on Japanese Modality Use by Morphosyntactic Environment

			Adverbial Clause				
					Reason		
	Question	Negation	Time	If	*node*	*kara*	Adversative
beki (obligation)	○	○	×	○	○	○	○
te-mo-ii (permission)	○	○	×	○	○	○	○
ka-mo-shire-nai (possibility)	×	×	×	×	○	○	○
ni-chigai-nai (certainty)	×	×	×	×	○	○	○
daroo (prediction)	×	×	×	×	×	○	○
yoo (volition)	×	×	×	×	×	×	×

Note. The table describes the co-occurrence restrictions on Japanese modality use by morphosyntactic environment. Adapted from "What is and is Not Language-specific about the Japanese Modal System? A Comparative and Historical Perspective," by T. Moriya and K. Horie, 2009, *In Japanese Modality,* p. 100.

Subjectivity is another important aspect to be considered as modality and

subjectivity sometimes are co-encoded (e.g., Horie, 1997; Kaufmann & Tamura, 2017; Larm, 2006; 2009; 2014; Masuoka, 2009; Moriya & Horie, 2009). Papagragou (2006) explained that with subjective modality "possible worlds in the conversational background are restricted to what the current speaker knows as of the time of the utterance (p.1965)" while with objective modality, such background "include(s) what is generally known to some community, or what the publicly available evidence is (p.1965)". Subjectivity factor group considers the use of the first, second, and third person as subject, and the first-person factor is believed to account for inaccurate modal auxiliary usage-subjective speculation is a more challenging task for L2 learners in spontaneous speech productions.

Lastly, prior context, especially modal auxiliary involvement in the prior text is tested as a factor group with the hypothesis that preceding modal uses in the conversation, primarily by the interviewer, may prepare the candidates in their following use of modal auxiliaries. Table 3.5 presents all potential factors and factor groups for L2 modal auxiliary variation.

Table 3.5

Factors and Factor Groups for Accurate/Inaccurate Modal Auxiliary Verb Use

Factor group	**Factor Levels (Coding)**
Modal	can, could, have to, might, shall, should, will, would, must, etc.
Politeness	P=Politeness (would, should, might, could) O=Others
Salience	F=Full form/Salient form only B=Full or Short form, both are okay
Modalities	E=Epistemic modality

	D=Deontic modality Y=Dynamic modality
Modal Functions	A=Advice B=Ability E=Prediction C=Certainty N=Necessity O=Obligation P=Possibility V=Volition Q=Permission (Request)
Morphosyntactic Environment	Q=Question (yes-no) N=Negation T=Adverbial clauses expressing time I=Hypothetical situations (if clauses) R=Reason clauses A=Adversative O=Others
Subjectivity	F=First person S=Second person T=Third person
Prior Context	A=Modal Agreement e.g., I: How can I help you? C: Can you repair the camera? M=Other Modal Verbs e.g., I: May I help you? C: Can you repair the camera? Q=Interrogative prior contexts (other than A and M) e.g., I: What is your plan? C: I plan to visit another store. N=Other contexts (other than A, M, and Q)

3.3.2 Nonlinguistic Factor Groups

This section will introduce nonlinguistic interview-related factor groups such as role play topics, gender, and role play properties that were considered important in predicting L2 morphosyntactic variation occurrences observed in the current data. Table 3.6 presents all nonlinguistic factors and factor groups as well as the corresponding factor levels within each factor groups that were considered for the

current analysis.

Table 3.6

Nonlinguistic Factor Groups and Factors

Factor group	**Factor Levels (Coding)**
Role Play Topics	18 different role play topics
Gender (Candidate)	F=Female M=Male
Gender (Interviewer)	F=Female M=Male
Gender (Candidate-Interviewer Agreement)	A=Gender Agreement (e.g., Male-Male or Female=Female) D=Gender Disagreement (e.g., Male-Female or Female=Male)
Role Play Levels	Levels of the role play (from 0 to 5)
Probing Action	P=Probing to a higher role play level D=Downgrade to a lower role play level N=No Probing Action
Probing Success	S=Success F=Failure N=No Probing Action

As described in the previous sections, 18 role play topics were used in the oral proficiency test and topics may be an important indicator for the current study. For example, Hinkel (1995, 2009) reported the usage of modal expressions of obligation and necessity being context-dependent, as L2 learners' usage differs significantly from that of native speakers on topics involving personal experiences such as family roles and relationships. Considering that individuals might be more or less familiar with one role play topic than the other, and certain role play topics might trigger certain variables more frequently than the other, we hypothesized that role play topics

might be systematically more or less challenging to specific L2 speakers, and role play topics were included as a factor group for decision analysis.

Second, gender has been a frequently debated issue that affects L2 learning and L2 learners' performance in language testing (e.g., Brown, 2003; Lumley & O'Sullivan, 2005; Markham, 1988; Motallebzadeh & Nematizadeh, 2011; Norton, 2005; O'Sullivan 2002; Reemann, Alas, & Liiv, 2013; Stumpf & Stanley, 1998). When assessing learners' oral abilities-often carried out by a native speaker in the form of oral interview-gender differences were found between male and female test-takers (e.g., Lumley & O'Sullivan, 2005; Norton, 2005; O'Sullivan 2002; Reemann, Alas, & Liiv, 2013; Stumpf & Stanely, 1998), as well as male and female interviewers (e.g., Porter, 1991; Porter & Shen, 1991; O'Loughlin, 2002; Motallebzadeh & Nematizadeh, 2011). O'Loughlin (2002), for example, suggested interviewer style differences across gender, as female interviewers are more "collaborative, symmetrical, cooperative, and supportive (p.170)" while male interviewers tend to conduct "controlling, asymmetrical, uncooperative, and unsupportive (p.170)" style of interview. Reemann and colleagues (2013) found that compared to male interviewers, female interviewers adhere more to the instructions, present interview elements more frequently, and are more willing to help the test-taker. Research on test-taker gender claimed that females perform better in verbal skills, recalling, and perceptual speed while males generally receive a higher score in delivery spatial description and mechanical reasoning (e.g., Norton, 2005; Stumpf & Stanley, 1998). Interestingly, Buckingham (1997) found an interaction effect between the gender of the interviewer and the gender of the test-taker: Japanese male test-

takers had a better performance when interviewed by a male interviewer and vice versa. It seems that gender effects on language learners' oral performance may depend on the cultural and language background of the learner. For Japanese learners of English, learners are assumed to be more sensitive to the roles played by the interviewer while expressing their thoughts (Davies & Ikeno, 2002; Iwazaki, 2009; O'Sullivan, 2000). To examine the role gender plays in the current oral proficiency interview with Japanese learners, three factor groups, *interviewer gender*, *candidate gender*, and *gender agreement* were examined.

Three additional factor groups were investigated as they demonstrate speaking task difficulty: role play levels (0 to 5), probing action, and probing success. As introduced in the previous Data section, role play levels are changing, as the interviewer probes to a higher level than the already detected level to identify the actual proficiency the candidate can sustainably demonstrate. Based on what Educational Testing Service (1999), MacGregor (2000), and Quinn (2000) has described, the level of role play is coded as a continuous factor to indicate the interviewer's judgment on task difficulty during the interview (level 1, 2, 3, 4, and 5). Following role play levels, two other factor groups, probing action and probing success were coded as supplementary evidence to the role play level factor group. As the interviewer probe to the next proficiency level, the task (i.e., role playing) becomes more challenging, which might or might not trigger inaccurate productions by the L2 speaker. Moreover, whether the probing is successfully retained depending on the candidate's performance is well qualified as an indicator of the candidate's actual level of proficiency. Since proficiency level itself is not available for the

current data set, role play levels, probing action, and probing success are well suited as a replacement of task difficulty and proficiency indication.

3.3.3 Paralinguistic Factor Groups

As discussed in Chapter 2, the paralinguistic factor groups involve speech-related features that serve as indicators of speech fluency and speech processing time. How speakers plan and produce their speech, especially the placement and timing of fluency indicators such as hesitations, pauses, fillers, etc. in the speech contains important information that relates to the speech planning and production processes (Chafe, 1980, Du Bois et al., 1992; Jenkins & Parra, 2003; Ullenius, 2015). Table 3.7 describes these paralinguistic factors and factor groups.

Speech processing time was coded as the total count of words in each utterance for each token. It is assumed that the longer the utterance, the harder it might be for a candidate to monitor and process his/her output, and therefore influence the accurate and inaccurate use of the morphosyntactic features. Speech processing time was coded as a continuous factor.

Speech fluency factor groups include modification, hesitation, repeat, repair, pause, and filler (filled pause) indicators. The placement and in some cases the duration of the fluency factor groups may intricately link to the linguistic variation and therefore these indicators were further divided into general and specific indicators. For example, while a general pause might be found anywhere in the token, a specific pause occurs in a specific place-it could only be found right before or during the occurrence of the variable. Example 3.7 and Example 3.8 below provide detailed examples of general factor groups and specific factor groups respectively.

The only factor group that is not presented in Example 3.7 and Example 3.8 is the modification factor group, a factor group that indicates any kind of fluency issue (hesitation, repeat, pause, repair, or filler) that happened in the token (general), right before or during the variable occurrences (specific).

Example 3.7 Speech Fluency Factor Groups (General)

6 I: good afternoon ma am may I help you?
7 C: yes please (.) umm::: (0.6) I'd like to um:: (1.2) I'd like to clean- I'd like kk (1.5) I like to clean:: my jacket

(a) Hesitation: *clean::* my jacket
(b) Repeat: *I'd like to* um:: (1.2) *I'd like to* clean- *I'd like*
(c) Repair: *I'd like* to clean- *I'd like* kk (1.5) *I like to*
(d) Filler: *Umm:::* (0.6) I'd like to *um::* (1.2) I'd like to clean-
(e) Pauses: Umm::: *(0.6)* I'd like to um:: *(1.2)* I'd like to clean- I'd like kk *(1.5)* I like to clean:: my jacket

Example 3.8 Speech Fluency Factor Groups (Specific)

03 I: ok can I help you?
04 C ok I'm looking for:: (0.5) some apartmen in vancouver
05 I: okay what kind of location would you like
06 C: well:: safe as possible
07 I: safe as possible okay
08 C: and and not too expensive
09 I: not too expensive (.) how much money are you willing to pay?
10 C: um:: what would would be the range for the uh::(0.8) move
11 (0.5) you know nice apartment then?

(a) Hesitation (specific): *ok I'm looking for:: (0.5) some apartmen in vancouver*
(b) Repeat (specific): *um:: what would would be the range*
(c) Repair (specific): *um:: what would would be the range for the uh:: (0.8) move*

(0.5) you know nice apartment then?

(d) Filler (general, specific): *um:: what would would be the range for the uh:: (0.8) move (0.5) you know nice apartment then?*

(e) Pauses (specific): *ok I'm looking for:: (0.5) some apartmen in vancouver*

In token (a), for example, the hesitation mark "::" for the preposition *for* is considered both a general hesitation in the token and a variable (preposition/particle)-specific hesitation. In token (b), the modal verb *would* repeated twice, and it therefore would be coded as both a general repeat and a variable (modal verb)-specific repeat. Token (c) marks a general and a variable (article)-specific repair: *nice apartment* is the repair for *the uh:: move*. Token (d) presents two specific fillers: *uh::* and *you know* before the missing article *a* for the noun phrase *nice apartment*. These two fillers would be counted as general fillers as well. And token (e) shows a pause marking *(0.5)* before the inaccurate use of the missing article *[an]*, a general and a variable (article)-specific pause.

Table 3.7

Paralinguistic Factor Groups and Factors

Factor Groups	**Factor Levels (Coding)**
Hesitation (General)	H=Hesitation O=None
Hesitation (Specific)	H=Hesitation O=None
Repeat (General)	R=Repeat O=None
Repeat (Specific)	R=Repeat O=None
Repair (General)	R=Repair

	O=None
Repair (Specific)	R=Repair O=None
Pause (General)	Timed and untimed break off in the intonation unit. S=Untimed pause (.), a break in the intonation unit or an audible inhalation, usually less than 300ms M=Timed pause less than 1 second (usually more than 300ms) L=Timed pause more than 1 second but less than 4 seconds X=Timed pause more than 4 seconds O=None
Pause (Specific)	Timed and untimed break off in the intonation unit. S=Untimed pause (.), usually less than 150ms, a break in phonation M=Timed pause less than 1 second (usually more than 300ms) L=Timed pause more than 1 second but less than 4 seconds X=Timed pause more than 4 seconds O=None
Filler (General)	Filled pause, including uh, unh, um, hm, huh, uh-huh, etc. F=1 filler V=2-3 fillers W=More than 3 fillers O=None
Filler (Specific)	Filled pause, including uh, unh, um, hm, huh, uh-huh, etc. F=1 filler V=2-3 fillers W=More than 3 fillers O=None
Modification (General)	M=Modification O=None
Modification (Specific)	M=Modification O=None
Speech Processing Time	Total count of words in each utterance for each token

Chapter 4: Analytical Strategies

The primary goal for this chapter is to demonstrate the analytical strategies after the transcription of the spoken data, careful examination of the linguistic variables of interest, and identification of the factor groups is completed. The next mission is to narrow the number of factor groups and find out within the factor groups selected, which factor level is most influential in classifying the variable of interest, i.e., differentiating inaccurate forms from the accurate forms in the observed variation. Classification trees, random forests, and variable rule analysis were applied to examine the relative importance and effect of factor groups on the outcome of interest for the current study. Additionally, implicational scales were designed to pick up the underlying properties which might be left out by statistical modeling.

4.1 Classification Trees and Random Forests

One of the major challenges for statistical modeling to generate an optimal result is having too many independent variables and too complex variable interactions. The same happen to the variable rule analysis-although the current research started in an exploratory way by attempting to examine any potential factor groups and factors on their hypothesized effect on the linguistic variable of interest, an overwhelming number of factor groups, factors, and interactions, when added into the model at the same time, would severely destabilize the model (Poplack & Tagliamonte, 2001; Tagliamonte & Baayen, 2012). In variable rule analysis, there is a general rule of thumb: the number of factor groups is advised to be no more than 6, and the number of tokens per cell to be sufficient for statistical significance is around 30 (Baayen,

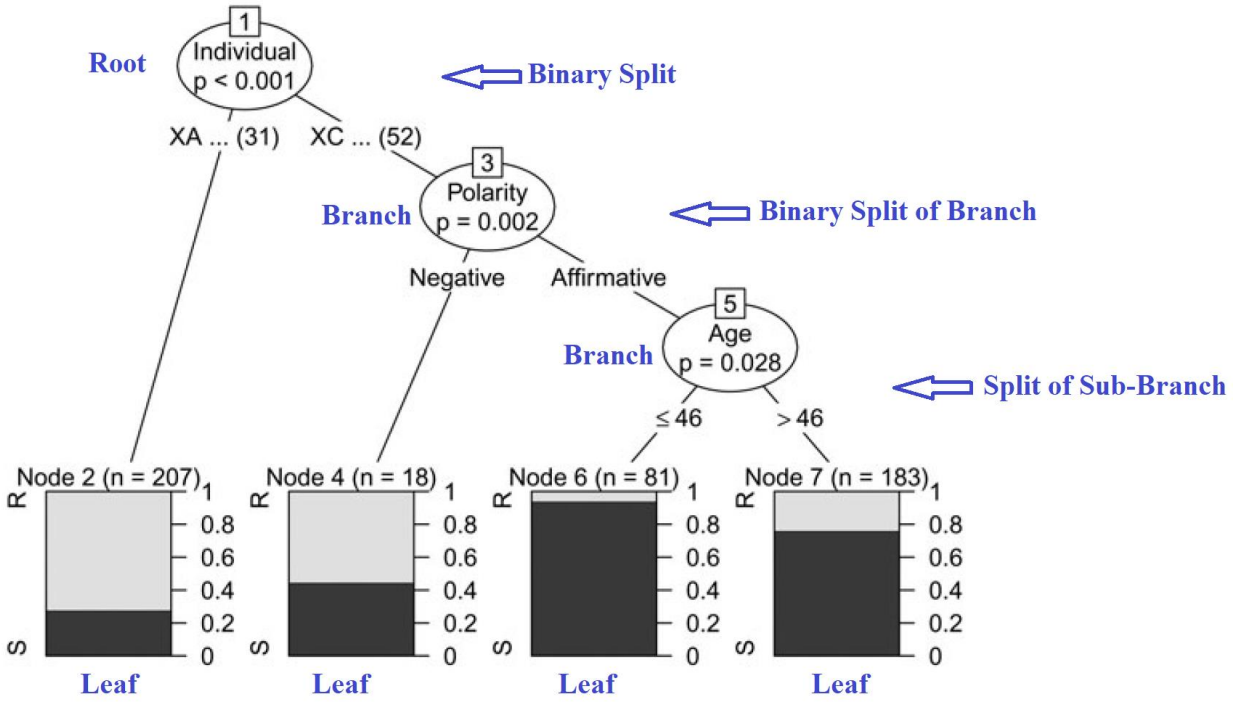
2008; Guy, 1980; also see Tagliamonte, 2011). In addition, variationists also suggested a type-token ratio (i.e., the total number of unique words divided by the total number of words in the token) should be assigned to avoid excessive numbers of one type of tokens by one speaker in the data. It is important to understand that these rules of thumb recommendations should be treated as guidelines, but not laws-the numbers were derived base on a limited number of models by research on different scenarios. For example, a type-token ratio is advisable for sociolinguistic research in which disproportional repeated type of tokens by one speaker may skew the data and compromise the analysis. However, it is justifiable for L2 speech, with the goal in exploring learner usage of the target native form, to have a free type-token ratio and include all tokens of similar or repeated forms (Tagliamonte, 2011).

Conditional inference trees (classification trees; Hothorn, Hornik, & Zeileis, 2006) and conditional inference forests (random forests; Hothorn, Bühlmann, Dudoit, Molinaro, & Van Der Laan, 2006; Strobl, Boulesteix, Zeileis, & Hothorn, 2007; Strobl, Boulesteix, Kneib, Augustin, & Zeileis, 2008) are used to evaluate whether the identified factor groups are useful in predicting the variant choice (i.e., choosing inaccurate forms over accurate forms) and rank those factor groups by importance. Both classification trees and random forests are operated in the open-source statistical programming environment R (R Core Team, 2018) in a package named *party* (Hothorn, Hornik, & Zeileis, 2006; Hothorn, Hornik, Strobl, & Zeileis, 2015). Unlike mix-effects models that can only take a limited number of variables and interactions, classification trees and random forests enjoy a clear advantage in which all factor groups could be included in the analysis at once and obviate

problems such as more variables than observations, predictor collinearity due to factor correlations, linearity constraints, and missing data (see a detailed discussion in Strobl, Malley & Tutz, 2009; Tagliamonte & Baayen, 2012).

Figure 4.1

Sample Classification Tree



Note. The Classification Tree Figure is an illustration of variable importance for a conditional reference tree analysis. Adapted from "Models, Forests, and Trees of York English: Was/were Variation as a Case Study for Statistical Practice," by S.A. Tagliamonte and R. H. Baayen, 2012, *Language Variation and Change*, *24*(2), p.16.

A conditional inference tree estimates the likelihood of the appearance of outcome variants (in this case, accurate and inaccurate morphosyntactic feature use) based on a series of binary classifications. Conditional inference trees (classification trees; Hothorn, Hornik, & Zeileis, 2006) use the *Ctree* function in the party package

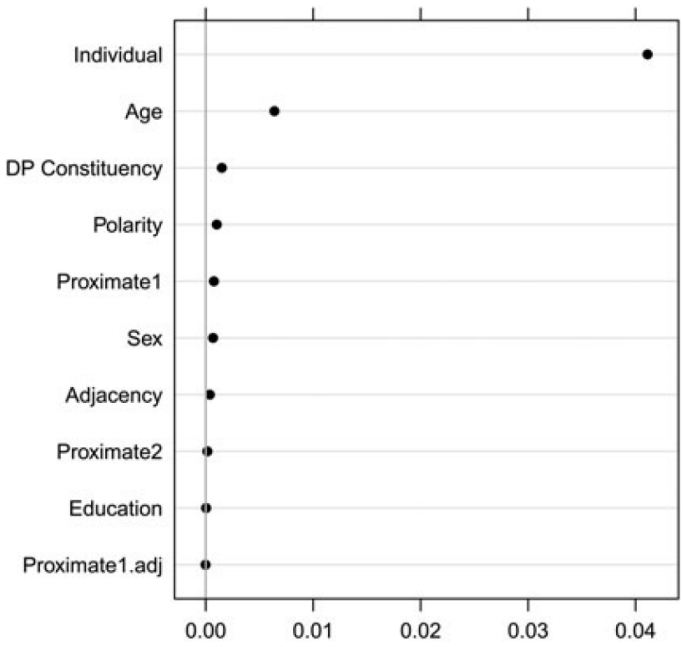
in R to evaluate how useful a factor group is for classifying the dependent variable (i.e., accurate and inaccurate forms) and identify a hierarchy of variable importance. The tree, see Figure 4.1 as an example, finds the most important factor group for the desired classification as the root and makes a binary split into a leaf (i.e., the endpoint of the classification) and a branch (i.e., the second most important variable, and further classification possible). The branch could be further binarily split into sub-branches, a leaf and a sub-branch, or two leaves. If more than one branches were identified, the one with the strongest association with the split is decided, with the p-value recorded (see Figure 4.1). The classification tree highlights the complex interaction between the root and branches and creates a hierarchy of branches whose structure (root-branch-sub-branch-leaf) is an indication of factor group importance relevant to the desired classification. One problem with conditional inference trees is that they do not generalize accurately to new data. To overcome this data-dependent shortcoming, researchers have introduced conditional inference forests.

Conditional inference forests (random forests; Hothorn et al., 2006; Strobl et al., 2007; 2008) creates random forests of a large number of classification trees. The random forests approach starts with randomly resampling the data and generating multiple sample combinations of (subsets of) factor groups and (subsets of) tokens to build multiple classification trees. Every time a classification tree was generated, the tokens (i.e., the dependent variable) left out of the random resampling create a comparison of the extent to which the model becomes worse, and the averaged influence of the tokens left out by each tree is the “out-of-bag accuracy”, a measure of random forest accuracy (Breiman, 2001). The random forest approach uses the

Cforest function in the *party* package in R (R Core Team, 2018), and once the random forest is generated, the *variamp* function in R evaluates the factor group importance in classifying the target morphosyntactic feature. Figure 4.2 below shows a sample variable importance dotplot in which the order of factor group importance is presented in the vertical axis (e.g., Factor Individual is the most important factor group, Age the second, and Proximate.adj the least). The horizontal axis presents the relative variable importance, and factor groups with a value of 0.00 or less are inconsequential.

Figure 4.2

Sample Variable Importance for a Random Forest



Note. The Variable Importance Dotplot is an illustration of variable importance for a conditional reference forest analysis. From "Models, Forests, and Trees of York English: Was/were Variation as a Case Study for Statistical Practice," by S.A. Tagliamonte and R. H. Baayen, 2012, *Language Variation and Change*, *24*(2), p.162.

To sum up, for L2 speech data which is normally unbalanced and with a limited number of tokens but a potentially overwhelming number of factor groups, the classification trees and random forests overcomes the limitation of mixed-effects models by allowing all factor groups to be examined at the same time. The classification trees and random forests, therefore, could be used as an exploratory tool in finding the probable factor groups based on their relative importance in differentiating outcomes of interest. Note that researchers have emphasized that classification trees and random forests should not be substituted for mixed-effects models, but as complement and guide the selection of factor groups (Tagliamonte & Baayen, 2012).

Once potential important factor groups were identified, the study then moved to a detailed analysis, i.e., a mixed-effects variable rule analysis, in finding how the factor groups work together and exploring the relative possibility of specific factor levels within a significant factor group on their co-occurrences with the morphosyntactic variation for the study.

4.2 Variable Rule Analysis

Variable rule analysis in the current days is different from the ones used back in the 1990s. As reviewed in Chapter 2, the analytical tool has experienced major technical changes from VARBRUL to *GoldVarb X* (e.g., (Cedergren & Sankoff, 1974; Pintzuk, 1988; Rand & Sankoff, 1991; Robinson, Lawrence, & Tagliamonte, 2001; Sankoff, Tagliamonte, & Smith, 2005) and to *Rbrul* (Johnson, 2009; 2017) to become more accurate in estimating subtle differences among internal and external effects. Over the

past 30 years, variable rule analysis has become a variationist field term of mixed-effect probabilistic modeling, a specifically designed set of computer programs in investigating multidimensional data obtained in linguistic variation studies (e.g., Johnson, 2009; Li, 2014; Regan, 2013; Sankoff, 1978; Tagliamonte, 2011; Williams, 2009; Young, 1996; 1998; Young & Bayley, 1996). The current study uses the Rbrul package (Johnson, 2009; 2017) which operates in the open-source programming environment R (R Core Team, 2018).

The mathematical underpinnings of variable rule method focused on the extent to which specific factor levels within each influential factor group increases the overall probability (p) of the occurrence of the target variant of the variable, according to the following formula (Young & Bayley, 1996, p.280; Bayley, 2002, p.126):

$$p = \frac{p_0 \times p_1 \times p_2 \times \cdots \times p_n}{[p_0 \times p_1 \times p_2 \times \cdots \times p_n] + [(1 - p_0) \times (1 - p_1) \times (1 - p_2) \cdots \times (1 - p_n)]}$$

In this formula, p_0 is the input probability irrespective of the presence and absence of any factor groups. The effect of each factor category within any influential factor groups included in the model are called the factor weights ($p_1,\ p_2, \cdots, p_n$). A value of 1.0 for a factor weight p_n indicates that if factor *n* is present, the dependent variable of interest (i.e., the inaccurate form) is predicted to co-occur with 100% certainty. Alternatively, if $p_n = 0$, the accurate form is predicted to co-occur with the factor *n* for sure. A factor weight of 0.5 for p_n indicates no effect on the dependent variable of interest by the factor *n*. Thus, a p_n larger than 0.5 but less than 1.0 favors the co-occurrence of the desired variant (i.e., the inaccurate form), and a p_n less than 0.5 but larger than 0 favors the co-occurrence of the alternative form (i.e., the accurate form)

as p_n decreases. To test the model significance, likelihood ratio tests (see the formula below) are used to compare models with and without factor groups of interest, and the more parsimonious model at the last stage of comparison becomes the final model:

$$\chi^2 \approx -2 \times (\log Likelihood_{Model1} - \log Likelihood_{Model2})$$

Additionally, the variable rule approach also integrates fixed effects (i.e., factor groups) and potential random effects (e.g., individuals, role play topics) in explaining the variation observed. The inclusion of random effect considers the current data in which multiple tokens are provided by one speaker and/or by one role play topic, hence different individuals may vary in their production of the morphosyntactic features, and so do different role play topics. However, whether individuals and role play topics vary across other fixed effects, i.e., should be considered as random intercepts, fixed effects, or random effects (slopes) need to be further evaluated through model fit comparisons.

4.3 Implicational Scales

Lastly, implicational scales are generated as supplementary discussion to the underlying properties in the observed variation that researchers are likely to ignore. Rickford (2002) adapted the definition of implicational scales in linguistic context from Gorden (1977) and Pavone (1980) as "depict(ing) hierarchical co-occurrence patterns in the acquisition or use of linguistic variables by individuals or groups (p.143)". As reviewed in Chapter 2, although with notable shortcomings, studies in second language acquisition have still favored the use of implicational scales to demonstrate interlanguage development and seek explanations for L2 processing (e.g., Bayley, 1999; Glahn et al., 2001; Nagy, Moisset, & Sankoff, 1996; Pienemann

& Mackey, 1993; Suherman, Indrayani, & Krisnawati, 2020; Trofimovich, Gatbonton, & Segalowitz, 2007).

The use of implicational scale has experienced radical style-shifting since its early appearance in the 1970s as the implicational power is closely related to scalability and discriminability (e.g., Pavone, 1980; Rickford, 2002; Dunn-Rankin, Knezek, Wallace, & Zhang, 2014). A recommended measure of scalability (IR, Index of Reproducibility) is 93% which approximates the statistical significance level at $\alpha = 0.05$ (Rickford, 2002; Dunn-Rankin et al., 2014). Furthermore, instead of the use of simple binary (+, -) and trinary scales (+, -, /) which may enjoy a better scale-data-fit, variationists have gradually shifted to more discriminatory scales such as frequency-valued or accuracy-valued scales (e.g., Anderson, 1978; Rickford, 2002). Cells with higher accuracies are moved to the upper left while cells with lower accuracies, the lower right. A perfect scalable implicational pattern (i.e., both the ranks and the accuracies are a perfect match in order across variables of interest) might be rare, and as reviewed in Chapter 2, might be criticized of over-regularity for potential implication patterns. An alternative consideration the present-day implicational analysis paradigm is in favor of are solid step-like line patterns (e.g., from upper right to lower left which separates cells with accuracy as a cutoff threshold). The step-like pattern, on one hand, have retained the systematicity of developmental indications by cutoff thresholds, on the other hand, it has also taken the fluctuating nature of development sequences into consideration, and hence would serve as more reliable evidence for the hypothesized acquisitional pattern and processability hierarchy (Glahn et al., 2001; Pienemann, 1998). Nevertheless, if alternative patterning was

found, further exploration and explanation are in need.

Chapter 5: Results

This Chapter presents analyses with four morphosyntactic features (i.e., preposition/particle, article, object pro-drop, and modal auxiliary verb) identified in Chapter 3 by using analytical strategies introduced in Chapter 4 (i.e., Classification tree, random forest, variable rule analysis, and implicational scales). A full list of factors and factor groups and their corresponding variable names entered into the statistical model could be found in Appendix S5.1. Situations of extreme invariance (i.e., certain factor level has 0% or 100% response proportions) which often give rise to overfitting problems are also discussed with potential solutions in this Chapter.

5.1 Preposition/Particle

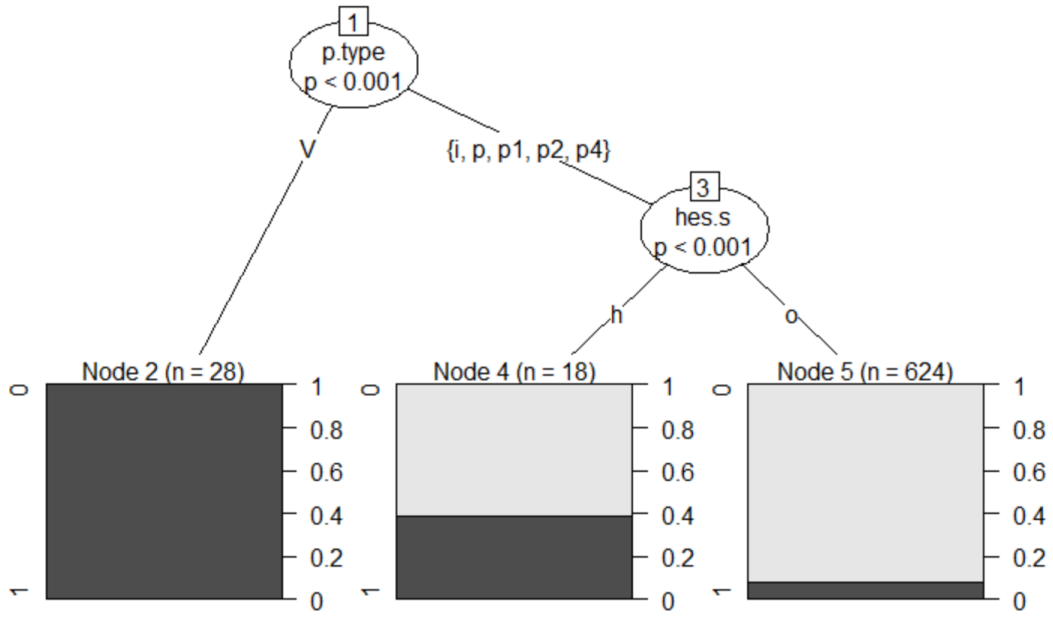
5.1.1 Conditional Inference Tree (Classification Tree)

A total of 27 factor groups were considered in the conditional inference tree analysis for accurate and inaccurate preposition/particle variation observed in the L2 speech data. Figure 5.1 presents the classification tree for the preposition/particle data: Factor group *phrase type* (e.g., phrasal verb, infinitive phrase, prepositional phrase with and without direct object, etc.) was recognized by the classification tree method as the most important factor group in classifying accurate and inaccurate use of preposition/particle. *Hesitation (specific)* was identified as the secondary sub-branch: after the binary split by the factor group *phrase type*, the second most important factor group in classifying the accurate and inaccurate use of preposition/particle is individuals' hesitation specifically happening right before or during the occurrence of the preposition/particle use in the speech (see also Figure 5.1). Generally, adding

more factor groups will lead to more splits (Tagliamonte, 2011), but for the current classification tree, no further splits were generated by the model.

Figure 5.1

Classification Tree: Inaccurate vs. Accurate Preposition/Particle Usage (Application Value = Inaccurate)

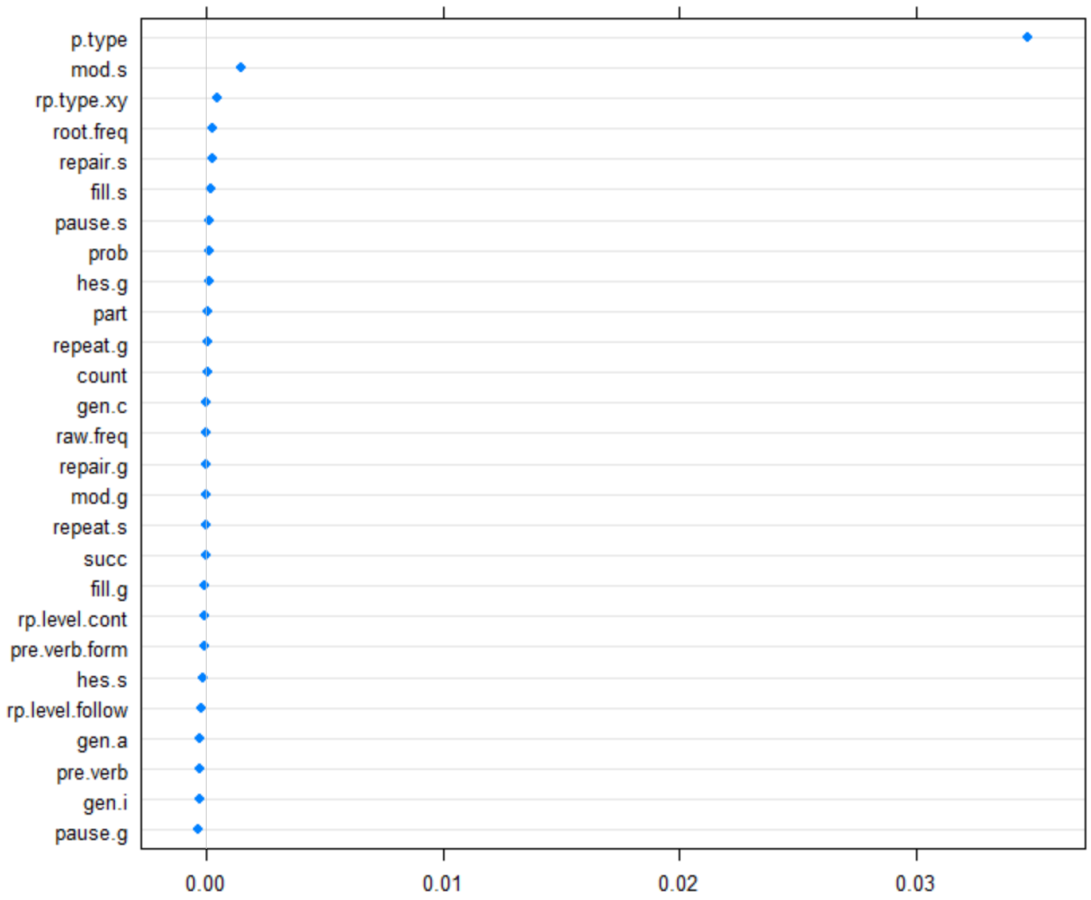


5.1.2 Conditional Inference Forest (Random Forest)

Since classification tree, random forest, and mixed-effects modeling are considered complement to each other in determining important factor groups in classifying the use of variants (Tagliamonte & Bayaan, 2012), the analysis moved to the random forest approach to extract the relative variable importance among the existing factor groups. Figure 5.2 displays the variable importance plot generated by the *varimp* function with the *cforest* analysis in R (R Core Team, 2018).

Figure 5.2

Variable Importance Plot: Inaccurate vs. Accurate Preposition/Particle Usage (Application Value = Inaccurate)

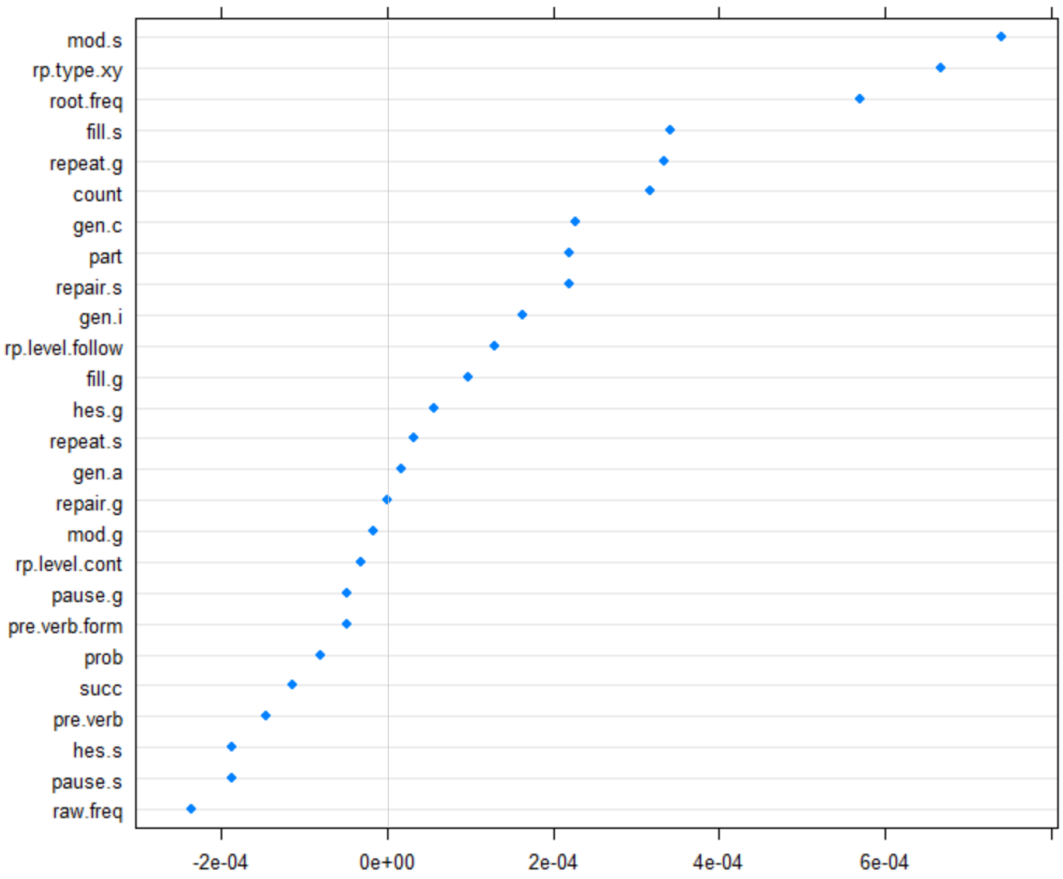


Similar to the results from the conditional inference tree (classification tree) analysis, the variable importance dot-plot shows that *phrase type* (p.type) is the most frequently identified top position across the subsampling of classification trees in the random forest. Second to *phrase type* is the speech fluency indicator *modification (specific)* which involves any specific hesitation(s), repeat(s), repair(s), pause(s), or filler(s) that happened right before or within the specific position where the

preposition/particle use occurs. The least important factor group identified by the random forest method is *pause (general)*. In addition, since *phrase type* in Figure 5.2 has an overwhelming advantage over other factors and factor groups, an adjusted variable importance dot-plot (see Figure 5.3) is generated to obtain a clearer understanding of the hierarchy of variable importance among the remaining factor groups. Interestingly, for the adjusted variable importance plot, the least relevant factor group is no longer *pause (general)*, but *raw frequency* (main verb).

Figure 5.3

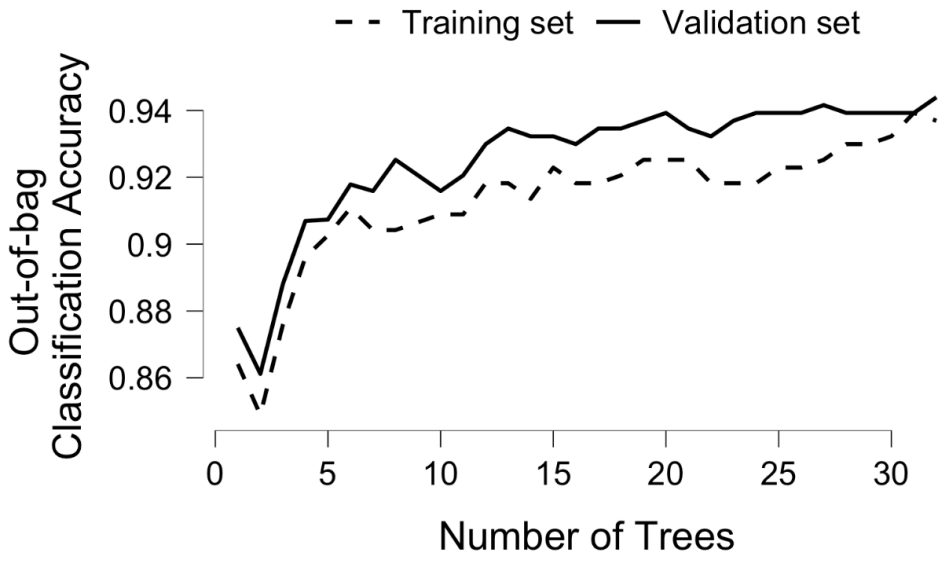
Variable Importance Plot (Adjusted): Inaccurate vs. Accurate Preposition/Particle Usage (Application Value = Inaccurate)



The out-of-bag (OOB) classification accuracy assessing the accuracy for the random forest is 91.79% for the random forest reported in Figure 5.2 (see Figure 5.4 for OOB by the number of trees using JASP; Han & Dawson, 2020), indicating that the tokens left out were at least 91.79% classifiable, a reasonable classification accuracy produced by the resampled classification trees.

Figure 5.4

Out-of-Bag Classification Accuracy by Number of Trees (Preposition/Particle)



5.1.3 Mixed-effects Logistic Regression Modeling

After all factor groups were screened by the classification tree and the random forest on their relative variable importance, tests of their systematic association with the desired outcome response (accurate/inaccurate) were carried out by the mixed-effect variable rule analysis. The variable rule approach detected a potential overfitting issue

for the *phrase type* factor group-the factor level "v" (i.e., verb only) within the factor group was coded to describe the erroneous addition situation where only verbs but no preposition/particles are needed (e.g., some waiter drop ***out*** the wine to my jacket). As a result, the factor group *phrase type* at factor level "v" co-occurs with only the inaccurate outcome invariantly-this cell is structurally invariant. By comparing the model with and without factor level "v" the pseudo-R-squared largely improved from 78.2% to 90.8% but the AIC (Akaike information criterion) increased from 352.372 to 354.372-an indication that the former (i.e., the model without factor level "v") has actually a better data-model fit. To address this potential overfitting issue, one could choose to either exclude the factor group *phrase type* or to exclude the factor level "v" from the *phrase type* factor group. The analysis then examined the AIC statistics for both models ($AIC_{\text{no p.type}} = 478.332$, $AIC_{\text{p.type without "v"}} = 352.372$), with the decision that phrase type "v" would be excluded from the analysis (i.e., the knockouts; Johnson, 2010). Table 5.1 presents the final model identified by the variable rule approach in which both step-up and step-down analysis confirmed a match in optimizing the data-model fit. The overall likelihood that an inaccurate preposition/particle use would occur, i.e., the input probability, is 12.4%.

Linguistic factor group *phrase type*, and paralinguistic factor groups *modification (specific)* and *repair (specific)* were identified as significant indicators influencing the desired dependent variable classification into inaccurate or accurate L2 preposition/particle production. No random intercept or random slope was found to further improve data-model fit.

Phrase Type. The results for *phrase type* indicate that prepositional phrases

without a main verb disfavored Japanese L2 speakers' accurate use of prepositions (p=0.603). Prepositional phrases without a direct object following the main verb (i.e., the "Verb +Preposition" structure) also disfavored the accurate use of prepositions (p=0.593). Phrasal verb, with a factor weight of 0.501, did not favor or disfavor the accurate/inaccurate particle use. Contrarily, infinitive phrases strongly favored Japanese L2 speakers' accurate use of the "to" particle (p=0.261). In addition, L2 speakers inaccurately used 10.8% of the prepositions in prepositional phrases without a main verb, and 9.9% of the prepositions in prepositional phrases that have no direct object after the main verb, compared to a 2.8% error rate when they misused the particle with an infinitive phrase.

Modification (specific). The results for *modification (specific)* supported that specific modification with preposition/particle use disfavored Japanese L2 speakers' accurate use of preposition/particle (p=0.668). Preposition/particle use with any type of specific modification (e.g., hesitation, repair, repeat, pause, and filled pause) were more likely to make a preposition/particle error (p=0.668) while tokens with no specific modification were less likely to do so (p=0.332).

Repair (specific). The results for specific repair particularly happened right before or within the occurrence of the preposition/particle usage show an opposite pattern than the factor group *modification (specific)*: inaccurate preposition/particle use did not co-occur with specific repair, with a factor weight less than 0.001, and the corresponding error rate is 0%. Interestingly, this 0% cell is unlikely to be structurally invariant but rather a result of a small token number (n=33) with extremely low probability (p<0.001). Still, with the 0% proportion response in mind, the analysis

examined whether it could give rise to potential overfitting issues. The model with the factor group *repair (specific)* is nested within the model without *repair (specific)*, therefore likelihood ratio test was performed to examine data-model fit: Loglikelihood $_{\text{with Repair (specific)}}$ = -169.186, Loglikelihood $_{\text{without Repair (specific)}}$ = -175.603- the model with *repair (specific)* is significantly better. The AIC statistics also support the decision to retain *repair (specific)*: AIC $_{\text{with Repair (specific)}}$ = 352.372, AIC $_{\text{without Repair (specific)}}$ = 363.206.

Table 5.1

Preposition/Particle Usage by Japanese L2 Learners of English: Inaccurate vs. Accurate Preposition/Particle Usage (Application Value = Inaccurate)

Factor Group	Log Odds	N	Proportion	Factor Weight
Phrase Type				
Prep Phrase #4 (no main verb)	0.418	187	0.108	0.603
Prep Phrase #1 (no direct object)	0.378	121	0.099	0.593
Prep Phrase #2 (with direct object)	0.242	67	0.090	0.560
Phrasal Verb	0.005	25	0.080	0.501
Infinitive	-1.043	142	0.028	0.261
Modification (specific)				
Modification	0.701	162	0.160	0.668
No Modification	-0.701	480	0.060	0.332
Repair (specific)				
No Repair	8.075	609	0.090	>.999
Repair	-8.075	33	0.000	<.001

Note. N: total token detected. Loglikelihood = -169.186. AIC = 352.372. R^2 = 78.2%. All factor groups significant at p<.05

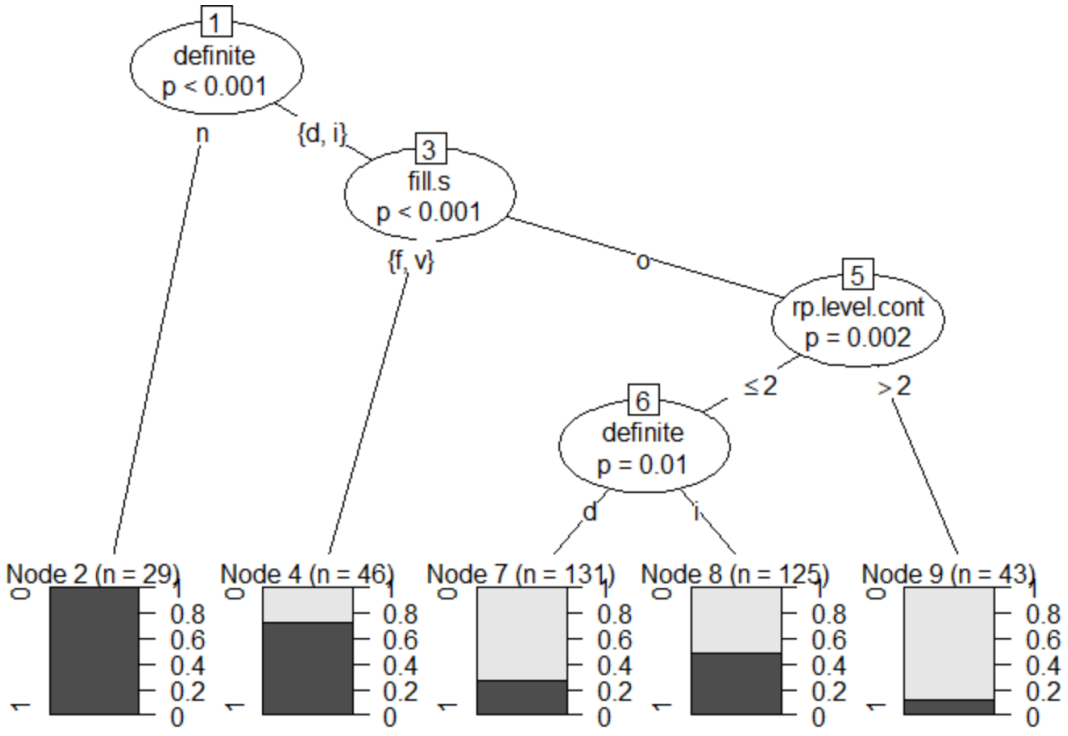
5.2 Article

5.2.1 Conditional Inference Tree (Classification Tree)

Similarly, the analysis starts with the conditional inference tree approach to identify a hierarchy of binary splits. *Definiteness* (definite article, indefinite article, zero article) was recognized as the root, followed by *filled pause (specific), role play level* (continuous, 1-4), and again *definiteness (definite, indefinite).*

Figure 5.5

Classification Tree: Inaccurate vs. Accurate Article Usage (Application Value = Inaccurate)



Note that again, for the case of erroneous addition, the factor group *definiteness* was coded as "n" for zero articles (e.g., please give me ***the*** ten dollars of them all) which co-occurs invariantly with only the inaccurate outcome

response. Indeed, an addition error has no accurate/inaccurate variation-another structurally invariant context similar to the "v" factor level in the preposition/particle analysis in 5.1.3. The classification tree could still use the same factor group *definiteness* again as the sub-branch (see Figure 5.5) when further splits find definite and indefinite subgroups favor the accurate and inaccurate use differently (see also Figure 5.5).

5.2.2 Conditional Inference Forest (Random Forest)

The conditional inference forest analysis also finds *definiteness*, *filled pause (specific)*, and *role play level* among the top important predictors of accurate and inaccurate article use. In addition, the forest recognizes *modification (specific)* a more important predictor than *filled (specific)* (see Figure 5.6). Similar to preposition/particle use, the variable importance analysis for article use in L2 speech also identified *pause (general)* and *hesitation (specific)* among the least important independent variables (cf. Figure 5.2, Figure 5.3).

The out-of-bag (OOB) classification accuracy assessing the accuracy for the random forest is 69.15% (see Figure 5.7 for OOB by the number of trees using JASP; Han & Dawson, 2020), indicating that the tokens left out were at least 69% classifiable, a reasonable classification accuracy produced by the resampled classification trees.

Figure 5.6

Variable Importance Plot: Inaccurate vs. Accurate Article Usage (Application Value = Inaccurate)

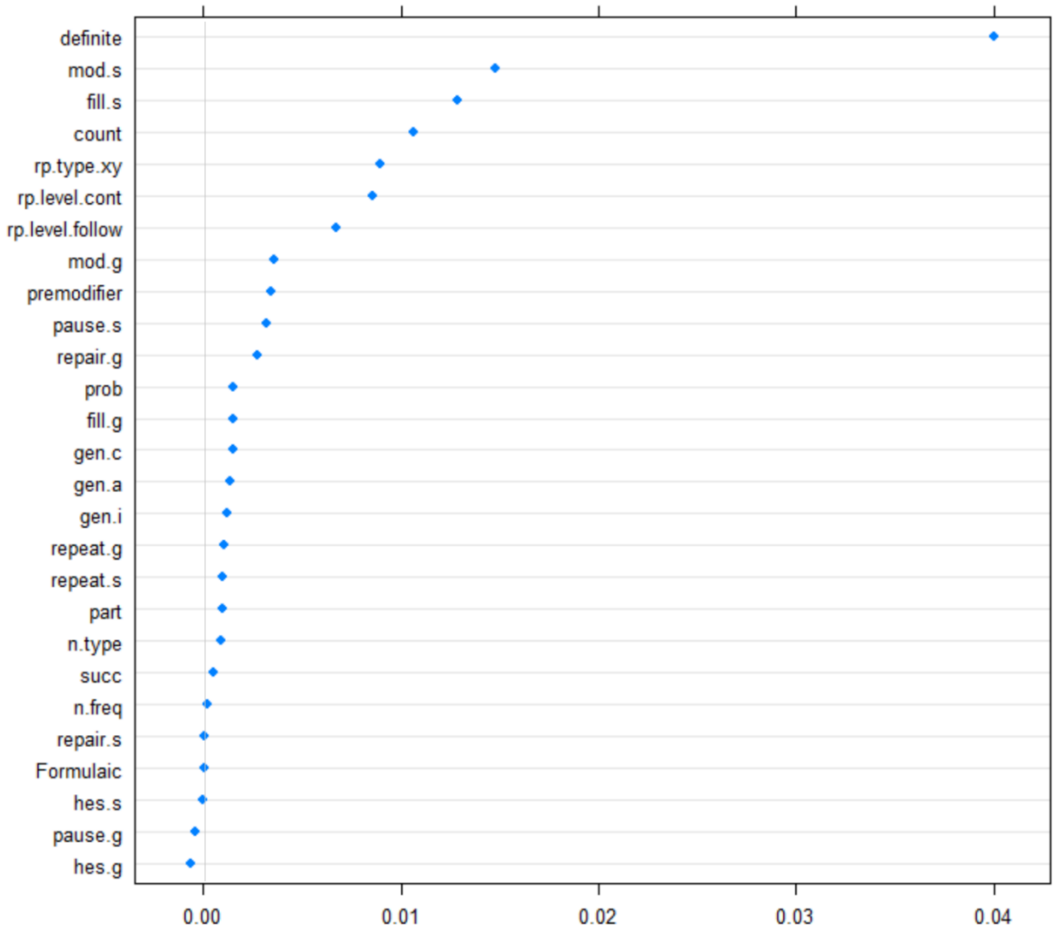


Figure 5.7

Out-of-Bag Classification Accuracy by Number of Trees (Article)

5.2.3 Mixed-effects Logistic Regression Modeling

As mentioned in the previous section, while taking the factor group *definiteness* into consideration, the erroneous addition situation (i.e., factor level "n" for zero articles) co-occurs with an invariant outcome response, i.e., inaccurate article uses. This again introduced a potential issue of overfitting: the model including data from factor level "i" (indefinite), "d" (definite), and "n" (zero articles) has a pseudo-R-squared statistics of 90.6% with an AIC of 399.289, but the model using only the "i" and the "d" factor level data has a significantly lower pseudo-R-squared of 40.5% with a lower AIC of 395.289-this model has a better data-model fit. To address the overfitting issue, AIC statistics were considered for either dropping the *definiteness* factor group completely or excluding the "n" factor level: $AIC_{\text{no definite}} = 455.28$, $AIC_{\text{definite without "n"}} = 395.289$. Smaller AIC indicates better data-model fit, therefore excluding the "n" factor level was favored by the fit statistics. This conclusion is consistent with what is reported in the previous conditional inference tree analysis (see 5.2.1. in this Chapter)-the factor group *definiteness* was identified as a sub-branch classifier again after the first split between "i", "d" and "n" (see Figure 5.5), indicating the binary definite/indefinite classification is still highly relevant to the accurate and inaccurate article usage classification. This study hence decided to keep the *definiteness* variable and exclude the "n" factor level for the variable rule analysis.

Table 5.2 displays the final model identified by the variable rule approach in which both step-up and step-down analysis confirmed a match in optimizing the data-model fit. The overall likelihood that an inaccurate article use would occur, i.e., the

input probability, is 28.1%. Linguistic factor groups *definiteness* and *premodifier*, paralinguistic indicators *modification (specific), repeat (specific), speech processing time,* as well as nonlinguistic predictors *role play level,* were detected as significant factors and factor groups influencing the dependent variable classification into inaccurate or accurate article usage in Japanese speakers' English role play performance.

Table 5.2

Article Usage by Japanese L2 Learners of English: Inaccurate vs. Accurate Article Usage (Application Value = Inaccurate)

Factor Group	Log Odds	N	Proportion	Factor Weight
Definiteness				
Indefinite	0.432	167	0.497	0.606
Definite	-0.432	178	0.281	0.394
Premodifier	-1.008 (log odds, Premodifier +1)			
Role Play Level	-0.504 (log odds, Level +1)			
Speech Processing Time	-0.0299 (log odds, Count +1)			
Premodifier*Modification (Specific)				
Modification	0.794 (log odds, Premodifier +1)			
No Modification	-0.794 (log odds, Premodifier +1)			
Modification (specific)				
Modification	0.475	98	0.571	0.617
No Modification	-0.475	247	0.312	0.383
Repeat (specific)				
Repeat	1.802	334	0.395	0.858
No Repeat	-1.802	11	0.091	0.142

Note. N: total token detected. Loglikelihood = -186.644. AIC = 395.289. R^2 = 40.5%. All factor groups significant at $p<.05$

Definiteness. For the current study, the variable rule analysis found indefinite

articles disfavored Japanese L2 speakers' accurate use of articles (p=0.606), while definite articles favored learners' accurate use of articles (p=0.394). L2 speakers inaccurately used 49.7% of the indefinite article and 28.1% of the definite article in the current data.

Modification (specific). Consistent with the preposition/particle analysis, the results for *modification (specific)* also support that specific modification with article use disfavored Japanese L2 speakers' accurate use of articles (p=0.617). Tokens without any specific modification (e.g., hesitation, repair, repeat, pause, and filled pause) were more likely to also have accurate article use (p=0.383).

Repeat (specific). The result for repeat particularly happened right before or within the occurrence of the article usage shows a similar pattern with *modification (specific)*: specific repeat co-occurred with inaccurate article use in the speech data, with a factor weight of 0.858, making *repeat (specific)* highly predictive of an article error. Approximately 39.5% of the tokens with a specific repeat have an inaccurate article use problem, while only 9.1% of the tokens without any specific repeat were found to associate with article usage problems (see Table 5.2).

Premodifier, and Modification (specific). The continuous factor *premodifier* (i.e., number of premodifiers) has an interactive effect with *modification (specific)* on accurate/inaccurate article use. In addition to the favoring effect specific modification has on inaccurate article use, as the number of premodifiers goes up by 1, the log odds of inaccurate article use go up by 0.794. Specifically, for head nouns without any premodifier, the likelihood of an inaccurate article use with specific modification is 0.617, and with one premodifier, the likelihood increased to 0.780, and with two

premodifiers, the likelihood would be 0.887. On the other hand, for tokens with no specific modification, the log odds of inaccurate use would decrease by 0.794 when premodifiers increases, and the corresponding likelihood are 0.322 (no premodifier), 0.177 (one premodifier), and 0.089 (two premodifiers).

Role Play Level (0-5). The continuously coded factor *role play level* also has a significant association with accurate/inaccurate article use variation. The reported logit coefficient for role play level is -0.504: as role play level goes up by 1, the log odds of an inaccurate article use go down by 0.504. Japanese L2 speakers are less likely to make an article use error as the role play level increases.

Speech Processing Time (Word Count per Token). The favoring and disfavoring effect of speech processing time factor *count* (i.e., word count per token) needs to be interpreted together with the random effect of individuals. Table 5.3 reports the random intercept and random slope (count) by candidate ID. Individuals vary systematically in their likelihood of inaccurately using an article in their L2 speech, ranging from 24.6% to 77.2%. The effect of *count* also differs across individuals, with a slope range from -0.0319 to 0.0290. Specifically, Candidate #cha has a negative random slope of -0.0319: as word count per token increases by 1, the log odds of an inaccurate article use will decrease by 0.0319, indicating the individual is less likely to make an article use error as the word count increases. Candidate #cfb, on the other hand, has a positive random slope of 0.0290, suggesting that the individual is more likely to make an error in article use as they produce longer tokens in their speech.

Table 5.3

Random Intercept and Random Slope (L2 Speaker): Inaccurate (Omission) vs. Accurate Article Usage (Application Value = Inaccurate)

Candidate#	Intercept	Slope (Count)	Tokens	Proportion	Factor Weight
cha	1.226	-0.0319	10	0.600	0.772
bgb	1.217	-0.0317	7	1.000	0.770
jhb	1.064	-0.0277	9	0.667	0.742
aha	0.904	-0.0236	10	0.800	0.710
ceb	0.861	-0.0224	5	0.800	0.702
bjb	0.842	-0.0219	27	0.481	0.698
cbb	0.730	-0.0190	6	0.667	0.673
abb	0.658	-0.0171	8	0.750	0.657
bbb	0.597	-0.0155	7	0.429	0.644
jha	0.532	-0.0139	9	0.667	0.629
acb	0.499	-0.0130	1	1.000	0.621
cca	0.351	-0.0092	4	0.750	0.585
cja	0.312	-0.0082	2	0.500	0.576
bea	0.302	-0.0079	8	0.625	0.573
dda	0.276	-0.0072	11	0.545	0.567
agb	0.254	-0.0066	8	0.375	0.562
aba	0.178	-0.0046	9	0.444	0.543
bha	0.126	-0.0033	5	0.400	0.530
jea	0.054	-0.0014	4	0.250	0.512
bcb	0.039	-0.0010	6	0.500	0.508
cgb	0.014	-0.0004	3	0.333	0.502
aeb	0.002	-0.0001	3	0.667	0.499
bga	-0.014	0.0004	4	0.500	0.495
aca	-0.042	0.0011	15	0.267	0.488
afb	-0.066	0.0017	10	0.300	0.482
bia	-0.156	0.0041	4	0.250	0.460
aib	-0.199	0.0052	4	0.500	0.449
cea	-0.199	0.0052	10	0.200	0.449
ahb	-0.222	0.0058	3	0.000	0.443
aea	-0.288	0.0075	9	0.556	0.427
bba	-0.311	0.0081	14	0.357	0.421
jeb	-0.353	0.0092	3	0.333	0.411
deb	-0.370	0.0096	5	0.400	0.407
bfa	-0.404	0.0105	2	0.000	0.399

cfa	-0.455	0.0119	2	0.000	0.387
cab	-0.490	0.0128	4	0.250	0.378
afa	-0.553	0.0144	5	0.000	0.364
bhb	-0.564	0.0147	6	0.333	0.361
caa	-0.642	0.0167	2	0.000	0.343
eab	-0.682	0.0178	10	0.100	0.334
ccb	-0.708	0.0184	1	0.000	0.329
bca	-0.932	0.0243	4	0.000	0.281
bfb	-0.988	0.0257	10	0.200	0.270
aia	-1.007	0.0262	49	0.082	0.266
cfb	-1.114	0.0290	7	0.143	0.246
Overall	1.020	0.0270	345	0.386	N/A

Note. Loglikelihood = -186.6. AIC = 395.289.

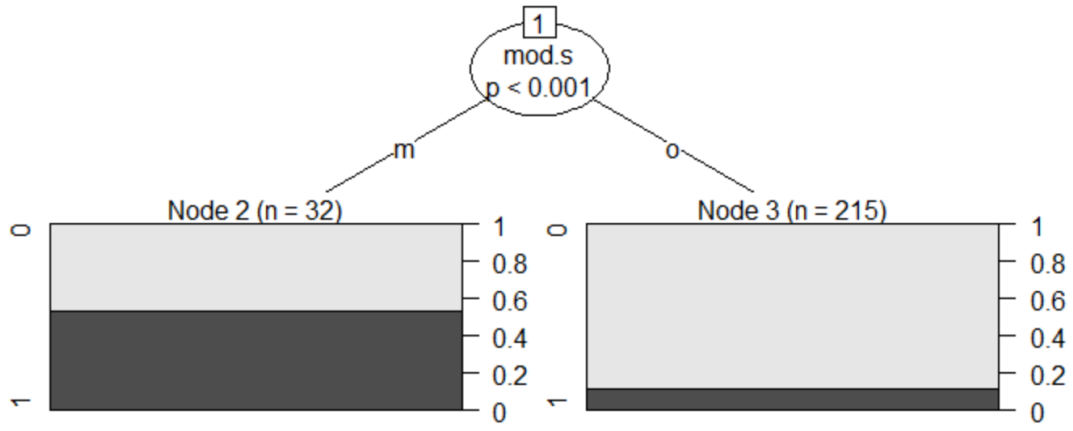
5.3 Object Pro-drop

5.3.1 Conditional Inference Tree (Classification Tree)

The conditional inference tree considered a total of 37 factors and factor groups, among which only *modification (specific)* was recognized as the root which then split into two leaves (m, o). No further splits were identified by the classification tree (see Figure 5.8).

Figure 5.8

Classification Tree: Accurate Object Pronoun Use vs. Inaccurate Object Pro-drop (Application Value = Inaccurate)



5.3.2 Conditional Inference Forest (Random Forest)

The conditional inference forest analysis recognized *modification (specific)* as the top important predictor of inaccurate object pro-drop in L2 speech. The random forest also provides a list of potential important factor groups such as *pause (specific)*, *pronoun*, *pronoun type*, *hesitation (general)*, *distance (to the most recent full form)*, and *distance (to the pro form mentioned in the last turn of conversation)*, etc. (see Figure 5.9). Additionally, unlike preposition/particle and article use, the variable importance analysis for object pro-drop identifies *role play level*, *repeat (general)*, *conversation turns from the most recent full form*, *speech processing time*, etc., as the least important predictors (cf. Figure 5.2, Figure 5.3, Figure 5.6).

The out-of-bag (OOB) classification accuracy assessing the accuracy for the random forest is 82.59% (see Figure 5.10 for OOB by the number of trees using JASP; Han & Dawson, 2020), indicating that the tokens left out were at least 82% classifiable, a reasonable classification accuracy produced by the resampled classification trees.

Figure 5.9

Variable Importance Plot: Inaccurate (Pro-drop) vs. Accurate Object Pronoun Usage (Application Value = Inaccurate)

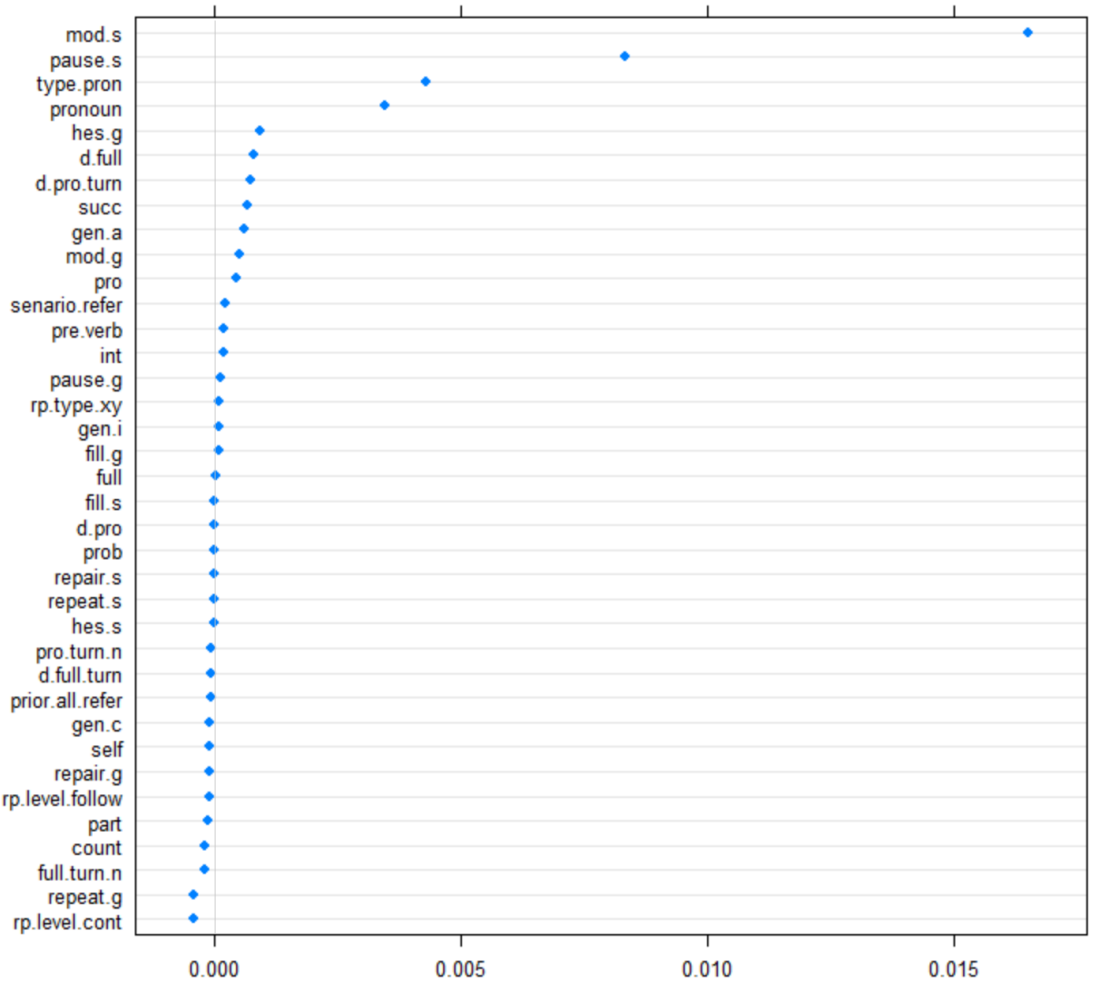
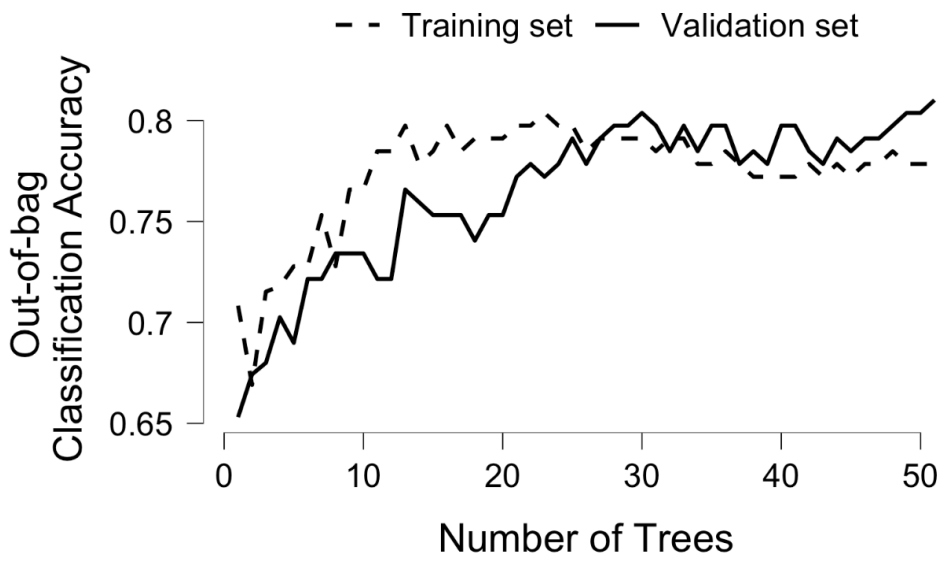


Figure 5.10

Out-of-Bag Classification Accuracy by Number of Trees (Object Pro-drop)



5.3.3 Mixed-effects Logistic Regression Modeling

Using variable rule analysis, linguistic factors such as *pronoun type* and *distance (to the pro-form mentioned in the last turn of conversation)*, as well as paralinguistic indicator *modification (specific)* and *hesitation (general)* were recognized as factor groups systematically associated with Japanese L2 English learners' object pro-drop in their speech. Table 5.4 presents the final model identified by the variable rule approach in which both step-up and step-down analysis confirmed a match in optimizing the data-model fit. The input probability is 3.45%.

Pronoun type. The factor group *pronoun type* contains an invariant context: factor level "f" representing reflexive pronouns has a response proportion=0% (i.e., the cell has zero object pro-drop). Although the cell is not a structurally invariant, the small token number (n=6) with a low probability may still give rise to the model overfitting problem. Data-model fit statistics were examined for models with and without the factor level "f": $AIC_{\text{with factor level "f"}} = 184.424$, and $AIC_{\text{without factor level "f"}} =$

182.424. Meanwhile, the corresponding pseudo-R-squared are 72.7% and 42.1%. The research therefore decided to exclude the invariant "f" (reflexive) level for the current analysis due to an increased AIC and a significantly higher pseudo-R-squared (see Table 5.4). The results for *pronoun type* indicate that impersonal pronoun (p=0.768) has a favoring effect for L1 Japanese L2 English learners' object pro-drop in their L2 speech. And demonstrative pronoun (p=0.353) and indefinite pronoun (p=0.263) has favoring effects for accurate object pronoun use. These L2 speakers inaccurately dropped 34.7% of the impersonal object pronoun, compared to a 5.9% proportion when they dropped an indefinite object pronoun in their speech.

Table 5.4

Object Pro-drop by Japanese L2 Learners of English: Inaccurate vs. Accurate Pronoun Usage (Application Value = Inaccurate)

Factor Group	Log Odds	N	Proportion	Factor Weight
Pronoun Type				
Impersonal	1.195	72	0.347	0.768
Relative	0.260	17	0.176	0.565
Personal	0.181	71	0.127	0.545
Demonstrative	-0.605	47	0.085	0.353
Indefinite	-1.031	34	0.059	0.263
Distance (to Pro-form)	0.000013 (log odds, Distance +1)			
Modification (specific)				
Modification	1.238	30	0.567	0.775
No Modification	-1238	211	0.123	0.225
Hesitation (general)				
No Hesitation	1.334	216	0.194	0.792
Hesitation	-1.334	25	0.040	0.208

Note. N: total token detected. Loglikelihood = -83.212. AIC = 182.424. R^2 = 42.1%. All factor groups significant at p<.05

Distance (to the most recent pro-form in the last turn of the conversation). Interestingly, the factor *Distance* (pro-form) was coded in three different ways: distance to the most recent pro-form in only the last turn of the conversation (*d.pro.turn*, see Figure 5.9), distance to the most recent pro-form in up to the previous 4 turns in the conversation (*d.pro*, see Figure 5.9), and distance to the most recent pro-form in the prior conversation regardless of turns (*prior.all.refer*, see Figure 5.9). Only the distance counted in the last turn of the conversation was found to be significantly associated with Japanese L2 English learners' object pro-drop performance. The reported log-odds for distance (pro-form) is 0.000013, which implies one unit (word) increase in the distance to the most recent pro-form in the last turn of the conversation, if available, would result in a 0.000013 unit change in the log of the odds that object pro-drop would co-occur. Consequently, if the pro-form is not available in the last turn of the conversation, the distance to the pro-form is not systematically associated with learners' object pro-drop problem.

Modification (specific). Similar to previous analyses with preposition/particle and article usage, the results for *modification (specific)* also support that specific modification with pronoun use favored Japanese L2 speakers' object pro-drop in their speech (p=0.775). Tokens without any specific modification (e.g., hesitation, repair, repeat, pause, and filled pause) were more likely to retain the necessary object accurately (p=0.225). Among tokens with specific modification, 56.7% observed an object pro-drop, while for the tokens without any specific modification, only 12.3% have object pro-drop issues.

Hesitation (general). For the first time, a general speech fluency indicator

popped up as a significant indicator in predicting the accurate/inaccurate use of a morphosyntactic feature. No hesitation in the token extracted has a favoring effect on Japanese L2 speakers' object pro-drop issue (p=0.792). On the other hand, if any hesitation was observed in the token extracted, there's a disfavoring effect for hesitation in general on Japanese L2 speakers' object pro-drop problem (p=0.208).

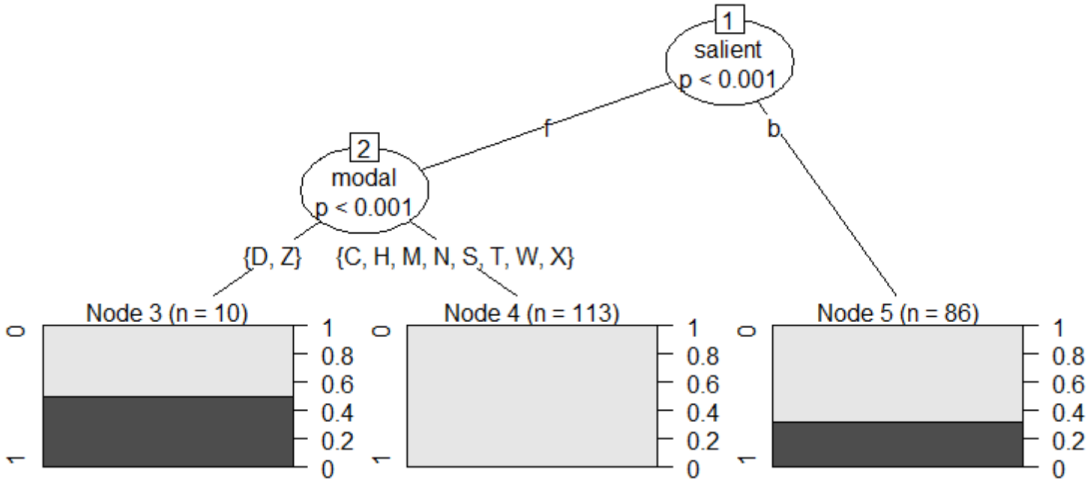
5.4 Modal Auxiliary Verb

5.4.1 Conditional Inference Tree (Classification Tree)

The conditional inference tree approach was conducted to analyze the effects of all 37 factors and factor groups identified in Chapter 3 for the modal auxiliary verb variable. *Salience* was recognized by the tree as the top-most node, i.e., the most important factor group in classifying the accurate and inaccurate use of modal verbs (see Figure 5.11). *Modal verb* was recognized as the sub-branch after the binary classification under level "f" (i.e., non-salient form) in which level "d" (could) and "z" (may) was suggested to be more likely to co-occur with inaccurate modal auxiliary verb usage.

Figure 5.11

Classification Trees: Inaccurate vs. Accurate Modal Verb Usage (Application Value = Inaccurate)



5.4.2 Conditional Inference Forest (Random Forest)

Figure 5.12 introduces the variable importance plot generated by the conditional inference forest analysis. The dot-plot shows that *modal auxiliary verb* is the most frequently identified top position across the trees in the random forest. Second to *modal verb* is *salience* and next, the speech fluency indicator *modification (specific)* which involves any specific modification(s), including hesitation(s), repeat(s), repair(s), pause(s), or filler(s) that happened right before or within the specific position where the variable occurs. The least important factor groups suggested by the random forest analysis include *repair (general)*, *filled pause (general)*, as well as *count*.

The out-of-bag (OOB) classification accuracy assessing the accuracy for the random forest is 83.73% (see Figure 5.13 for OOB by the number of trees using JASP; Han & Dawson, 2020): the tokens left out were at least 83% classifiable, a reasonable classification accuracy produced by the resampled classification trees.

Figure 5.12

Variable Importance Plot: Inaccurate vs. Accurate Modal Verb Usage (Application Value = Inaccurate)

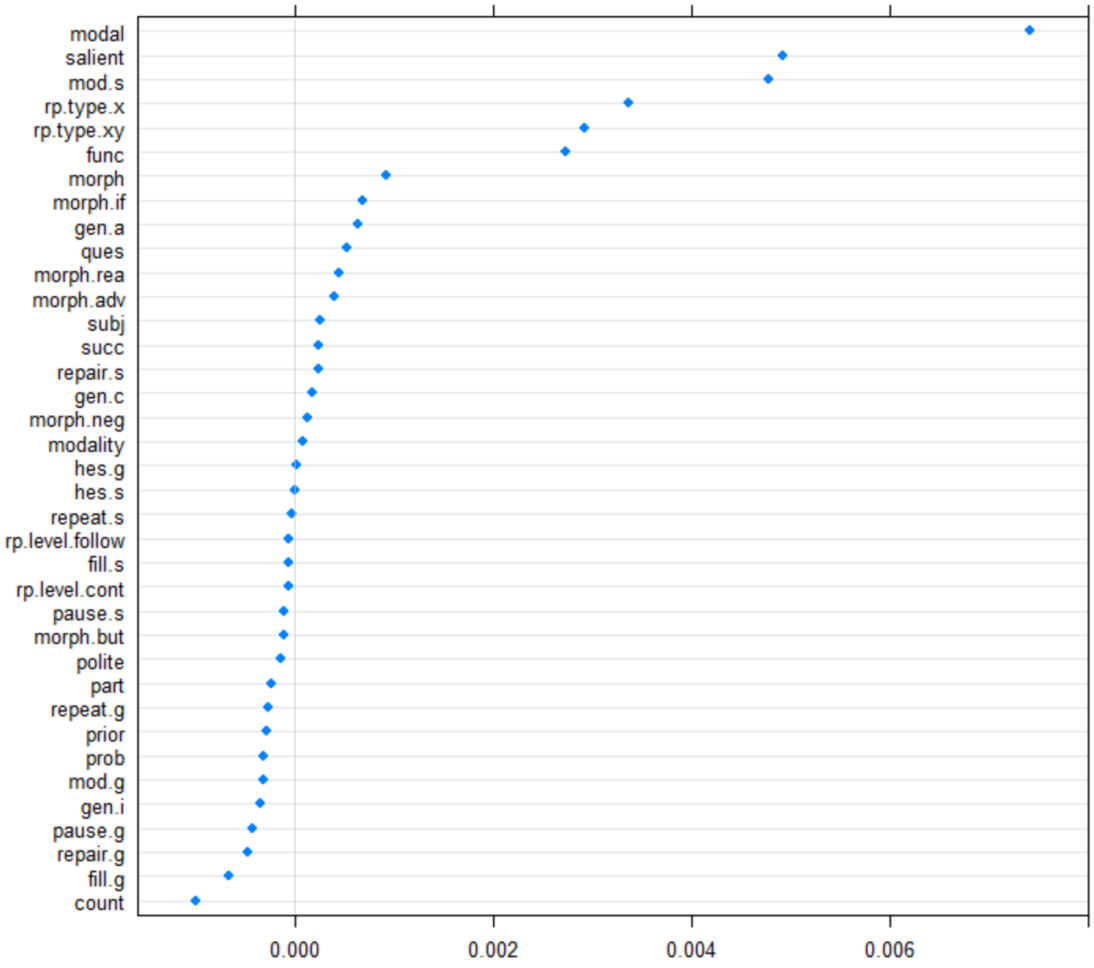
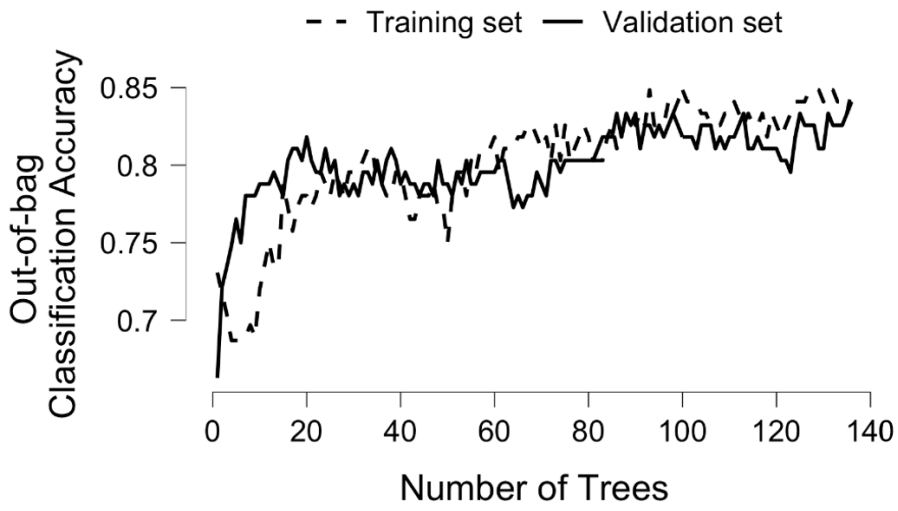


Figure 5.13

Out-of-Bag Classification Accuracy by Number of Trees (Modal Auxiliary Verb)



5.4.3 Mixed-effects Logistic Regression Modeling

Once the factor groups were screened by the classification tree and the random forest on their variable importance, tests of their systematic association with the co-occurrences of the outcome of interest were carried out by the variable analysis. Importantly, although both classification tree and random forest have supported the relative importance of the factor group *modal verb*, multiple invariant contexts were observed for factor levels within this factor group: factor level “c” (can), “s” (shall), “h” (have to), “m” (might), and “n” (must). All have a response probability of 0%, while factor level “z” (may) is a cell that has a response probability showing 100%. Notice that token numbers for some of the factor levels are relatively small (e.g., n=1 for factor level “z”, and n=2 for factor level “n” and “s”), these factor levels are unlikely to be structurally invariant, but the study had to "knockout" the factor group *modal verb* for the possibility of overfitting and seek for alternative modeling solutions.

Table 5.5 presents the final model identified by the variable rule approach in

which both step-up and step-down analysis confirmed a match in optimizing the data-model fit. Linguistic factor groups including *salience* and *morphosyntactic environment* (question) and speech fluency indicators *modification (specific)*, *repair (specific)*, and *hesitation (specific)* were identified as significant factor groups influencing the dependent variable classification into inaccurate or accurate L2 modal verb production. The overall likelihood that an inaccurate modal verb usage would occur, i.e., the input probability, is 60.5%.

Table 5.5

Modality Auxiliary Verb Usage by Japanese L2 Learners of English: Inaccurate vs. Accurate Modal Verb Usage (Application Value = Inaccurate)

Factor Group	Log Odds	N	Proportion	Factor Weight
Salience				
Non-salient form	1.128	86	0.326	0.755
Salient form	-1.128	123	0.049	0.245
Morphosyntactic Environment (Yes-No Question)				
Non-question	0.769	151	0.205	0.683
Question (Yes-No)	-0.769	58	0.052	0.317
Modification (specific)				
No modification	0.956	91	0.242	0.722
Modification	-0.956	118	0.102	0.278
Repair (specific)				
Repair	1.601	18	0.278	0.832
No repair	-1.601	191	0.152	0.168
Hesitation (specific)				
Hesitation	2.335	4	0.500	0.912
No hesitation	-2.335	205	0.156	0.088

Note. N: total token detected. Log likelihood = -64.5. AIC = 143.009. R^2 = 64.0%. All factor groups significant at $p<.05$

Salience. The results for *salience* support that non-salient forms of modal verbs (e.g., will='ll; would='d) favored (p=0.755) and salient forms disfavored (p=0.245) Japanese L2 speakers' inaccurate use of modal auxiliaries. L2 speakers inaccurately used 32.6% of the modal verbs that have a non-salient form, compared to a 4.9% error proportion with that of the salient forms.

Morphosyntactic Environment (Yes-No Question). Considering the factor group *morphosyntactic environment*, the highly constrained morphosyntactic environment in L1 Japanese including yes-no question, negation, adverbial clause for time, adverbial reason clause, if clause, and adversative clause were examined. Only the *yes-no question* factor group detected a significant effect on their co-occurrences in differentiating accurate and inaccurate forms of modal verb usage. Morphosyntactic environment involving no yes-no questions was likely to co-occur with an inaccurate use of modal verbs (p=0.683), while yes-no question sentences were not very likely to have such observation (p=0.317).

Modification (specific). For the factor group *modification (specific)*, factor level "n" (no specific modification) has a favoring effect on inaccurate modal verb usage: the variable rule analysis found that tokens without any type of specific modification were more likely to have inaccurate modal verb usage (p=0.722), while tokens with specific modifications favored accurate modal verb usage (p=0.278).

Repair (specific) and Hesitation (specific). The results for repair and hesitation particularly happened right before or within the occurrence of the modal verb usage both have an opposite favoring effect compared to that of *modification (specific)*. As shown in Table 5.5, the likelihood of an inaccurate modal verb use is

reported to systematically co-occur with specific repair (p=0.832) and specific hesitation (p=0.912).

In addition, the variable rule analysis found each *role play topics* differ significantly in their effect on the likelihood of inaccurately using modal auxiliary verbs in L2 speech, ranging from 15.5% to 83.0%. Table 5.6 presents the random intercepts and their factor weights by *role play topics*.

Table 5.6

Random Intercept (Role Play Topic): Inaccurate vs. Accurate Modal Verb Usage (Application Value = Inaccurate)

Role Play Topic	Intercept	Tokens	Proportion	Factor Weight
Sydney-Wallet	1.663	23	0.261	0.830
Chicago-Doctor	1.503	9	0.333	0.806
Chicago-Flight	1.299	3	0.667	0.772
London-Suit Jacket	1.033	43	0.163	0.722
Wind-down	0.786	34	0.294	0.670
California-Gift	0.743	5	0.600	0.661
Washington-Check	-0.003	1	0.000	0.480
California-Dinner	-0.020	5	0.200	0.476
Travel Suggestions	-0.044	3	0.000	0.470
Boss talk	-0.071	5	0.000	0.463
Lost wallet	-0.121	1	0.000	0.451
New York-Stomachache	-0.170	5	0.000	0.438
Seattle-Clothes	-0.220	2	0.000	0.426
San Francisco-Lunch	-0.434	8	0.000	0.375
Teacher-daughter	-0.447	4	0.000	0.372
Warm-up	-0.638	7	0.000	0.328
Vancouver-Apartments	-0.724	7	0.000	0.310
Atlanta-Taxi	-1.047	26	0.077	0.245
Oxford-Camera	-1.619	18	0.000	0.155
Overall	1.379	209	0.163	N/A

Note. Log likelihood = -64.5. AIC = 143.009. R^2 = 64.0%.

5.5 Implicational Scales

Table 5.7 below presents accuracy-valued implicational scales with both rankings and accuracies listed against speakers as rows in portraying cross-sectional interlanguage development for the current research. The overall candidate rankings were calculated by aggregating all four morphosyntactic features (see column "Overall Rank"), and the order of morphosyntactic features were decided based on the averaged error rate across candidates: Preposition/particle use has the lowest error rate (p=11.3%), while article use has the highest error rate (p=44.7%). Cells with higher accuracies were moved to the upper left to construct potential step-like line patterns (e.g., from upper right to lower left which separates cells with accuracy as a cutoff threshold). A clear step-like cutoff threshold of 0% error rate were observed across four features and the pattern aligns well with the averaged error rate from low to high (see Table 5.7, the dotted line in blue). Besides the 0% error rate threshold, no other cut-off values could be found with perfect step-like implicational patterns. Nevertheless, from Table 5.7, it could be generally concluded that the lower right section is filled with more red cells, especially for the article feature.

Table 5.7

Accuracy Scale for Morphosyntactic Features Ranked by Candidate ID#

ID#	Overall Rank	Preposition/ Particle		Object Pronoun		Modal Verb		Article	
		Rank	Err%	Rank	Err%	Rank	Err%	Rank	Err%
bca	4	1	0.000	1	0.000	1	0.000	1	0.000
ccb	4	1	0.000	1	0.000	1	0.000	1	0.000

cfa	17	1	0.000	1	0.000	N/A	N/A	15	0.333
bhb	18	1	0.000	1	0.000	1	0.000	15	0.333
bga	26	1	0.000	1	0.000	1	0.000	23	0.500
cja	26	1	0.000	1	0.000	1	0.000	23	0.500
bfa	36	1	0.000	1	0.000	33	0.250	1	0.000
cab	39	1	0.000	1	0.000	25	0.143	12	0.250
eab	40	1	0.000	1	0.000	29	0.200	9	0.182
bfb	41	1	0.000	29	0.200	1	0.000	10	0.200
aca	41	25	0.120	1	0.000	1	0.000	14	0.267
cea	41	29	0.154	1	0.000	1	0.000	10	0.200
caa	42	1	0.000	39	0.429	1	0.000	1	0.000
ahb	48	45	0.364	1	0.000	1	0.000	1	0.000
cfb	50	40	0.250	1	0.000	1	0.000	8	0.143
aia	54	18	0.086	28	0.184	1	0.000	7	0.135
ceb	65	21	0.111	1	0.000	1	0.000	42	0.800
afa	70	35	0.200	1	0.000	33	0.250	1	0.000
deb	75	21	0.111	34	0.333	1	0.000	19	0.400
dda	79	24	0.118	1	0.000	25	0.143	29	0.583
jhb	82	17	0.048	26	0.143	1	0.000	38	0.727
abb	83	1	0.000	1	0.000	42	0.667	39	0.750
bba	83	29	0.154	1	0.000	33	0.250	20	0.438
aeb	84	1	0.000	1	0.000	43	1.000	39	0.750
jeb	86	1	0.000	41	0.500	29	0.200	15	0.333
bgb	86	40	0.250	1	0.000	1	0.000	44	1.000
bia	92	19	0.091	32	0.250	29	0.200	12	0.250
bbb	92	29	0.154	39	0.429	1	0.000	23	0.500
aba	94	1	0.000	29	0.200	36	0.333	28	0.545
aib	94	26	0.125	44	0.667	1	0.000	23	0.500
aea	96	32	0.167	1	0.000	29	0.200	34	0.667
jha	96	33	0.174	1	0.000	28	0.167	34	0.667
bea	97	38	0.211	26	0.143	1	0.000	32	0.625
bha	97	44	0.333	29	0.200	1	0.000	23	0.500
cca	100	19	0.091	41	0.500	1	0.000	39	0.750
jea	100	26	0.125	1	0.000	36	0.333	37	0.700
agb	106	43	0.308	1	0.000	41	0.500	21	0.444
cgb	110	21	0.111	34	0.333	25	0.143	30	0.600
aha	112	35	0.200	34	0.333	1	0.000	42	0.800
cha	115	28	0.132	33	0.286	24	0.133	30	0.600
cbb	115	39	0.222	41	0.500	1	0.000	34	0.667
bcb	118	42	0.273	1	0.000	43	1.000	32	0.625
acb	119	1	0.000	38	0.400	36	0.333	44	1.000

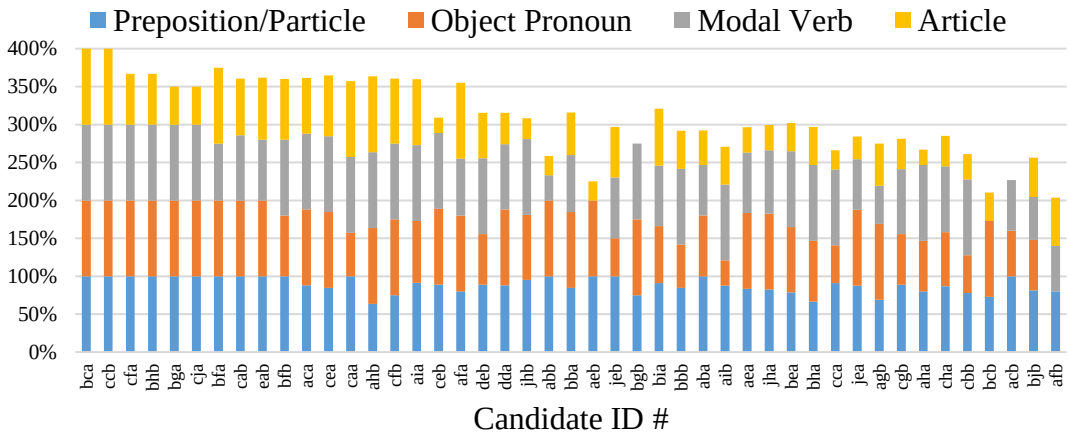
bjb	130	34	0.186	34	0.333	40	0.438	22	0.481
afb	137	35	0.200	45	1.000	39	0.400	18	0.364
Averaged Error Rate			0.113		0.164		0.166		0.447

Note. The table illustrates the color-coded accuracy scale for L2 morphosyntactic features examined in the current study. Cell colors from green to white to red indicate an increased tendency of error rate (i.e., decreased accuracy).

To better address the accuracy implicational patterns among these morphosyntactic features, the accuracy data for Table 5.7 was also used to generate a trajectory stacked column chart to demonstrate the visual pattern of cross-sectional L2 interlanguage development (see Figure 5.14). All four features seem to have maintained an overall steady trend of accuracy (left: higher, and right: lower). Still, it could be observed in Figure 5.14 that some individuals have a relatively fluctuating patterning and hence a relatively mixed combination of accuracies.

Figure 5.14

Supplementary to Implicational Analysis: A Trajectory Stacked Column Chart



Chapter 6: Discussions

Using an expanded tool kit, the analyses in Chapter 5 have shown that complementary techniques such as mixed-effects models, classification trees, and random forests have opened various analytical possibilities for data exploration in variation studies. By examining inter-learner variation and cross-sectional intra-learner variability with linguistic, paralinguistic, and nonlinguistic factors and factor groups, this dissertation further explored the L2 use of preposition/particle, article, object pronoun, as well as modal auxiliary verb in demonstrating variation in the group and the individual (See Table 6.1 for a summary). Furthermore, the findings have yielded important theoretical and empirical implications for the current understanding of second language acquisition from a variationist perspective.

Table 6.1

Results Summary

Linguistic Variables	Linguistic Factor Groups	Paralinguistic Factor Groups	Nonlinguistic Factor Groups
Preposition /Particle	Phrase Type	Modification (specific) Repair (specific)	
Article	Definiteness Premodifier* Modification (specific)	Premodifier* Modification (specific) Repeat (specific) Speech processing time (random slope, by ID)	Role Play Level ID (random intercept)
Object Pronoun	Pronoun Type Distance (to Pro-form)	Modification (specific) Hesitation (general)	
Modal Auxiliary Verb	Salience Morphosyntactic environment (yes-no question)	Modification (specific) Repair (specific) Hesitation (specific)	Role Play Topics (random intercept)

6.1 Statistical Implications

Recent development in statistics has introduced conditional inference trees and machine learning tools such as random forest which provides complimentary assistance to the mixed-effects modeling method. Both conditional inference trees and random forests are still new to variationist SLA. To address RQ1, this dissertation has provided a cross-sectional example to evaluate how variation analysis in SLA could benefit from these new tools.

6.2.1 The Application of Classification Tree and Random Forest Approach

Classification trees straightforwardly present how multiple factor groups operate in a hierarchical order on the observed variation and can help generate the most important hierarchy of independent variables which co-occur with the desired binary classification (i.e., the accurate/inaccurate variation). In the analysis in Chapter 5, the top nodes of each tree (i.e., *phrase type* for preposition/particle use, *definiteness* for article use, *specific modification* for object pro-drop, and *salience* for modal verb use) were also recognized by the mixed-effects variable rule analyses as significant predictors. However, although it is generally understood that adding more independent variables will generate more splits (Tagliamonte, 2011; Tagliamonte & Baayen, 2012), the current research did not find a direct association between variable numbers and sub-nodes of the classification trees. The object pro-drop classification tree with 37 independent variables has a one-root classification structure, while the article classification tree with 27 independent variables has generated up to 4 sub-

nodes.

The current research also observed that the number of nodes and sub-nodes identified by the classification tree is generally less than the actual number of significant factors and factor groups identified by the mixed-effect variable rule approach: the trees detect 2, 4, 1, and 2 nodes and sub-nodes for preposition/particle, article, object pronoun, and modal verb use respectively, while the corresponding numbers of actual independent variables detected by the variable rule analysis are 3, 6, 4, and 5. This indicates a potential shortcoming of the classification tree method: further splits are possible only when a hierarchical tree structure is available (e.g., potential effects on a subgroup of the binary splits). For the current study, the classification tree may stop producing sub-branches because of the large number of binary-coded factor groups that constrain the availability of further binary splits (see while in Tagliamonte and Baayen, 2012, splits are available firstly to non-binary variables).

As reviewed in Chapter 4, the classification tree approach has also been criticized for its generalizability to new data, and random forests using ensemble learning method could overcome this short-coming and rank the relative importance among various independent variables. By randomly resampling the original data with a subset of variables to replicate the process of classification tree analysis multiple times, the random forest approach is a more comprehensive tool for researchers who are interested in knowing the whole picture. Unlike the classification tree approach which stops when further binary classification is not available, the random forest approach provides a plot of variable importance that all independent variables entered

into the model will be aligned together by their relative importance.

Despite the shortcomings, the classification tree structure is more stable, i.e., if repeating the classification tree methods for multiple times, it is likely that the trees generated are identical in structure, however, if the same procedure is repeated for the random forest, or one or more (less relevant) independent variables are removed from the random forest analysis, forest with slightly different variable importance structures may be generated, especially for the variables that are "less" important (see Figure 5.2, and Figure 5.3 as an example). Indeed, random forest applies the machine learning method and depending on the size of the data, the number of trees generated, the testing and validation sets, etc., the random forest may give different recommendations of variable importance.

6.2.1 Variable Selection Comparisons

In addition, by comparing the decision suggestions made by classification trees and random forests against the actual model, agreement on variable choice was found for most of the top-ranking variables (see Table 6.2 for details): For preposition/particle use, the actual model includes *phrase type, modification (specific),* and *repair (specific)*. Both tools (i.e., classification tree and random forest) agree that *phrase type* is most influential factor group detected in accounting for the observed variation, and *modification* (specific) is the second most important variable recommended by the random forest-the inclusion of both factor groups was confirmed by the mixed-effects model. Another significant factor group in the final model, *repair* (specific), also ranked 5th in the variable importance plot.

Table 6.2

Variable Rule Analysis, Classification Tree, and Random Forest Model Comparison

Morphosyntactic Feature	Factor Groups Identified by Variable Rule Analyses	Classification Tree Ranking	Random Forest Ranking
Preposition /Particle (N=27)	Phrase Type	Root (1)	1
	Modification (specific)		2
	Repair (specific)		5
Article (N=27)	Definiteness	Root (1) Sub-branch (4)	1
	Modification (specific)		2
	Speech Processing Time		4
	Role Play Level	Sub-branch (2)	6
	Premodifier		9
	Repeat (specific)		18
Object Pronoun (N=37)	Modification (specific)	Root (1)	1
	Pronoun Type		3
	Hesitation (general)		5
	Distance (to Pro-form)		7
Modal Auxiliary Verb (N=37)	Salience	Root (1)	2
	Modification (specific)		3
	Role Play Topics		5
	Morphosyntactic Environment (yes-no question)		10
	Repair (specific)		15
	Hesitation (specific)		20

Note. n = total number of variables considered.

Interestingly, unlike variationist sociolinguistic studies on native speaker data using the expanded tool kit which yields surprisingly similar results in variable identification, the current analysis also observed heterogeneous outcomes on variable identification in model building. *Filled pause (specific)*, for example, was recognized by both the classification tree and the random forest method to be the second most important variable for accurate/inaccurate article use classification. The actual model, however, did not detect a significant contribution from *filled pause (specific)* in the

model building procedure. One explanation for the difference may lie in the multicollinearity between a number of factor groups such as *modification (specific)* or *pause (specific)*-the former was significant in the actual model. In fact, multicollinearity is not simply a shortcoming but a two-sided sword: On one hand, random forest outcomes on variable importance became imprecise when variables highly correlated with each other ranked close to each other, and researchers have to test out the relative variable importance by applying the actual model. Alternatively, the random forest method provides a potential solution to coding dilemmas that linguistic researchers often came across when no golden rule is available. In the current research, for the object pro-drop error observed in the L2 speech data, different coding methods of the distance to the most recent pro-form mentioned earlier in the conversation makes a significant difference in mixed-effects modeling. When given the raw data, there is no rule of thumb to decide the availability of the pro-form when it is mentioned several turns of conversation away. In the end, the factor was coded in three different ways (e.g., one conversation turns, up to four conversation turns, and any prior mentioning). All were entered into the random forest at the same time. Pro-form appeared only at the last turn of conversation ranked 7th among the 37 factors and factor groups in the variable importance plot (see Figure 5.9), while pro-form available in the last 4 turns of conversation ranked 21st, with the factor or any prior mentioning ranked 28th. The mixed-effects variable rule analysis confirmed this recommendation: only pro-form available in the last round of conversation is significantly associated with the observed object pro-drop variation. This is a valuable finding especially for linguistic research dealing with heavy work

of data annotation: researchers no longer need to worry about the best coding method-they can code the same variable in many ways and ask random forest to find the best one for them.

Still, collinearity alone cannot explain that unimportant factor groups such as specific repeat (ranked #18/27, article), specific repair (ranked #15/37, modal verb) morphosyntactic environment-question-(ranked #10/37, modal verb), and specific hesitation (ranked #20/37, modal verb) identified by random forest indeed contributed significantly to explaining the observed variation in the mixed-effects models. Another potential explanation is that random forest considers only the relative importance of classifying the observed variation. In contrast, mixed models consider a joint effect of multiple fixed and random effects. As a result, random effects may override some of the relatively important factor groups in the mixed-effects models. In the current study, random intercept and random slope by *candidate ID* were identified for L2 article use, and random intercept by *role play topics* was found to systematically contribute to L2 modal verb use-the two morphosyntactic features whose variation were found to co-occur with less important factor groups identified by the corresponding random forests.

To sum up, this dissertation has provided a handful of examples that variationist SLA, especially when data size is relatively small with a large number of potential predictors, can substantially benefit from the expanded tool kit of classification trees and random forests. Both the classification tree and random forest method have their advantages and disadvantages when analyzing the observed variation, and should be considered as complementary suggestions for the actual

model-building procedure. With disagreement between different methods across different morphosyntactic features for the current study in mind, future researchers need to pay close attention when applying classification tree and random forest to variation data, especially when potential random effects are involved.

6.2 External Effects: Linguistic, Paralinguistic, and Nonlinguistic Aspects

To address RQ2 and RQ3, the multivariate analyses for the current study have examined related factor and factor groups with regard to their relative contribution on L1 Japanese L2 English leaners' variable use of English preposition/particle, article, object pronoun, and modal verb. A fair number of linguistic, paralinguistic, and nonlinguistic factors and factor groups were found to systematically co-occur with the observed accurate/inaccurate variation, indicating an inherently consistent relationship between individual and group patterns (RQ4).

6.2.1 Linguistic Factors and Factor Groups

For linguistic contexts, prepositional phrases without a main verb have been found to be likely to co-occur with inaccurate preposition usage. By carefully screening the data, "at" phrases indicating time (e.g., my flight is [at] 7 pm) and quantifiers including "of" (e.g., a few [of] my friends, a little bit [of] [a] problem, a little bit [of] fever) are the two most frequently observed inaccurate preposition use under this factor level. A possible explanation is interlingual errors with L1 transfer (e.g., Chelli, 2013; James, 2013): no preposition is needed in Japanese when indicating time, and the reciprocal quantifier expression in Japanese has no preposition. Furthermore, prepositional phrases with a "main verb + preposition" structure, which oftentimes

confuses L2 learners due to its superficial structural similarity with the phrasal verb "verb + particle" structure (e.g., Matsumoto, 1996; Miyajima, 1984; Neagu, 2007; Talmy, 1985, 1991; Yasuda, 2010), also have favoring effects for inaccurate preposition use with a factor weight of 0.593. Prepositional phrase with a direct object after the main verb favored inaccurate preposition use less as compared with the structure without a direct object, which may suggest that a direct object after the main verb could serve as a clue for learners to tell the structure from phrasal verbs. Phrasal verb structure, on the other hand, does not favor accurate or inaccurate particle use ($p=0.5$). Considering the underuse of phrasal verb ($n=25$) as compared with prepositional phrases ($n=232$ for "main verb + preposition" structure), it is likely that Japanese L2 learners of English may have avoided the use of phrasal verb in their L2 speech since they are less aware of the necessity of the particle components (Yasuda, 2010).

For L2 article use, more definite articles were used than indefinite articles, and Japanese L2 speakers for the current study are more likely to make an indefinite article error than a definite article error. The finding is consistent with what has been reported by Nomura (2010, 2012). For a language like Japanese that lacks an article system, this may indicate that rules governing the use of indefinite article may be more complicated than definite articles for Japanese L2 speakers.

For object pro-drop error, the result of the current research on ungrammatical object pro-drop has supported Yuan's conclusion (1997) that ungrammatical null object in L2 is difficult to realize and maybe a product of L1 transfer. Consistent with research which have proposed that pro-drop in radical pro-drop languages such as

Japanese is licensed at the pragmatic level by a preceding topic recovered in the discourse (e.g., Cote, 1996; Huang, 1984; 1989; Ohara, 2007; O'Grady, Yamashita, & Cho, 2008; Tomioka, 2003), this study also found distance to the most recent pro-form in the conversation is predicative of ungrammatical object pro-drop. Interestingly, the distance that permits the object pro-drop is different across L1 and L2: in Japanese, object pro-drop could happen if the preceding topic, in either full form (i.e., the jacket) or pro-form (i.e., it), by oneself or by the interlocutor, could be recovered earlier in the conversation. As long as the conversation is around the topic, and the missing argument has been mentioned earlier, Japanese native speakers are allowed to drop the object pronoun in their L1. On the other hand, Japanese L2 learners of English unlearned the default L1 parameter setting (i.e., null object) by constraining the availability of the preceding topic: if the topic is in pro-form and is within the very last turn of the conversation, the closer the pro-form, the less likely an object pro-drop error would occur. In addition, this research also confirmed one of Yuan's observation (1997) that inanimate null object sentences are classified as grammatical for even very advanced L1 Chinese L2 English learners: for the current study with Japanese L2 learners, impersonal pronoun "it" also has a strong favoring effect for inaccurate object pro-drop.

Lastly, the modal verb usage analysis gives evidence that L2 Japanese speakers are more likely to omit non-salient forms of modal verbs-salience was the most important and most relevant factor group regarding L2 modal auxiliary variation and was found to improve the data-modal-fit to the greatest extent among all factor groups. As the effect of salience and related discussions has been well documented in

second language acquisition research, the current finding further supports the idea that L2 learners are more likely to acquire physically, psychologically, and acoustically salient input (e.g., Brown, 1973; Bybee, 2000; Ellis, 1996; 2006; Goldschneider & DeKayser, 2001; Herron & Bates, 1997; Zuraw, 2003). Secondly, a potential L1 transfer was observed in the analysis of L2 modal verb use. The results for the current analysis support that L2 modal auxiliary use was systematically associated with the morphosyntactic environment *yes-no question*-an environment that constrains the use of modality in L1. On the other hand, not all L1-associated morphosyntactic environments were found to affect the linguistic variable of interest, leaving further examination and explanation for the current and future research.

6.2.2 Paralinguistic Factors and Factor Groups

A number of speech fluency indicators were found to show varied effects on the observed variation across different morphosyntactic features.

Overall, modification (specific) is a well-recognized predictor of accurate/inaccurate morphosyntactic feature variation. The factor group was recognized by all four features as a significant independent variable. Specifically, for preposition/particle, article, and object pronoun usage, specific modification happened right before or during the occurrence of the corresponding morphosyntactic feature has a favoring effect for inaccurate use, while for modal auxiliary verb, specific modification has a disfavoring effect for inaccurate modal verb use. This may indicate that for preposition/particle, article, and object pronoun, learners are more aware of the inaccurate uses-they either hesitated, repaired the speech, repeated the morphosyntactic structure, paused, or inserted filler words in their speech with added

uncertainty, which ultimately co-occured with their inaccurate use of the identified morphosyntactic feature. While for modal auxiliary verb, since 94% of the modal verb error are omission errors (i.e., missing modal verbs), learners might not be aware of the mistakes they made, and hence this seemingly contradicting effect across different features are in fact not impossible.

In addition to specific modification, specific hesitation, specific repair, specific repeat that happened right before or during the variable occurrence, as well as general hesitation that happened anywhere in the token were found to be systematically associated with the inaccurate use of the morphosyntactic features for the current study. Particularly, contradictory patterns of favoring and disfavoring effects were observed: Repair (specific) favored accurate preposition/particle use and inaccurate modal verb use, repeat (specific) favored inaccurate article use, hesitation (general) disfavored object pro-drop, and hesitation (specific) favored inaccurate modal verb use. Effects of paralinguistic factors and factor groups are still new to the field of variationist SLA, more fine-tuned studies are needed to explain these paralinguistic patterns.

Lastly, speech processing time (i.e., total word count per token extracted) was found to have individual-based random effects on accurate/inaccurate article use, i.e., different individuals have a different favoring effect for inaccurate article use (see Table 5.3). As the word count increases, Candidates (e.g., #cha, #bgb, #jhb) with a larger intercept (i.e., a higher probability of making an article error) will have a reduced likelihood of inaccurate article use, and candidates with a smaller intercept (e.g., #cfb, #bia, #bfb) will have an increased likelihood of inaccurate article use.

Overall, this dissertation research has found a systematic association between paralinguistic factors and accurate and inaccurate morphosyntactic usage variation in L2 speech. The study would be able to conclude that learner language is systematically governed by certain paralinguistic rules, most likely specific modifications. Still, further explorations and explanations are in need to explain why certain contradictory patterns were observed with speech fluency indicators favoring accurate and inaccurate L2 uses across different morphosyntactic features.

6.2.3 Nonlinguistic Factors and Factor Groups

This study has also found that nonlinguistic factors and factor groups are likely to co-occur with accurate and inaccurate L2 article use variation. Role play level (from 1-5) is negatively associated with the log odds of inaccurate article use: the higher the role play level given by the interviewer, the less likely a candidate will make an article usage error. Reasonable as the observation is, the effect of role play level is not universal for the other three morphosyntactic features. A potential explanation is that article use might be a good indicator of L2 learners' proficiency level, as article has been consistently recognized as a morpheme in natural order and L2 acquisition order studies (see Luk & Shirai 2009 for a summary). Preposition/particle, object pronoun, and modal auxiliary verbs might be surface forms for a range of phenomenon whose accurate and inaccurate uses are composed of multiple aspects involving skills from multiple acquisitional stages. This explanation could be further tested by including other natural order/L2 acquisitional order morphemes, but due to the face-to-face role play situation nature of the current study, not enough tokens could be extracted for morphemes such as regular and irregular past tense, third-person singular, and

possessive -s. Future studies with larger data sets or with other written and spoken data could further examine the association between proficiency level indicated by the task difficulty and variation observed in the use of these acquisition order morphemes.

In sum, the exploration of linguistic, paralinguistic, and nonlinguistic conditionings on variation in L1 modal auxiliary verb usage shows that learner language is not entirely random. More than one significant linguistic, paralinguistic, and nonlinguistic factor groups were recognized by variable rule analyses for the current research, indicating individuals do mirror group patterns on certain aspects. The current analysis has justified an empirical and a theoretical variationist-based view of the relationship between the group and the individual: individual performances do match group variation patterns, and Japanese learners of English do acquire their second language similarly on several aspects (e.g., Bayley & Longman, 2002; 2004; Regan, 2013; Young, 1989; 1991).

6.3 The Importance of Random-Effect Factors

To extend the discussion on RQ4 about individual and group patterns, the mixed-effects models nevertheless assessed the importance of random-effect factors. The results support that although the learner language system is governed by a number of linguistic, paralinguistic, and nonlinguistic "grammar" (i.e., factor groups), learners and role play topics that were involved have individualized baseline and grammar that are not shared by the group.

For L2 article use, individuals as random intercepts significantly improved data-model fit, indicating that although certain grammar are subject to inter-lingual

factors, individuals can have different baseline preferences for native-like versus non-native-like use of L2 article use. Adding individuals as random intercepts allows the baseline to be adjusted for each learner separately, while the effect of definiteness, premodifier, role play levels, and paralinguistic factor groups are the same across all individuals. Additionally, individuals have a speech processing time-based favoring effect (i.e., random slopes) for inaccurate article use, suggesting that besides linguistic and paralinguistic aspects that are subject to shared interlanguage factors, individuals do have their own paralinguistic "grammar".

For accurate and inaccurate L2 modal verb use variation, role play topics as random intercepts significantly improved data-model fit, indicating that different role play topics have different baselines of preference for native-like/nonnative-like modal auxiliary verb usage. Interestingly, the magnitude of role play topics varied from 0.830 to .155, indicating a relatively large effect size of role play topics on modal auxiliary verb usage. The results that role play topics differ significantly (from 15.5% to 83%) in promoting L2 speakers to make a modal verb usage error in their L2 speech supported that modal verb usage is highly sensitive to the situation given: Candidates facing the likelihood of 83% situation *Sydney-wallet,* for example, were asked to borrow some money from the interviewer to go to lunch. In the 15.5% situation *Oxford-Camera*, however, candidates were asked to have a store clerk (the interviewer) repair his/her camera. The two role play topics differ in the degree of modality involved, the former requires the candidate to play with request, permission, possibility, certainty, prediction (to pay the money back), etc., while the latter is more about making simple requests.

6.4 L2 Development

Lastly, to address the last research question (RQ5), the dissertation generated an accuracy-based implicational scale based on Japanese L2 learners' performance across different morphosyntactic features. The study also reported a potential L2 morphosyntactic development sequence from preposition/particle to object pronoun usage, modal auxiliary verb, and finally article use, based on a cross-sectional step-like patterning with a cutoff threshold of performance accuracy.

It is not surprising that perfect scales are not available, and individuals have inconsistent proficiency rankings (accuracy score) across different morphosyntactic features. One could generally indicate an overall accuracy pattern by observing the color-coding feature of the implicational scales generated (see Table 5.7), but whether the observed accuracy-based scale indicate a pre-determined or fixed L2 developmental order is under debate for a number of reasons. On one hand, taken external and internal effects discussed in the previous section into consideration, the differences between the four structures seem to reflect the difficulty level of these structures due to a set of L1-specific factors. For example, preposition/particle use may reflect the unavailability of satellite-framed verb structure in Japanese, article use may reflect the unavailability of the article system in Japanese, object pro-drop is associated with the nature that Japanese is a radical pro-drop language, while modal auxiliary verb is constrained by the morphosyntactic environment in learners L1. On the other hand, external factors like salience, definiteness, pronoun type, indicated task difficulty (i.e., role play level), etc., that are reported in previous studies with other groups of L2 learners have also been found to independently contribute the

observed variation for the four different structures. The interpretation of the implicational scales and indicated patterns need to consider the multi-dimensional nature of L2 morphosyntactic, and no easy conclusion could be made without a careful examination of not only the cross-sectional trajectory, but also the longitudinal individual trajectory. With the current result, the dissertation is only able to conclude that the accuracy-based differences and patterns indicated by the implicational scales may reflect an overall earlier establishment of the accurate use of L2 preposition/particle, object pronoun, modal verb than article use for Japanese learners of English in oral production.

Importantly, the use of implicational scales in linguistic variation research should not be satisfied to fulfill only the statistical purpose for finer data ordering and discriminability but, as Rickford (2002) suggests, should always seek empirical and theoretical groundings for the similarities and differences between the observed data and the hypothesized patterns.

Chapter 7: Limitations and Conclusions

As ambitious as the current study in providing novel practice of modern analytical tools in understanding L2 variation and L2 development, limitations of the dissertation need to be mentioned. Firstly, since the L2 spoken data were manually transcribed, only 50 role plays from 46 L2 speakers were analyzed in the current analysis, which resulted in small token numbers for some conditions, and more extremely, overfitting cases that are not structurally invariant but had to be "knocked-out" due to invariability. There is a clear need for future study to analyze a larger data set in order to not only confirm the conclusions reported in the current dissertation, but also examine some of the potential factors and factor groups that might be neglected due to limited sample size. Secondly, although the current cross-sectional study reported intra-learner variability by comparing L2 competence in multiple morphosyntactic structures within each learner, it would still be viewed as debatable and indirect evidence about L2 development, and due to the multi-causality nature of the observed variation, one can hardly generalize to empirical and theoretical conclusions about L2 developmental stages. Longitudinal studies are still the best option to investigate individual developmental trajectory in order to make a conclusion about the existence of a pre-determined L2 acquisitional order. Statistically, it is worth mentioning that the current data set were collected around 2003 and 2004, therefore some of the conclusions for the current analysis may not make as much sense as to Japanese L2 learners nowadays as compared to learners 20 years ago. Future research may want to apply the modern tools used in the current

dissertation with a set of data collected more recently, and compare how the results might have changed for the past 20 years.

Despite these limitations, this dissertation research observed, studied and analyzed cross-sectional morphosyntactic variation in L2 speech to provide novel understandings of L2 processing as well as interlanguage development. The study has presented a detailed flow with modern analytical strategies on how the variationist paradigm, treating native-like and nonnative-like L2 morphosyntactic uses as variants of invariant native speaker forms, has extended our understanding of the ongoing discussion on L2 competence and development.

Importantly, the study examined various linguistic, paralinguistic, and nonlinguistic factors and factor groups in support of the idea that learner language is inherently systematic. Particularly, this dissertation is one of the first to systematically analyze paralinguistic conditionings (e.g., hesitation, repeat, repair, pause, filler, modification) on variation in SLA. Specific speech modification(s) was found to be systematically associated with the accurate/inaccurate L2 use across all four morphosyntactic features examined in this study. Moreover, taking random effects by individuals and role play topics into consideration, this dissertation has implicated that variation in learner speech is not simply subject to shared inter-lingual and intra-learner factors. In fact, individuals may have their own inter-lingual "grammar" regarding speech processing time, and learners' modal auxiliary verb use may also be context-sensitive. Furthermore, implicational patterns have unveiled important information that may help current and future researchers to understand the underlying processing hierarchy of Japanese learners in acquiring English as a second

language. The author hope that future research may find this dissertation a good starting point to explore the possibility of applying the modern variationist paradigm to quantitatively describe and analyze learner language and interlanguage development in answering important questions about internal and external patterning of variation and variability in learner language.

Appendices

Appendix S3.2 List of Scenarios

Atlanta-Taxi

You're staying at a hotel in Atlanta, and you have to catch a flight to buffalo early tomorrow morning. The interviewer will play the role of the concierge, ask her to order you a taxi.

Boss' talk

You and your boss are having a meeting about a project you are working on with the boss' son. The son is not doing a good job. Convince your boss that the son should be removed from the project. The interviewer, that's me, will play the role of the boss.

California-Dinner

You are going out to dinner in a lovely Californian city of Monterey. And you want to make sure that everything goes well. Call the restaurant and make the reservations. I will play the role of the restaurant employee.

California-Gift

You are in a gift shop in Monterey and Monterey is in California. And it's right down the Ocean, Monterey. And you'd like to take some souvenirs back home to your family. The interviewer will be the role of a salesperson. Tell her what kinds of gifts you want and buy them

Chicago-Doctor

You are staying in a hotel in Chicago when you suddenly feel sick. The interviewer will play the role of the front desk clerk. Tell the clerk about your problem, and ask

for a doctor to be called.

Chicago-Flight

Due to a long running business meeting in Chicago, you arrived at the airport late, and discovered you have missed your flight to Narita. Go to the airline clerk and book another flight to get to your destination as soon as possible. The interviewer, that's me, will play the of the clerk.

Dormitory room

I'd like you to describe the physical layout of your dormitory room. Say I'm standing of the door, I'm looking around, what do I see?

Find lost wallet

This time you are shopping in a store and accidentally left your wallet on the counter. The interviewer will play the role of a clerk. Explain your problem.

London-Suit Jacket

You are in London on business, and you notice that your suit jacket is dirty. The interviewer will play the role of a dry-cleaner. Please ask to have your jacket cleaned.

New York-Stomachache

You are in New York， and you have a stomachache. Ask the pharmacist for some medicine. The interviewer will play the part of the pharmacist.

Oxford-Camera

You're sightseeing in England, in Oxford, and your camera is broken. Interviewer will play the role of a store clerk. Ask to have your camera repaired.

San Francisco-Lunch

You are in San Francisco on business, and you want to have lunch with a prospective

business partner. Call me on the phone and arrange a lunch appointment. I will play the role of the business partner.

Seattle-Clothes

You're the clothing store in Seattle. You need to buy some clothes for colder weather. I'm going to play the role of a store clerk. Talk with me and arrange to buy what you need.

Sydney-Wallet

You are working in Sydney, and you are about to go to lunch and find that you forgot your wallet. The interviewer will play the role of a co-worker, borrow some money.

Teacher-daughter

Your child has brought home her report card and her grade are much lower than you expected, and you are quite upset. The interviewer will play the role of a teacher, that's me, I'm a teacher. Discuss the situation with the teacher.

Travel safety suggestions

(Followed by a conversation on travel safety) As someone who has travel experience, what advice would you offer to your compatriots who're going to a long trip overseas?

Vancouver-Apartments

In this role play you will come to Vancouver to work, and you need a place to live. The interviewer, that's me, will play the role of a real-estate agent, ask to see some apartments.

Washington-Check

You are a visitor in Washington and need to change some money. The interviewer

will play the role of the bank teller. Change some traveler's checks.

Appendix S5.1 Variable List

Aspects	Factor groups	Variable Coding
Linguistic Factors and Factor Groups	Preceding Verb	prev.verb prev.verb.form (inflective or root form)
	Verb Frequency	raw.freq
	Verb Frequency (Root Form)	root.freq
	Phrase type	p.type
	Definiteness	definite
	Syntactic type (Head Noun)	n.type
	Word Frequency (Head Noun)	n.freq
	Premodifier	premodifier
	Formulaic Language	formulaic
	pronoun	pronoun
	Pronoun Types	pron.type
	Preceding Verb	pre.verb
	Prior	prior.all.refer
	Scenario	senario.refer
	Pro-Form	pro
	Full Form	full
	Candidate	self
	Interviewer	int
	Distance (Pro-Form)	d.pro (pro-form mentioned in the last 4 turns of conversation) d.pro.turn (pro-form mentioned in only the last turn of conversation) pro.turn.n (number of turns for pro-form)
	Distance (Full Form)	d.full (full-form mentioned in the last 4 turns of conversation) d.full.turn (full-form mentioned in only the last turn of conversation) full.turn.n (number of turns for full-form)
	Modal	modal
	Politeness	polite
	Salience	salient
	Modalities	modality

	Modal Functions	func
	Morphosyntactic Environment	morph (morphosyntactic environment coded 7 different levels) ques (binary coded; yes-no question, others) morph.adv (binary coded; adverbial clauses expressing time, others) morph.rea (binary coded; reason clause, others) morph.if (binary coded; if clause, others) morph.neg (binary coded; negation, others) morph.but (binary coded; adversative, others)
	Subjectivity	subj
	Prior Context	prior
Nonlinguistic Factors and Factor Groups	Role Play Topics	rp.type.xy (warm-up and wind-down were coded as two different category) rp.type.x (warm-up and wind-down were coded as one category) part (warm-up, role play, wind-down were coded as three distinct categories only)
	Gender (Candidate)	gen.c
	Gender (Interviewer)	gen.i
	Gender (Candidate-Interviewer Agreement)	gen.a
	Role Play Levels	rp.level.cont (warm-up and wind-down were coded as level2) rp.level.follow (warm-up was coded as level 2 and wind-down was coded as the level of the preceding conversation)
	Probing Action	prob
	Probing Success	succ
Paralinguistic Factors and Factor Groups	Hesitation (General)	hes.g
	Hesitation (Specific)	hes.s
	Repeat (General)	repeat.g
	Repeat (Specific)	repeat.s

Repair (General)	rep.g
Repair (Specific)	rep.s
Pause (General)	pause.g
Pause (Specific)	pause.s
Filler (General)	fill.g
Filler (Specific)	fill.s
Modification (General)	mod.g
Modification (Specific)	mod.s
Speech Processing Time	count

References

Abe, M. (2003). A corpus-based contrastive analysis of spoken and written learner corpora: The case of Japanese-speaking learners of English. *Proceedings of the 2003 Corpus Linguistics Conference, Lancaster*, 1-9.

Adamson, H. D., & Regan, V. M. (1991). The acquisition of community speech norms by Asian immigrants learning English as a second language: A preliminary study. *Studies in Second Language Acquisition, 13*(1), 1-22.

Adamson, H. D., Fonseca-Greber, B., Kataoka, K., Scardino, V., & Takano, S. (1996). Tense marking in the English of Spanish-speaking adolescents. *Second Language Acquisition and Linguistic Variation, 10*, 121-134.

Adamson, H. D. (2009). *Interlanguage Variation in Theoretical and Pedagogical Perspective*. New York: Routledge.

Aijmer, K. (2002). Modality in advanced Swedish learners' written interlanguage. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* (pp. 55-76). Amsterdam: John Benjamins Publishing Company

Al-Khresheh, M. H. (2015). A review study of interlanguage theory. *International Journal of Applied Linguistics and English Literature, 4*(3), 123-131.

Allan, K. (1980). Nouns and countability. *Language, 56*(3), 541-567.

Allerton, D. J. (1975). Deletion and proform reduction. *Journal of linguistics, 11*(2), 213-237.

Altman, R. (1990). Giving and taking advice without offense. In R. C. Scarcella, E. S.

Andersen, & S. D. Krashen (Eds.), *Developing Communicative Competence in a Second Language* (pp. 95-106). New York : Newbury House Publishers

Andersen, R. W. (1978). An implicational model for second language research. *Language Learning*, *28*(2), 221-282.

Anderson, R. C., Reynolds, R. E., Schallert, D. L., & Goetz, E. T. (1977). Frameworks for comprehending discourse. *American Educational Research Journal*, *14*(4), 367-381.

Baayen, H. (2008) *Analyzing Linguistic Data: A Practical Introduction to Statistics*. Cambridge: Cambridge University Press.

Back, J., & Lee, E. (2017). Perception and production of English modal verbs by Korean university students in written academic discourse. *The Journal of Linguistic Science, 81,* 133-159.

Barrett, N. E., & Chen, L. M. (2011). English article errors in Taiwanese college students' EFL writing. *International Journal of Computational Linguistics & Chinese Language Processing, Volume 16, Number 3-4, September/December 2011*.

Basham, C., & Kwachka, P. (1989). Variation in modal use by Alaskan Eskimo student writers. *Variation in Second Language Acquisition, 1,* 129-143.

Bayley, R. (1994). Interlanguage variation and the quantitative paradigm: Past tense marking in Chinese-English. *Research Methodology in Second Language Acquisition*, 157-181.

Bayley, R. (1996). Competing constraints on variation in the speech of adult Chinese learners of English. *Second Language Acquisition and Linguistic*

Variation, 97-120.

Bayley, R. (1999). The primacy of aspect hypothesis revisited: Evidence from language shift. *Southwest Journal of Linguistics*, *18*(2), 1-22.

Bayley, R. (2002). The quantitative paradigm. In J. K. Chambers & N. Schilling (Eds), *The Handbook of Language Variation and Change* (pp. 117-141). New Jersey: Wiley-Blackwell

Bayley, R., & Langman, J. (2004). Variation in the group and the individual: Evidence from second language acquisition. *International Review of Applied Linguistics in Language Teaching (IRAL)*, *42*(4), 303-318.

Bayley, R., & Preston, D. R. (Eds.). (1996). *Second Language Acquisition and Linguistic Variation* (Vol. 10). Amsterdam: John Benjamins Publishing.

Beebe, L. M. (1980). Sociolinguistic variation and style shifting in second language acquisition 1. *Language Learning*, *30*(2), 433-445.

Beebe, L., & Zuengler, J. (1983). Accommodation theory: An explanation for style shifting in second language dialects. In N. Wolfson & E. Judd (Eds), *Sociolinguistics and Language Acquisition* (pp. 195-213). Rowley, MA: Newbury House

Berlin, L., & Adamson, H. D. (2009). The acquisition of English irregular past tense by Chinese-speaking children. In A. D. Adamson (Ed.), *Interlanguage Variation in Theoretical and Pedagogical Perspective* (pp. 61-74). New York: Routledge.

Biber, D., Johansson, S., Leech, G. N., Conrad, S., & Finegan, E. (1999). *Grammar of Spoken and Written English*. London: Pearson Education.

Bickerton, D. (1973). The nature of a creole continuum. *Language*, 640-669.

Bitchener, J., Young, S., & Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing, 14*(3), 191-205.

Blondeau, H., & Nagy, N. (2003). A comparison of gender non-marking in Montreal L1 & L2 French. *Conference on New Ways of Analyzing Variation (Vol. 32).*

Boncea, I. J. (2013). Hedging patterns used as mitigation and politeness strategies. *Annals of the University of Craiova*, *2*, 2-25.

Bond, F., & Ikehara, S. (1996, August). When and how to disambiguate? Countability in machine translation. *International Seminar on Multimodal Interactive Disambiguation: MIDDIM-96 Seminar*, 29-40.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5-32.

Brown, A. (2003). Interviewer variation and the co-construction of speaking proficiency. *Language Testing*, *20*(1), 1-25.

Brown, R. (1973). *A First Language: The Early Stages.* Cambridge, MA: Harvard University Press.

Buckingham, A. (1997). *Oral language testing: Do the age, status and gender of the interlocutor make a difference.* [Unpublished MA dissertation]. University of Reading.

Butler, Y. G. (2002). Second language learners' theories on the use of English articles: An Analysis of the Metalinguistic Knowledge Used by Japanese Students in Acquiring the English Article System. *Studies in Second Language Acquisition*, *24*(3), 451-480.

Bybee, J. L., Perkins, R. D., & Pagliuca, W. (1994). *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World* (Vol. 196). Chicago: University of Chicago Press.

Byram, M., & Feng, A. (2006). *Living and Studying Abroad*. Blue Ridge Summit: Multilingual Matters.

Cadierno, T. (2008). Learning to talk about motion in a foreign language. In P. Robinson & N. C. Ellis (Eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition* (pp. 239-275). New York: Routledge.

Camacho, J. A. (2013). *Null Subjects*. Cambridge: Cambridge University Press.

Cancino, H., Rosansky, E., & Schumann, J. (1978). The acquisition of English negatives and interrogatives by native Spanish speakers. In E. Marcussen Hatch (Ed.), *Second Language Acquisition: A book of Readings* (pp. 207-230). New York: Newbury House Pub.

Cappelle, B. (2005). *Particle patterns in English: A comprehensive coverage.* [Doctoral dissertation]. Université Catholique de Louvain.

Cedergren, H. J., & Sankoff, D. (1974). Variable rules: Performance as a statistical reflection of competence. *Language,* 333-355.

Chafe, W. (1980). Some reasons for hesitating. *Temporal Variables in Speech*, 169-180.

Chafe, W., & Danielwicz, J. (1987). Properties of spoken and written language. *Technical Report No. 5.*

Chambers, J. K. (2002). Studying language variation: An informal epistemology. *The Handbook of Language Variation and Change*, *11*, 3.

Chambers, J. K. (2009). *Sociolinguistic Theory: Linguistic Variation and Its Social Significance* (Language in Society, 22). Malden, MA: Blackwell.

Chapman, M. (2003). The role of the TOEIC in a major Japanese company. *Proceedings of the 2nd annual JALT Pan-SIG conference* (Vol. 79). Retrieved from: https://hosted.jalt.org/pansig/2003/HTML/Chapman.htm

Chapman, M. (2004). An interview with Kazuhiko Saito. *Shiken: JALT Testing & Evaluation SIG Newsletter, 8*(2) 10-12. Retrieved from: http://jalt.org/test/sai_cha.htm

Chelli, S. (2013). Interlingual or intralingual errors in the use of prepositions and articles. *Communications Internationales*. Retrieved from: http://archives.univ-biskra.dz/handle/123456789/3571

Cheshire, J. (2005). Syntactic variation and beyond: Gender and social class variation in the use of discourse-new markers 1. *Journal of Sociolinguistics*, *9*(4), 479-508.

Childs, M. (1995). Good and bad uses of TOEIC by Japanese companies. *1995*, 66-75.

Chodorow, M., Gamon, M., & Tetreault, J. (2010). The utility of article and preposition error correction systems for English language learners: Feedback and assessment. *Language Testing*, *27*(3), 419-436.

Choi, S. (1995). The development of epistemic sentence-ending modal forms and functions in Korean children. In J. L. Bybee & S. Fleischman (Eds.), *Modality in Grammar and Discourse* (pp. 165-204). Amsterdam: John Benjamins.

Coates, J. (1983). *The Semantics of the Modal Auxiliaries*. New York: Routledge.

Cook, W. A. (1978). Semantic structure of the English modals. *TESOL Quarterly, 12*(1), 5-16.

Cooper, T. C. (1999). Processing of idioms by L2 learners of English. *TESOL quarterly*, *33*(2), 233-262.

Corder, S. P. (1981). *Error Analysis and Interlanguage*. Oxford University Press.

Cote, S. A. (1996). *Grammatical and discourse properties of null arguments in English*. [Doctoral Dissertation]. University of Pennsylvania.

Cunningham, C. (2002). *The TOEIC test and communicative competence: Do test score gains correlate with increased competence.* [Unpublished master's thesis]. University of Birmingham, Birmingham, United Kingdom.

Davies, R. J., & Ikeno, O. (Eds.). (2002). *The Japanese Mind: Understanding Contemporary Culture*. North Clarendon, VT: Tuttle Publishing.

De Bot, K., Lowie, W., & Verspoor, M. (2007). A dynamic systems theory approach to second language acquisition. *Bilingualism: Language and Cognition*, *10*(1), 7-21.

DeCarrico, J. S. (1986). Tense, aspect, and time in the English modality system. *TESOL Quarterly, 20*(4), 665-682.

DeKeyser, R. M. (2005). What makes learning second-language grammar difficult? A review of issues. *Language Learning*, *55*(S1), 1-25.

Derwing, T. M., & Munro, M. J. (2013). The development of L2 oral language skills in two L1 groups: A 7-year study. *Language Learning*, *63*(2), 163-185.

Derwing, T. M., Thomson, R. I., & Munro, M. J. (2006). English pronunciation and fluency development in Mandarin and Slavic speakers. *System*, *34*(2), 183-

193.

Dewaele, J. M. (1992). L'omission du ne dans deux styles oraux d'interlangue française. *Interface. Journal of Applied Linguistics*, *7*(1), 3-17.

Dewaele, J. M., & Regan, V. (2001). The use of colloquial words in advanced French interlanguage. *EUROSLA yearbook*, *1*(1), 51-67.

Dewaele, J. M., & Regan, V. (2002). Maîtriser la norme sociolinguistique en interlangue française: le cas de l'omission variable de 'ne'. *Journal of French Language Studies*, *12*(2), 123-148.

Dickerson, L. (1975). Internal and External Patterning of Phonological Variability in the Speech of Japanese Learners of English: Toward a Theory of Second-Language Acquisition. *Dissertation abstracts international. A. The humanities and social sciences*, *35*(11), 7285-7286.

Dickerson, L. J. H. (1974). *Internal and external patterning of phonological variability in the speech of Japanese learners of English: Toward a theory of second-language acquisition*. [Dissertation]. University of Illinois at Urbana-Champaign.

Dickerson, L., & Dickerson, W. (1977). Interlanguage phonology: Current research and future directions. *The Notions of Simplification, Interlanguages, and Pidgins and Their Relation to Second Language Pedagogy, Actes du 5 une colloque de Linguistique Appliqueede Neufchatel. AIMAV/Didier, Paris*.

Downes, W. (1984). *Language and Society.* London: Fontana.

Drummond, R. (2012). Aspects of identity in a second language: ING variation in the speech of Polish migrants living in Manchester, UK. *Language Variation and*

Change, *24*(1), 107-133.

Du Bois, J. W. (1991). Transcription design principles for spoken discourse research. *Pragmatics*, *1*(1), 71-106.

Du Bois, J. W., Cumming, S., Schuetze-Coburn, S., & Paolino, D. (1992). *Santa Barbara Papers in Linguistics, Volume 4: Discourse Transcription.* Santa Barbara: University of Santa Barbara.

Dulay, H., & Burt, M. (1975). A new approach to discovering universal strategies of child second language acquisition. *Georgetown University Round Table on Languages and Linguistics*, 209-233.

Dunn-Rankin, P., Knezek, G. A., Wallace, S. R., & Zhang, S. (2014). *Scaling Methods*. Hove, East Sussex, United Kingdom: Psychology Press.

Dutra, D. P. (1998). *The acquisition of English root modality by non-native speakers*. [Doctoral Dissertation]. University of Florida.

Educational Testing Service. (1999). *TOEIC® Language Proficiency Interview Manual*. Princeton, NJ: Author.

Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, *18*(1), 91-126.

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, *24*(2), 143-188.

Ellis, N. C. (2006). Language acquisition as rational contingency learning. *Applied Linguistics*, *27*(1), 1-24.

Ellis, R. (1987). Interlanguage variability in narrative discourse: Style shifting in the

use of the past tense. *Studies in Second Language Acquisition*, *9*(1), 1-19.

Ellis, R., & Ellis, R. R. (1994). *The Study of Second Language Acquisition*. Oxford: Oxford University.

Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning*, *54*(2), 227-275.

Eskildsen, S. W. (2012). L2 negation constructions at work. *Language Learning*, *62*(2), 335-372.

Eskildsen, S. W. (2015). What counts as a developmental sequence? Exemplar-based L2 learning of English questions. *Language Learning*, *65*(1), 33-62.

Fauziati, E. (2011). Interlanguage and error fossilization: a study of Indonesian students learning English as a foreign. *Indonesian Journal of Applied Linguistics, 1*(1), 23-38.

Foulkes, P., Docherty, G. J., & Watt, D. (2005). Phonological variation in child-directed speech. *Language*, *81*(1), 177-206.

Freeman, D. E. L. (1975). The acquisition of grammatical morphemes by adult ESL students. *TESOL quarterly*, 409-419.

Gatbonton, E. (1978). Patterned phonetic variability in second-language speech: A gradual diffusion model. *Canadian Modern Language Review*, *34*(3), 335-347.

Gibbs, R. W. (1990). Psycholinguistic studies on the conceptual basis of idiomaticity. *Cognitive Linguistics, 1*(4).

Gibbs, D. A. (1990). Second language acquisition of the English modal auxiliaries can, could, may, and might. *Applied Linguistics, 11*(3), 297-314.

Glahn, E., Håkansson, G., Hammarberg, B., Holmen, A., Hvenekilde, A., & Lund, K. (2001). Processability in Scandinavian second language acquisition. *Studies in Second Language Acquisition*, *23*(3), 389-416.

Goldschneider, J. M., & DeKeyser, R. M. (2001). Explaining the "natural order of L2 morpheme acquisition" in English: A meta-analysis of multiple determinants. *Language Learning*, *51*(1), 1-50.

Gorden, R. L. (1977). *Unidimensional Scaling of Social Variables: Concepts and Procedures*. New York: Free Press.

Gries, S. T., & Ellis, N. C. (2015). Statistical measures for usage-based linguistics. *Language Learning*, *65*(S1), 228-255.

Guy, G. (1980). Variation in the group and the individual: The case of final stop deletion. In W. Labov (Ed.), *Locating Language in Time and Space* (pp. 1-36). New York: Academic Press.

Guy, G. R. (2005). *Language Variation and Linguistic Theory*. Oxford: Blackwell.

Hakuta, K. (1975). Learning to speak a second language: What exactly does the child learn. In D. Dato (Ed.), *Developmental Psycholinguistics: Theory and Applications* (pp. 193-207). Georgetown University Press.

Hakuta, K. (1976). A case study of a Japanese child learning English as a second language. *Language Learning*, *26*(2), 321-351.

Han, H., & Dawson, K. J. (2020, June 3). *JASP (Software)*. https://doi.org/10.31234/osf.io/67dcb

Han, N. R., Chodorow, M., & Leacock, C. (2004, May). Detecting errors in english article usage with a maximum entropy classifier trained on a large, diverse

corpus. *Proc. 4th International Conference on Language Resources and Evaluation, Lisbon, Portugal*, 1625-1628.

Han, N. R., Chodorow, M., & Leacock, C. (2006). Detecting errors in English article usage by non-native speakers. *Natural Language Engineering*, *12*(2), 115-129.

Hansen Edwards, J. G. (2008). Social factors and variation in production in L2 phonology. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and Second Language Acquisition* (pp. 251–279). Amsterdam: Benjamins.

Hanse Edwards, J. G. (2011). Deletion of /t, d/and the acquisition of linguistic variation by second language learners of English. *Language Learning*, *61*(4), 1256-1301.

Hansen, J. G. (2001). Linguistic constraints on the acquisition of English syllable codas by native speakers of Mandarin Chinese. *Applied Linguistics*, *22*(3), 338-365.

Hansen, J. G. (2006). *Acquiring a Non-native Phonology: Linguistic Constraints and Social Barriers*. London: Continuum International Publishing Group Ltd

Henderson, M. M. (1985). The interlanguage notion. *Journal of Modern Language Learning*, *21*, 23-27.

Herron, D. T., & Bates, E. A. (1997). Sentential and acoustic factors in the recognition of open-and closed-class words. *Journal of Memory and Language*, *37*(2), 217-239.

Herschensohn, J., & Young-Scholten, M. (2013). *The Cambridge Handbook of Second Language Acquisition.* Cambridge: Cambridge University Press.

Hiki, M. (1991). *A study of learners' judgments of noun countability*. [Doctoral dissertation]. Indiana University.

Hilles, S. L. (1989). *Access to Universal Grammar in second language acquisition*. [Doctoral dissertation]. University of California, Los Angeles.

Hinkel, E. (1995). The use of modal verbs as a reflection of cultural values. *TESOL quarterly*, *29*(2), 325-343.

Hinkel, E. (2009). The effects of essay topics on modal verb uses in L1 and L2 academic writing. *Journal of Pragmatics*, *41*(4), 667-683.

Hirakawa, M., & Suzuki, K. (2010). Learnability and modality restrictions on conditionals in L2 Japanese and English. *Proceedings of the Second language research Forum 2010*.

Hoffman, M. F., & Walker, J. A. (2010). Ethnolects and the city: Ethnic orientation and linguistic variation in Toronto English. *Language Variation and Change*, *22*(1), 37-67.

Horie, K. (1997). Form-meaning interaction in diachrony: A case study from Japanese. *English Linguistics*, *14*, 428-449.

Hothorn, T., & Zeileis, A. (2015). partykit: A modular toolkit for recursive partytioning in R. *The Journal of Machine Learning Research*, *16*(1), 3905-3909.

Hothorn, T., Bühlmann, P., Dudoit, S., Molinaro, A., & Van Der Laan, M. J. (2006). Survival ensembles. *Biostatistics*, *7*(3), 355-373.

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical*

statistics, *15*(3), 651-674.

Hothorn, T., Hornik, K., Strobl, C., & Zeileis, A. (2015). *party: A Laboratory for Recursive Partytioning*. URL: http://CRAN.R-project.org/package=party. R package version 1.0-20.

Howard, M. (2002). Prototypical and non-prototypical marking in the advanced learner's aspectuo-temporal system. *Eurosla Yearbook*, *2*(1), 87-113.

Howard, M. (2004). On the interactional effect of linguistic constraints on interlanguage variation: The case of past time marking. *IRAL: International Review of Applied Linguistics in Language Teaching*, *42*(4), 319-344.

Howard, M., Lemée, I., & Regan, V. (2006). The L2 acquisition of a phonological variable: The case of /l/ deletion in French. *Journal of French Language Studies*, *16*(1), 1-24.

Hu, C., & Li, X. (2015). Epistemic modality in the argumentative essays of Chinese EFL learners. *English Language Teaching*, *8*(6), 20-31.

Huang, C.T.J. (1989). Pro-drop in Chinese: a generalized control theory. In O. A. Jaeggli & K. J. Safir (Eds.), *The Null Subject Parameter. Studies in Natural Language and Linguistic Theory, vol 15* (pp. 185-214). Dordrecht/Boston/London: Kluwer Academic Publishers.

Huang, C. T. J. (1984). On the distribution and reference of empty pronouns. *Linguistic Inquiry*, 531-574.

Huebner, T. (1985). System and variability in interlanguage syntax. *Language Learning*, *35*(2), 141-163.

Hurtado, L. M., & Estrada, C. (2010). Factors influencing the second language

acquisition of Spanish vibrants. *The Modern Language Journal, 94*(1), 74-86.

Hyltenstam, K. (1977). Implicational patterns in interlanguage syntax variation. *Language Learning, 27*(2), 383-410.

Iwasaki, N. (2009). Stating and supporting opinions in an interview: L1 and L2 Japanese speakers. *Foreign Language Annals, 42*(3), 541-556.

Izumi, E., & Isahara, H. (2004, December). Investigation into language learners' acquisition order based on an error analysis of a learner corpus. *IWLeL 2004: an Interactive Workshop on Language e-learning*, 63-71.

Izumi, E., Uchimoto, K., Saiga, T., Supnithi, T., & Isahara, H. (2003, July). Automatic error detection in the Japanese learners' English spoken data. *The Companion Volume to the Proceedings of 41st Annual Meeting of the Association for Computational Linguistics*, 145-148.

James, C. (2013). *Errors in Language Learning and Use: Exploring Error Analysis*. New York: Routledge.

Jenkins, S., & Parra, I. (2003). Multiple layers of meaning in an oral proficiency test: The complementary roles of nonverbal, paralinguistic, and verbal behaviors in assessment decisions. *The Modern Language Journal, 87*(1), 90-107.

Jia, L., & Bayley, R. (2008). The (re) acquisition of perfective aspect marking by Chinese heritage language learners. In A. W. He & Y. Xiao (Eds.), *Chinese as a Heritage Language: Fostering Rooted World Citizenry* (pp. 205-224). Honolulu, HI: National Foreign Language Resource Center.

Johnson, D. E. (2009). Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass, 3*(1),

359-383.

Johnson, D. E. (2010). *Rbrul manual*. http://www.danielezrajohnson.com/Rbrul_manual.html

Johnson, D. E. (2017). Progress in regression: Statistical and practical improvements to Rbrul, *Paper Presented at NWAV 46 (2-5 November), Madison (Wisconsin).*

Jordan, G. (2004). *Theory Construction in Second Language Acquisition* (Vol. 8). Amsterdam: John Benjamins Publishing.

Kasper, G. (1979). Communication strategies: Modality reduction. *Interlanguage Studies Bulletin*, 266-283.

Kaufmann, M., & Tamura, S. (2017). Japanese modality-possibility and necessity: Prioritizing, epistemic, and dynamic. *The Handbook of Japanese Semantics and Pragmatics*, 1-33.

Kennedy, G. (2002). Variation in the distribution of modal verbs in the British National Corpus. *Using Corpora to Explore Linguistic Variation*, 73-90.

Khor, S. Y. (2013). A Corpus Based Study in Morpheme Acquisition Order of Young Learners of English: A comparison of Swedish students in grade 6 and grade 7. [Dissertation]. Uppsala University. Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-195862

Koeneman, O. N. C. J., & Zeijlstra, H. (2019). Morphology and pro drop. In M. Aronoff (Ed.), *Oxford Research Encyclopedias of Linguistics* (pp. 1-37). Oxford: Oxford University Press.

Kovac, C., & Adamson, H. D. (1981). Variation theory and first language

acquisition. *Variation Omnibus*, 403-410.

Krashen, S. (1977). Some issues relating to the monitor model. *On TESOL*, *77*, 144-158.

Krashen, S. D. (1982). Principles and practice in second language acquisition. *Learning*, *46*(2), 327-69.

Kurotaki, M. (2005). *Deontic kara Epistemic e no Fuhensei to Sootaisei: Modariti no Nichieigo Taishoo Kenkyuu* [Universality and Relativity from Deontic to Epistemic Modality: Contrastive Studies of English Modality and Japanese Modality]. Tokyo: Kurosio.

Kwachka, P., & Basham, C. (1990). Literacy acts and cultural artifacts: on extensions of English modals. *Journal of Pragmatics*, *14*(3), 413-429.

Kwon, E. Y. (2005). The "natural order" of morpheme acquisition: a historical survey and discussion of three putative determinants. *Studies in Applied Linguistics and TESOL*, *5*(1), 1-21.

Labov, W. (1963). The social motivation of a sound change. *Word*, *19*(3), 273-309.

Labov, W. (1964). *The social stratification of English in New York City*, [Doctoral dissertation]. Columbia University.

Labov, W. (1966). *The Social Stratification of English in New York City*. Washington D.C.: Center for Applied Linguistics.

Labov, W. (1969). Contraction, deletion, and inherent variability of the English copula. *Language*, 715-762.

Labov, W. (1972). *Sociolinguistic Patterns* (No. 4). University of Pennsylvania Press.

Labov, W. (2001). *Principles of Language Change: Social Factors*. Malden, MA:

Blackwell.

Labov, W., Ash, S., Baranowski, M., Nagy, N., Ravindranath, M., & Weldon, T. (2006). Listeners' sensitivity to the frequency of sociolinguistic variables. *University of Pennsylvania Working Papers in Linguistics*, *12*(2), 10.

Lakshmanan, U. (1991). Morphological uniformity and null subjects in child second language acquisition. In L. Eubank (Ed.), *Point Counterpoint: Universal Grammar in The Second Language* (pp. 389-410). Amsterdam: John Benjamins.

Lardiere, D. (2013). Nativelike and non-nativelike attainment. In J. Herschensohn & M. Young-Scholten (Eds.), *The Cambridge Handbook for Second Language Acquisition* (pp. 670-691). Cambridge University Press.

Larm, L. (2009). West meets East: a Kindaichian approach to subjective modality. *Japanese Modality*, 56-86.

Larm, L. (2014). Modality packaging in Japanese: the encoding of modal meanings and subjectivity. *Studies in Pragmatics*, *16*, 20-46.

Larm, L. (2006). *Modality in Japanese.* [DPhil thesis]. University of Oxford.

Langman, J., & Bayley, R. (2002). The acquisition of verbal morphology by Chinese learners of Hungarian. *Language Variation and Change, 14*(1), 55-77.

Larsen-Freeman, D. E. (1976). An explanation for the morpheme acquisition order of second language learners. *Language Learning*, *26*(1), 125-134.

Leacock, C., Chodorow, M., Gamon, M., & Tetreault, J. (2010). Automated grammatical error detection for language learners. *Synthesis Lectures on*

Human Language Technologies, *3*(1), 1-134.

Lee Amuzie, G., & Spinner, P. (2013). Korean EFL learners' indefinite article use with four types of abstract nouns. *Applied Linguistics*, *34*(4), 415-434.

Leech, G., & Svartvik, J. (2013). *A Communicative Grammar of English*. New York: Routledge.

Lehrer, A. (1970). Notes on lexical gaps. *Journal of Linguistics*, *6*(2), 257-261.

Leli, L. P. (2017) *An error analysis in using modal auxiliary verbs of the eight grade students at SMP Muhammadiyah 05 Medan*. [S.Pd Thesis]. University of Muhammadiyah Sumatera Utara Medan.

Lemée, I. (2002). Acquisition de la variation socio-stylistique dans l'interlangue d'apprenants hibernophones de français L2: le cas de on et nous. *Marges linguistiques*, *4*, 56-67.

Li, X. (2014). Variation in subject pronominal expression in L2 Chinese. *Studies in Second Language Acquisition*, *36*(1), 39-68.

Liceras, J. (1988). Syntax and stylistics: more on the pro-drop parameter. *Learnability and Second Languages*, 71-93.

Lightfoot, A. R. (1998). *Japanese second language learner and English article system: a study in error analysis.* Retrieved from: http://ardle.net/linguistics.html.

Lowie, W. M. (2013, May). L2 phonological development: A plea for a dynamic, process-based methodology. *Presentation at the International Symposium on the Acquisition of Second Language Speech (New Sounds 2013)*.

Lowie, W., & Verspoor, M. (2011). The dynamics of multilingualism: Levelt's

speaking model revisited. In M. S. Schmid & W. Lowie (Eds.), *Modeling bilingualism: from structure to chaos: in honor of Kees de Bot* (pp. 267-287). Amsterdam: John Benjamins Publishing Company.

Lowie, W., & Verspoor, M. (2015). Variability and variation in second language acquisition orders: A dynamic reevaluation. *Language Learning, 65*(1), 63-88.

Luk, Z. P. S., & Shirai, Y. (2009). Is the acquisition order of grammatical morphemes impervious to L1 knowledge? Evidence from the acquisition of plural-s, articles, and possessive's. *Language Learning, 59*(4), 721-754.

Lumley, T., & O'Sullivan, B. (2005). The effect of test-taker gender, audience and topic on task performance in tape-mediated assessment of speaking. *Language Testing, 22*(4), 415-437.

MacGregor, L. (2000). An introduction to the TOEIC® LPI test. *JALT Testing & Evaluation SIG Newsletter, 4*(2), 2-5. Retrieved from: https://hosted.jalt.org/test/mac_3.htm

Major, R. C. (2004). Gender and stylistic variation in second language phonology. *Language Variation and Change, 16*(3), 169-188.

Markham, P. L. (1988). Gender and the perceived expertness of the speaker as factors in ESL listening recall. *TESOL Quarterly, 22*(3), 397-406.

Masuoka, T. (2009). Modality from a Japanese perspective. *Japanese Modality*, 36-55.

McCrostie, J. (2010). The TOEIC in Japan: A scandal made in heaven. *JALT Testing and Evaluation Sig Newsletter, 14*(1), 2-10.

McDouall, A. (2012). A corpus based investigation into the use of English modal

auxiliaries by adult korean L2-learners. *KLING, 6*, 33-44.

McEnery, T., & Kifle, N. A. (2002). Epistemic modality in argumentative essays of second-language writers. *Academic Discourse*, 182-195.

McFerren, K. A. (2015). *Order of acquisition: a comparison of L1 and L2 English and Spanish morpheme acquisition.* [Thesis]. Liberty University.

Meisel, J. (1983). Strategies of second language acquisition: More than one kind of simplification. In R. W. Andersen (Ed.), *Pidginization and Creolization as Language Acquisition* (pp. 120-157). Rowley, MA: Newbury House Publishers, Inc.

Meisel, J. M. (1983). Transfer as a second-language strategy. *Language & Communication*, *3*(1), 11-46.

Meyerhoff, M., & Schleef, E. (2012). Variation, contact and social indexicality in the acquisition of (ing) by teenage migrants. *Journal of Sociolinguistics*, *16*(3), 398-416.

Miyata, R. (2004). A Review of TOEIC: A critical analysis of the structure, methods, and purpose of TOEIC, including its effects on teaching and learning English as a second or foreign language. *Bullentin of Beppu University Junior College*, 59-69.

Molenaar, P. C. (2008). On the implications of the classical ergodic theorems: Analysis of developmental processes has to focus on intra-individual variation. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, *50*(1), 60-69.

Molenaar, P., Huizenga, H. M., & Nesselroade, J. R. (2003). The relationship

between the structure of interindividual and intraindividual variability: A theoretical and empirical vindication of developmental systems theory. In U.M. Staudinger & U. Lindenberger (Eds.), *Understanding Human Development* (pp. 339-360). MA: Kluwer Academic Publishers.

Monteiro, R. B., & Hansen, N. M. (2013). Aegisub. *Section*, *3*, 51-73.

Moriya, T., & Horie, K. (2009). What is and is not language-specific about the Japanese modal system? A comparative and historical perspective. *Japanese Modality*, 87-114.

Motallebzadeh, K., & Nematizadeh, S. (2011). Does gender play a role in the assessment of oral proficiency?. *English Language Teaching*, *4*(4), 165-172.

Mougeon, R., & Dewaele, J. M. (Eds). (2004). Preface. *In International Review of Applied Linguistics*, *42*(4), 295-301.

Mougeon, R., & Rehner, K. (2001). Acquisition of sociolinguistic variants by French immersion students: The case of restrictive expressions, and more. *The Modern Language Journal*, *85*(3), 398-415.

Mougeon, R., Nadasdi, T., & Rehner, K. (2002). État de la recherche sur l'appropriation de la variation par les apprenants avancés du FL2 ou FLE. *Acquisition et Interaction en Langue Étrangère*, (17), 7-50.

Mougeon, R., Nadasdi, T., & Rehner, K. (2010). *The Sociolinguistic Competence of Immersion Students*. . Blue Ridge Summit: Multilingual Matters.

Mougeon, R., Rehner, K., & Nadasdi, T. (2004). The learning of spoken French variation by immersion students from Toronto, Canada. *Journal of Sociolinguistics*, *8*(3), 408-432.

Murakami, A. (2014). *Individual variation and the role of L1 in the L2 development of English grammatical morphemes: Insights from learner corpora*, [Doctoral dissertation]. University of Cambridge.

Murakami, A., & Alexopoulou, T. (2016). L1 influence on the acquisition order of English grammatical morphemes: A learner corpus study. *Studies in Second Language Acquisition, 38*(3), 365-401.

Murata, M., & Nagao, M. (1994). Determination of Referential Property and Number of Nouns in Japanese Sentences for Machine Translation into English. *Proceedings of the Fifth Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*. 218-225.

Nadasdi, T., Mougeon, R., & Rehner, K. (2003). Emploi du 'futur'dans le français parlé des élèves d'immersion française. *Journal of French Language Studies, 13*(2), 195-219.

Nagy, N., Blondeau, H., & Auger, J. (2003). Second language acquisition and "real" French: An investigation of subject doubling in the French of Montreal Anglophones. *Language Variation and Change*, *15*(1), 73-103.

Nagy, N., Moisset, C., & Sankoff, G. (1996). On the acquisition of variable phonology in L2. *University of Pennsylvania Working Papers in Linguistics*, *3*(1), 9.

Nakamura, M. (1987). Japanese as a pro language. *The Linguistic Review, 6*(4), 281-296. https://doi.org/10.1515/tlir.1987.6.4.281

Nakayama, S. (2021). Modal auxiliary verbs in Japanese EFL learners' conversation: a corpus-based study. *Asia Pacific Journal of Corpus Research, 2*(1), 23-34.

Nance, C., McLeod, W., O'Rourke, B., & Dunmore, S. (2016). Identity, accent aim, and motivation in second language users: New Scottish Gaelic speakers' use of phonetic variation. *Journal of Sociolinguistics, 20*(2), 164-191.

Neagu, M. (2007). English verb particles and their acquisition: A cognitive approach. *Revista Española de Lingüística Aplicada,* (20), 121-138.

Neeleman, A., & Szendrői, K. (2005). Pro drop and pronouns. *Proceedings of the 24th West Coast Conference on Formal Linguistics* (Vol. 24), 299-307.

Neeleman, A., & Szendrői, K. (2007). Radical pro drop and the morphology of pronouns. *Linguistic Inquiry, 38*(4), 671-714.

Neeleman, A., & Szendröi, K. (2008). Case morphology and radical pro-drop. In T. Biberauer (Ed.), *The Limits of Syntactic Variation* (Vol. 132) (pp. 331-348). Amsterdam: John Benjamins.

Nippold, M. A. (1998). *Later Language Development: The School-age and Adolescent Years*. Austin: Pro-Ed.

Nippold, M. A., & Taylor, C. L. (1995). Idiom understanding in youth: Further examination of familiarity and transparency. *Journal of Speech, Language, and Hearing Research, 38*(2), 426-433.

Noels, K. A. (2014). Language variation and ethnic identity: A social psychological perspective. *Language & Communication, 35*, 88-96.

Nomura, M. (2012). Variability in articles in Japanese EFL learners' spoken and written production: A learner corpus-based approach. *Corpus-based Language Education Research Report, 8*, 281-302.

Norton, B. (2000). *Identity and Language Learning: Gender, Ethnicity and*

Educational Change. London: Longman Publishing Group.

Norton, J. (2005). The paired format in the Cambridge Speaking Tests. *ELT Journal*, *59*(4), 287-297.

Nozawa, Y. (2010). *An analysis of the use of modal verbs in EFL textbooks in terms of politeness strategy of English*. Waseda University Repository.

O'Loughlin, K. (2002). The impact of gender in oral proficiency testing. *Language Testing*, *19*(2), 169-192.

Odlin, T. (1989). *Language Transfer* (Vol. 27). Cambridge: Cambridge University Press.

O'Grady, W., Yamashita, Y., & Cho, S. (2008). Object drop in Japanese and Korean. *Language Acquisition*, *15*(1), 58-68.

Ohara, M. (2007). Object drop in English and in Japanese. *Annals of Foreign Studies in Shimane University*, *67*, 69-80.

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, *17*(4), 273-281.

O'Sullivan, B. (2000). Exploring gender and oral proficiency interview performance. *System*, *28*(3), 373-386.

Othman, J. (2003). Variation in interlanguage: Theoretical perspectives and empirical evidence. In H. Abd-Rahim & S. Manan (Eds.), *Current Issues in Linguistics, Literature and Language Teaching* (pp. 139-191). Malaysia, Penang: USM.

Palmer, F. R. (1990). *Modality and the English Modals*. New York: Longman.

Palmer, F. R. (2001). *Mood and Modality*. Cambridge: Cambridge University Press.

Papafragou, A. (2006). Epistemic modality and truth conditions. *Lingua*, *116*(10),

1688-1702.

Park, H. (2004). A minimalist approach to null subjects and objects in second language acquisition. *Second Language Research*, *20*(1), 1-32.

Parrish, B. (1987). A new look at methodologies in the study of article acquisition for learners of ESL. *Language Learning*, *37*(3), 361-384.

Pavone, J. (1980). *Implicational scales and English dialectology*, [Ph.D. Dissertation]. Indiana University.

Pemberton, C. (2020). In my opinion: Modality in Japanese EFL learners' argumentative essays. *Asia Pacific Journal of Corpus Research*, *1*(2), 57-72.

Pertiwi, D. P., & Adityarini, H. (2019). *An error analysis on the use of modal auxiliary made by English education students at Muhammadiyah University of Surakarta*, [Doctoral dissertation]. Universitas Muhammadiyah Surakarta.

Phinney, M. (1987). The pro-drop parameter in second language acquisition. In T. Roeper & E. Williams (Eds.), *Parameter Setting* (pp. 221-238). Dordrecht: D. Reidel.

Pica, T. (1983). Adult acquisition of English as a second language under different conditions of exposure. *Language Learning*, *33*(4), 465-497.

Pienemann, M. (1998). *Language Processing and Second Language Development: Processability Theory* (Vol. 15). Amsterdam: John Benjamins Publishing.

Pienemann, M. & Mackey, A. (1993). An empirical study of children's ESL development. In National Languages and Literacy Institute of Australia (Ed.), *Development: Language and Literacy in Schools* (pp. 115-259).

Pintzuk, S. (1988). *VARBRUL programs.* Philadelphia: University of Pennsylvania

Department of Linguistics.

Poplack, S., & Tagliamonte, S. (2001). *African American English in the Diaspora*. New Jersey: Wiley-Blackwell.

Porter, D. (1991) Affective factors in the assessment of oral interaction: Gender and status. In S. Anivan, (Ed.), *Current Developments in Language Testing. Anthology Series 25* (pp. 92-102). Singapore: SEAMEO Regional Language Center.

Porter, D., & Shen, S. (1991). Sex, status and style in the interview. *The Dolphin*, *21*(2), 117-128.

Preston, D. R. (1989). *Sociolinguistics and Second Language Acquisition*. Oxford: Blackwell.

Quinn, K. (2000). A description and evaluation of the TOEIC LPI. *Journal of Nagoya Women's University. Humanities ・Social Science*, *46*, 225-234.

R Core Team (2018). R: A language and environment for statistical computing. *R Foundation for* Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/.

Rand, D., & Sankoff, D. (1990). *GoldVarb: A Variable Rule Application for the Macintosh, Version 2.* Montreal: Centre de recherches mathématiques, Université de Montréal.

Rand, D., & Sankoff, D. (1991). *GoldVarb: A Variable Rule Application for the Macintosh (Version 2.1)* [Computer Program]. Montreal: Centre de recherches mathématiques, Université de Montréal.

Reemann, E., Alas, E., & Liiv, S. (2013). Interviewer behaviour during oral

proficiency interviews: A gender perspective. *Eesti Rakenduslingvistika Ühingu aastaraamat, 9*, 209-226.

Regan, V. (1995). The acquisition of sociolinguistic native speech norms. In B. F. Freed (Ed.), *Second Language Acquisition in a Study Abroad Context* (pp. 245-267). Amsterdam: John Benjamins.

Regan, V. (1996). Variation in French interlanguage: A longitudinal study of sociolinguistic competence. In R. Bayley & D. Preston (Eds.), *Second Language Acquisition and Linguistic Variation* (pp. 177-202). Amsterdam: John Benjamins.

Regan, V. (1997). Les apprenants avancés, la lexicalisation et l'acquisition de la compétence sociolinguistique: une approche variationniste. *Acquisition et Interaction en Langue étrangère*, (9), 193-210.

Regan, V. (2004). The Relationship between the Group and the Individual and the Acquisition of Native Speaker Variation Patterns: A Preliminary Study. *International Review of Applied Linguistics in Language Teaching (IRAL), 42*(4), 335-348.

Regan, V. (2013). Variation. In J. Herschensohn & M. Young-Scholten (Eds.), *The Cambridge Handbook for Second Language Acquisition* (pp. 290-291). Cambridge University Press.

Regan, V., & Bayley, R. (2004). Introduction: The acquisition of sociolinguistic competence. *Journal of Sociolinguistics*, *8*(3), 323-338.

Regan, V., & Nestor, N. (2010). French Poles, language and identity. *Language Practices and Identity Construction by Multilingual Speakers of French L2:*

The Acquisition of Sociostylistic Variation, *80*, 145.

Regan, V., Howard, M. & Lemée, I. (2009). *The Acquisition of Sociolinguistic Competence in a Study Abroad Context*. Bristol, Blue Ridge Summit: Multilingual Matters. https://doi.org/10.21832/9781847691583

Rehner, K. A. (2003). *The development of aspects of linguistic and discourse competence by advanced second language learners of French*. [Ph.D. dissertation]. University of Toronto.

Rehner, K., Mougeon, R., & Nadasdi, T. (2003). The learning of sociolinguistic variation by advanced FSL learners: The case of nous versus on in immersion French. *Studies in Second Language Acquisition*, *25*(1), 127-156.

Richards, J. C. (1971). A non-contrastive approach to error analysis. *English Language Teaching, 25*(3), 204-219.

Richards, J. C., & Schmidt, R. W. (2013). *Longman Dictionary of Language Teaching and Applied Linguistics*. New York: Routledge.

Rickford, J. R. (2002). Implicational scales. In J. K. Chambers, P. Trudgill & N. Schilling (Eds.), *The Handbook of Language Variation and Change* (pp. 142-167). New Jersey: Wiley-Blackwell.

Robberecht, P., & Van Peteghem, M. (1982, June). A functional model for the description of modality. *Fifth International Conference on Contrastive Projects, Jyväskylä, Finland.*

Robinson, J., Lawrence, H., & Tagliamonte, S. (2001). *GoldVarb 2001: A Multivariate Analysis Application for Windows.* Nova York: University of York.

Rosansky, E. J. (1976). Methods and morphemes in second language acquisition research. *Language Learning*, *26*(2), 409-425.

Sankoff, D. (1978). *Linguistic Variation. Models and Methods.* New York: Academic Press.

Sankoff, D. (1988). Sociolinguistics and syntactic variation. *Linguistics: the Cambridge Survey*, *4*, 140-161.

Sankoff, D., & Labov, W. (1979). On the uses of variable rules. *Language in Society*, *8*(2-3), 189-222.

Sankoff, D., Tagliamonte, S., & Smith, E. (2005). *Goldvarb X: A variable rule application for Macintosh and Windows.* Department of Linguistics, University of Toronto.

Sato, C. J. (1984). Task variation in interlanguage phonology. *University of Hawai'i Working Papers in English as a Second Language 3(1)*, 21-50.

Sax, K. J. (2003). *Acquisition of stylistic variation in American learners of French*. [Published Dissertation]. Indiana University ProQuest Dissertations Publishing.

Schleef, E., Meyerhoff, M., & Clark, L. (2011). Teenagers' acquisition of variation: A comparison of locally-born and migrant teens' realisation of English (ing) in Edinburgh and London. *English World-wide*, *32*(2), 206-236.

Schmitt, N. (Ed.). (2004). *Formulaic Sequences: Acquisition, Processing, and Use* (Vol. 9). Amsterdam: John Benjamins Publishing.

Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the full transfer/full access model. *Second Language Research*, *12*(1), 40-72.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics in Language Teaching, 10*(1-4), 209-232. https://doi.org/10.1515/iral.1972.10.1-4.209

Selinker, L. (1974). Error analysis: Source, cause, and significance. In J. C. Richards (Ed.), *Error Analysis: Perspective on Second Language Acquisition* (pp. 189-215). New York: Routledge.

Selinker, L., & Douglas, D. (1985). Wrestling with 'context' in interlanguage theory. *Applied Linguistics, 6*(2), 190-204.

Sembiring, R. A., Sibarani, A. S. M., & Mangaraja, J. E. (2021). Error analysis of using auxiliary verbs in writing descriptive text. *Journal of English Education and Teaching*, *5*(4), 556-566.

Shirato, J., & Stapleton, P. (2007). Comparing English vocabulary in a spoken learner corpus with a native speaker corpus: Pedagogical implications arising from an empirical study in Japan. *Language Teaching Research*, *11*(4), 393-412.

Slobin, D. I. (1997). Mind, code, and text. In J. L. Bybee, J. Haiman, & S. A. Thompson (Eds.), *Essays on Language Function and Language Type: Dedicated to T. Givón* (pp. 437-467). Amsterdam: John Benjamins.

Stepahny, U. (1995). Function and form of modality in first and second language acquisition. In A. G. Ramat & G. C. Galèas (Eds.), *From Pragmatics to Syntax: Modality in Second Language Acquisition* (Vol 405) (pp. 105-120). Tübingen: Gunter narr verlag.

Strobl, C., Boulesteix, A. L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC*

Bioinformatics, *9*(1), 1-11.

Strobl, C., Boulesteix, A. L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, *8*(1), 1-21.

Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods*, *14*(4), 323.

Stumpf, H., & Stanley, J. C. (1998). Stability and change in gender-related differences on the College Board Advanced Placement and Achievement Tests. *Current Directions in Psychological Science*, *7*(6), 192-196.

Suherman, R., Indrayani, L. M., & Krisnawati, E. (2020). Portraying the English morphological development of Indonesian-English interlanguage learners. *Indonesian Journal of English Language Teaching and Applied Linguistics*, *5*(1), 83-96.

Sweetser, E. (1990). *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure* (Vol. 54). Cambridge: Cambridge University Press.

Tagliamonte, S. A. (2006). *Analysing Sociolinguistic Variation*. Cambridge: Cambridge University Press.

Tagliamonte, S. A. (2011). *Variationist Sociolinguistics: Change, Observation, Interpretation*. Hoboken, New Jersey: John Wiley & Sons.

Tagliamonte, S. A., & Baayen, R. H. (2012). Models, forests, and trees of York English: was/were variation as a case study for statistical practice. *Language*

Variation and Change, *24*(2), 135-178.

Tajika, H. (1999). *Variable patterns of tense/aspect marking in interlanguage*. [Unpublished Doctoral dissertation]. The University of Minnesota, Minnesota.

Takayama, Y. (2002). *Nihongo Modaritii no Shiteki Kenkyuu*. [Historical Research on Japanese Modality]. Tokyo: Hitsuji Shoboo.

Talmy, L. (1991, July). Path to realization: a typology of event conflation. *Annual Meeting of the Berkeley Linguistics Society* (Vol. 17, No. 1), 480-519.

Talmy, L. (1985). Lexicalization pattern: Semantic structure in lexical form. In T. Shopen (Ed.), *Language Typology and Syntactic Description* (pp. 36-149). Cambridge University Press.

Tarone, E. (1979). Interlanguage as chameleon. *Language Learning*, *29*(1), 181-191.

Tarone, E. (1983). On the variability of interlanguage systems. *Applied Linguistics*, *4*(2), 142-164.

Tarone, E. (1988). *Variation in Interlanguage*. London: Hodder Arnold.

Tarone, E. (1990). On variation in interlanguage: A response to Gregg. *Applied Linguistics*, *11*(4), 392-400.

Tarone, E. (1994). Interlanguage. In R.E. Asher (Ed.), *The Encyclopedia of Language and Linguistics 4* (pp. 1715-1719). Amsterdam: Elsevier Science.

Tarone, E. (2001). Interlanguage. In R. Mesthrie (Ed.), *Concise Encyclopedia of Sociolinguistics* (pp. 475-481). Oxford: Pergamon.

Tarone, E. (2012). Interlanguage. In C. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (pp. 1-7). Chichester, West Sussex, UK: Wiley-Blackwell,

Tarone, E. E. (1982). Systematicity and attention in interlanguage. *Language*

Learning, *32*(1), 69-84.

Tarone, E. E. (1985). Variability in interlanguage use: A study of style-shifting in morphology and syntax. *Language Learning*, *35*(3), 373-403.

Tarone, E., & Parrish, B. (1988). Task-related variation in interlanguage: The case of articles. *Language Learning*, *38*(1), 21-44.

Templer, B. (2004). High-stakes testing at high fees: Notes and queries on the international English proficiency assessment market. *Journal for Critical Education Policy Studies, 2*(1), 1-8.

Tetreault, J., & Chodorow, M. (2008, August). The ups and downs of preposition error detection in ESL writing. *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, 865-872.

Thomas, M. (1989). The acquisition of English articles by first-and second-language learners. *Applied Psycholinguistics*, *10*(3), 335-355.

Ting, S. H., Mahadhir, M., & Chang, S. L. (2010). Grammatical errors in spoken English of university students in oral communication course. *GEMA Online Journal of Language Studies*, *10*(1).

Tomasello, M. (2005). *Constructing a Language: A Usage-based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.

Tomioka, S. (2003). The semantics of Japanese null pronouns and its cross-linguistic implications. *The Interfaces: Deriving and Interpreting Omitted Structures*, *61*, 321.

Tono, Y., & Aoki, M. (2000). A computer learner corpus-based analysis of the acquisition of irregular verbs: optimizing the learning list and monitoring the

learning process. *The Bulletin of the Kanto-Koshin-etsu English Language Education Society*, *14*, 103-122.

Trofimovich, P., Gatbonton, E., & Segalowitz, N. (2007). A dynamic look at L2 phonological learning: Seeking processing explanations for implicational phenomena. *Studies in Second Language Acquisition*, *29*(3), 407-448.

Ullenius, S. (2015). *L2 and L1 repairs: Speech production in a comparative perspective.* [BA Dissertation]. Stockholm University. Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-126086

Uritescu, D., Mougeon, R., Rehner, K., & Nadasdi, T. (2004). Acquisition of the internal and external constraints of variable Schwa deletion by French immersion students. *International Review of Applied Linguistics in Language Teaching (IRAL)*, *42*(4), 349-364.

van Geert, P. (2014, March). L2 acquisition and the study of change in complex systems. In K. de Bot & D. Larsen-Freeman (Chair), *From Universality to Variability in Second Language Development. Colloquium Conducted at the Meeting of the American Association for Applied Linguistics*.

Weinreich, U., Labov, W., & Herzog, M. (1968). *Empirical Foundations for a Theory of Language Change*. University of Texas Press.

Weitze, M., McGhee, J., Graham, C. R., Dewey, D. P., & Eggett, D. L. (2011). Variability in L2 acquisition across L1 backgrounds. *Selected Proceedings of the 2009 Second Language Research Forum*, 152-163.

White, L. (1985). The "pro-drop" parameter in adult second language acquisition. *Language Learning*, *35*(1), 47-61.

White, L. (2003). *Second Language Acquisition and Universal Grammar*. Cambridge: Cambridge University Press.

White, L. (2016). Pro-drop then and now: Changing perspectives on null subjects in second language acquisition. In A. A. de la Fuente, E. Valenzuela, & C. M. Sanz (Eds.), *Language Acquisition Beyond Parameters: Studies in Honour of Juana M. Liceras* (pp. 17-35). Amsterdam: John Benjamins.

Williams, L. (2009). Sociolinguistic variation in French computer-mediated communication: A variable rule analysis of the negative particle ne. *International Journal of Corpus Linguistics*, *14*(4), 467-491.

Wolfram, W. (1985). Variability in tense marking: A case for the obvious. *Language Learning*, *35*(2), 229-253.

Wolfram, W., & Beckett, D. (2000). The role of the individual and group in earlier African American English. *American Speech*, *75*(1), 3-33.

Woodford, P. E. (1982). An introduction to TOEIC: The initial validity study. *TOEIC Research Summary*.

Yasuda, S. (2010). Learning phrasal verbs through conceptual metaphors: A case of Japanese EFL learners. *TESOL Quarterly*, *44*(2), 250-273.

Young, R. (1986). The acquisition of a verbal repertoire in a second language. *Working Papers in Educational Linguistics (WPEL)*, *2*(1), 4.

Young, R. (1988). Variation and the interlanguage hypothesis. *Studies in Second Language Acquisition*, *10*(3), 281-302.

Young, R. (1991). *Variation in Interlanguage Morphology* (Vol. 1). New York: Peter Lang Pub Incorporated.

Young, R. (1996). Form-function relations in articles in English interlanguage. *Second Language Acquisition and Linguistic Variation, 10*, 135.

Young, R. F. (1989). *Variation in Interlanguage Morphology:(s) Plural-marking in the Speech of Chinese Learners of English* [Doctoral dissertation]. University of Pennsylvania.

Young, R., & Bayley, R. (1996). VARBRUL analysis for second language acquisition research. In R. Bayley & D. R. Preston (Eds.), *Second Language Acquisition and Linguistic Variation* (pp. 253-306). Amsterdam: John Benjamins.

Yuksel, G. (2007). *Grammatical errors in the compositions written by Turkish learners of English.* [Unpublished MA Thesis]. Ataturk University: Institute of Social Sciences.

Yuldashev, A., Fernandez, J., & Thorne, S. L. (2013). Second language learners' contiguous and discontiguous multi-word unit use over time. *The Modern Language Journal*, *97*(S1), 31-45.

Zobl, H. (1982). A direction for contrastive analysis: The comparative study of developmental sequences. *TESOL quarterly*, *16*(2), 169-183.

Zobl, H. (1990). Evidence for parameter-sensitive acquisition: A contribution to the domain-specific versus central processes debate. *Interlanguage Studies Bulletin (Utrecht)*, *6*(1), 39-59.

Zuraw, K. (2003). Probability in Language. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic Linguistics* (pp. 139-176). The MIT Press.