

# Subject and Object Markings in Conversational Korean

by

Taeho Kim

August 19, 2008

A dissertation submitted to the  
Faculty of the Graduate School of  
the State University of New York at Buffalo  
in partial fulfillment of the requirements for the  
degree of

Doctor of Philosophy  
Department of Linguistics

Copyright by  
Taeho Kim  
2008

## Acknowledgements

During the past four and a half years of my stay in Buffalo, I have had many encounters. Though there are some I would rather have not experienced, many others made me feel extremely lucky.

First and foremost, I am so lucky to have Dr. Mitsuaki Shimojo as my dissertation advisor. Since I started this dissertation, he has always been supportive in every aspect. From the bottom of my heart, I would very much like to thank him for his guidance, care and encouragement not only as an academic advisor but also as a mentor. He willingly shared much of his precious time during his sabbatical year with me for the discussion of this dissertation. His insightful and penetrating comments have always guided me along the right track. Not only his academic diligence and achievement but also his sincere care for his students have been an example of the teacher and linguist that I eventually want to be. He always encouraged me when I was overwhelmed and frustrated by surrounding pressures. Without his tremendous help, I could not have completed this dissertation. No matter what I could say or with how many words, it is certainly not enough to describe my deepest gratitude for his help. I sincerely admire him in every aspect.

I am also greatly indebted to the rest of my dissertation committee members, especially Dr. David A. Zubin. His guidance was extremely helpful, particularly in developing my idea into this dissertation topic. Since the initial stages, his comments on every draft of this dissertation led it to be a much better finished thesis. In addition to my gratitude for his advice on this thesis, I would also like to thank him for his empathy toward me when I was having difficult times. He is surely a warm-hearted teacher who understands the true meaning of professorship, something that is rarely encountered, especially these days. I sincerely admire him not only as my teacher but also as my mentor. I am also grateful to Dr. Eunhee Lee for her perceptive comments on this

dissertation, as well as for helping me to complete the dissertation as scheduled.

I would like to acknowledge faculty members at the department of Linguistics, Karin Michelson, Jean-Pierre Koenig, Matthew Dryer, Jeri J. Jaeger, David Fertig, Wolfgang Wölck, Robert D. Van Valin, Jr., Tsan Huang and Douglas Ronald for having helped me to be fond of this field, as well as for having guided me along the right track.

I am also very thankful to my best friend, Youngjo You for his sincere friendship. He has always been just one step away from me, and was always there whenever I needed his moral support. He has shown me a sincere friendship, which I can never forget. I can only hope our friendship will last forever.

I am thankful to Jinho Choi for his perceptive comments on this dissertation that were very helpful, as well as for his assistance on various things. Thanks also go to my friend, Justin Boffemmyer, for his friendship during my entire stay at the University at Buffalo, as well as for his comments on early drafts of this thesis. I would also like to acknowledge my colleagues at the Linguistics department, especially Robert Painter, Carlos Castillo-Garsow, Hui-Chen Sabrina Hsiao and Paul Heider for their friendship and encouragement. I also want to thank the ESL interns, particularly Mollie McCabe, for continuously checking the English grammar. I often asked her to read chapters under very short time constraints, but she always completed the jobs in a timely manner without fail. I would also like to acknowledge Anne Montgomery for proofreading the final draft of my dissertation. I am grateful to the Linguistic department secretaries, Jodi Reiner, Sharon Sell and Patricia Wieclaw for their assistance. I would also like to sincerely express my deep gratitude to Prof. Junyeong Kim and his family for their hospitality. They often invited me and my wife to many dinners at their house during my stay in Buffalo. I would like to thank Jaehee Bak at the University of Toronto for his time discussing my dissertation, and for sharing his resources with me. I am indebted to the Graduate Student Association at the University at Buffalo for its financial support

through Mark Diamond Research Fund. My sincere thanks also go to the 20 participants who provided the data for this dissertation. I especially wish to thank Han-gyu Lee, my former professor of the department of English language and literature at Kyunghee University, who has always been my mentor, having supported me since I was his student and who has always been like my dear uncle. He always cared for his students, cried for his students, guided them when they are astray and never hesitated to express his joy for his students' success as if it was his own. He is surely a warm-hearted teacher who understands that being a teacher is beyond the contents of a textbook. I admire him very much for his sincerity. He has taught me a lot that cannot be taught by others. I can only hope to see more teachers like him. I would also like to express my sincere gratitude to Prof. Jong-Bok Kim for providing me with many motivations to pursue Linguistics. I am also grateful to Prof. Myeongah Shin for inspring me to study abroad, and also to Profs. Kunsoo Lee, Jaehak Yoon, Byeongsoo Park, Sangchul Anh and Kyu-Hyun Kim from Kyunghee University who helped to teach me much of what I know today.

Finally, I most deeply thank my family in Korea for their patience and endless love: my father and mother, my father-in-law and mother-in-law, my brother and sister-in-law, my sister and brother-in-law. My parents and my brother have always supported me morally as well as financially. Last but not least, I would likely to thank my dearest wife, Sunjung Ha, for her patience and love. Without her patience and love, this dissertation project would have been much more difficult. I dedicate this dissertation to her.

There are many more people I wish to thank, though they are not listed here.

# Contents

Acknowledgements . . . . .	ii
Table of Contents . . . . .	v
List of Tables . . . . .	xi
The Yale System of Korean Romanization . . . . .	xii
Abbreviation . . . . .	xiii
Abstract . . . . .	xv
<b>1 Introduction</b>	<b>1</b>
<b>2 Previous Studies and Terms</b>	<b>8</b>
2.1 Outline . . . . .	8
2.2 Transitivity . . . . .	8
2.3 Grammatical relations . . . . .	9
2.4 Topic marking in Korean . . . . .	10
2.5 Focus marking in Korean . . . . .	15
2.6 Contrastive marking in Korean . . . . .	18
2.7 Word order and post-predicative encoding . . . . .	23
2.8 Case markers . . . . .	24
2.9 Zero particle and overt particles . . . . .	30
2.10 Saliency, givenness, and activation . . . . .	33

2.10.1	Saliency . . . . .	33
2.10.2	Givenness . . . . .	34
2.10.3	Activation . . . . .	36
2.11	Referential distance and referential persistence . . . . .	37
2.11.1	Referential distance . . . . .	38
2.11.2	Referential persistence . . . . .	40
<b>3</b>	<b>Conversational Korean Data</b>	<b>42</b>
3.1	Outline . . . . .	42
3.2	Data collection procedures . . . . .	42
3.3	Methodology for data analysis . . . . .	46
3.3.1	Identification of clausal units . . . . .	46
3.3.2	Identification of subjects . . . . .	53
3.3.3	Identification of objects . . . . .	58
3.4	Overview of tokens . . . . .	60
<b>4</b>	<b>Subject Marking</b>	<b>68</b>
4.1	Outline . . . . .	68
4.2	Subject encoding types . . . . .	69
4.3	Intransitive and transitive subjects . . . . .	71
4.4	Sentence types . . . . .	73
4.5	Factors . . . . .	76
4.5.1	Negation . . . . .	77
4.5.2	Animacy . . . . .	80
4.5.3	Word order . . . . .	89
4.5.4	Definiteness . . . . .	99
4.5.5	Length . . . . .	108

4.5.6	Repair . . . . .	119
4.5.7	Anaphoric saliency . . . . .	124
4.5.8	Cataphoric saliency . . . . .	140
4.5.9	Contrast . . . . .	148
4.5.10	Summary of Chapter 4 . . . . .	159
<b>5</b>	<b>Object Marking</b>	<b>167</b>
5.1	Outline . . . . .	167
5.2	Object encoding types . . . . .	168
5.3	Sentence types . . . . .	172
5.4	Factors . . . . .	176
5.4.1	Negation . . . . .	176
5.4.2	Animacy . . . . .	182
5.4.3	Word order . . . . .	190
5.4.4	Verb adjacency . . . . .	200
5.4.5	Definiteness . . . . .	204
5.4.6	Length . . . . .	212
5.4.7	Repair . . . . .	223
5.4.8	Anaphoric Saliency . . . . .	226
5.4.9	Cataphoric saliency . . . . .	234
5.4.10	Contrast . . . . .	240
5.4.11	HA ‘do’ constructions . . . . .	250
5.4.12	Nominative-marked objects: split case marking . . . . .	252
5.5	Summary of Chapter 5 . . . . .	259
<b>6</b>	<b>Concluding Remarks</b>	<b>265</b>
6.1	Outline . . . . .	265



6.2	Subject and object markings . . . . .	266
6.3	Findings of the present study . . . . .	277
6.4	Issues for future study . . . . .	280
	<b>Bibliography</b>	<b>281</b>

# List of Tables

1	Table of Yale Romanization . . . . .	xii
2.1	The topic accessibility scale (Lambrecht 1994:165) . . . . .	11
3.1	List of Participants for the Conversation Recording . . . . .	44
3.2	Scale of Naturalness . . . . .	45
3.3	Types of Predicates . . . . .	61
3.4	Sentence Types . . . . .	61
3.5	Word Order . . . . .	62
3.6	Negation . . . . .	63
3.7	Subject Forms . . . . .	63
3.8	Subject Marking Particles . . . . .	65
3.9	Object Forms . . . . .	65
3.10	Object Marking Particles . . . . .	67
4.1	Subject Encoding Types . . . . .	69
4.2	Transitivity and Subject Encoding Types . . . . .	72
4.3	Sentence Types and Subject Marking . . . . .	73
4.4	Negation and Subject Marking . . . . .	78
4.5	Subject Animacy and Subject Marking . . . . .	82
4.6	Subject Interlocutor and Subject Marking . . . . .	84

4.7	Object Animacy and Subject Marking . . . . .	87
4.8	Word Order and Subject Marking . . . . .	93
4.9	Subject Definiteness and Subject Marking . . . . .	104
4.10	Monosyllable and Subject Marking . . . . .	110
4.11	Monosyllabic Subject Type and Subject Marking . . . . .	112
4.12	Modifiers for Monosyllabic Subjects and Subject Marking . . . . .	114
4.13	Subject Modifiers and Subject Marking . . . . .	116
4.14	Subject Repair and Subject Marking . . . . .	120
4.15	Subject Repair with Transitivity and Subject Marking . . . . .	121
4.16	Subject Anaphoric Recency (RD) and Subject Marking . . . . .	127
4.17	Subject RD (1-5, 6-NPM) and Subject Marking . . . . .	128
4.18	Subject Anaphoric Frequency (AF) and Subject Marking . . . . .	136
4.19	Subject AF (0-3, 4-10) and Subject Marking . . . . .	137
4.20	Subject Cataphoric Saliency-uninterruption (RP) and Subject Marking . . . . .	142
4.21	Subject RP (0-3, 4-10) and Subject Marking . . . . .	143
4.22	Subject Cataphoric Saliency-frequency (RP- <i>f</i> ) and Subject Marking . . . . .	145
4.23	Subject RP- <i>f</i> (0-3, 4-10) and Subject Marking . . . . .	146
4.24	Subject Contrast and Subject Marking . . . . .	154
4.25	Sub-typed Subject Contrast and Subject Marking . . . . .	157
5.1	Object Encoding Types . . . . .	168
5.2	Transitivity and Subject Encoding Types-repeated . . . . .	170
5.3	Sentence Types and Object Marking . . . . .	173
5.4	Negation and Object Marking . . . . .	178
5.5	Object Animacy and Object Marking . . . . .	184
5.6	Object Interlocutor and Object Marking . . . . .	186
5.7	Subject Animacy and Object marking . . . . .	188

5.8	Word Order Variations and Object Marking . . . . .	195
5.9	Verb Adjacency and Object Marking . . . . .	202
5.10	Object Definiteness and Object Marking . . . . .	209
5.11	Monosyllabic Objects and Object Marking . . . . .	215
5.12	Monosyllabic Object Type and Object Marking . . . . .	216
5.13	Monosyllabic Object Modifiers and Object Marking . . . . .	218
5.14	Object Modifiers and Object Marking . . . . .	220
5.15	Object Repair and Object Marking . . . . .	225
5.16	Object Anaphoric Recency (RD) and Object Marking . . . . .	229
5.17	Object RD (1-5, 6-NPM) and Object Marking . . . . .	230
5.18	Object Anaphoric Frequency (AF) and Object Marking . . . . .	232
5.19	Object AF (0-3, 4-10) and Object Marking . . . . .	233
5.20	Object RP and Object Marking . . . . .	235
5.21	Object RP (0-3, 4-10) and Object Marking . . . . .	236
5.22	Object RP- <i>f</i> and Object Marking . . . . .	237
5.23	Object RP- <i>f</i> (0-3, 4-10) and Object Marking . . . . .	238
5.24	Object Contrast and Object Marking . . . . .	245
5.25	Detailed Object Contrast and Object Marking . . . . .	248
5.26	<i>hata</i> -V and Object Marking . . . . .	251
5.27	Case Alternation in Split-case Marking . . . . .	256
5.28	Split-Case Marking and Mean RD;AF;RP;RP- <i>f</i> . . . . .	257

# The Yale System of Korean Romanization

Yale romanization has been adopted for transcribing Korean examples in this dissertation. Table 1 is based on Sohn (1999).

Vowels			Consonants		
Yale	Phonemic value in IPA	Phonetic value in IPA	Yale	Phonemic value in IPA	Phonetic value in IPA
a	/a/	[a]	p	/p/	[p, b]
ya	/ja/	[ja]	ph	/ph/	[p <sup>h</sup> ]
e	/ə/	[ə]	pp	/pʰ/	[pʰ]
ye	/jə/	[jə]	t	/t/	[t, d]
o	/o/	[o]	th	/th/	[t <sup>h</sup> ]
yo	/jo/	[jo]	tt	/tʰ/	[tʰ]
wu	/u/	[u]	s	/s/	[s]
yu	/ju/	[ju]	ss	/sʰ/	[sʰ]
u	/i/	[i]	c	/c/	[c, ɟ]
i	/i/	[i]	ch	/ch/	[c <sup>h</sup> ]
wi	/wi/	[wi]	cc	/cʰ/	[cʰ]
ey	/e/	[e]	k	/k/	[k, g]
yey	/je/	[je]	kh	/kh/	[k <sup>h</sup> ]
wey	/we/	[we]	kk	/kʰ/	[kʰ]
oy	/we, ø/	[we, ø]	m	/m/	[m]
ay	/ɛ/	[ɛ]	n	/n/	[n]
yay	/jɛ/	[jɛ]	ng	/ŋ/	[ŋ]
way	/wɛ/	[wɛ]	l	/l/	[l, ɾ]
we	/wə/	[wə]	h	/h/	[h]
wa	/wa/	[wa]			
uy	/ij/	[i(j), i, e]			

Table 1: Table of Yale Romanization

# Abbreviation

An asterisk [\*] indicates that the sentence is clearly ungrammatical, and the symbol [#] shows that the sentence is grammatical but is clearly inappropriate in the context in question. The symbols [?] or [??] are used to indicate an awkward sentence, due to either grammatical constraints or discourse-pragmatic constraints. The following abbreviations are used throughout this dissertation.

ACC: accusative marker

AF: anaphoric frequency

CONN: connective

COP: copular

DCL: declarative

DP: discourse particle

DR: directional particle

FUT: future

LOC: locative marker

NC: numeral classifier

NQ: numeral quantifier

NMLZ: nominalizer

NOM: nominative marker

PL: plural marker

PST: past

Q: question

REL: relative

RD: referential distance

SEM: sentence ending marker

RP: referential persistence (*uninterruption*)

RP-*f*: referential persistence (*frequency*)

SEM: sentence ending marker

TOP: topic marker

## Abstract

This dissertation investigates the case markers *ka* and *lul* in Korean from an empirical perspective. Korean is a well-known nominative-accusative language in which the subject and object are marked by the post-nominal particles *ka* and *lul* respectively. As has long been observed, the subject and object are not always overtly marked by these particles, particularly in informal spoken Korean. In fact, the presence or absence of these particles has instigated much debate as to whether or not they are case markers, and this is still disputable. Despite many studies examining the structural properties of the case markers, there are very few studies investigating them from an empirical perspective. The case markers cannot be explained solely from structural perspectives, and discourse-based observations should also be considered in the discussion of the structural properties of the case markers. I propose two discourse-pragmatic factors, namely processing load and informational prominence, for their analysis.

The earlier part of the dissertation introduces several issues that are related to the so-called case markers *ka* and *lul*, and presents the issues that are pertaining to this dissertation, e.g., when the case markers do and do not overtly occur in a clause. The later part of the dissertation discusses the data and the factors used for the analysis of the case markers in Korean. Throughout the dissertation, the observations made for the case markers in Korean were compared to those in Japanese in an attempt to delineate the differences between the two languages.

In this dissertation, approximately five hours of informal two-party conversations were used for the investigation of the presence and absence of the case markers in Korean. I proposed several individual factors that are related to either processing load or informational prominence or to both. The two discourse-pragmatic factors that are used for the discussion of the case markers well capture the presence and absence of such markers, i.e., *ka* and *lul* tend to overtly occur in a clause when the subject and

object would represent processing load, informational prominence or both, and they tend not to overtly occur in a clause when neither of the two factors is represented by the subject and object.

The approach I take for the discussion of the case markers in Korean is original, and effectively describes their presence and absence. Along with other studies of the case markers from an empirical perspective, the present study contributes empirical evidence depicting cross-linguistic differences between Korean and Japanese case markers, which still need to be further delineated, and sheds light on the discussion of the structural properties of the case markers in Korean from an empirical perspective.



# Chapter 1

## Introduction

Korean is a language with a nominative-accusative case marking system (Sohn 1999). In Korean, the particle ‘ka’, often referred to as a nominative marker, is used to mark subjects, and the particle ‘lul’, often referred to as an accusative marker, is used to mark objects. Subject and object case markers in Korean have been discussed in many previous studies (e.g., Min 1982; Kang 1986; Lee & Thompson 1987; Kim 1990; Hong 1991; Hong *et al.* 1998; Sohn 1999; Kim 2000; Ko. S 2000; Jun 2003; Kim 2003, *inter alia*), but they have been investigated vastly from syntactic or semantic perspectives with little attention to empirical data. Discourse-based observations should also be considered in the discussion of the structural properties of the case markers. In other words, case markers should be discussed in an empirical perspective as well as a structural perspective.

The goal of this dissertation is to examine discourse-pragmatic properties of the subject and object marking particles, *ka* and *lul*,<sup>1</sup> respectively. More specifically, this

---

<sup>1</sup>The particle *i* is a morphological variation of the particle *ka* in Korean. The particle *i* follows a consonant, whereas the particle *ka* follows a vowel. The particles *l* and *ul* are morphological variations of the particle *lul*. The particle *ul* appears after a consonant, the particles *l* and *lul* after a vowel (Sohn 1999). Throughout this dissertation, the particles *ka* and *lul* represent the subject marking particle and the object marking particle, respectively.

dissertation investigates when these particles explicitly occur and when they do not, based on the data of informal Korean conversation, and it claims that the occurrence and non-occurrence of these particles are related to processing load and informational prominence. Based on the occurrence and non-occurrence of these particles, this dissertation attempts to capture discourse-pragmatic properties of post-nominal markers for subject and object (e.g., *ka* and *lul*).

Generally speaking, a speaker and a hearer strategically interact with each other (Givón, 1995). The speaker makes assumptions about the addressee's knowledge when (s)he utters, and the sentences used reflect on such assumptions. The hearer seeks to interpret the utterances accordingly based on the available information. Based on this general property, I propose that the occurrence and non-occurrence of subject and object markers, *ka* and *lul* respectively, appear overtly when they are needed to properly process sentences, but they do not appear overtly when they are not needed. More specifically, they are utilized when there are processing load and informational prominence expected, but they are not exercised when there are not.

Let us look at several examples in English as well as in Korean. The basic word order in English is generally restricted to SVO (Subject+Verb+Object), so it is relatively easy to identify a subject or an object in a sentence. The subject appears before the object, and thus, the first argument in the sentence is likely to be the subject and the second one to be the object, as shown in the examples below.

(1.1) John met Sue yesterday.

(1.2) Sue met John yesterday.

(1.3) I met him yesterday.

(1.4) \*Him met I yesterday.

(1.1) and (1.2) are both grammatical, but they are differently interpreted. In (1.1), the first entity *John* is the subject, and the second one *Sue* is the object. In (1.2), the first entity *Sue* is the subject, and the second one *John* is the object. Simply put, the utterances are interpreted based on the employed word order. (1.3) is grammatical, as the first argument *I* is the subject, and the second one *him* is the object of the clause. On the other hand, (1.4) is ungrammatical because the object *him* precedes the pronominal subject *I* in the sentence.

Unlike English, the word order in Korean is said to be SOV (Subject+Object+Verb), but it appears to be relatively flexible, particularly in informal spoken Korean (Sohn 1999). Therefore, word order is not always useful in the identification of subjects and objects of in the sentence in Korean. There is, however, another means available in Korean to help with the identification of subjects and objects in the sentence, namely subject and object marking particles. That is to say, different forms of particles are utilized for marking the subject and the object, and they aid in the identification of subjects and objects as shown in the examples below.

(1.5) *Junsu-ka ku chayk-ul ilk-ess-ta.*  
 Junsu-NOM that book-ACC read-PST-DCL  
 ‘Junsu read that book.’

(1.6) *ku chayk-ul Junsu-ka ilk-ess-ta.*  
 that book-ACC Junsu-NOM read-PST-DCL  
 ‘Junsu read that book.’

The word order in (1.5) is SOV. The first entity, *Junsu*, is the subject, and the second one, *ku chayk*, ‘that book’ is the object of the sentence. On the other hand, the word order in (1.6) is OSV. The first entity, *ku chayk*, ‘that book’ is the object, and the second one, *Junsu*, is the subject. Although the most common and therefore unmarked word order, namely SOV, is not exercised in (1.6), the subject and the object

are identified by the attached particles, *ka* for the subject and *lul* for the object. Still, these particles are not always helpful for the identification of subjects and objects because these particles are not always available or do not always appear explicitly in the sentence, as shown (1.7) through (1.9).

(1.7) *Junsu-Ø<sup>2</sup> onul hakkyo-ey ka-ss-ni?*  
 Junsu-( ) today school-LOC go-PST-Q  
 ‘Did Junsu go to school today?’

(1.8) *Junsu-ka ecey ku sakwa-Ø mek-ess-ta.*  
 Junsu-NOM yesterday that apple-( ) eat-PST-DCL  
 ‘Junsu ate that apple yesterday.’

(1.9) *Junsu-Ø onul achim-Ø mek-ess-ni?*  
 Junsu-( ) today breakfast-( ) eat-PST-Q  
 ‘Did Junsu eat breakfast today?’

In (1.7), the subject *Junsu* is not marked with *ka*, namely the nominative marker, and the object *ku sakwa* ‘that apple’ in (1.8) is not marked with *lul*, namely the accusative marker. In (1.9), neither the subject *Junsu* nor the object *achim* ‘breakfast’ occurs with its case marking particle. Yet all these examples are acceptable as they would appear in conversational Korean, regardless of the particles which are required in formal written Korean. It is noted that the subject and object marking particles may be omitted when the canonical word order, namely SOV, is exercised (Sohn 1999). These particles, however, are not often expressed overtly even when the employed word order is not considered canonical or unmarked, as seen in (1.10).

(1.10) *onul achim-Ø mek-ess-ni, ne-Ø?*  
 today breakfast-( ) eat-PST-Q 2SG-( )  
 ‘Did you eat breakfast today?’

---

<sup>2</sup>Throughout this dissertation, the symbol  $\emptyset$  refers to a zero particle, which indicates that an associated particle is deleted due to (relatively) easy identification (Sohn 1999) or it is not realized overtly in a surface form (Min 1982).

In (1.10), the object, *achim* ‘breakfast’, precedes the subject, *ne* ‘you’, in the sentence, and the word order of this sentence is OVS. Both the subject and the object in (1.10) occur without their particles, and this example is still acceptable, at least in certain discourse contexts. As a matter of fact, it would sound unnatural if the subject and the object in (1.10) are overtly marked with their particles. Furthermore, these particles are often replaced with other particles (e.g., *nun* TOP, *eykey* DAT, etc.), which may be combined not only with subjects but also with objects, and thus contribute very little to the identification of the subject and the object in the sentence. This suggests that the (unmarked) word order is not the only factor which determines the occurrence and non-occurrence of such particles.

What has been thus far discussed indicates that subjects and objects may be identified by something other than the (unmarked) word order or the case markers, *ka* and *lul*, respectively. It also suggests that the case markers do not always appear overtly in the sentence, and case marking per se may not be the only function encoded in these markers. Then, when do the post-nominal particles *ka* and *lul* appear and not appear overtly in the sentence? If the post-nominal particles *ka* and *lul* have functions other than marking subjects and objects, what other functions are encoded in such particles? As an attempt to answer these questions, I examine these post-nominal particles in conversational Korean from discourse-pragmatic perspectives. I argue that processing load and informational prominence are two factors which influence when these post-nominal particles appear overtly in the sentence and when they do not. Based on this, I claim that the post-nominal particles *ka* and *lul* not only mark the subject and the object, respectively, in the sentence, but also represent functional properties closely related to processing load as well as informational prominence. Simply put, the post-nominal particles *ka* and *lul* are attached to the subject and the object when the arguments would benefit from being highlighted under certain processing load or informational

prominence.

This dissertation continues as follows: Chapter 2 introduces and summarizes previous claims and observations that are relevant to this dissertation, namely previously claimed properties of the post-nominal particles *ka* and *lul*, as well as a zero particle. Chapter 2 also introduces and defines some terminologies that are used throughout this dissertation.

Chapter 3 presents an overview of the informal Korean conversation data used in this dissertation, which includes a description of the data collection procedure as well as the preparation for data analysis. Chapter 3 also discusses the methodology that is utilized to analyze the data, and it further presents the overall token counts in terms of different forms of particles.

Chapter 4 discusses the subject encoding types in informal conversational Korean, particularly in terms of processing load as well as informational prominence. More specifically, it discusses how a speaker chooses one particular subject form over another, largely based on two different but related factors, i.e., processing load and informational prominence, which I propose are useful notions to describe the occurrence and non-occurrence of the case marker *ka*. Factors used for the discussion of subject encoding types are related to either processing load, informational prominence, or both, and they are: negation, animacy, word order, definiteness, length, repair, anaphoric saliency, cataphoric saliency, and contrast. In Chapter 4, the speaker's selection of one subject form over another is discussed in the conversation data of the present study based on individual factors, and then based on multi-factors when they are relevant. In this chapter, I argue that the speaker's selection of one particular subject encoding type in opposition to other encoding types is not an arbitrary but a systematic process, and the selectional process is based on processing load and informational prominence. For instance, I argue that the speaker tends to select an explicit particle, e.g., *ka*, for coding

subjects representing an indefinite referent since the indefinite referent is a deviation from the norm for the subject NPs, hence increased processing load is expected.

Chapter 5 is devoted to the in-depth discussion of the object encoding types in conversational Korean, also based on processing load and informational prominence. That is to say, it discusses how a speaker chooses one object encoding type over another based on the two different but related factors, i.e., processing load and informational prominence, which are used to describe the occurrence and non-occurrence of the case marker *lul*. Factors that are used for the discussion of object encoding types are: negation, animacy, word order, verb adjacency, definiteness, length, repair, anaphoric saliency, cataphoric saliency, as well as contrast. These factors are all related to either processing load, informational prominence, or both in one way or another. In Chapter 5, I argue that the speaker's selection of one object form in an opposition to other forms is also not an arbitrary but systematic process, and the selectional process is based on processing load and informational prominence. For example, I claim that the speaker tends to choose an overt particle, e.g., *lul*, for encoding objects representing a definite referent because the definite referent is a deviation from the norm for the object NPs, thus increased processing load is anticipated. In addition to these factors, the choice of object encoding types is also discussed with regard to *hata*-verb and split-case marking as well.

Chapter 6 concludes this dissertation by summarizing the findings, as well as discusses some issues that remains to be studied in the future.

# Chapter 2

## Previous Studies and Terms

### 2.1 Outline

Chapter 2 introduces and defines terminologies that are used throughout this dissertation. It also discusses previous observations as well as claims that are relevant to this dissertation, particularly focusing on post-nominal particles used for marking subjects and objects in conversational Korean.

### 2.2 Transitivity

While clauses are usually considered to be either transitive or intransitive, transitivity is a complex phenomenon which involves a variety of explanations from both semantic and syntactic perspectives. Transitivity has been characterized by the ‘prototype transitivity event’ (Hopper & Thompson 1980; Givón 2001b; Thompson & Hopper 2001). There is also a syntactic prototype of transitive clauses, which is observed for most languages: clauses and verbs that have either a covert or overt ‘direct object’ are syntactically transitive, and all others are syntactically intransitive (Givón 2001a:109). Yet “the syntactic definition of transitive clauses—and thus of transitive verbs—is subjected



to a certain measure of cross-language typological variation” (Givón 2001a:109).

The ‘transitivity’ judgment of clauses used in this dissertation is based on Givón’s (2001a) syntactic prototype of transitive clauses. If clauses have a direct object, in either a covert or overt form, these clauses are transitive, and they are intransitive otherwise. In this dissertation, the notion ‘transitive’ is used to refer to clauses with a transitive verb in Korean. The notion ‘intransitive’ is used to refer to clauses with an adjectival predicate, a nominal predicate, as well as an intransitive verb in Korean, unless otherwise specified.

## 2.3 Grammatical relations

Grammatical relations have been one of the central issues in linguistics. In general, “participants in states or events, in whatever semantic role, can also assume some grammatical relation in the clause” (Givón 2001a:108). They are, for example, subject, direct object, and indirect object, among others. Subject, direct object, and indirect object may all be defined from syntactic, semantic, as well as pragmatic perspectives.

Chomsky (1965), from a syntactic point of view, defines a subject as an NP which is immediately dominated by S’ in the phrase structure tree, and a direct object as an NP immediately dominated by the VP. These configurational bases of grammatical relations are proposed mainly for configurational languages like English and French. There were also attempts at mapping from semantic roles to grammatical relations, and some generalizations emerged from such attempts. The following generalizations are taken verbatim from (Givón 2001a:108):

“Constraints on mapping from semantic roles to grammatical relations in simple clauses

- a. An agent can only be the subject.

- b. A patient can only be a subject or direct object.
- c. A dative can be a subject, direct object or indirect object.
- d. All other semantic roles can only be indirect objects.”

In Role and Reference Grammar (henceforth, RRG), the generalized semantic roles called ‘semantic macroroles’ are mapped with syntactic arguments. Specifically, the actor, ‘the generalized agent type role’, is associated with subject, and the undergoer, ‘the generalized patient type role’, with object (Van Valin & Lapolla 1997:141).

In languages with a fixed word order, grammatical relations can be determined by word order. Case markers can also be used to decide grammatical relations for languages with a case system, such as Korean and Japanese. In this dissertation, word order, as well as case markers, are used to identify subjects and objects in Korean. Specifically, the first argument in the clause was identified as the subject, and the second one as the object by default, unless the case marking, if available, suggests otherwise; Namely, the argument marked with the post-nominal particle *ka* was viewed as the subject, and the one marked with the post-nominal particle *lul* was seen as the object. If overt case marking was not available, and/or the arguments were ellipsed, arguments were identified by their semantics and retrieved from the context if necessary.

## 2.4 Topic marking in Korean

The term ‘topic’ has been an interesting issue. Topic may be viewed as a part of a sentence that denotes what the sentence is about (Reinhart 1981; Gundel 1985; Trask 1999, *inter alia*).

From a discourse-pragmatic perspective, it has generally been assumed that a topic denotes given information, whereas a comment conveys new information, in terms of its information content.<sup>1</sup> In previous studies on information structure, the notion ‘topic’

---

<sup>1</sup>Trask (1999) notes that some linguists often confuse the term ‘topic’ with the term ‘focus’, but

is often regarded as synonymous with the ‘given’ or ‘presupposed’. On the other hand, Lambrecht (1994) claims that what is presupposed is not a topic itself, though the topic appears to play a role in a given proposition due to its given discourse status. He further notes a correlation between the pragmatic state of the topic referent and its acceptability as a topic. More specifically, it is likely that the more accessible the topic referent of an utterance is, the less processing effort to interpret that utterance is needed (Lambrecht 1994). This correlation is summarized as the Topic Acceptability Scale in Table 2.1. Chafe (1987) also states that it requires low cognitive effort to interpret a discourse-active referent.

Table 2.1: The topic accessibility scale (Lambrecht 1994:165)

Active	Most acceptable
Accessible	
Inactive	↕
Brand-new anchored	
Brand-new unanchored	Least acceptable

It should be pointed out that the topic is often associated with contrastiveness. If the topic is contrastive, it is categorized as a contrastive topic, in contrast with the ‘regular’ topic, which is often called a ‘continuing topic’. The contrastive topic is sometimes referred to as a ‘shifted topic’ (Herring 1990; Aissen 1992). These two types of topics are further elaborated in the subsequent section with examples in Korean.

Korean, like Japanese, is a language where a topic is grammatically encoded by a so-called topic marker (i.e., *nun*<sup>2</sup> for Korean and *wa* for Japanese). For this reason, they are clearly distinct from each other.

<sup>2</sup>The forms *un* and *n* are phonological variations of *nun*, and the choice between these is determined phonologically. The form *nun* appears after a vowel, and the form *un* appears after a consonant. The form *n* is a contracted form of *nun*. For the sake of convenience, the form *nun* represents a topic marker in this study.

a brief overview of how the term ‘topic’ is characterized in Korean is provided in this section.

Thus far, the notion ‘topic’ in Korean, like Japanese, has frequently been investigated from semantic and syntactic points of view (e.g., Kuno 1973; Whitman 1989; Kim. K 1990; Shibatani 1991; Choi 1996; Choi 1997; Han 1998; Wee 1998; Han 1999; Sohn 1999, among others). There have also been studies that discuss topic from discourse-pragmatic perspectives (Maynard 1980; Kim 1993; Choi & Shimojo 2001; Shimojo 2005, *inter alia*). The former approach has focused on semantic characterizations of the topic marker, mostly based on constructed examples in isolated sentences, whereas the latter attempts to characterize the interactional functions of the topic marker based on naturally occurring conversation by examining how a speaker uses a topic marker in actual discourse. For instance, the speaker uses a topic marker when (s)he initiates a story or negates a previous utterance (Kim 1993).

As previously mentioned, it is well established that the so-called topic marker (*nun*) is used to mark the topic of a sentence in Korean (Sohn 1999). It is also believed that the topic marked NPs can not only encode a topic reading but also a contrastive reading depending on the given contexts, as seen in (2.1) below.

- (2.1) *Junsu-nun Suji-lul cohahanta.*  
 Junsu-TOP Suji-ACC like  
 ‘As for Junsu, (he) likes Suji.’ [topic]  
 ‘Junsu (but not others) likes Suji.’ [contrastive]

The *nun*-marked subject *Junsu* in (2.1) predominantly receives a (sentence) topic reading, and the subject *Junsu* is interpreted as what the sentence is ‘about’, namely the topic or theme of the sentence. On the other hand, the subject *Junsu* receives a contrastive (topic) reading when it is implied that “Junsu, but not others, likes Suji.”

It is a predominantly accepted view that the topic of the sentence is encoded by the so-called topic marker *nun* in Korean. There are also claims, though not a widely

accepted view, such that the topichood is encoded not by the topic marker but by scrambling (Choi 1997). Choi (1997) further claims that *nun* is not a topic marker but a contrastive marker, in that a *nun*-marked NP unit, either subject or object, in the base position only denotes a contrastive reading, and it gets a topic reading when it is in a scrambled position. The utterances in (2.2), which are taken from Choi (1997:549), exemplify *nun*-marked objects in the base position, i.e., (2.2a), and scrambled positions, i.e., (2.2b) and (2.2c).

- (2.2) a *Mary-ka ecey John-un manna-ss-ta.*  
 Mary-NOM yesterday John-TOP meet-PST-DCL  
 ‘Mary met John yesterday (but nobody else).’
- b *Mary-ka John-un ecey manna-ss-ta.*  
 Mary-NOM John-TOP yesterday meet-PST-DCL  
 ‘Mary met John yesterday (and Bill today).’
- c *John-un Mary-ka ecey manna-ss-ta.*  
 John-TOP Mary-NOM yesterday meet-PST-DCL  
 ‘As for John, Mary met him yesterday (and as for Bill, Jane met him today).’

The in-situ object *John* in (2.2a) gets a contrastive reading, but the scrambled object as in (2.2b) and (2.2c) receives a contrastive topic reading, as can be seen in English translation. That is to say, unlike the in-situ object *John* in (2.2a), the scrambled object in (2.2b) and (2.2c) is not only “[t]he topic of the sentence in the sense that it is what the sentence is ‘about’, but also is ‘contrastive’ in the sense that it implies that the claim that the sentence is making need not be true of something else or that some other claim may be true of something else” (Choi 1997:549-50). Based on this, Choi (1997) claims that *nun*, known as a topic marker in Korean, is not a topic marker but a contrastive marker. Her claim may sound plausible in one hand, but it is untenable on the other hand. For instance, her claim that scrambling, but not the topic marker *nun*

itself, is what encodes the topichood does not explain the topichood for the *nun*-marked subject in-situ because the in-situ subject is already in the default topic position for a topic. As a matter of fact, all in-situ subjects are topics. It is thus too extreme to claim that scrambling, but not the topic marker *nun* itself, is what encodes the topic of a sentence, though scrambling may be related to the topichood encoding because the sentence initial position is a default topic position for the topic of the sentence. For this reason, the view that *nun* is not a topic marker in Korean is not accepted in the present study.

Han (1998) points out that the so-called topic marker *nun* conveys not only a topic reading and a contrastive reading, but also a contrastive focus reading as well. This view is against Lee's (2003) argument in which *nun* is viewed as a contrastive topic marker, and *ka* as a contrastive focus marker in Korean. Han (1998) proposes the formal semantic characterization<sup>3</sup> for *nun* and presents the three different properties of *nun* (i.e., a topic reading; a contrastive reading; a contrastive focus reading). She also claims that the particular readings for *nun* are determined by the syntactic environment in which it occurs. Specifically, a VP-external *nun*-marked NP results in a topic or contrastive topic reading, and a VP-internal *nun*-marked NP receives a contrastive focus reading. She further claims that the function of the so-called topic marker *nun* is not merely to mark the topic of the sentence, but its primary function is to introduce a presuppositional set and the marker functions as a topic marker only in certain syntactic environments.

In summary, there have been many studies investigating the properties of the post-

---

<sup>3</sup>The marker *-(n)un* presupposes a non-empty set (including a singleton set), and proposes the following semantics.

“presupposition of  $\alpha$ -*(n)un*, where  $\alpha$  is an individual and  $X$  is a set variable over individuals:

$$\exists x[(\alpha \in X) \wedge ([X] \geq 1)]$$

When the set presupposed by *-(n)un* is a singleton set, only a topic reading is available. When the set contains more than one element, topic reading, contrastive topic reading and contrastive focus reading are all available” (Han 1998:5).

nominal particle *nun* in Korean (Kim 1990; Choi 1997; Han 1998; Lee 2003, among others). Some claim that *nun* is a topic marker (Sohn 1999; Lee 2003, inter alia), while others claim that *nun* is a contrastive marker (Choi 1997; Han 1998; Lee 2003, among others). Still others attempt to define the properties of the post-nominal particle *nun* merely based on the notion ‘old/new’, in that *nun* encodes old information, in contrast to the nominative *ka*, which usually encodes new information (Sohn 1999). One certain function of the so-called topic marker *nun* is to mark the topic of the sentence, i.e., what the sentence is about, but it may also encode contrastiveness as well as old information, depending on the context. In short, the post-nominal particle *nun* cannot be defined by a single factor alone, though each factor may describe the functions of *nun* in one way or another.

In the present study, it is discussed how the post-nominal particle *nun* is used in conversational Korean, and based on the usage of *nun* in actual discourse, I attempt to disclose the discourse-pragmatic properties of the particle *nun*.

## 2.5 Focus marking in Korean

Just like the notion ‘topic’, the term ‘focus’ has frequently been discussed. Trask (1999:95-6) defines the term ‘focus’ as a way of “singling out some particular element of a sentence or an utterance as representing the most important new information.” Shimojo (2005:16) states that “focusing is a mechanism to select a particular piece of information to pay attention to; hence, the information being selected for focusing is in one’s cognitive attention.” The focus<sup>4</sup> is generally believed to be distinct from the topic, and it is even regarded as an opposite notion in terms of informational status. Specifically, topic conveys ‘old’ information, and focus conveys ‘new’ information (Choi

---

<sup>4</sup>Trask (1999:96) points out that “[f]ocus must be clearly distinguished from topic, though even professional linguists have been known to confuse these terms.”

1997). The term ‘focus’ often refers to “[t]he status of certain sentence constituents that systematically differ from topic expressions in their pragmatic function and in their formal expression, and the focus is defined as the ‘complement of topic.’” (Lambrecht 1994:206). The notion ‘focus’ can also be defined in terms of phonetic prominence (Wee 1998).

It is noted that just like ‘topic’, ‘focus’ is not a uniform notion. It has been suggested that there are two types of focus, namely PRESENTATIONAL FOCUS and CONTRASTIVE FOCUS<sup>5</sup> (Herring 1990:164). It is claimed that the presentational focus differs from contrastive focus in that entities presented for the first time are completely new, whereas contrastively focused entities are already explicitly or implicitly present in the discourse context (Herring 1990:164). Similarly, Rochemont (1986) and Rochemont and Culicover (1990) argue, using the notion of C-CONSTRUABILITY<sup>6</sup>, that a ‘presentational’ focus is discourse-new, but a ‘contrastive’ focus is not. Kuno (1972; 1973) views the notion ‘focus’ in two different types of focus domain: ‘neutral description’ reading and ‘exhaustive listing’ reading.

There are several means of encoding a particular part of an utterance as a focus of that utterance. In English, stress can be used to mark the focus of an utterance, whereas in Basque, any element that immediately precedes the verb is focused (Trask 1999). In still other languages, such as Korean and Japanese, a morphological device (i.e., *ka* and *lul* in Korean; *ga* and *o* in Japanese) is used to encode the focus of the sentence (Kuno 1973; Sohn 1999).

Focus can be viewed as a complement of ‘presupposition’ according to Jackendoff (1972). Jackendoff (1972), whose analysis builds on those of Halliday and Chomsky (1970), defines the ‘presupposition of a sentence’ as “the information in the sentence

---

<sup>5</sup>Szabolcsi (1981:18) notes that in rather informal terms, “[t]he common feature distinguishing ‘topic’ from ‘focus’ from neutral constituents is that only the former may be contrastive.”

<sup>6</sup>The notion of C-CONSTRUABILITY is defined to be ‘under discussion’ (Rochemont 1986:47) or to ‘have a semantic/pragmatic antecedent in the discourse’ (Rochemont and Culicover 1990:20).



that is assumed by the speaker to be shared by him and the hearer” and the ‘focus of a sentence’ as “the information in the sentence that is assumed by the speaker not to be shared by him and the hearer” (1972:230). For Jackendoff, the focus is thus the complement of the presupposition in a sentence.

As noted above, it is generally believed that in Korean, the focus of the sentence is marked by the so-called focus marker. In other words, *ka* or *lul*-marked elements are considered to be focused, as shown in (2.3).

- (2.3) *Junsu-ka Suji-lul sakwi-ko issta.*  
 Junsu-NOM Suji-ACC date-CONN be.DEC  
 ‘There is Junsu dating Suji.’ [presentational]  
 ‘It is Junsu (not others) who is dating Suji.’ [contrastive focus]

In (2.3), we can receive, at least, two different readings. The first reading in (2.3) is a presentational (focus) reading which involves the introduction of a referent into a discourse, and the whole sentence is ‘new’. The second reading in (2.3) is a (contrastive) focus reading that does not necessarily denote new information, though it can. The most prominent function of the (contrastive) focus reading is to “highlight a constituent in a sentence in contradistinction to a parallel phrase” (Herring 1990:164). Based on this, it can be said that in the second reading in (2.3), *Junsu* is interpreted as a (contrastive) focus.<sup>7</sup> As was previously mentioned, the topic/focus distinction is often considered the same as the old/new distinction. Yet here we learn that there are cases where the element in focus is old in terms of information status, although, in general, the element in focus is new. It is therefore obvious that the topic/focus distinction should be defined in a more precise way than simply by the old/new distinction.

It is also well-known that *nun* is a topic marker and *ka* is a focus marker in Korean (Lee 2003). However, the so-called focus marker *ka* can also encode the topic of the

---

<sup>7</sup>In Kuno (1972; 1973), the first reading in (2.3) is interpreted as a ‘neutral description’ reading, and the second reading as an ‘exhaustive listing’ reading.

sentence, as the so-called topic marker *nun* is often used as a contrastive focus marker, as addressed in Choi (1997). As a matter of fact, Korean differs from Japanese in this matter, in that *ka*-marked elements in Korean can be a (continuing) topic of the sentence (Choi 1997), while *ga*-marked entities in Japanese is not regularly used in the same situation and it cannot be a marker of the continuation of the same topic (Shimojo & Choi 2001).

Thus far, there have been many studies that attempted to define the function of the so-called nominative marker *ka*, as well as the accusative marker *lul* in Korean (Kim 1990; Choi 1998; Sohn 1999; Lee 2003, among others). For example, some claim that *ka* encodes new information, thus a focus marker, while *nun* conveys old information, hence a topic marker (Sohn 1999). Others argue that *ka* is a contrastive focus marker, and *nun* is a contrastive topic marker (Lee 2003). I argue that *ka*, as well as *nun*, cannot simply be defined by a single factor, though each factor may describe the functions of *ka*, as well as *nun*, in one way or another.

In this dissertation, it will be discussed that the notion of focus per se may not be a good means to distinguish *ka* from *nun*. Instead, this dissertation will attempt to define the usage of *ka* and *nun* from other viewpoints such as processing load and informational prominence. For instance, I propose that *ka* is used to mark subjects that represent processing load whereas *nun* would not be used for such a purpose, i.e., encoding processing load, but is used for a purpose such as marking contrastiveness.

## 2.6 Contrastive marking in Korean

Many linguists have attempted to define the notion ‘contrastiveness’ in different ways (Halliday 1967; Chafe 1976; Clancy & Downing 1987; Lambrecht 1994, among others).

For Halliday (1967), the term ‘contrastive’ is defined as “contrary to some predicted

or stated alternative.” In other words, if there is a specific alternative available in mind, in which a particular element can be replaced, the element is regarded as ‘contrastive’. For Chafe (1976), the term ‘contrastiveness’ involves three factors: (i) a background knowledge that is shared by both a speaker and an addressee; (ii) a set of possible candidates for the role played by the element that is being contrasted; and (iii) the assertion of which of these possible candidates is the correct one. Clancy & Downing (1987) argue that the notion of contrastiveness should be defined in a more careful way due to varying degrees of contrastivity encoded in the elements of the sentence, and they divide it into two sub-categories based on directiveness: directly contrastive and indirectly contrastive. Due to the problems in defining the notion ‘contrastive’, Lambrecht (1994) views the notion ‘contrastive’ not as a category of grammar, but as a gradient notion, following Bolinger (1961). Simply put, any element in the clause can be contrastive in a broad sense. More specifically, there can be a contrast between a particular element in the clause and anything else, though there is no specific contrast with that particular element. Of course, the degree of contrastivity becomes stronger as the alternatives are narrowed down. Instead of viewing the notion ‘contrastiveness’ as a category of grammar, Lambrecht (1994) suggests that the impression of contrastiveness arises from particular inferences that are drawn on the basis of given conversational contexts. Lambrecht (1994) also points out that anything unusual can be potentially interpreted as contrasting with a more usual alternative.

The term ‘contrastive’ is difficult to define due to the fact that anything can inherently be contrastive in a certain sense, thus it is more appropriate to define the term ‘contrastive’ in a gradient approach (rather than single-lined term). Therefore, contrastivity should be understood not as a category of grammar but rather in a non-grammatical sense, as the result of the general cognitive processes also referred to as ‘conversational implicatures’.

Along with the term ‘contrast’, the notion ‘exhaustive-listing’ (Kuno 1972; 1973), also known as ‘narrow focus’ (Lambrecht 1994), has often been discussed, because an element that is referred to as ‘exhaustive-listing’ is considered to contrast with things other than the given element itself. Hence, the present study includes tokens of an exhaustive-listing into the entire body of contrastive elements. That is to say, the present study follows Clancy & Downing’s (1987) and Kuno’s (1972; 1973) descriptions given for the term ‘contrast’. More specifically, the term ‘contrast’ is categorized largely into three different sub-types: ‘directly contrastive’, ‘exhaustive-listing’, as well as ‘directly contrastive and exhaustive-listing’. Following approaches used by Clancy & Downing (1987), the ‘directly contrastive’ type of contrastiveness is classified more precisely, based on its intended description, into two different sub-types: ‘parallel activities/states’ and ‘action/state reaction’.

The ‘parallel activities/states’ type of directly contrastive occurs in the format of “A does/is X, but B does/is Y.” The ‘action/state reaction’ type of directly contrastive typically appears “when the speaker is developing a plot sequence containing successive activities involving different participants” (Clancy & Downing 1987:37). Shimojo (2005:176) also points out that “[t]he paired elements in ‘action/state reaction’, unlike those in ‘parallel activities/states’, do not appear in parallel; but the first element is presented, and then the second element is (re)introduced into the state of affairs with regard to the first element.” The following exemplifies two subtypes of directly contrastive.

(2.4) *ccacangmyen-un an mek-ko, kunayng pap mek-ess-ta, olaynnamaney.*  
 black.noodle-top NEG eat-and normal meal eat-PST-DCL longtime  
 ‘(I) didn’t eat black noodles, but (I) just ate a normal rice meal, (which I haven’t had for a while).’

(2.5) A1: [*ne-Ø ppalli chayk-Ø kac-ko wase kongpwu-Ø hay*]<sub>1st</sub>  
 2SG-( ) right.away book-( ) take-and come.and study-( ) do

‘You bring your books (to the library) and study right away.’

A2: [*hyeng-Ø ecey kongpwu-Ø hay-ss-e*]<sub>2nd</sub>.  
 brother-( ) yesterday study-( ) do-PST-SEM  
 I studied yesterday (lit. Brother studied yesterday).’

The utterances in (2.4) show an instance of ‘parallel activities/states’ type of contrast. More specifically, ‘did NOT eat a black noodle’ and ‘did eat a normal rice meal’ in (2.4) are regarded as parallel activities and they represent the contrastive opposition. The utterances in (2.5) exemplify an ‘action/state reaction’ type of contrast, in the sense that the utterance A2, which is in the second set of brackets, is made in a reaction to the utterance A1, which is in the first set of brackets.

It has been widely accepted that in Korean, as well as in Japanese, the contrastivity of a particular element in the clause is marked by a morphological device, i.e., *nun* in Korean, *wa* in Japanese. In (2.6), *Junsu* in the first clause is perceived as contrasted with *na* ‘I’ in the second clause, and the post-nominal particle *nun* marks their contrastivity.

(2.6) *Junsu-nun Seoul-ey sal-ko, na-nun Buffalo-ey san-ta.*  
 Junsu-TOP Seoul-LOC live-and 1SG-TOP Buffalo-LOC live-DCL  
 ‘Junsu lives in Seoul, and I live in Buffalo.’ [contrastive]

In (2.6), the subject, i.e., *Junsu*, is clearly in contrastive relation with another element *na* ‘I’, in that *na* ‘I’ plays a role as a possible alternative that is accessible to both a speaker and an addressee at the time the speaker makes this utterance. Maynard (1980) also points out, based on her study of the post-nominal particle *wa* in colloquial Japanese, that the contrastive effect arises when the NP-*wa* appears in a context where alternatives (i.e., the other pole of the contrastive relation) are identifiable. Nonetheless, it should still be pointed out that the distinction between contrastive and non-contrastive is not clear-cut, but a matter of degree (Clancy & Downing 1987). This

is because contrast inherently assumes a presupposition, so any element can potentially be perceived as contrastive in a certain sense.

- (2.7) a *John-i Seoul-ey ka-ss-ni?*  
 John-NOM Seoul-LOC go-PST-Q  
 ‘Was it John who went to Seoul’
- a’ *John-Ø Seoul-ey ka-ss-ni?*  
 John-( ) Seoul-LOC go-PST-Q  
 ‘Did John go to Seoul’
- b *Junsu-ka Seoul-ey ka-ss-eyo.*  
 Junsu-NOM Seoul-LOC go-PST-SEM  
 ‘It was Junsu (not John) who went to Seoul.’
- b’ *John-Ø Seoul-ey ka-ss-eyo.*  
 John-( ) Seoul-LOC go-PST-SEM  
 ‘John went to Seoul.’
- b” *??Junsu-Ø Seoul-ey ka-ss-eyo.*  
 Junsu-( ) Seoul-LOC go-PST-SEM  
 ‘Junsu went to Seoul.’

In Korean, the contrastivity of an element may be marked by the post-nominal particle *nun* as seen in (2.6), but it can also be marked by *ka* as shown in (2.7). The subject *Junsu* in (2.7b), which is given as an answer to (2.7a), contrasts with *John* in (2.7a) as can be seen in English translations. On the other hand, (2.7b”) is not acceptable if it is given as an answer to (2.7a) because the contrastivity of the subject *Junsu* in (2.7b”) is not encoded by a zero particle. Still, the contrastivity marked by the nominative *ka* is distinguished from that marked by *nun*. The former is generally viewed as a contrastive focus (or an exhaustive-listing), and the latter as a contrastive topic (Lee 2003), although Choi (1997) claims that the topicality of an argument is encoded not by the post-nominal particle *nun* but by scrambling, and she views the post-nominal particle *nun* not as a topic marker but a contrastive marker. Simply put,

the contrastivity of an element in the clause can be marked not only by *nun*, namely a contrastive topic marker, but also by *ka*, a so-called contrastive focus marker in Korean.

In summary, in the present study, the notion of contrast is categorized into four different subtypes, and the choice of subject and object encoding types is discussed based on these four different subtypes, with an assumption that contrastive elements represent informational prominence.

## 2.7 Word order and post-predicative encoding

In general, a sentence consists of one predicate and one or more arguments that are required by the predicate. The argument can be a word, phrase or clause, which functions as the subject, direct object, or indirect object in a sentence.

Korean is typologically an SOV language, as it has the basic word order of subject, object, and verb (Sohn 1999). It is also well-known that, as in Japanese, Korean is a language that allows pre-predicative scrambling (Cho 1994; Choi 1996; Choi 1997). Word ordering of pre-predicative entities such as subject and object is flexible, and an object can precede a subject as long as the clause ends with a predicate. In other words, post-predicative entities are normally not regarded as a part of the preceding clause.

Non-verbal elements can appear after the predicate in colloquial speech, and they are commonly viewed as as a ‘afterthought’, not as a post-predicative argument (Kuno 1978). However, not all post-predicative entities should be precluded, because some often convey an essential part of the proposition, and they can also be the arguments of the preceding clause, as was discussed in Shimojo (2005). For this reason, the instances of post-predicative entities are distinguished from those of ‘after-thoughts’ in this dissertation, mainly based on the presence or absence of a significant pause between the predicates and those post-predicative entities.

Post-predicative encoding<sup>8</sup> is often viewed as the result of speakers' production-based motivation, namely 'important information first' proposed by Simon (1989: 189), who writes that:

“[p]ostposings are simply results of important or urgent information coming to the speaker's mind first and thus being vocalized first, especially under time pressure... The phenomenon under consideration there is more like a case of 'syntactic metathesis', so to speak, in which a certain element appears post-verbally instead of in its normal position; it is simply the way a sentence comes out.”

However, Shimojo (2005) views this production-based motivation approach to post-predicative encoding as problematic. Instead, he proposes comprehension-based arguments for post-predicative encoding.

It is a possible assumption that the properties of subject and object markings in conversational Korean cannot be considered separately from word ordering. Furthermore, it may be assumed that subjects and objects are marked with overt particles more frequently in so-called non-canonical word order (e.g., OSV, OVS, etc.) than in canonical word order (i.e., SOV) for the sake of processing load as well as informational prominence. For instance, the identification of subject and object in a clause tends to be more complex in processing with non-canonical word order than with canonical word order. For this reason, the present study examines whether there is a correlation between subject/object markings and word ordering in conversational Korean.

## 2.8 Case markers

Korean is a nominative-accusative language with morphological case particles. The subject is marked with a case particle conventionally known as the 'nominative' marker,

---

<sup>8</sup>In general, post-predicative encoding is distinguished from scrambling, which only allows pre-predicative mutation of the canonical SOV word order (Kuno 1978; Choi 1996; Yamashita 2002).



whereas the object is marked with another case particle called the ‘accusative’ marker (Hong 1991; Sohn 1999). However, there have been several debatable issues in this regard (Han 1999; Ko. E 2000; Ko. S 2000; Jun 2003; Bak 2004), as summarized below.

These issues are rooted in the fact that the post-nominal particles which mark subjects and objects, *ka* (nominative marker) and *lul* (accusative marker) respectively, often appear more than once in a simple clause. This phenomenon in Korean is known as ‘double nominative and accusative constructions’<sup>9</sup> (Yang 1999; Kim 2000; Schütze 2001; Um 2003; Yoon to appear, inter alia). Many studies (e.g., Han 1999; Yang 1999; Ko. S 2000; Yoon 2007, inter alia) have attempted to explain this phenomenon in various ways. Some argue that Korean is a language which allows multiple subjects and objects in a simple clause (Yoon 2007). Others argue that they are not case markers, but pragmatic markers (e.g., focus marker) in Korean (Han 1999; Lee 2003). Still others argue that they are determiners, also referred to as delimiters, just like *man* ‘only’ and *to* ‘also’ (Mok 1998).

(2.8) *Junsu-ka cha-ka manhta.*  
 Junsu-NOM car-NOM be.many  
 ‘Junsu has many cars.’

(2.9) *Junsu-eykey cha-ka manhta.*  
 Junsu-DAT car-NOM be.many  
 ‘Junsu has many cars.’

(2.8) exemplifies the ‘double nominative construction’ in Korean, and its proposed meaning is basically the same as that of (2.9) as seen in its English translation.<sup>10</sup> This suggests that the noun *Junsu* in (2.8) and (2.9) has the same grammatical relation

---

<sup>9</sup>Bak (2007) claims that the ‘double nominative’ construction differs from the ‘multiple nominative’ construction in several ways, and that they should not be treated the same. They are, however, not distinguished in this dissertation.

<sup>10</sup>Some may argue that there is a slight difference in meaning between these examples, in that *Junsu* in (2.8) gets a (contrastive) focus whereas it only receives a neutral interpretation in (2.9).

with the adjectival predicate *manhta* ‘be many’. This raises several interesting issues which need to be discussed. First, why are two different particles utilized to mark the same argument? Second, the noun *Junsu* in (2.8) may not be the subject of the clause, even though the nominative marker *ka* is attached to it. If the noun *Junsu* is not the subject, what is the property of the particle *ka* in (2.8), which has been claimed to mark the subject? Does the particle *ka* have properties other than marking a subject in a clause? Let us now look at examples with the double accusative marker.

(2.10) *Junsu-ka na-lul cha-lul cwuessta.*  
 Junsu-NOM 1SG-ACC car-ACC gave  
 ‘Junsu gave me a car.’

(2.11) *Junsu-ka na-eykey cha-lul cwuessta.*  
 Junsu-NOM 1SG-DAT car-ACC gave  
 ‘Junsu gave me a car.’

(2.10) demonstrates the ‘double accusative construction’ in Korean. Its meaning is the same as that of (2.11) as seen in the English translation.<sup>11</sup> This raises the same issues which were presented above: why different particles are used to mark the same argument, and what other properties the particle *lul* has than marking the object in a clause. This issue extends to the following examples.

(2.12) *Junsu-ka cha-ka cohta.*  
 Junsu-NOM car-NOM be.nice  
 ‘Junsu’s car is nice.’

(2.13) *Junsu-uy cha-ka cohta.*  
 Junsu-GEN car-NOM be.nice  
 ‘Junsu’s car is nice.’

(2.14) *Junsu-ka na-lul son-ul capassta.*  
 Junsu-NOM 1SG-ACC hand-ACC grabbed  
 ‘Junsu grabbed my hand.’

---

<sup>11</sup>Again, some may say that there is a slight difference in meaning between these examples.

- (2.15) *Junsu-ka na-uy son-ul capassta.*  
 Junsu-NOM 1SG-GEN hand-ACC grabbed  
 ‘Junsu grabbed my hand.’

The sentences in (2.12) through (2.15) are additional instances of double nominative and double accusative constructions, but they differ from the sentences in (2.8) through (2.11) in several respects. The underlined NPs in (2.8) through (2.11) are more closely related to their associated predicates, whereas those in (2.12) and (2.15) are more closely related to their associated nominal entities, as seen in the English translations. In other words, the underlined NPs in (2.12) and (2.14) do not have direct grammatical relations with the predicates, yet the same particle (i.e., *ka* and *lul*) is used in both cases. The nominative particle *ka* in (2.12) and the accusative particle *lul* in (2.14) can be replaced with the genitive marking particle, as shown in (2.13) and (2.15). This phenomenon is well discussed in Cho (2003). This suggests that the particles *ka* and *lul* have properties other than to mark a subject and an object in a clause, since the genitive marked NP can be neither a subject nor an object in the clause.

As to these phenomena, Nam (1996) claims that *ka* and *lul* are not case markers, but particles which have specific meanings. Ko. S (2000) also argues that they are not syntactic case markers. Instead, he attempts to account for *ka* and *lul* in semantic and pragmatic contexts only. Unlike Nam (1996) and Ko. S (2000), Jun (2003) claims that the case marking system exists in Korean, but it should be accounted for from both syntactic and semantic perspectives. Han (1999) also claims that they are case markers, but that they function both as semantic and pragmatic case markers. Similar to Han (1999), Choi (1996) and Lee (2003) argue that they are contrastive focus markers. Hence, it is untenable to claim that the post-nominal particles *ka* and *lul* are not case markers, or that there is no case marking system in Korean, because there are cases which support the view that they are case markers.<sup>12</sup> Furthermore, Schütze (2001)

---

<sup>12</sup>Blake (2001) reports, based on his typological study of 100 languages, that 46 languages have a

suggests that the particle *ka* may not be restricted to a case marking system. In an extension to Schütze's (2001) suggestion, I claim that the case markers *ka* and *lul* not only mark the subject and object, but they are also used when processing load and informational prominence are expected, which is represented by the subject and object. Let us look at more examples.

(2.16) *Junsu-Ø cha-Ø cohta.*  
 Junsu-( ) car-( ) be.nice  
 'Junsu's car is nice.'

(2.17) *Junsu-ka Jisu-Ø son-Ø capassta.*  
 Junsu-NOM Jisu-( ) hand-( ) grabbed  
 'Junsu grabbed Jisu's hand.'

As was mentioned earlier, it is well-known that case markers in Korean, as well as in Japanese, are not often expressed overtly, depending on contexts. Such a phenomenon is demonstrated in (2.16) and (2.17). Some may take this phenomenon as supporting evidence for the argument that the particles *ka* and *lul* are not case markers because they are not on the surface to mark subject and object. Yet it may simply be the case that case markers in Korean are optional, and that the occurrence or non-occurrence of these markers may depend on contexts.

Most of the previous studies (e.g., Kim 1990; Hong 1991; Han 1999; Ko. S 2000; Jun 2003, inter alia) have been done only to show whether the particles *ka* and *lul* are case markers or what case features they have. Although it is still under much debate, whether they are case markers or not is not of much concern in this dissertation. What is of more concern in the present study is when these markers are and are not overtly expressed.

The following are additional examples of case marking in Korean.

---

case system, and the other 54 languages do not. Blake (2001) classifies Korean as a language with a case system.

- (2.18) a. *na-nun photo-lul mek-ko siphta.*  
 1SG-TOP grape-ACC eat-CONN want  
 ‘I want to eat grapes.’
- b. *na-nun photo-ka mek-ko siphta.*  
 1SG-TOP grape-NOM eat-CONN want  
 ‘I want to eat grapes.’
- (2.19) a. *na-nun tosisaynghwal-ul cohaha-ko siphta.*  
 1SG-NOM urban.life-ACC like-CONN want  
 ‘I want to like city life.’
- b. *?na-nun tosisaynghwal-i cohaha-ko siphta.*  
 1SG-NOM city.life-NOM like-CONN want  
 ‘I want to like urban life.’

It is notable that in Korean, there are certain constructions in which the so-called nominative marker *ka* appears where the so-called accusative marker *lul* is expected. (2.18) and (2.19) generally demonstrate the same construction. However, both nominative and accusative markers are natural for the nominal *photo* ‘grape’ in (2.18), whereas only the accusative marker is natural for the nominal *tosisaynghwal* ‘urban life’ in (2.19). This data brings up several interesting issues.

First, the particle *ka* can sometimes be used to mark the object,<sup>13</sup> and it can happen only with those transitive verbs that can be combined with certain auxiliary verbs such as *siphta* ‘want’ (Um 2003). Yet why do only limited types of predicates allow the so-called nominative case particle to mark the object in the clause?

Second, if (2.18) and (2.19) are the same construction, why is the nominative marker *ka* allowed only in (2.18), but not in (2.19)? Are there constraints which make such a distinction, and so, what would they be?

Third, what are the functions of the nominative marker *ka* when it is used to mark the object? Although a substantial amount of work has been done on double nominative

---

<sup>13</sup>This phenomenon is generally called ‘case alternation’ or ‘case shifting’, distinguished from other ‘double nominative constructions’ in Korean (Um 2003; Yoon 2004, *inter alia*).

constructions in Korean, not enough attention seems to have been paid to the fact that the particle *ka* can sometimes mark the object, and that this phenomenon is limited to particular types of transitive verbs. A similar phenomenon is observed in Japanese, and it is discussed in Kuno (1973:79-95).

At this moment, I cannot provide plausible justifications for these issues, but I propose that the use of *ka* for marking the object may correlate with certain pragmatic factors (e.g., contrast, focus, etc.). In Chapter 5, this phenomenon is discussed in further detail.

Finally, it is worth emphasizing that such double nominative and accusative constructions do not entail double subjects and objects, though it is often claimed that there exist multiple subject constructions in Korean (Yoon 2007; Yoon to appear). In this dissertation, only one nominative and one accusative NP are identified as the grammatical subject and object in a clause.<sup>14</sup> For example, only the first nominal, *na* ‘I’, in (2.18) is a subject, and the second one, *photo* ‘grape’, in (2.18) is an object, though the latter is marked with the nominative *ka*. This issue, as well as the identification of subject and object, is discussed later in further detail.

## 2.9 Zero particle and overt particles

As shown in (2.16) and (2.17), subjects and objects in Korean often are not explicitly marked with their particles. The general explanation for this phenomenon is that case markers in Korean may be omitted when they can be derived from the context (Sohn 1999). In other words, case markers may be deleted when the case relations of their associated arguments are obvious. Another possibility of accounting for this phenomenon is that case markers are not expressed overtly when they are unnecessary

---

<sup>14</sup>Keenan (1976) labels this grammatical subject as ‘basic subject’.

for the identification of grammatical relations of their associated arguments (Min 1982; Kim 2006). Some may accept neither account for this phenomenon, but may argue that a speaker will choose either zero-marking or overt marking for subjects and objects depending on the given contexts. In short, there are three ways of viewing the same phenomenon depending on what they take as a default: case markers may be deleted; case markers may appear overtly; either an overt case marker or a zero particle (or called a covert case marker) may be selected. Nonetheless, it is not of great concern in the present study if the case markers are omissible, expressible or selectable. What is of more concern is when subjects and objects are marked with a zero particle, and when they are explicitly marked with overt particles (e.g., *ka*, *lul*).

Some note that not all zero particles in Korean, as well as in Japanese, should be treated the same. More specifically, there are cases where the zero particle is optional, and cases in which it is obligatory. The former is called the ‘optional zero particle’, and the latter is called the ‘obligatory zero particle’ (Min 1982; Shibatani 1990; Ko. S 2000; Shimojo 2005, *inter alia*). See the examples below.

- (2.20) a. *ne, pap-Ø mek-ess-ni?*  
 2SG meal-( ) eat-PAST-Q  
 ‘Did you eat breakfast (or lunch/dinner)?’
- b. *e, na-Ø/??nay-ka/#na-nun pap-Ø mek-ess-e.*  
 yeah 1SG-( )/1SG-NOM/1SG-TOP meal-( ) eat-PAST-SEM  
 ‘Yes, I had breakfast (or lunch/dinner).’

(2.20b) is most natural, as an answer to (2.20a), when pronominal *na* ‘I’ occurs with a zero particle. On the other hand, the use of *ka* or *nun* makes (2.20b) an unnatural answer to (2.20a), unless focus or contrastive reading is allowed. Again, the general analysis for this kind of example is that the case marker in Korean is optional, so it is not always required, particularly when it is easily retrieved from the given context (Sohn 1999). Yet there is a problem with this analysis. If the case marker is simply

optional, this example should sound natural with the use of either a zero particle or an overt particle. However, (2.20b), as an answer to (2.20a), seems to be natural only with the use of a zero particle, when no contrast or focus reading is encoded in the utterance. For this reason, some argue that the optional zero particle should be distinguished from the obligatory zero particle in Korean (Min 1982; Ko. S 2000). A similar claim has also been addressed in several studies pertaining to a phenomenon similar to this matter in Japanese (Shibatani 1990; Hasegawa 1993; Lee 2002; Shimojo 2005), in that the zero particle is obligatory in some cases, whereas it is optional in other cases. The following examples are taken verbatim from Shimojo (2005:37).

- (2.21) a. *Tanaka-san*       $\emptyset/ga/wa$       *inakunarimasita.*  
 Tanaka-Mr./Ms. ( )<sup>15</sup>/NOM/TOP disappeared  
 ‘Mr./Ms. Tanaka has disappeared.’
- b. *watasi*  $\emptyset/*ga/?wa$       *bikkurisimasita.*  
 I ( )/NOM/TOP was.surprised  
 ‘I was surprised.’

Shimojo (2005:37-8) illustrates that in general, (2.21a) is fine with any of the three particles. Yet in (2.21b), the zero particle is the best choice, and the other two particles are either unacceptable or less preferred. However, there is an inherent problem with analyzing the use of zero particle simply as either optional or obligatory because the optional/obligatory distinction in the use of zero particle is not crystal clear, as Shimojo (2005:38) points out that “the choice of a particle is dependent upon how compatible the pragmatic function of the given utterance is with the function associated with the particular particle in question.” In other words, (2.20b) and (2.21b) are also acceptable with overt particles (e.g., *ka* in (2.20b); *ga* in (2.21b)) in certain discourse contexts.

In summary, it is clear that the zero particle is preferred to the overt particle in certain cases, while the preference is the other way around in other cases, though it

---

<sup>15</sup>The notation ‘( )’ indicates a zero particle.



may still be debatable that the optional zero particle should be distinguished from the obligatory one. In this dissertation, I investigate when the zero particle is preferred and when it is not preferred to mark subjects and objects in conversational Korean, regardless of whether the zero-marking is obligatory or optional, mainly from the viewpoints of processing load as well as informational prominence.

## 2.10 Saliency, givenness, and activation

The terms ‘saliency’, ‘givenness’, and ‘activation’ have been used in many discourse studies (e.g., Lee & Thompson 1989; Chafe 1994; Dryer 1996; Fujii & Ono 2000; Shimojo 1995; Shimojo 2005; Yamasaki 2005). These notions are expected to play important roles in examining the subject and object markings in conversational Korean, and this section is devoted to the illustration of these terms.

### 2.10.1 Saliency

The term ‘saliency’ has been utilized with several different meanings. For instance, Chafe (1994) uses ‘saliency’ in the sense of identifiability, in that a referent is salient if it is distinctive linguistically or even extra-linguistically in a certain context. Lambrecht (1994) uses the term ‘saliency’ associated with the topicality of referents, in that active referents, compared to inactive or semi-active referents, are most acceptable as a topic. In some studies, the term ‘saliency’ is illustrated in the sense of activation, and in fact ‘salient’ is often viewed as synonymous with ‘activated’ (Gundel *et al.* 1993; Chafe 1994; Dryer 1996).

It is generally assumed that a referent is continuously activated if the particular referent continues to be focused. The referent is assumed to be ‘salient’ if it is continuously activated in one’s cognitive focus of attention. The ‘salient’ referent can be

viewed as the ‘prominent or important’ referent, based on such an assumption that the important information tends to be continuously talked about, thus continuously activated and salient.

I argue that the notion of ‘saliency’, in the sense of informational prominence, is relevant to the discussion of the subject and object markings in conversational Korean. More specifically, explicit particles are preferred for marking subjects and objects that represent salient referents, due to the informational prominence, which is encoded in the referents. In this dissertation, the saliency of entities is measured by referential persistence, which is discussed in later sections.

### **2.10.2 Givenness**

Utterances are usually produced not in isolated context, but some context which involves what has previously been said, or what is known to the speaker and the hearer. Hence, some part of an utterance denotes given information, while another part introduces new information into the discourse. The given information represents the part that is already known to the hearer in one way or another, but the new information serves as a main part of the utterance in terms of information content. It is noted that instead of terms *given* and *new*, the terms *theme* and *rheme*, or *topic* and *comment*, are sometimes used as equivalent notions (Trask 1999).

Given information is generally viewed as the information that is predictable or recoverable from the preceding context (Chafe 1994; Shimojo 1995; Dryer 1996, *inter alia*). It is also viewed as the shared knowledge, which is usually regarded as presupposition (Lambrecht 1994:52). The given information is often illustrated in the sense of the activation and deactivation of information. More specifically, given information is already active in the addressee’s consciousness at the time of an utterance, whereas new information is newly activated in the addressee’s consciousness at the time of an

utterance (Chafe 1994:72). Prince (1981:236) notes, “a discourse entity is Inferrable if the speaker assumes the hearer can infer it, via logical—or, more commonly, plausible—reasoning, from discourse entities already Evoked or from other Inferrables.” Although the notion ‘givenness’ has been described in several different ways, one thing that seems to be common is that given information is more accessible to both a speaker and an addressee than new information.

Lee & Thompson (1989) note that the greater the sharedness between communicators, the less necessity to specify grammatical relations. I also argue that the term ‘givenness’ is relevant to the discussion of the subject and object markings in conversational Korean, particularly in that subject and object NPs with shared, thus given, information tend to be marked with a zero particle. More specifically, subjects and objects which denote given information are more accessible, thus more easily processed, to both the speaker and the hearer, so the zero particle is preferred as their marker, unless other factors influence their marking. Furthermore, an NP tends to be de-emphasized or defocused when the NP represents shared information between the speaker and the hearer (Masunaga 1988). Based on this, I hold that the NP with shared, and therefore given, information is not considered to be prominent, and thus it tends to be marked with a zero particle.

I further argue that the notion ‘givenness’ is relevant to the discussion of the subject and object markings in conversational Korean in two respects, *accessibility* and *prominence*, more specifically in that the subject and object entities with given information prefers to be marked with a zero particle, because given information is not only more accessible but also less prominent than new information. In this dissertation, the givenness of entities is determined primarily based on how recent as well as how frequently an entity is mentioned in anaphoric discourse. The recency of referents is measured by the referential distance (RD), and the frequency of referents is measured by the anaphoric

frequency (AF), which are both discussed in later sections.

### **2.10.3 Activation**

Just like the notions ‘saliency’ and ‘givenness’, the notion ‘activation’ can refer to several different notions: active concepts, semi-active concepts, and inactive concepts in Chafe’s (1994) terms; focus of attention, activated, semi-activated, and non-activated in Dryer’s (1996) terms. Nevertheless, they all refer to attention.

Yamasaki (2005:39) states that the “recent cross-disciplinary studies in experimental psychology and cognitive linguistics center around the notion of ‘attention’.” She notes that the concept of ‘cognitive attention’, or simply ‘attention’, parallels the linguistic concept of ‘activation’. She further states that linguistic processes are instantiations of more general cognitive processes. Dryer (1996:482) attempts to map certain linguistic forms to certain cognitive processes. He points out that the activation is in fact the status of cognitive entities, and its relevance to the field of linguistics “lies in the fact that certain linguistic choices (like pronoun vs. noun, or active vs. passive, or position of focal accent) may be systematically related to, if not determined by, the activation status of cognitive entities.” He also points out that in some cases, it may be the activation status of cognitive entities in the mind of the speaker that are relevant, and in other cases, it may be the speaker’s assumption about the activation status of corresponding cognitive entities in the mind of the hearer that are relevant.

Dryer (1996) notes that “the status of activation of an entity can be measured by examining recency, frequency, or inferability of its previous mention in anaphoric discourse.” For example, the more frequently an entity is mentioned in the previous discourse, the more prominently the referent of the entity is registered in the speaker’s and the hearer’s consciousness, and is thus more activated. Similarly, the more recently the entity is mentioned, the fresher its referent is in the speaker’s and the hearer’s mind,

and thus the more activated it is (Chafe 1994; Tomlin 1995; Dryer 1996).

Dryer (1996) further notes that the activation is a continuous notion, not a binary value of active and nonactive. An entity becomes activated when it is mentioned for the first time. Once the entity gets activated, it may become deactivated and then fade away, or may remain activated. The entity eventually decays if it is not mentioned further. Dryer (1996) proposes, based on the assumption that activation is a continuum, that there are entities which are less activated than others, but yet more activated than deactivated or decayed entities, and he labels those intermediate entities as ‘semi-deactivated’ entities.

Notions of activation and deactivation can be useful for the discussion of subject and object markings in informal conversational Korean, because a speaker’s selection of one particular post-nominal particle over another is related to the activation status of the linguistic entities to which that particle is attached. More specifically, I argue that the zero particle is not preferred for semi-deactivated or deactivated entities, which are generally assumed to be less accessible, if not completely inaccessible. In this dissertation, the activation status of an entity is measured by the referential distance (RD) as well as anaphoric frequency (AF) which are discussed in the following section.

## **2.11 Referential distance and referential persistence**

The notions ‘referential distance’ as well as ‘referential persistence’ are expected to be related to the speaker’s selection of one particular particle over another marking a subject and an object in conversational Korean. Hence, an overview of these terms is provided in this section. In this dissertation, referential distance is used to measure anaphoric saliency (e.g., givenness, activation status), and referential persistence is used to measure cataphoric saliency (e.g., informational prominence) of the referent.

### 2.11.1 Referential distance

The term ‘referential distance’ (henceforth, RD) is introduced as one of the quantitative measurements which is extensively discussed in Givón (1993) and frequently utilized in subsequent text analyses in many languages (Shimojo 2005, *inter alia*). RD is used to indicate the linguistic distance in clausal units, and it is measured by counting clausal units backward to the most recent representation of the coreferential expression which usually includes those of zero anaphor (Givón 1983). For instance, an RD 1 indicates that the most recent representation of the current referent in question is made in the immediately preceding clausal unit, and an RD 2 indicates that the most recent coreferential expression is represented in two clausal units previous. In short, RD is the distance to the most recent co-referential expression.

RD measurement has been used for individual referents, and can be an appropriate method to examine the level of activation associated with the given referent (Shimojo 1995, 2005). For example, if a speaker refers to a particular referent in conversation, the referent becomes activated in a hearer’s consciousness. As the hearer, as well as the speaker, processes the information, the referent that was activated moments ago starts to decay and finally becomes deactivated (Dryer 1996). Thus, a referent of RD 1 may be assumed to be more activated than a referent of RD 10 at a given point in discourse, provided that all other things are considered equal.

However, as Shimojo (2005) notes, RD only provides an estimating measure for the level of activation. RD is simply a heuristic way of measuring the level, and RD by itself may not be of great cognitive significance (Givón 1993). There are other factors, which also influence the level of activation. For example, referents can be activated by way of other activated referents, if they are associated with each other in one way or the other (Chafe 1994; Shimojo 2005). Entities that are visually available in the conversational context can be activated more easily than those that are not available. In spite of all

the potential factors which can affect the activation level of referents, RD seems to be the only quantifiable anaphoric measurement currently available, as Shimojo (2005) points out. Thus, RD is used to measure the anaphoric saliency, namely the activation status of a referent in this dissertation.

The activation status of a referent is usually measured by examining the recency of a referent, but it can also be measured by examining how frequently it is mentioned in anaphoric context, as Dryer (1996) states. Hence, the anaphoric saliency of the referent, in the present study, is measured by not only counting clausal units backward to the most recent representation of the coreferential expression, but also by counting the number of coreferential expressions in the anaphoric context. In order to distinguish the latter case (*frequency*) from the former case (*recency*), the ‘frequency of coreferential expressions’ is called ‘anaphoric frequency’ and is labeled as AF in this dissertation. AF1 indicates that there are one coreferential expression within a given anaphoric context, and AF5 means that there are five coreferential expressions available in the anaphoric context. In general, the more frequently a referent is previously mentioned, the more likely it is activated. Thus, a referent of AF5 can be assumed to be more activated than a referent of AF1 at a given point in discourse, provided that all other things are considered equal.

The RD, as well as AF, measurements can also be used for propositions, as well as for referents (Shimojo 2005). In this dissertation, RD and AF are utilized only for the referents. When RD is measured, 20 preceding clausal units are examined for coreferential expressions, following previous studies (e.g., Givón 1983; Shimojo 2005). When AF is measured, 10 preceding clausal units are looked at for coreferential expressions. It is notable that to the best of my knowledge, there have been no previous attempts made to quantify the level of activation or the accessibility by measuring the frequency of coreferential expressions in the anaphoric context, though it has been suggested by

Dryer (1996).

It is generally assumed that processing requires an activation cost, and the cost depends on the activation status of information prior to the processing. Information which is activated in one's consciousness can be processed at little activation cost, while deactivated and semi-activated information seems to require a greater activation cost. Based on this, the RD and AF measurements are expected to play an important role in examining the speaker's choice of one particular particle over others marking subjects and objects in conversational Korean. More specifically, I propose that the use of a zero particle is not preferred for marking subjects and objects when RD is greater and AF is lower.

### **2.11.2 Referential persistence**

The notion of 'referential persistence' (henceforth, RP) is introduced to measure the degree of decay of information in the cataphoric context (Givón 1983). Information which does not persist in the cataphoric context loses its saliency in the subsequent context. Givón (1983:15) states that RP is basically a way of measuring importance in the sense that "[m]ore important discourse topics appear more frequently in the register, i.e., they have a higher probability of persisting longer in the register after a relevant measuring point."

RP can be measured by counting "the number of clauses to the right from the locus of study that the same referent remains an argument of the predication" (Hinds 1983:58). RP can also be measured by counting the number of coreferential expressions within the following ten clausal units, as used in previous studies (e.g., Shimojo 2005). For instance, RP1 indicates that there is only one clausal unit containing the coreferential expression within the range of the following ten clausal units. RP5 indicates that there are five clausal units within the range. When RP is measured in this disserta-



tion, following Watanabe (1989), a coreferential expression is counted, regardless of its grammatical status, if it is overtly present. However, if it is not overtly present, only those expressions that are considered arguments (i.e., zero anaphor) are counted.

Shimojo (2005:99) points out that the simple frequency measurement of coreferential expressions within ten cataphoric clausal units “does not directly reflect saliency in the sense of cognitive focus of attention, i.e. continued activation in mind.” Simply put, RP1 may or may not reflect continued activation. For instance, RP1 could mean that the coreferential expression is found in the immediately following clausal unit without interruption. Yet it could also mean that the coreferential expression is not found until the last clausal unit within the specified cataphoric range. For this reason, Shimojo (ibid.) argues that referential persistence should be “measured more precisely if the RP value indicates uninterrupted cataphoric representation of a referent in question.” For example, RP indicates the distance in clausal units for which a particular referent continues to be present without interruption in the cataphoric context. In order to distinguish the latter case from the former case, he labels ‘uninterrupted persistence’ as RP, and ‘persistence in frequency’ as RP-*f*. Shimojo’s argument is valid, thus such a distinction is made when the cataphoric saliency is measured in this dissertation. In other words, cataphoric saliency is measured not only by RP (*uninterrupted persistence*), but also by RP-*f* (*frequency*) in this study.

Measurement of RP, as well as RP-*f*, is assumed to be important in this dissertation, in the sense that the cataphoric saliency of a particular referent is closely related to the degree of importance of the referent. More specifically, overt particles are assumed to be used for marking subjects and objects in conversational Korean when RP and RP-*f* values are high. Based on this assumption, the present study examines the speaker’s selection of subject and object encoding types with relation to RP and RP-*f* in conversational Korean.

# Chapter 3

## Conversational Korean Data

### 3.1 Outline

This chapter presents an overview of the informal Korean conversational data that is used in this dissertation. It includes a description of the procedure used for data collection and preparation for data analyses. It also discusses the methodology that is utilized to analyze the data used in this dissertation, and further presents the overall token counts in terms of the forms of subject and object markings found in informal conversational Korean.

### 3.2 Data collection procedures

The discourse data used in this dissertation is comprised of two-party informal conversations by ten pairs of Korean native speakers, which were audio-recorded. Each pair's conversation is 31-35 minutes long, and the entire volume of discourse data consists of about five hours of informal conversation. The overall procedures for data collection are based on Shimojo (2005:55-6) and are summarized as follows.

The recordings were all done at the University at Buffalo, Buffalo, New York, from

November through December of 2006. As summarized in Table 3.1, all of the participants were native speakers of Korean, who were born, raised, and lived most of their lives in Korea. The age of participants ranged from 20 to 28 years old. The longest duration of their stay abroad at the time of the recording was 15 months at most, but most of them had stayed abroad only for three to seven months. The paired participants were mutual friends, and many considered their partner ‘a good friend’ or ‘a very good friend’ as seen in Table 3.1.

The nature of all the conversations is considered to be informal, and the casual nature of the conversation is indicated by the predicate forms and sentence endings (e.g., *-e*, *-yo*), which were utilized in each pair’s conversation. In order to eliminate the potential effect that may have been caused by gender differences, the population of participants was controlled. There were four pairs of only female participants, three pairs of only male participants, and three pairs of male and female participants, as shown in Table 3.1. Most participants, except two who were born and raised in the southern region, were from Seoul or the greater Seoul area, and they are all considered to speak the Seoul dialect. In this dissertation, those participants from the southern area are also considered to speak the Seoul dialect, which they acquired while attending colleges in Seoul for several years. Therefore, the regional difference as a possible affecting factor is eliminated in this dissertation.

All conversations that are used in this dissertation took place in a small classroom, and they were all recorded with a professional voice recorder (Marantz CDR300 Portable CD Field Recorder). During the recording, the voice recorder was unattended, and participants were left by themselves in the classroom, in order to create a comfortable environment for casual conversation. The participants were asked to sit close together, facing each other, so that they could see each other during the conversation. At the beginning of each recording session, participants were asked to converse naturally

Table 3.1: List of Participants for the Conversation Recording

Pair	Speaker	Gender	Age	Hometown	Duration in U.S.	Relationship
1	A	Female	21	Anyang, Korea	4 months	a very good friend
	B	Female	22	Seoul, Korea	10 months	a friend
2	A	Female	22	Seoul, Korea	4 months	best friend
	B	Female	22	Pohang, Korea	4 months	a very good friend
3	A	Female	22	Kwangju, Korea	9 months	a good friend
	B	Female	23	Seoul, Korea	6 months	a good friend
4	A	Female	21	Sungnam, Korea	6 months	best friend
	B	Female	21	Seoul, Korea	6 months	best friend
5	A	Male	28	Seoul, Korea	3 months	a very good friend
	B	Male	26	Seoul, Korea	3 months	best friend
6	A	Male	21	Seoul, Korea	10 months	a very good friend
	B	Male	26	Seoul, Korea	15 months	a good friend
7	A	Male	25	Seoul, Korea	7 months	a very good friend
	B	Male	23	Seoul, Korea	7 months	best friend
8	A	Male	24	Seoul, Korea	3 months	a good friend
	B	Female	20	Seoul, Korea	3 months	a good friend
9	A	Male	25	Seoul, Korea	3 months	a friend
	B	Female	22	Seoul, Korea	3 months	a friend
10	A	Male	24	Seoul, Korea	3 months	a good friend
	B	Female	22	Seoul, Korea	4 months	a good friend

with their partner about any topic that they could think of until the researcher came back to the classroom, approximately 35 minutes later, and asked them to finish the conversation. They were also asked to try not to pay attention to the recording equipment or to the fact that their conversations were recorded. They were asked to turn off

their pagers and cellular phones during the recording. They were told that everything, including timing, would be controlled by the researcher, and that they would not need to watch the time during the conversation. Following Maynard (1989) and Shimojo (2005:56), the participants were asked after the recording session how naturally they were able to carry on the conversation in terms of the scale given in Table 3.2.

Table 3.2: Scale of Naturalness

Scale	Description	Percent
1	I could talk completely naturally without being conscious of being recorded at all.	30 %
2	I was reminded of being recorded from time to time, but otherwise I talked naturally.	65 %
3	I was conscious of being recorded all the time, but I think I talked fairly naturally.	5 %
4	Since I was being recorded, I was not able to relax and could not talk too naturally.	0 %
5	I was conscious of the recorder considerably; this conversation participation was very different from what I normally engaged in.	0 %

As Maynard (1989) points out, the discourse data recorded in a (somewhat) controlled situation would not be identical to those from a natural every-day conversation. Yet based on the participants' own reflections on their conversation, described in Table 3.2, it appears to be rather safe to assume that the recorded discourse data acquired for this dissertation are natural enough, if not completely natural, and therefore they are considered as discourse data of daily natural conversations for the purpose of this dissertation.

## 3.3 Methodology for data analysis

### 3.3.1 Identification of clausal units

All of the recorded conversations that are used in this dissertation are two-party conversations, and they were transcribed into texts for further analysis. Then, all of the utterances were initially analyzed on a clausal basis, instead of a sentential basis, based on the coding for clauses that Thompson & Hopper (2001:30-1) proposed, as well as the definition that is given in Trask (1999:35-6). Throughout this dissertation, the following procedures are adopted for data analysis.

First, as was mentioned above, the transcribed discourse data were categorized into clausal units. Throughout this dissertation, a simple sentence, also called an independent clause, is considered to consist only of a single clause, which contains a subject and a predicate. Both a compound and a complex sentence consist of two or more clauses<sup>1</sup>, and they were divided into simple clausal units accordingly. In this dissertation, nominalized clauses were also counted as separate clausal units. However, relative clauses, which were considered modifiers, were not counted as separate clausal units. The following are examples of several different types of clauses, and relevant clauses are bracketed. More specifically, (3.1) is an example of a clausal complement, in that the clause is a complement of the predicate, a verbal predicate in this case. (3.2) is an example of a nominalized clause in that the (embedded) clause is nominalized, and is utilized as a subject of the predicate, an adjectival predicate in this case. (3.3) is an example of an adverbial clause, which is a subordinate clause, in that the bracketed clause is used like an adverb. (3.4) is an example of a relative clause.

---

<sup>1</sup>A compound sentence contains two independent clauses joined by a coordinator, and a complex sentence has an independent clause, i.e., main clause, joined by one or more dependent clauses, i.e., subordinate clause.

- (3.1) [*ne phulinthu etinci*]-Ø *alci?*  
 2SG Flint where.be-( ) know  
 ‘You know where the Flint (apartment) is, don’t you?’
- (3.2) *kunyang [honca taniki]-to kulehkwu.*  
 well alone travel.around-also be.so  
 ‘Well, traveling around alone is also a concern.’
- (3.3) [*hato ka-ka manhakacikwu*], *samsipsa pen-ka-eyse thukpyelhan*  
 very street-NOM be.many.because 34 NC-street-LOC special  
*ke-n mos nukkyesse.*  
 thing-TOP NEG felt  
 ‘Because there are so many avenues, I couldn’t see anything special on 34th street (in Manhattan).’
- (3.4) *kuke-n cincca [kongpwu-Ø hako siph-un] salam-i ha-nun ken-tey.*  
 that-TOP really study-( ) do like-REL person-NOM do-REL that-be  
 ‘That’s something that someone who really likes to study would do.’

Second, verbal compounds are often included in the complex predicates. In this dissertation, however, they are viewed as simple predicates because they behave just like other simple predicates in Korean. For instance, just like simple predicates, verbal compounds in Korean have one subject or object, as demonstrated in (3.5). Therefore, clauses that occur with such verbal compounds are counted as single clausal units in this dissertation.

- (3.5) *kyay, yayki-Ø [tuleponikka], tayman-eyse sangtanghi nolasstentey.*  
 him, story-( ) hear.and.see Taiwan-LOC significantly hung.out  
 ‘As for him, according to what I heard, (he) hung out a lot in Taiwan.’

Similarly, predicates connected with *-ko*, *-hako* or *-se*—the commonly used connective forms in Korean—are also regarded as single clausal units, unless disjoint subject reference results in ungrammaticality. This is often referred to as ‘a serial verb construction’, which is a string of verbs or verb phrases within a single clause that express

simultaneous or immediately consecutive actions, and that have a single grammatical subject (Larson 1991; Han 1999; Givón 2001). (3.6) is an example of a serial verb construction in Korean in which two verbs, ‘carry’ and ‘move around’ in this case, are connected together.

- (3.6) *ike-lul [tul-ko tani]-myense,*  
 this-ACC carry-and move.around-while  
 ‘While (we) are carrying this around,’

Third, clauses with incomplete predicates are not counted as acceptable clausal units in this study because the incomplete predicates would make the utterance incomplete, thus unanalyzable. Thus, only those with complete predicates are counted as clausal units in this dissertation. For example, the predicate *ttwiekass-* ‘to run’ in (3.7) is not considered complete because a proper predicate inflection is missing, and in fact, the incomplete predicate makes the first utterance incomplete, hence unanalyzable. Thus, the first clause, that is, *palo hwacangcillo twiekass,* is not counted as a clausal unit in this dissertation, but the second clause, that is, *nanun thohalci alko,* in which the predicate is in its complete form, is counted as an appropriate clausal unit.

- (3.7) *palo hwacangcil-lo [ttwiekass], na-nun thohalci alko,*  
 immediately restroom-to ran 1SG-TOP vomit thought  
 ‘(I) ra(n) to the restroom because I thought (I) was going to vomit.’

Fourth, clauses that are co-constructed by both a speaker and an addressee are not counted as clausal units regardless of incomplete or complete predicates, and they are excluded from the analyzed data. (3.8) is an example of a co-constructed utterance, in that the speaker A starts the utterance with the subject, and the speaker B completes the utterance by adding the predicate.

- (3.8) A: *[honcase,*  
 by.oneself  
 ‘You, by yourself,’



B: *ilehkey mos hantey?*  
 like.this NEG do  
 ‘(You, by yourself) cannot do like this?’

Fifth, back-channeling utterances are not counted as clausal units in this dissertation, because they “are in general propositionally empty and are given by the hearer during the speaker’s conversation turn” (Shimojo 2005:58). Yngve (1970:568) notes that a back-channeling utterance is when “the person who has the turn receives short message such as ‘yes’ and ‘uh huh’ without relinquishing the turn.” Maynard (1993:58) discusses back-channel utterances in further detail by saying “back channel is a short message (including nonverbal behavior) which the listener sends during the interlocutor’s speaking turn, and the short messages for which the interlocutor shows a reaction of relinquishing the speaking turn are not regarded as back-channel.” In short, the back-channel utterances are those given by the hearer while the speaker still holds the conversational turn. Utterances such as *ung* ‘yeah’, *kulehci* ‘right’, and *cengmal* ‘really?’ among others are examples of back-channel utterances that are found in the discourse data of this dissertation. (3.9) demonstrates a back-channeling utterance, in that the speaker A passes his conversational turn to the speaker B by releasing an utterance which has no propositional content. For the purpose of this dissertation, back-channel utterances are not viewed as clausal units, and they are excluded from the analyzed data.

(3.9) B: *kulemyen kyayney-tul-un incey te cemcem caki casin-i*  
 then 3PL-PL-TOP now more gradually self self-NOM  
*pichamhaycimyense ike-n mwonka siphunkeci.*  
 miserable.become this-TOP what.be like  
 ‘Then, they wonder what that is about while they gradually become even more miserable.’

A: [*kulehci.*]  
 right  
 ‘Right.’

However, those that are clearly the speaker's turn, not as the hearer's, and are given in response to the preceding utterance are counted as separate clausal units because those utterances convey propositions that are not considered empty. In short, the conversational turn changes from one speaker to another speaker when these utterances are given. Such utterances are exemplified in (3.10) in that the second utterance made by speaker B is given as a short and quick answer in response to speaker A's question. These utterances are counted as clausal units in this dissertation.

- (3.10) A: *tamim-i taykay pwulphyenhaysskeyssta.*  
class.teacher-NOM very was.uncomfortable  
‘Your teacher must have been in a very uncomfortable situation.’
- B: *[kuchi].*  
be.so  
‘Yes (, he was).’

Sixth, there are many utterances in the conversational discourse data which contain repairs. When a repair is made during an utterance, only the item that is reproduced, instead of the interrupted one, is considered when the data is analyzed because interrupted entities are assumed not to be important, thus can be ignored. Interrupted entities are disregarded and excluded from further analysis. (3.11) shows a repaired construction, in that the initial utterance *pongkocha thayksi* ‘van taxi’ is reproduced into *pongko* ‘van’, so *pongko* ‘van’, not *pongkocha thayksi* ‘van taxi’, is only taken as a subject.

- (3.11) *pongkocha thayksi, [pongko]-ka te pissa.*  
van taxi van-NOM more expensive  
‘A van-taxi, a van-(taxi) is more expensive.’

Seventh, an utterance that is regarded as a simple repetition of the preceding utterance is not counted as a separate clausal unit, unless there are significantly long pauses<sup>2</sup>

---

<sup>2</sup>The length of the pause indicates the time duration from the end of one clause unit to the beginning

placed after the preceding clause. (3.12) shows a simple repetition without a noticeable pause between utterances. Thus, the entire utterance is counted as a single clausal unit even though the utterance consists of two clauses on the surface.

- (3.12) *way ttena, [way ttena.]*  
 why leave, why leave  
 ‘Why (do I have to) leave?, why (do I have to) leave?’

On the other hand, when the utterance is repeated after a significant pause or it is made to clarify the previous utterance as shown in (3.13), it is counted as a separate clausal unit, even though the repeated utterance is very similar to, if not exactly the same as, the previous one. In (3.13), utterance A2 is repeated in order to clarify the previous utterance which speaker B did not understand. Thus, utterance A2 is counted as a separate clausal unit.

- (3.13) A1: *salam-tul-i tto nosukholia yok-Ø haysskwuna.*  
 people-PL-NOM again North.Korea curses-( ) did  
 ‘People spoke ill of North Korea again, didn’t they?’

B1: *ung?*  
 uh  
 ‘What?’

A2: *nosukholia tto yok-Ø haysskwuna.*  
 North.Korea again curses-( ) did  
 ‘(People) spoke ill of North Korea again, didn’t they?’

Eighth, independent noun phrases are also considered as separate clausal units, but they are not counted toward the 9,249 total clausal units that occur with overt predicates. Unlike clausal units with overt predicates, these NPs are considered only for the purpose of referential distance (RD) and referential persistence (RP) measurement.

---

of the immediately following clausal unit. The actual time for the pause was measured using ‘Multi-Trans’.

An NP clausal unit is exemplified by the utterance B in (3.14), and there are 183 such units in total.

(3.14) A: *kulemyen chalali thapsu-eyse sa.*  
 then rather Tops-LOC buy  
 ‘Then, buy (water) at Tops.’

B: [*thaps?*]  
 Tops  
 ‘Tops?’

Finally, Korean is classified as an SOV (Subject+Object+Verb) language (Sohn 1999), although this canonical word order is not always observed in spoken Korean, as was discussed earlier. A predicate normally appears at the end of the clause, but subjects or objects may also be placed after predicates in certain cases. Those subjects or objects are often called post-predicative arguments or nominals if there is no significant pause<sup>3</sup> put between the predicate and the post-predicative argument. For the purpose of this dissertation, those post-predicative arguments are viewed as parts of the planned utterances. On the other hand, when there is a significant pause placed between predicates and arguments, those arguments are viewed as ‘afterthoughts’, not as post-predicative arguments, and counted as separate clausal units, but not as part of the planned utterances. The post-predicative argument construction is exemplified in (3.15).

(3.15) *ne, nyuyok-kkaci elmana haysse, pihayngkiphyo-Ø?*  
 2SG, NYC-to how.much cost plane.ticket-( )  
 ‘Hey, how much was your plane ticket to NYC?’

The procedures outlined thus far produced approximately 800 to 1,200 clausal units for each conversation, excluding the NP ‘clausal’ units which are counted at 183 tokens in total, and 9,249 clausal units for all conversations in total.

---

<sup>3</sup>Again, the actual time for the pause is measured using ‘Multi-Trans’, and any pause longer than one second is considered significantly long in this study.

### 3.3.2 Identification of subjects

This section discusses how subjects are determined in this dissertation. Subject tokens include all 9,249 clausal units that are utilized in this dissertation. They consist of clausal units with adjectival predicates, nominal predicates,<sup>4</sup> as well as intransitive and transitive verbs, and they are all expected to have either explicit or non-explicit subjects. As was briefly described in Section 2.3, there are several ways of determining grammatical relations such as subject, object, etc. In this dissertation, case marking, word order and the generalized semantic roles called ‘semantic macroroles’, among other means, are used as primary measures to identify subjects in a clause.

First, it is assumed that all clausal units have a subject, which can appear in either explicit or non-explicit form, and that explicit subjects may be marked either with a zero particle or overt particles. (3.16) shows a subject in a clause with an adjectival predicate, and (3.17) represents a subject in a transitive clause.

(3.16) [*ne meli*] *cikum nemwu kile*.  
2SG hair now too be.long  
‘Hey, your hair is too long now.’

(3.17) *yok-un [caki]-ka ta mekunikka*.  
curses-TOP self-NOM all eat.because  
‘Because (he) himself gets all the curses,’

In (3.16), *ne meli* ‘your hair’ in brackets appears as the subject of this clause, and it is the only argument available for this clause. In (3.17), *caki* ‘self’ in brackets occurs as the subject of the clause, in that it is not only marked with *ka* but it is also an actor in RRG terms.

Second, not all NPs marked with the nominative *ka* are subjects, though the nominative *ka* is usually used to mark subjects in clause. There are cases in which the

---

<sup>4</sup>Adjectival predicates and nominal predicates are often classified simply as non-verbal predicates.

nominative *ka* is mistakenly used, and those cases are regarded as grammatical errors. There are also cases of ‘case alternation’, in which objects are marked with the nominative *ka*. Those cases are ignored when subjects are identified, and they are exemplified in (3.18) and (3.19).

- (3.18) *[kasum]-i mak wa taha?*  
 heart-NOM actively come touch  
 ‘Do (his words) really touch (your) heart?’ (lit. Do (his words) go right to (your) heart?)

- (3.19) *nemwu [chopap-ilang sayngsen]-i nemwu mekko siphunkeyeyyo.*  
 very sushi-and fish-NOM very eat.and want  
 ‘(I) wanted to eat sushi and fish so much.’

It is rather clear that the bracketed NP *kasum* ‘heart’ in (3.18), though marked with the particle *i*,<sup>5</sup> is not a subject, because its semantic role is a goal, which is not to be assigned to subjects for this predicate. The subject of this intransitive clausal unit is in fact realized as a zero anaphor<sup>6</sup> as seen in the English translation provided. In (3.19), the NP *chopap-ilang sayngsen* ‘sushi and fish’ in brackets is marked with the nominative *i*, however, it is not a subject but rather an object because it is an undergoer, not an actor. Furthermore, the nominative *i* can be replaced with the so-called accusative marker *lul* without affecting the grammaticality of the clause in this case. Thus, these cases are excluded from the total counts of subject tokens. In fact, the subject of the transitive clausal unit in (3.19) is realized as a zero anaphor, which is the speaker herself, as shown in the English translation.

It is notable that objects in Korean are normally marked with *lul*, an accusative marker, yet sometimes they are also marked with *ka*, a nominative marker, in unusual

---

<sup>5</sup>Recall that the particle *i* is a morphophonological variation of the particle *ka*, and *ka* represents the nominative marker throughout the study.

<sup>6</sup>In this dissertation, the notion ‘zero anaphor’, often referred to as ‘zero pronoun’, refers to the NPs that occur in their covert forms, i.e., NPs that are realized as an ellipsis.

cases like the one discussed above. Seemingly little attention has been brought to the latter case, which is further discussed in Chapter 5.

Third, subjects in Korean are often marked with particles other than the nominative *ka*. For example, the bracketed NP *haksaynghoy* ‘student association’ in (3.20) occurs with the locative marking particle *eyse*, but it is identified as a subject because it is an actor, and also because it can occur with the nominative marking particle *ka* without having its grammaticality affected.<sup>7</sup> The bracketed NP *yeciney pang* ‘Yeji’s room’ in (3.21) is marked with an adverbial particle *to* ‘also’, but its grammatical relation is the subject because it is the only argument available for this clause with an adjectival predicate, and also because it can also occur with the nominative marking particle *i* without affecting the grammaticality of the clause. Thus, these NPs have been identified as subjects in this dissertation.

- (3.20) *wuli-nun [haksaynghoy]-eyse phwulecwununtey, cokpo-lul keuy.*  
 we-TOP student.association-LOC give.out solution-ACC mostly  
 ‘In our case, the student association usually gives out the solution.’

- (3.21) *[yeciney pang]-to etwuptela.*  
 Yeci’s room-also dark  
 ‘Yeji’s room is also dark.’

Fourth, ‘double or multiple nominative constructions’ are claimed to exist in Korean (Kim 2000; Schütze 2001; Bak 2004; Yoon 2004, *inter alia*), but double or multiple subjects are not allowed in Korean. Various subject-sensitive phenomena such as ‘honorification’, ‘binding’, ‘control’, ‘plural marker copying’, and ‘reflexivization’ suggest that in double or multiple nominative constructions, the nominative marked NP that immediately precedes the predicate, among other nominative marked NPs, carries the canonical properties of the grammatical subject (Kim 2000). Furthermore, particles for

---

<sup>7</sup>The use of the locative marker *eyse* for subjects is a well-known example of the ‘case alternation’ or ‘case shifting’ phenomenon observed in Korean (Um 2003; Yoon 2004, *inter alia*).

those non-subject nominative marked NPs can often be replaced by other particles such as the genitive marker *uy* or the dative marker *eykey* without having the grammaticality of the clause affected. Therefore, the viewpoint that there is only one grammatical subject available in a simple clause is accepted in this study. In the so-called double or multiple nominative constructions, only one nominative NP is identified as a subject for a clause, and all other nominative NPs are ignored for the identification of the subject in this dissertation. The double or multiple nominative constructions are exemplified in (3.22) and (3.23).

(3.22) *ku-ccok ay-tul-un nemwu [kayincwuy]-ka simhaykackwu.*  
 that-side folks-PL-TOP too.much individualism-NOM be.too.excessive  
 ‘As for those folks, the individualism is too excessive.’

(3.23) *yosay kalswulok ay-tul-i [ssakaci]-ka epsnuntey.*  
 these.days as.time.goes.by kid-PL-NOM manners-NOM not.exist  
 ‘These days, kids have less manners as the time goes by.’ (lit. the kids’ manners do not exist.)

In (3.22), *kayincwuy* ‘individualism’, which is in brackets, is the grammatical subject of the clause, in that it is the immediate preverbal nominative marked NP, as well as its particle cannot be replaced by other particles such as *uy* GEN or *eykey* DAT.<sup>8</sup> In (3.23), *ssakaci* ‘manners’, which is in brackets, is identified as the subject since it is the immediate preverbal nominative marked NP, and its particle cannot be replaced by other particles, whereas the other nominative marked NP *aytul* ‘kids’ can also occur with the genitive marking particle *uy* without resulting in the ungrammaticality of the clause.

Finally, a numeral quantifier is generally classified as a determiner or an adverb that describes a referent’s quantity. It usually functions as a modifier of a noun, but it can

---

<sup>8</sup>Some may consider *kayincwuy* ‘individualism’ as the subject of the adjectival predicate *simhata* ‘be too excessive’, and that this subject and predicate form a predicate unit, which has *ku-ccok ay-tul* ‘those folks’ as its subject.



also function as a pronoun. In Korean, a numeral quantifier itself can be a numeral noun or pronoun, and it behaves just like other nouns or pronouns (Sohn 1999). For instance, numeral quantifiers are often marked with post-nominal particles, such as nominative, accusative, or plural markers, and they may be identified as the subject or the object in the clause. In this dissertation, a numeral quantifier is identified as a subject if it is not a modifier, as well as if there are no associated host nouns available explicitly in the clause. There are also cases in which a numeral quantifier occurs together with its host noun in the clause. In those cases, the quantifier is seen as a modifier of its host noun, and only its host noun is counted as a subject. This is shown in (3.24) through (3.26).

(3.24) [ta]-Ø *talle?*  
 all-( ) be.different  
 ‘Do all differ (from one other)?’

(3.25) [hyenkwanmwun-ilang hwacangcilmwun]-Ø *ta namwu-lo toy-n*  
 front.door-and bathroom.door-( ) all wood-with be-REL  
*keyessnuntey.*  
 was.thing  
 ‘The front door and bathroom door were all made of wood.’

(3.26) [kyengchal]-i *sey-tay ttak oteni.*  
 police-NOM three-NC you.know come  
 ‘Three police (cars) came by, you know.’

The numeral quantifier *ta* ‘all’ or ‘everything’ in (3.24) is viewed not as a modifier but as a pronoun, which refers to the noun *cokpo* ‘solution’ in the preceding clause, and there is no associated host noun available overtly. Thus, it is identified as a subject marked with a zero particle in this dissertation. In (3.25), the quantifier *ta* ‘all’ is seen as a modifier, while its associated host noun in brackets, *hyenkwanmwun-ilang hwacangcilmwun* ‘the front and bathroom doors’, is viewed as a subject that is marked

with a zero particle. Similarly, the numeral quantifier *sey-tay* ‘three (cars)’ occurs with its host noun *kyengchal* ‘police’ in (3.26), so it is viewed as a modifier, while the host noun *kyengchal* ‘police’ is identified as a subject marked with a nominative *i*.

### 3.3.3 Identification of objects

This section is devoted to the discussion of how objects are determined in this dissertation. There are only 3,692 clausal tokens in total, because only transitive predicates, not intransitive predicates, are concerned with objects. For the purpose of this dissertation, the following procedures are utilized to identify objects.

First, accusative marked NPs are generally seen as objects, but not all accusative marked NPs are identified as objects in this study. It is noted that double or multiple accusative constructions exist in Korean, but there should be only one grammatical object allowed in a simple clause. Thus, only those accusative marked NPs whose grammatical relations are objects in terms of the objecthood discussed in Section 2.3 are identified as grammatical objects in this dissertation. Observe the following example:

- (3.27) *ce thipi pangyeng yensel-ul sippwun [hwalyong]-ul hayse.*  
 that TV broadcasting speech-ACC fully utilizing-ACC do  
 ‘(He) fully utilized that TV broadcasting speech.’ (lit., (He) did the full utilization of that TV broadcasting speech.)

There are two accusative marked NPs in (3.27), and they are *ce thipi pangyeng yensel* ‘that TV broadcasting speech’ and *hwalyong* ‘utilization’. Yet only the latter is identified as the grammatical object of the transitive predicate *hata* ‘do’, because it is the immediate preverbal NP, and its particle may not be replaced with another particle. On the other hand, the first accusative marked NP is viewed as a modifier, not only because it is not the immediate preverbal NP, but also because its particle may be replaced with another particle such as *uy* GEN to modify its head noun *hwalyong*

‘utilization’ without affecting the grammaticality of the clause. Hence, only the latter accusative marked NP, and not the former, is identified as an object in this dissertation.

Second, a numeral quantifier in Korean can be a modifier, numeral noun or pronoun, as was pointed out earlier in Section 3.3.2. Just like numeral quantifiers functioning as subjects, numeral quantifiers may function as objects as well. In this dissertation, numeral quantifiers are identified as objects when they are not modifiers, and their head nouns or any other nominals that are considered objects of the predicates do not occur explicitly in the clause. On the other hand, when a numeral quantifier explicitly occurs together with its host noun, only the host noun is identified as an object, and the numeral quantifier is viewed as a modifier. Observe the following examples:

(3.28) *manyakey [hana]-lul allyecwumyenun,*  
           in.cases   one-ACC   teach.if  
           ‘For example, if I teach (him) one thing (about Korean),’

(3.29) *[yenge]-Ø twu-kay tulumyen,*  
           English-( ) two-NC   take.if  
           ‘If (I) take two English (related courses),’

(3.30) *il nyen-ul [hyuhak]-ul hayse.*  
           one year-ACC   leave.of.absence-ACC   do  
           ‘(He) took a leave of absence for a year.’

In (3.28), the numeral quantifier *hana* ‘one’ is identified as an object, because it is a pronoun, not a modifier, as well as because its head nominal does not appear overtly in the clause. On the other hand, the numeral quantifier *twukay* ‘two’ in (3.29) is viewed only as a modifier, not as an object, because its head noun *yenge* ‘English’ appears as an object of the transitive predicate *tutta* ‘take’. In (3.30), there are two accusative marked NPs, *il nyen* ‘one year’ and *hyuhak* ‘leave of absence’; This is an instance of a ‘double accusative construction’. The first accusative marked NP is simply a numeral quantifier which modifies its head noun. On the other hand, the latter one, *hyuhak*

‘leave of absence’, is the head noun, and it is marked with the particle *lul*. Thus, only the latter is identified as a grammatical object of this clause.<sup>9</sup>

Finally, objects in Korean are generally marked with the accusative *lul*, but they are sometimes also marked with other particles. In this dissertation, they are identified as objects so long as their grammatical relations are clearly objects. Such a case is exemplified in (3.19) repeated here as (3.31).

(3.31) *nemwu [chopap-ilang sayngsen]-i nemwu mekko siph-unkeyeyyo.*  
 very sushi-and fish-NOM very eat.and want-you.know  
 ‘(I) wanted to eat sushi and fish so much, you know.’

In (3.31), the bracketed NP *chopap-ilang sayngsen* ‘sushi and fish’ is marked with the particle *i*, a so-called nominative marker, but the particle *i* can be replaced with *lul*, a so-called accusative marker, without resulting in ungrammaticality. Furthermore, it is an undergoer, not an actor. Therefore, it is appropriate to identify this NP, *chopap-ilang sayngsen* ‘sushi and fish’, as the object of the clause, though it is marked as nominative.

### 3.4 Overview of tokens

This section presents a brief overview of tokens that are used in this dissertation, though the detailed discussion of the data used is provided in later chapters. The data set used in this dissertation contains a total of 9,249 clausal units, excluding NP clausal units that occur without explicit predicates.

All clausal tokens are divided into two different groups of intransitive and transitive, based on their predicate types. Table 3.3 shows all of the tokens in terms of both intransitive and transitive predicates. The intransitive predicates consist of a total of

---

<sup>9</sup>The entire NP (‘one year’s leave of absence’) may be seen as an object in this clause.

5,557 tokens, or 60 percent of the total number, while the transitive predicates provide a total of 3,692 tokens, or 40 percent of the total number.

Table 3.3: Types of Predicates

PREDICATES	NO.	%
intransitive	5,557	60
transitive	3,692	40
TOTAL	9,249	100

All clausal tokens are also divided into three different groups in terms of their sentence types: declarative, interrogative, and directive. In this study, the sentence type is decided based on the predicate endings, and Table 3.4 shows all tokens in terms of declarative, interrogative, as well as directive sentence type distinction. Sentences of declarative types consist of a total of 8,213 tokens, or 89 percent of the total, and sentences of interrogative types provide 1,000 tokens, or 11 percent of the total, while sentences of directive types count only 36 tokens.

Table 3.4: Sentence Types

SENTENCE TYPES	NO.	%
declarative	8,213	89
interrogative	1,000	11
directive	36	0
TOTAL	9,249	100

As was mentioned earlier, Korean is generally classified as a verb-final language, and word order in the sentence is relatively free as long as the sentence ends with the predicate (Sohn 1999).

Table 3.5: Word Order

WORD ORDER VARIATIONS		NO.	%	CANONICAL	¬CANONICAL
intransitive	V	2,700	29	✓	
	SV	2,757	30	✓	
	VS	100	1		✓
transitive	OV	1,918	21	✓	
	SOV	715	8	✓	
	V	710	8	✓	
	SV	237	3	✓	
	OSV	44	0		✓
	VO	21	0		✓
	OVS	18	0		✓
	VS	14	0		✓
	SVO	14	0		✓
	VSO	1	0		✓
TOTAL		9,249	100	98%	2%

However, word order patterns which do not end with the predicate are also observed in the conversational discourse data of this study, though the tokens of those patterns are far smaller. The various patterns of word order in conversational Korean are shown in Table 3.5, and SV, V, OV and SOV are the four most common patterns observed. Based on this, it can be claimed that SV, V, OV and SOV are unmarked or canonical word order patterns in Korean, although they are not the only options, or that clauses in conversational Korean normally end with the predicate, classifying Korean as a predicate-final language.

There are two types of negation in Korean: sentential negation and constituent negation (Sohn 1999:388-394). Sentential negation in Korean can be further divided

Table 3.6: Negation

NEGATION	NO.	%
affirmative	8,056	87
negative	1,193	13
TOTAL	9,249	100

into short-form negation and long-form negation (Sohn 1999). The short form negation is often called pre-verbal negation, and the long form negation is called post-verbal negation (Kim 2000). The number of affirmative constructions (87%) is far more than that of negative constructions (13%) reported in our data, as shown in Table 3.6.

Table 3.7: Subject Forms

SUBJECTS	NO.	%
zero anaphor	5,349	58
<i>full</i> NP	2,673	29
<i>bare</i> NP	1,227	13
TOTAL	9,249	100

Table 3.7 presents all tokens in terms of their encoding types. Subjects in Korean, as well as in Japanese, are often not explicitly expressed in a clause when they can be retrieved from a linguistic or non-linguistic context (Kuno 1978; Sohn 1999). According to Table 3.7, subjects in conversational Korean are most likely to occur as zero anaphors (58%). A similar phenomenon is also observed in informal spoken Japanese, which reveals many linguistic similarities to Korean (Ono *et al.* 2000; Shimojo 2005, *inter alia*). On the other hand, if subjects overtly occur in the clause, they tend to occur more frequently as a full NP (29%), than as a bare NP (13%), regardless of the particle

forms. This distributional pattern is interesting, because it was initially expected that subjects would occur more frequently with a zero particle than with an overt particle, due to the lack of necessity of an overt marking. More specifically, subjects, particularly those of transitive clauses, are generally definite, thus identifiable, and the overt marking is not necessary. In fact, subjects appear with a zero particle more frequently than with an overt particle in colloquial Japanese (Fujii & Ono 2000; Shimojo 2005).

The difference in distributional patterns observed between these two languages may suggest to us that the overt particle may function somewhat differently in these languages. This difference, as well as subject forms, is discussed in Chapter 4 in further detail.

For the purpose of this study, all subject tokens that appear overtly in the clause were sub-divided based on the particles used to mark them. Table 3.8 presents the distributional pattern of the particles used. According to Table 3.8, the nominative *ka* is the particle which occurs most frequently with a subject. This tendency observed in conversational Korean contrasts with that observed in colloquial Japanese. Specifically, the use of the particle *ga* is relatively uncommon in colloquial Japanese (Shimojo & Choi 2001), whereas the nominative *ka* is the most frequently used particle for marking subjects in colloquial Korean. This distributional difference between these languages is interesting and is further discussed in Chapter 4.

Other particles (e.g., *nun*<sup>10</sup> TOP, *to*<sup>11</sup> ‘also’) may also be attached to subjects in conversational Korean as shown in Table 3.8. Adverbial particles *to* ‘also’, *man* ‘only’ or *cochato* ‘even’ are used to mark subjects only when such lexical meanings need to be coded (Lee & Thompson 1989). In other words, such lexical meanings are not encoded

---

<sup>10</sup>The particle *nun* is called a topic marker in Korean, and particles *un* and *n* are morphophonological variations of the particle *nun*. *Nun* is used to represent topic marker in Korean in this study.

<sup>11</sup>Adverbial particles *to* ‘also’, *man* ‘only’, *cochato* ‘even’, etc. are sometimes called ‘pragmatic particles’ (Lee & Thompson 1989), or ‘delimiter’ (Ko. S 2000) for different purposes. The particle *to* represents such adverbial particles in the present study.



in the proposition unless these adverbial particles are utilized to mark the subjects. Hence, it is assumed that such adverbial particles would not occur as a zero particle. Particles labeled as ‘etc.’ in Table 3.8 consist of any particles (e.g., *ilato* ‘(even) with’ or ‘any’, *kkaci* ‘even’, etc.) other than those three types (i.e., *ka* NOM, *nun* TOP, and *to* ‘also’) which could all be attached to the subjects.

Table 3.8: Subject Marking Particles

PARTICLES	NO.	%
<i>ka</i> (NOM)	1,557	40
$\emptyset$ <sup>12</sup>	1,237	32
<i>nun</i> (TOP)	644	16
<i>to</i> (‘also’)	411	11
<i>etc.</i>	55	1
TOTAL	3,904	100

As was previously described, object tokens consist only of the transitive clausal units, and this results in 3,692 object tokens, approximately 40% of the entire data, to analyze in this dissertation. Table 3.9 presents all of the object tokens in terms of their forms. Unlike subjects, objects tend to occur overtly (74%), particularly without an overt particle (46%).

Table 3.9: Object Forms

OBJECTS	NO.	%
<i>bare</i> NP	1,688	46
<i>full</i> NP	1,043	28
zero anaphor	961	26
TOTAL	3,692	100

<sup>12</sup>The symbol  $\emptyset$  refers to a zero particle, as indicated earlier.

This distributional tendency is opposite to that of the subjects in conversational Korean. This distributional pattern somewhat complies with that reported in previous studies in colloquial Japanese (e.g., Shimojo 2005), in that objects tend to appear overtly in the clause in both languages. However, they also differ from each other, in that the zero anaphor is the most common form of object in Japanese (Shimojo 2005), whereas among three options, it is the least common form of object in Korean. These similarities and differences are interesting, and are discussed further in Chapter 5.

For the purpose of this study, all object tokens are also sub-categorized in terms of the particles that are used for marking objects. Table 3.10 presents all tokens of objects either with a zero particle or with overt particles. Overt object NPs are most likely to be marked with a zero particle, and they consist of approximately 62 percent of all overt object tokens. When objects are marked with an overt particle, the particle *lul* is more preferred to other particles. The particle *lul* is used for a total of 699 tokens, or takes up about 26 percent of the total particles used to mark objects. The particle *nun* consists of a total of 110 tokens, or about 4 percent of the total particles used for marking objects. The adverbial particles (e.g., *to* ‘also’) that are used to mark objects are counted at 200 tokens, and this accounts for approximately 7 percent of the entire particles used for object marking. Finally, 34 tokens of object marking particles are particles other than those three main forms of particles (i.e., *lul* ACC, *nun* TOP, and *to* ‘also’).

The use of a zero particle for marking objects is quite high (62%) in conversational Korean, and this distributional pattern attests to what is reported in previous studies of Japanese (Fujii & Ono 2000; Shimojo 2005). However, this pattern does not comply with Kim’s (2006) finding, in that objects are most frequently marked with the accusative *lul*, instead of a zero particle, in his study on the Korean accusative marker. Yet such difference in the distributional pattern seems to be attributed to the differ-

Table 3.10: Object Marking Particles

PARTICLES	NO.	%
$\emptyset$	1,688	62
<i>lul</i> (ACC)	699	26
<i>nun</i> (TOP)	110	4
<i>to</i> ('also')	200	7
etc.	34	1
TOTAL	2,731	100

ence of the genre of the data used, as is suggested by Ko. E's (2000) findings. In fact, the discourse data used in Kim's (2006) study is narrative, while the data used in this dissertation is conversational spontaneous discourse.

Chapter 3 has provided the overview of the informal conversational Korean data used in this dissertation, and it has also discussed how the data was acquired as well as how it was analyzed, and the overall token distribution in our data, comparing those with what has been reported in previous studies. The detailed discussion of subject and object markings in conversational Korean continues in the following chapters.

# Chapter 4

## Subject Marking

### 4.1 Outline

Chapter 4 is devoted to the in-depth discussion of the subject encoding types in conversational Korean. More specifically, it discusses how a speaker chooses one particular subject encoding type over another, largely based on two different factors: processing load and informational prominence. The speaker's selection of a particular subject marking particle is discussed here by both individual factors and combined factors. Factors that are used for the discussion of subject marking particles are related to either processing load, informational prominence, or both. Individual factors that are used for the discussion of subject marking in conversational Korean are: negation, animacy, word order, definiteness, length, repair, anaphoric saliency, cataphoric saliency, and contrast. I claim that these factors are related to the use of particular subject markings and the relationship may be described in terms of processing load or informational prominence in one way or another.

## 4.2 Subject encoding types

As was previously noted, each clausal unit, regardless of transitivity or intransitivity, has a subject, which can be defined in several different ways. Subjects in Korean may or may not appear explicitly in a clause. Furthermore, if they are expressed explicitly in a clause, they can occur with or without the use of a particle to mark them. In short, subjects in Korean are realized largely in three different NP forms: a zero anaphor, a *bare* NP, and a *full* NP with an overt post-nominal marker.

What needs to be pointed out here is that the selection of a particular NP form is not random, but instead shows the interaction between the grammatical role and the information status of the NP in a clause, as was noted in Bak (2004). Ko. E (2000), in her study of Korean conversation, also notes that the selection of NP forms is based on the information status of the NPs in a clause.

In the present chapter, I examine how subject NPs are realized in conversational Korean. Table 4.1 presents their distributional pattern.

Table 4.1: Subject Encoding Types

SUBJ.ET	Zero-Anaphor <sup>1</sup>	NP-Ø	NP-ka	NP-nun	NP-to	NP-etc.	SUB-TOTAL	TOTAL
NO.	5,349	1,227	1,564	645	408	56	2,673	9,249
%	58	13	17	7	4	1	29	100

As presented in our data in Table 4.1, subjects were most likely to be unexpressed, i.e., realized as a zero anaphor (58%), and if explicitly expressed, they occurred more frequently with an explicit particle (29%) than with a zero particle (13%) (zero anaphor

<sup>1</sup>Recall that the term ‘zero anaphor’ refers to the NPs that occur in their covert forms (i.e., null NPs), and the symbol ‘Ø’ indicates the use of a zero particle. The notation ‘SUB-TOTAL’ indicates the sum of all overt particles used, i.e., the number of *full* NPs marked with an overt post-nominal particle.

> NP-*ka* > NP- $\emptyset$ ). This distributional pattern can be attributed to the general characteristics of subjects, which Fry (2003) pointed out, in that subjects tend to convey definite information, and thus are accessible to both a speaker and a hearer. In other words, referents of a subject are accessible to both interlocutors, so they often do not appear overtly in a clause.

The occurrence of subjects marked with a zero particle is also expected to be higher than those with an explicit particle because the referents of a subject tend to be identifiable, which would induce an attenuated NP form. However, the opposite tendency is reported in conversational Korean. One may wonder why the ratio for subjects marked with a zero particle is not higher than the one for those marked with an explicit particle, if subjects are expressed or unexpressed simply based on how accessible they are to a speaker and a hearer. This distributional tendency suggests that in conversational Korean, the occurrence and non-occurrence of subject marking particles may also depend not only upon the accessibility of subjects but also on other properties. In fact, I argue that informational prominence is another relevant factor that influences a speaker's selection of a particular subject marking over another in conversational Korean. More specifically, the speaker selects the post-nominal particle *ka* over a zero particle to mark subjects when the subjects represent informational prominence.

Interestingly, the distributional pattern reported for subject encoding types in colloquial Japanese (Ono *et al.* 2000) differs from that of conversational Korean reported in the present study. In spoken Japanese, subjects are most likely to be unexpressed, and if explicitly expressed, they occur more frequently with a zero particle than with an explicit particle (zero anaphor > NP- $\emptyset$  > NP-*ga*<sup>2</sup>) (Ono *et al.* 2000). Is this because informational prominence does not facilitate the use of an explicit particle in Japanese, whereas it does in Korean? In fact, it was previously reported that the informational

---

<sup>2</sup>Just like in Korean, in Japanese, other particles (e.g., *wa*) may be used to mark subjects, but the particle *ga* represents the use of an explicit particle in this study.

importance does also facilitate the use of an explicit particle in colloquial Japanese (Ono *et al.* 2000).

One may wonder why the distributional pattern observed for subjects in conversational Korean differs from what has been reported in colloquial Japanese, if the same principle (i.e., easy accessibility) applies to both languages, which share many linguistic characteristics. As for the differences observed between these two languages, I propose that the way informational prominence plays a role in the selection of subject marking particles may differ in both languages. While the set level in which informational prominence instigates the use of overt marking may be different in these languages, it may be the case that informational prominence is simply a stronger factor than accessibility for the selection of subject marking in conversational Korean, in contrast with Japanese.

In this chapter, subject encoding types are discussed mostly based on individual factors as well as several factors together, all of which are related to either processing load or informational prominence, or both. I will discuss subject encoding types in the conversation data of the present study, based on individual factors, and then subject encoding types based on multi-factors when they are relevant.

### **4.3 Intransitive and transitive subjects**

In Section 4.2, I showed that in conversational Korean, a zero anaphor was the most commonly observed encoding type for subject NPs, an overtly marked NP was the second most common type for subject nominals, and then a bare NP for subject NPs: zero anaphor > an overtly marked NP > a bare NP (see Table 4.1). One may wonder whether subject encoding types may differ depending on the transitivity of clauses. In order to see whether or not the difference in transitivity may result in the difference

in subject encoding types, I examined the entire body of subject tokens based on the transitivity of clauses where they appear. Table 4.2 presents distributional patterns of subject encoding types with respect to the transitivity of clauses in which the subjects occur.

Table 4.2: Transitivity and Subject Encoding Types

SUBJ.ET TRANS.		ZERO- ANAPHOR	NP-Ø	NP-ka	NP-nun	NP-to	NP-etc.	SUB- TOTAL	TOTAL
intran.	NO.	2,700	921	1,183	410	314	29	1,937	5,557
	%	<b>49</b>	<b>17</b>	<b>21</b>	7	6	1	<b>35</b>	101
tran.	NO.	2,649	306	381	235	94	27	736	3,692
	%	<b>72</b>	<b>8</b>	<b>10</b>	6	3	1	<b>20</b>	100
TOTAL	NO.	5,349	1,227	1,564	645	408	56	2,673	9,249
	%	58	13	17	7	4	1	29	100

According to Table 4.2, subjects are explicitly expressed more often in intransitive clauses (17+35=52%) than transitive clauses (8+20=28%). This distributional pattern may simply be attributed to the universal tendency to express at most one argument per clause, which Fry (2003) points out. On the other hand, this tendency may indicate the ergative-absolutive pattern, in that subjects tend to be overtly expressed in intransitive clauses, whereas objects tend to be overtly expressed in transitive clauses (Dixon 1994, Thompson & Hopper 2001). This tendency is also consistent with informational structure; in transitive clauses, objects tend to represent new information, hence are overtly expressed, while in intransitive clauses, subjects encode new information, hence are overtly expressed.

Table 4.2 also shows that the particle *ka* occurs more frequently with S, i.e., intransitive subject (21%) than with A, i.e., transitive subject (10%). A similar distributional pattern is also observed with the so-called nominative particle *ga* in spoken Japanese (Ono *et al.* 2000). This issue will be discussed again in Chapter 5.



## 4.4 Sentence types

It has been addressed in the studies of Japanese that in general, an explicit post-nominal marking is not preferred and often omitted in questions, and the use of an overt particle after the argument is even unnatural in questions (Makino & Tsutsui 1986; Backhouse 1993). However, no justification has been provided for such a tendency.

There have been some studies done with regard to subject markings in Korean (e.g., Ko. S 2000), but no studies have been done to further investigate subject markings with respect to the variation of sentence types. For this reason, this dissertation examines whether subject markings are realized differently depending upon types of sentences in which subject NPs are engaged. In order to investigate whether subject markings are in correlation with sentence types in conversational Korean, I divided all clausal units into three different sub-groups based on their sentence types: declarative, interrogative and directive sentences. Table 4.3 presents the distributional patterns for subject markings with relation to sentence types.

Table 4.3: Sentence Types and Subject Marking

SUBJ.MARK		NP-Ø	NP-ka	NP-nun	NP-to	NP-etc.	SUB-TOTAL	TOTAL
CL.TYPE								
declarative	NO.	1,038	1,448	587	374	51	2,460	3,498
	%	<b>30</b>	<b>41</b>	17	11	1	<b>70</b>	100
interrogative	NO.	182	114	58	34	5	211	393
	%	<b>46</b>	<b>29</b>	15	9	1	<b>54</b>	100
directive	NO.	7	2	0	0	0	2	9
	%	<b>78</b>	<b>22</b>	0	0	0	<b>22</b>	100
TOTAL	NO.	1,227	1,564	645	408	56	2,673	3,900
	%	31	40	17	10	1	69	100

Table 4.3 shows that an overt marking is more common with subject NPs of ‘declarative’ sentences (70%) than with subject NPs of ‘interrogative’ (54%) or ‘directive’

sentences (22%). More specifically, the occurrence of *ka*-marking is much less common for subject NPs that occur in ‘interrogative’ (29%) as well as ‘directive’ (22%) sentences than for those occurring in ‘declarative’ sentences (41%). The occurrence of zero-marking is instead high for coding subjects occurring in either interrogative (46%) or directive (78%) sentences. This distributional pattern is consistent with what has been reported in previous studies on particle ellipsis in spoken Japanese (Makino & Tsutsui 1986; Backhouse 1993), in that an explicit particle (*ka* in Korean; *ga* in Japanese) is not preferred for marking subjects of interrogative sentences in either language.

As was mentioned above, no justification has been provided for why the occurrence of zero-marking, instead of an overt marking (*ka* for Korean; *ga* for Japanese), is frequent in interrogative sentential constructions. In this dissertation, I explore why the use of zero-marking is more common. The following are examples of subjects in interrogative sentences excerpted from the data of the present study.

(4.1) A1: *ayin han pen te ola kulayyaci.*  
 boy/girl.friend once NC more come say  
 ‘You wanna tell your boyfriend to come, right?’

B1: *an kulayto cikum nyuyok-ey iss-e.*  
 not be.so now NewYork-LOC be-SEM  
 ‘In fact, (he) is now in New York.’

A2: *ah, kulay?*  
 uh, be.so  
 ‘Oh, is that right?’

B2: *yehayng-Ø ha-ko iss-e.*  
 travel-( ) do-and be-SEM  
 ‘(He) is traveling now.’

A3: *honcase? kyay-Ø hyuhak-sayng-i-ya?*  
 alone 3SG-( ) leave.of.absence-student-be-Q  
 ‘By himself? Is he on a leave of absence?’

(4.2) *cihachel-pi-Ø* *elmaya, nyuyok-uy?*  
 subway-fare-( ) how.much NewYork-GEN  
 ‘What is the fare for the subway in New York?’

(4.3) *ne-Ø* *yekise siss-ullyekwu?*  
 2SG-( ) here wash-going.to.do  
 ‘Are you going to take a shower here?’

(4.1) through (4.3) exemplify subjects appearing in an interrogative sentence, and relevant subjects are underlined. More specifically, in (4.1), the subject *kyay* ‘he’ is marked with a zero-particle. Both subjects *cihachelpi* ‘subway fare’ in (4.2) and *ne* ‘you’ in (4.3) are also zero-marked. In (4.1), the underlined NP *kyay* ‘he’ has been talked about, so it has already been activated before the utterance. Therefore, the use of zero-marking for the subject is not surprising. In (4.2), the underlined NP ‘*cihachelpi* ‘subway fare’ is newly introduced into the conversation, but interlocutors have been talking about a trip to New York City. Therefore, it can be said that the underlined NP, ‘*cihachelpi* ‘subway fare’, has been made accessible by previously given concepts that are related to New York City. In (4.3), the underlined NP *ne* ‘you’ is an interlocutor, i.e., a hearer, and thus it is already accessible to both interlocutors. In summary, the element that is being asked about seems to have been activated at the time of the utterance, and thus an explicit marking is not common for subjects of interrogative sentences.

The utterance in (4.4), on the other hand, is an instance which shows a subject occurring in a directive sentence, and the underlined subject, *ne* ‘you’, appears with a zero-marking.

(4.4) *ne-Ø* *ittaka ta ssa tulkwusenun naylye-wala.*  
 2SG-( ) later all pack carry.and come.down-DIR  
 ‘You pack everything and come here later!’

In directive sentences, the referents of subjects are most likely to be the hearer, which is accessible to both the speaker and the hearer at the time of utterance. Since they are highly accessible, hence there is a minimal processing load in referent identification, they are most frequently unexpressed. Yet when they are overtly expressed, they are commonly marked with a zero particle. For example, in (4.4), the underlined NP, *ne* ‘you’, is the hearer, and its referent is highly accessible to both interlocutors at the time of utterance. Therefore, the use of zero-marking is not unexpected for coding the underlined subject, which is highly accessible.

To summarize, in interrogative sentences, the referents of subjects which are being asked about, i.e., the focus of question, are accessible to both interlocutors at the time of utterance, and thus they are often zero-marked, as are exemplified in utterances in (4.1) through (4.3) above. Similarly, subjects in directive sentences are highly accessible to both the speaker and the hearer, and thus they frequently occur with a zero particle as is exemplified in the utterance in (4.4) above.

## 4.5 Factors

Earlier in this dissertation, I indicated that in conversational Korean, the choice of subject encoding types would correlate with factors such as negation, animacy, word order, and length, among others, which are related to processing load, informational prominence or both in one way or another. For the purpose of this dissertation, it was assumed that a deviation from the norm or the default pattern would indicate (an increase in) processing load, and that extra information to process, i.e., informationally loaded, would represent informational prominence. In this section, I discuss how these factors influence the speaker’s selection of one particular subject form over another in conversational Korean.

### 4.5.1 Negation

It has been widely agreed that there are two major types of negation in Korean. One is negation by a morphological negative (*an* ‘not’), and the other is negation by a negative auxiliary (*ci anhta* ‘be not’) (Kim 2000a). The former is often called preverbal or short form negation, and the latter is called post-verbal or long form negation (Sohn 1999; Song 2005). (4.5) and (4.6) are examples of pre-verbal negation and post-verbal negation, respectively.

(4.5) *na-nun, kuntey cikum, mom-i nemwu an cohase,*  
1SG-TOP you.know now body-NOM too NEG be.well  
‘Because I am not well these days,’

(4.6) *hankwuk ke-n kulehkey cop-ci anh-ass-ess-nuntey,*  
Korea thing-TOP so small-COMP NEG-PST-PST-SEM  
‘The Korean (police) car was not that small,’

In addition to these two types of negation (i.e., preverbal and post-verbal negations), there are other ways of encoding negativity in Korean (Sohn 1999; Kim 2000a). For instance, lexical items such as adverbs or negative predicates (e.g., *mos* ‘not’, *moruta* ‘don’t know’) can also be used to code the negativity of a given utterance. (4.7) and (4.8) exemplify utterances in which negativity is encoded by a lexical item.

(4.7) *ne-hanthey cenhwa-lul mos hayssta.*  
2SG-DAT phone-ACC not did  
‘(I) couldn’t call you,’

(4.8) *ku taumey ttalun ke-n molukeyssko,*  
that next other thing-TOP not.know  
‘And then, (I) don’t know other things,’

It is generally assumed that negative constructions are conceptually more complex than their affirmative counterparts. I assume that the additional complexity would lead

to an increase in processing load, which would facilitate the use of an explicit particle for marking subjects. Furthermore, the number of negative sentences is much smaller than that of affirmative sentences (see Table 3.6), so negative sentences can be viewed as a deviation from the norm, which would result in an increase in processing load. In order to see if the negativity of sentences would influence subject encoding types, I divided all clausal units into two different sub-categories based on their polarity: negative and affirmative. Then, I examined whether the choice of subject encoding types is related to the negativity of sentences in which subjects occur. The distributional patterns of subject markings with respect to the negativity of sentences are summarized in Table 4.4.

Table 4.4: Negation and Subject Marking

SUBJ.MARK		NP-Ø	NP-ka	NP-nun	NP-to	NP-etc.	SUB-TOTAL	TOTAL
NEGATION								
affirmative	NO.	1,059	1,361	530	326	53	2,270	3,329
	%	<b>32</b>	41	<b>16</b>	<b>10</b>	2	68	100
negative	NO.	168	203	115	82	3	403	1,193
	%	<b>29</b>	36	<b>20</b>	<b>14</b>	1	71	100
TOTAL	NO.	1,227	1,564	645	408	56	2,673	3,900
	%	31	40	17	10	1	69	100

In Table 4.4, I do not find a significant correlation between the negativity of sentences and subject markings (zero-marking vs. overt marking:  $\chi^2 = 1.29$ ,  $P = 0.256$ ). That is to say, the distributional patterns shown in Table 4.4 do not suggest that subjects occurring in negative sentences are more frequently coded with an overt marking than those occurring in affirmative sentences, although the overall occurrence of an explicit particle is slightly higher for marking subjects occurring in negative sentences (71%) than for marking subjects occurring in their affirmative counterparts (68%).

Furthermore, the occurrence of zero-marked subjects is slightly lower in negative sentences (29%) than in their affirmative counterparts (32%), yet the occurrence of subjects coded with the nominative *ka* is also lower in negative sentences (36%) than in their affirmative counterparts (41%). Simply put, it seems that in conversational Korean, the negativity of sentences does not correlate with the choice between *ka*-marking and zero-marking for subjects ( $\chi^2=0.008$ ,  $P=0.9266$ ).

Table 4.4 also shows that the occurrence of subjects coded with the post-nominal particle *nun* is slightly higher in negative sentences (20%) than in affirmative sentences (16%). I assume that this is because already-established, thus old, information rather than unestablished, thus new, information is usually in the focus of negation, so the post-nominal particle *nun*, which has been claimed to encode old information (Lee 2003, *inter alia*), is more common for coding subjects occurring in negative sentences than for those occurring in affirmative sentences (*ka*-marking vs. *nun*-marking:  $\chi^2=8.717$ ,  $P < 0.01$ ).<sup>3</sup>

Lastly, the occurrence of *to*-marking is higher in negative sentences (14%) than in affirmative sentences (10%) (zero marking vs. *to*-marking:  $\chi^2=9.701$ ,  $P < 0.01$ ). This tendency can be attributed to the fact that the post-nominal particle *to* ‘even’ in Korean often functions as a negative polarity item (NPI), which is commonly used to mark entities in negative sentences (Sells 2006). Let us see the utterance in (4.9), which exemplifies a *to*-marked subject in a negative sentence.

- (4.9) *simcie*            [*nay*            *rwummeyithu*]-*to*    *molununkeya*.  
           worst.of.all 1SG.GEN roommate-even    don’t.know  
           ‘Worst of all, even my roommate doesn’t know that.’

In (4.9), the subject in brackets, *nay rwummeyithu* ‘my roommate’, appears in the negative sentence, and it is marked with the particle *to* ‘even’, which commonly

---

<sup>3</sup>See Section 2.4 for the discussion of the properties of the post-nominal particle *nun* in Korean.

functions as an NPI.

To summarize, unlike my initial expectation, the negativity of a sentence is not linked with the use of an explicit particle (e.g., *ka*) for coding subjects. That is to say, the overall occurrence of an overt marking is not any greater for coding subjects occurring in negative sentences than for those occurring in affirmative sentences. More specifically, the use of the nominative *ka* is less common for subjects in negative sentences than in affirmative sentences. I attribute this to the assumption that the negation would generally increase processing load due to the conceptual complexity as well as its marked status, but at the same time, subjects that are in the focus of negation usually occur as already-established, thus old, information, hence easily identifiable, so the use of *ka* is less common for coding subjects in negative sentences than in affirmative sentences, as shown in Table 4.4. Contrastively, the use of *nun*-marking is more common for subjects in negative sentences than in affirmative sentences, and this is attributed to the fact that an entity in the focus of negation often contrasts with another entity, so it is commonly coded with *nun*-marking. Lastly, the use of the particle *to* ‘even’, which is often referred to as an NPI, is common for subjects in negative sentences.

#### **4.5.2 Animacy**

The notion ‘animacy’ has often been discussed in previous studies (e.g., Keenan 1976; Hopper & Thompson 1980; Comrie 1989; Dixon 1994; Givón 2000a; 2000b; Fry 2003, among others). In general, human beings and any other living creatures are viewed as animate (Keenan 1976; Hopper & Thompson 1980). In this dissertation, the animacy of a given NP was determined based on whether the referent of the given NP was a human or living creature, not simply based on morphological form of the NP. In other words, the animacy was decided not based on the signifier but on the signified. Following are examples of animate and inanimate subject NPs used in this dissertation.



- (4.10) *kapcaki kemun cha-ka na-hanthey wayo, ilehkey.*  
 suddenly black car-NOM 1SG-DAT come like.this  
 ‘Suddenly, a black car is moving toward me, like this.’
- (4.11) *kyengchal-i na-l ttak poteni,*  
 police-NOM 1SG-ACC DP see  
 ‘A police (officer) saw me, and then ...’
- (4.12) *ne-Ø ku ttay ku sanghwang-selmyeng-ul hay-ss-e?*  
 2SG-( ) that time that situation-explanation-ACC do-PST-Q  
 ‘Did you explain the situation at that time?’
- (4.13) *na-nun syophing-Ø pyello an cohaha-y.*  
 1SG-TOP shopping-( ) little NEG like-SEM  
 ‘I do not like shopping much.’
- (4.14) *pam-mata chencang-eyse cwi-ka kulk-e.*  
 night-per ceiling-LOC rat-NOM scratch-SEM  
 ‘Every night, rats make scratches from the ceiling.’

In (4.10), the underlined NP, *kemun cha* ‘black car’, is viewed as an inanimate subject because its referent, which is the car itself, not the driver, is inanimate. In (4.11), the underlined NP, *kyengchal* ‘police’, is counted as an animate subject because it refers to a police officer, not the police in general. In other words, the morphological form, namely the signifier, for *kyengchal* ‘police’ is inanimate but its referent, namely the signified, is animate. For this reason, (4.11) is counted as an example of an animate subject in this dissertation. (4.12) through (4.14) are also examples of animate subjects that are excerpted from the data used in this dissertation.

Fry (2003) claims that in Japanese, a subject particle ellipsis has a strong correlation with the animacy of a subject referent. More specifically, referents of subjects are generally animate, and subject particle ellipsis occurs more frequently with a subject whose referent is animate. Keenan (1976) and Comrie (1989) make similar claims in this regard. Neither of them, however, provides an explanation for such a tendency.

More importantly, there have been, to the best of my knowledge, no studies investigating whether there is a correlation between the choice of subject markings and the animacy of the subject NP in conversational Korean. In an attempt to see whether a speaker's selection of one particular subject form over another is related to the animacy of a subject NP, I divided the entire body of subjects into two different subtypes based on the animacy of their referents: subjects representing an animate referent and subjects representing an inanimate referent. Then, I examined all subject NPs with respect to the animacy of their referents, as well as to their marking. The results are summarized in Table 4.5.

Table 4.5: Subject Animacy and Subject Marking

SUBJ.MARK		NP-Ø	NP-ka	NP-nun	NP-to	NP-etc.	SUB-TOTAL	TOTAL
SUBJ.ANIMACY								
inanimate	NO.	594	876	232	209	24	1,341	1,935
	%	<b>31</b>	<b>45</b>	<b>12</b>	11	1	69	100
animate/-human	NO.	18	28	9	8	0	45	63
	%	<b>29</b>	<b>44</b>	14	13	0	71	100
human	NO.	615	660	404	191	32	1,290	1,905
	%	<b>32</b>	<b>35</b>	<b>21</b>	10	2	68	100
TOTAL	NO.	1,227	1,564	645	408	56	2,673	3,900
	%	31	40	17	10	1	69	100

The distributional pattern shown in Table 4.5 indicates that as is in Japanese, a subject marking particle ellipsis is related to the animacy of a subject referent in conversational Korean, although the occurrence of zero-marked subjects is almost the same across all subject encoding types (31% for inanimate; 29% for non-human animate; 32% for animate). More specifically, the occurrence of *ka*-marking is high for coding subjects that represent an inanimate (45%) or non-human animate referent (44%), while it is relatively low for coding subjects that represent a human referent (35%) (inanimate vs. human:  $\chi^2=16.973$ ,  $P < 0.001$ ).

As for the relatively high occurrence of *ka*-marking for subjects that represent an inanimate or non-human animate referent, I argue that referents of subject NPs are most likely to be animate, so subjects representing an inanimate or non-human animate referent are considered a deviation from the norm, which result in processing load.<sup>4</sup> Due to processing load, the occurrence of *ka*-marking is relatively high for coding subjects representing an inanimate or non-human animate referent, as exemplified in (4.14). Contrastively, the occurrence of *ka*-marking is relatively low for coding subjects whose referents are animate, particularly human beings, which is considered unmarked animacy for subjects, as seen in (4.12). Simply put, the occurrence of *ka*-marking is a little higher for coding subjects that represent marked subject animacy, i.e., inanimate, than for coding subjects that represent unmarked subject animacy, i.e., animate, due to processing load, although the occurrence of zero-marking is about the same regardless of the animacy.

In Table 4.5, I also find that the occurrence of *nun*-marking is high when referents of subject NPs are human beings (21%). As was noted above, in general, subjects that represent a human referent are not explicitly identified. However, when they are overtly identified, they are often coded with *nun*-marking as shown in (4.13). In (4.13), the subject *na* ‘I’ is introduced into the utterance as an opposition to previous utterances. More specifically, before the utterance in (4.13), they were talking about shopping during Thanksgiving, and that everyone should at least go window-shopping. However, the subject *na* ‘I’, the speaker, in (4.13) said he did not like shopping in general, in opposition to previous utterances that say everyone should go at least window-shopping. Simply put, subjects with a human referent are normally not identified with *ka*, but they are often coded with *nun* for a discourse function such as marking the contrast

---

<sup>4</sup>The universal tendency that subject referents are generally animate (Hopper & Thompson 1980) is not reflected in the data in Table 4.5. I attribute this to the high occurrence of unexpressed subjects, i.e., zero-anaphor, that represent an animate referent, particularly the speaker.

encoding in subjects.

Thus far, I have said that referents of subject NPs are usually human beings, so subjects that represent a human referent are rarely identified with an explicit particle. Based on this tendency, I speculate that subjects with an interlocutor referent (i.e., the speaker or the hearer) are much less likely to be coded with an explicit marking because an interlocutor referent is not only a human being, thus unmarked subject animacy, but it can also be easily identified. In short, I speculate that subjects with an interlocutor referent are much easier to process. For this reason, I examined subject markings further, based on whether subject referents are interlocutors or not. The results are summarized in Table 4.6.

Table 4.6: Subject Interlocutor and Subject Marking

SUBJ.INTERLOCUTOR		SUBJ.MARK	NP-Ø	NP-ka	NP-nun	NP-to	NP-etc.	SUB-TOTAL	TOTAL
¬interlocutor	NO.		836	1,257	399	315	51	2,022	2,858
	%		<b>29</b>	<b>44</b>	<b>14</b>	11	2	71	100
interlocutor	NO.		391	307	246	93	5	651	1,042
	%		<b>38</b>	<b>29</b>	<b>24</b>	9	0	62	100
TOTAL	NO.		1,227	1,564	645	408	56	2,673	3,900
	%		31	40	17	10	1	69	100

In Table 4.6, I find that the use of a zero particle is more common for marking subjects whose referents are interlocutors (i.e., the speaker or the hearer) (38%), and the use of *ka*-marking is less common for encoding subjects with an interlocutor referent (29%). The distributional patterns shown in Table 4.6 suggest that in conversational Korean, the selection of subject markings correlates much more strongly with an interlocutor referent than with a mere human referent (zero-marking vs. *ka*-marking:  $\chi^2=54.901$ ,  $P < 0.001$ ). I attribute this tendency to the assumption that in general, an interlocutor referent, which is animate, is considered the default animacy for a subject

and is highly accessible, so it is rarely identified with *ka*. That is to say, the occurrence of *ka*-marking (29%) is low for coding subjects whose referents are interlocutors because they are easily identified; the use of a zero particle (38%) is instead more common.

In Table 4.6, I also find that the occurrence of *nun*-marking is relatively high for coding subjects when their referents are interlocutors (24%). Again, subjects that represent an interlocutor referent are normally not coded with *ka*, but they are often coded with *nun* for a discourse function such as marking the contrast, as was discussed above along with (4.13).

- (4.15) *kinyemphwumkakey-eyse na-Ø sasip pwul-Ø ssesseyo.*  
 souvenir.shop-LOC 1SG-( ) 40 dollar-( ) spent  
 ‘I spent \$40 at a souvenir shop.’

As has been discussed thus far, it is a universal tendency that referents of subject NPs are normally animate, and referents of object NPs are usually inanimate (Thompson & Hopper 2001; Aissen 2003). The utterance in (4.15) exemplifies a subject representing an animate referent together with an object representing an inanimate referent, in that the referent of subject *na* ‘I’ is animate, and the referent of object *sasip pwul* ‘40 dollars’ is inanimate. As repeatedly stated, the ‘animate’ is viewed as the norm for the referent of a subject NP, and the ‘inanimate’ is considered the norm for the referent of an object NP, so there is no particular need to disambiguate the object from the subject. Therefore, the occurrence of zero-marking is expected to be more common for coding subjects when the referent of an object is inanimate, as exemplified in (4.15).

- (4.16) *kulaykacikwu taykay poswucek-i-ya. wancen nay-ka kulayse*  
 so very conservative-be-SEM completely 1SG-NOM thus  
*appa-lul talmassnapa.*  
 dad-ACC resemble  
 ‘So (he) is very conservative. I seem to resemble my dad very much.’

On the other hand, the utterance in (4.16) is an example of both a subject and object that represent an animate referent, in that the referents of the subject *na* ‘I’ and the object *appa* ‘dad’ are animate. As stated above, the ‘animate’ is regarded as a deviation from the norm for the referent of an object NP, which would result in processing load, while it is considered the default animacy for the referent of a subject NP. Due to processing load, the use of a zero particle is expected to be less common for marking subjects when the referent of an object is animate, and the use of an overt particle is assumed to be more common for marking subjects instead, as exemplified in (4.16).

As repeatedly stated, it is a universal tendency that referents of subject NPs are generally animate, while those of object NPs are normally inanimate (Hopper & Thompson 1980; Givón 2000a; Aissen 2003). Hence, it is reasonable to assume that the choice of subject encoding types may be related to the animacy of object NPs. More specifically, the use of *ka*-marking may be more common for subjects when referents of objects are animate rather than when they are inanimate, in that both subject and object’s being animate is considered a deviation from the norm, hence processing load, resulting in the tendency for subjects to be overtly identified. In an attempt to more closely examine the relationship between the animacy and subject markings, I looked at the choice of subject markings with relation to the animacy of object referents. The results are summarized in Table 4.7.

According to Table 4.7, the token frequency for animate objects (69 tokens, approximately 9% of the total) is far less than that for inanimate objects (723 tokens, roughly 91% of the total).<sup>5</sup> As a matter of fact, this strongly supports the general tendency of objects, that is, referents of objects are usually inanimate.

---

<sup>5</sup>The token counts in Table 4.7 reflects only those clausal units that include both an overt subject and an overt object.

Table 4.7: Object Animacy and Subject Marking

SUBJ.MARK OBJ.ANIMACY		NP-Ø	NP-ka	NP-nun	NP-to	NP-etc.	SUB- TOTAL	TOTAL
inanimate	NO.	227	260	166	50	20	496	723
	%	<b>31</b>	<b>36</b>	<b>23</b>	7	3	<b>69</b>	100
animate	NO.	8	29	27	4	1	61	69
	%	<b>12</b>	<b>42</b>	<b>39</b>	6	1	<b>88</b>	100
TOTAL	NO.	235	289	193	54	21	557	792
	%	30	36	24	7	3	70	100

In Table 4.7, I find a noticeable correlation between the animacy of object referents and the choice of subject markings in that the occurrence of zero-marking is less common for coding subjects when referents of objects are animate than when they are inanimate (zero-marking vs. overt marking:  $\chi^2=14.803$ ,  $P < 0.001$ ). More specifically, when referents of object NPs are animate, the occurrence of zero-marking is relatively low (only 12%) for coding subject NPs, but the occurrence of *ka*-marking is relatively high (42%) for coding subject NPs. On the other hand, when referents of objects are inanimate, the occurrence of zero-marking is relatively high (31%) for subject NPs, and the occurrence of *ka*-marking is relatively low for subject NPs (zero-marking vs. *ka*-marking:  $\chi^2=8.682$ ,  $P < 0.01$ ). I attribute this tendency to the assumption that if both subject NPs and object NPs represent animate referents, which is considered a deviation from the norm, hence an increase in processing load, there is a tendency for the high occurrence of *ka*-marking for subject NPs. The data in Table 4.7 also shows that the occurrence of *nun*-marking (39%) is relatively high for subjects which occur with an object whose referent is animate (zero-marking vs. *nun*-marking:  $\chi^2=15.813$ ,  $P < 0.001$ ). As was noted above, subjects are sometimes identified with *nun* for a discourse function such as marking the contrast. This will be discussed further in detail

later in Section 4.5.9.

To summarize, in conversational Korean, the choice of subject markings does not correlate with the simple animacy of subject NPs. That is to say, although the use of *ka*-marking is less common for coding subjects whose referents are human beings than for coding those whose referents are inanimate or non-human animate, the use of a zero particle is almost equally low for marking subjects regardless of their animacy (see Table 4.5). In fact, this observation does not comply with previous findings in spoken Japanese (e.g., Keenan 1976; Comrie 1989; Fry 2003), that is, subject particle ellipsis occurs more frequently with a subject whose referent is animate. I also examined the choice of subject markings with relation to whether referents of subjects are interlocutors (i.e., the speaker or the hearer), as well as with relation to the animacy of objects. By examining the subject markings with regard to the animacy in a more precise way, I was able to show a correlation between the choice of subject markings and the animacy. More specifically, the occurrence of *ka*-marking is relatively low for coding subjects when the subject referents are interlocutors (see Table 4.6) because subjects representing an interlocutor referent can easily be identified, hence low processing load. Furthermore, the occurrence of zero-marking was relatively low (12%) and the occurrence of *ka*-marking was relatively high (42%) for coding subjects when referents of objects are animate (see Table 4.7) because it is considered a deviation from the norm when subject NPs and object NPs both represent animate referents, hence an increase in processing load. Based on the results summarized in Table 4.5 through Table 4.7, I argue that the occurrence of *ka*-marking is relatively high and the occurrence of zero-marking is relatively low when a deviation from the norm, thus an increase in processing load, is expected, and the opposite tendency is expected when a deviation from the norm does not occur. The fact that the correlation is clear when subject markings were examined with relation to the animacy of object NPs actually



suggests that there may be more than a single factor playing a role in the selectional process of subject markings in conversational Korean. For instance, the choice of subject markings is related not only to the animacy of subject NPs but also to the animacy of object NPs.

Lastly, there have been some previous studies (Fry 2003, among others) investigating subject markings with regard to the animacy of subject referents in a language like Japanese, but to the best of my knowledge, no studies have examined subject markings with regard to animacy in conversational Korean. More importantly, this study is not only more comprehensive than previous studies but also suggests the correlation between the choice of subject markings and the animacy more precisely.

### **4.5.3 Word order**

When word order is discussed, it is often an issue of whether post-predicative entities should be regarded as parts of the planned utterance. For example, some (e.g., Kuno 1978) claim that only pre-predicative entities should be considered when the word order is examined, and any post-predicative entities should be treated independently from the planned utterance. They view post-predicative entities as ‘after-thoughts’, which are normally not regarded as a part of the planned utterance. Others (e.g., Shimojo 2005) argue that not all post-predicative entities should be precluded for the determination of word order variations because some convey an essential part of the proposition, and they may be the (core) arguments of the planned utterance. Shimojo (2005) argues that entities are postposed for specific discourse functions such as defocusing. For this reason, Shimojo (*ibid.*) distinguishes instances of postposed arguments from those of simple ‘after-thoughts.’ The latter approach is valid, and it is taken in this dissertation; when the word order is determined, instances of postposed arguments are included,

whereas those of ‘after-thoughts’ are excluded in the present study.<sup>6</sup>

As was previously pointed out in Chapter 3, word order in Korean, similar to Japanese, is somewhat flexible. More specifically, the canonical or unmarked word order in Korean has been claimed to be SOV, but such a word order is not always observed, especially in spoken Korean (Sohn 1999; Ko. S 2000; Kim 2006). The data used in this study supports this view, in that, although SOV and SV are two most commonly observed word order patterns<sup>7</sup>, there are other word order variations (e.g., OSV) reported as well. It is also notable that Korean, as well as Japanese, has been claimed to be a predicate-final language (Kuno 1972; Cho 1994; Choi 1996; Sohn 1999), but not all clauses used in this study end with a predicate. In subsequent texts are given examples that show word order patterns found in the data of the present study.<sup>8</sup>

First, the following are examples that show the characteristics of a predicate-final language; (4.17) through (4.19) are examples which all end with a predicate.

(4.17) *kulen tancem-i isstela, kunkka.*  
such shortcoming-NOM exist you.know  
‘There was such a shortcoming, you know.’

(4.18) *akka ttak sihem-Ø kkuthna-ko,*  
while.ago DP exam-( ) end-and  
‘A while ago, after the exam was over, and...’

(4.19) *nay-ka mwullon kulehkey nay-ko*  
1SG-NOM of.course so pay-and  
‘Of course, I can pay (that much money), and ...’

---

<sup>6</sup>As previously noted in Section 3.3.1, for the distinction between after-thoughts and postposed arguments, I relied on whether there is a significant pause placed between the predicate and the post-predicative entity. If there is a significant pause, i.e., longer than 1 second, it is regarded as an after-thought. If there is no significant pause, it is viewed as a postposed argument, which is part of the planned utterance.

<sup>7</sup>In my data, SOV and SV together constitute 95% of the clausal units in total, excluding clausal units with unexpressed subjects.

<sup>8</sup>In the present study, only those clauses containing an explicit subject are considered for the determination of word ordering.

(4.17) and (4.18) are examples of SV word ordering in intransitive clauses (labeled as  $SV_i$  in Table 4.8). The subject in (4.17), *kulun tancem* ‘such a shortcoming’, is marked with the nominative *ka*, whereas the subject in (4.18), *sihem* ‘the exam’, is coded with a zero-particle. (4.19) is an example of SV word ordering in a transitive clause (labeled as  $SV_t$  in Table 4.8). The predicate in (4.19), *nayta* ‘pay’, is a transitive verb which requires an object, either in an implicit or explicit form. (4.19) is thus classified as an example of  $SV_t$  word ordering, which includes an implicit object.

Second, there are also many examples where a predicate does not appear at the end of the clause.

(4.20) *cincca eps-ta, phalsipo nyen sayng-Ø*  
 really not.exist-DCL 85 year birth-( )  
 ‘Wow, there is essentially no one who was born in 1985.’

(4.21) *chainathawun-ey ka-to eps-tela ccacangmyen-un.*  
 China.town-LOC go-though not.exist-SEM black.noodle-TOP  
 ‘Although (I) went to the China town, there was no black noodle.’

(4.22) *yulichang-eyta ilehkhey nakse-lul hay non-ta, salam-tul-i.*  
 window-on you.know scribble-ACC do place-DCL person-PL-NOM  
 ‘People scribbled on the window, you know.’

(4.20) and (4.21) are instances that exemplify VS word ordering in intransitive clauses (labeled as  $VS_i$  in Table 4.8). The subject in (4.20), *85 nyensayng* ‘the one born in 1985’, is marked with a zero particle, while the subject in (4.21), *ccacangmyen* ‘black noodle (dish)’, is coded with *nun*-marking. (4.22) is an example of an OVS word ordering where its subject, *salamtul* ‘people’, is marked with *i*, namely the nominative marker.

Lastly, there are also examples where a predicate appears at the end of the clause but an object appears before a subject.

(4.23) *yehayng-un*    *nay-ka*    *kunyang*    *an*    *ha-llyekwu*.  
           traveling-TOP    1SG-NOM    just            NEG    do-will  
           ‘As for the travel, I just won’t do (it).’

In (4.23), the predicate, *hata* ‘do’, appears in the final position of the clause, yet the object, *yehayng* ‘travel’, is topicalized with *nun*, and precedes the subject, *nay* ‘I’. In other words, (4.23) is an example of OSV word ordering, and its subject, *nay* ‘I’, is coded with *ka*-marking.

In the present study, the examination of word order variation is important in the discussion of subject markings in conversational Korean in two respects: processing load (incurred from a deviation from the norm) and informational prominence. More specifically, an increase in processing load can be assumed when subjects appear in a non-canonical, thus marked, word order variation, which is considered a deviation from the norm. It is further assumed that the increased processing load invites the use of *ka*-marking for coding subjects occurring in a marked word ordering. On the other hand, it has been claimed that in general, entities are postposed for the purpose of defocusing, and postposed NPs represent less important information than non-postposed NPs (Simon 1989; Shimojo 2005). As a matter of fact, Shimojo (2005:203) points out that in spoken Japanese, zero-marking is the encoding type that fits best for post-predicative arguments. I expect this general tendency to be observed for subjects in conversational Korean. That is to say, postposed subjects are assumed to be informationally less prominent than non-postposed subjects, so the use of zero-marking would be common for such postposed subjects in conversational Korean.

There have been several studies discussing word ordering with regard to subject markings in spoken Japanese, yet, to the best of my knowledge, there have been no studies discussing a relationship between subject markings and word orderings in conversational Korean. With an assumption that subject markings correlate with word order variations, I examined the choice of subject markings with respect to word or-

derings. For the purpose of this dissertation, I subcategorized all of the observed word order patterns based on the transitivity of sentences,<sup>9</sup> and the results are presented in Table 4.8.

Table 4.8: Word Order and Subject Marking

SUBJ. MARK		NP-Ø	NP-ka	NP-nun	NP-to	NP-etc.	SUB-TOTAL	TOTAL	
WORD	ORDER								
intran.	SV <sub>i</sub>	NO.	887	1,149	381	313	27	1,870	2,757
		%	<b>32</b>	<b>42</b>	<b>14</b>	11	1	68	100
	VS <sub>i</sub>	NO.	34	34	29	1	2	66	100
		%	<b>34</b>	<b>34</b>	<b>29</b>	1	2	66	100
tran.	SOV	NO.	214	255	181	51	14	501	715
		%	30	36	<b>25</b>	7	2	70	100
	SV <sub>t</sub>	NO.	65	90	39	37	6	172	237
		%	27	<b>38</b>	16	16	3	73	100
	OSV	NO.	11	23	5	1	4	33	44
		%	25	<b>52</b>	11	2	9	75	100
	OVS	NO.	5	7	4	0	2	13	18
		%	<b>28</b>	<b>39</b>	<b>22</b>	0	11	72	100
	VS <sub>t</sub>	NO.	6	2	3	3	0	8	14
		%	<b>43</b>	<b>14</b>	<b>21</b>	21	0	57	100
	SVO	NO.	4	4	3	2	1	10	14
		%	29	29	21	14	7	71	100
	VSO	NO.	1	0	0	0	0	0	1
		%	100	0	0	0	0	0	100
TOTAL	NO.	1,227	1,564	645	408	56	2,673	3,900	
	%	31	40	17	10	1	69	100	

— SV<sub>i</sub> and VS<sub>i</sub>:

In Table 4.8, I find that unlike Japanese, in conversational Korean, the occurrence of zero-marking is not particularly high for coding postposed subject NPs. In fact, the oc-

<sup>9</sup>There were also many instances of word order variations (e.g., V, OV, VO, etc.) where subjects were not explicitly expressed, and those instances with unexpressed subjects were precluded for the discussion of subject markings. Only those sentences with expressed subjects are considered for the examination of the relationship between word order variations and subject markings.

currence of zero-marking is almost equally low for coding both pre-predicative subjects (32%) and post-predicative subjects (34%). Therefore, such a claim that zero-marking is the encoding type that fits best for post-predicative arguments is not supported by the data in Table 4.8. Contrastively, the occurrence of *ka*-marking is relatively low for coding subject NPs occurring in the  $VS_i$  word ordering (34%), compared to those appearing in the  $SV_i$  word ordering (42%) (zero-marking vs. *ka*-marking:  $\chi^2=1.107$ ,  $P=0.2928$ ). Earlier in the present section, it was pointed out that post-predicative entities are usually defocused, i.e., unimportant, whereas important or urgent information is fronted or is placed before the less important or less urgent information (Simon 1989; Shimojo 2005). It was also pointed out that in conversational Japanese, post-predicative subjects are viewed as thematically unimportant, and zero-marking is a basic encoding type for post-predicative subjects due to the lack of thematic prominence (Shimojo 2005). However, in conversational Korean, post-predicative subjects do not necessarily imply thematically unimportant information, i.e., the lack of thematic prominence. In short, there is a striking difference between the two languages; in Japanese, post-predicative subjects represent thematically unimportant information, whereas in Korean, post-predicative subjects do not necessarily represent thematically unimportant information.<sup>10</sup>

On the other hand, word ordering of  $VS_i$  is considered a deviation from the norm, which would lead to an increase in processing load. That is to say,  $VS_i$  word ordering is regarded as marked word ordering, hence an increase in processing load, which would call for the use of *ka*-marking for post-predicative subjects. According to Table 4.8, the occurrence of *ka*-marked subjects is, however, lower in  $VS_i$  word ordering (34%) than in  $SV_i$  word ordering (42%). Furthermore, the percentage of *ka*-marked subjects (34%) is

---

<sup>10</sup>Unlike in Japanese, the cataphoric saliency, either by RP or by RP-*f*, which is used to measure thematic prominence (Shimojo 2005), does not play a role in the choice of subject encoding types in conversational Korean, as is discussed in Section 4.5.8.

the same as the percentage of zero-marked subjects (34%) in  $VS_i$  word ordering. This rather unexpected tendency may be due to the assumption that despite the marked word ordering, a subject is the only argument available in an intransitive sentence, so it is the only argument to identify. For example, despite the marked word ordering, the postposed subject in (4.20), *phalsipo nyensayng* ‘the one born in 1985’, is the only argument of the intransitive sentence, and it is coded with a zero-particle, . Unlike (4.20), there are two arguments to identify (i.e., the subject *salamtul* ‘people’ and the object *nakse* ‘scribble’) in (4.22), and the postposed subject, *salamtul* ‘people’, is coded with *ka*-marking.

This tendency may also be explained based on the accessibility of the postposed subjects. That is to say, the postposed subject, *phalsipo nyensayng* ‘the one born in 1985’, in (4.20) had been mentioned before it was reintroduced into the utterance in (4.20), though previous utterances were not spelled out in (4.20). Due to its being highly accessible, it is coded with a zero particle. On the other hand, the postposed subject in (4.22), *salamtul* ‘people’, is newly introduced into this utterance, so it had not been activated, i.e., it is inaccessible. Due to its being inaccessible, it is coded with *ka*-marking. In fact, this suggests that there can be more than a single factor influencing the selectional process of subject markings in conversational Korean.

Furthermore, based on previous claims for Japanese, which is a verb final language, that is, the post-predicative arguments are defocused, i.e., degraded prominence (Simon 1989; Shimojo 2005, *inter alia*), it can be assumed for Korean, which is also a verb final language, that the post-predicative arguments somehow represent attenuated informational prominence. That is to say, although the post-predicative subject, *phalsipo nyensayng* ‘the one born in 1985’, in (4.20) is in marked position, hence high processing load, it also represents attenuated informational prominence. Therefore, despite the marked position, the motivation for *ka*-marking for the subject in (4.20) is downgraded.

Based on this tendency, I argue that both factors of processing load and informational prominence often interplay as in word order variations, and they both play roles in the selection of subject markings.

The data in Table 4.8 also shows that the occurrence of *nun*-marking is higher for coding subjects in  $VS_i$  word order (29%) than for coding subjects in  $SV_i$  word order (14%). For instance, the post-predicative subject in (4.21), *ccacangmyen* ‘black noodle’, is coded with *nun*-marking, while the pre-predicative subject in (4.17), *kulen tancem* ‘such a shortcoming’, is coded with *ka*-marking. A possible explanation of this tendency can be that subjects in a sentence-final position happen to be a grammatical topic of a given sentence, which is marked by the post-nominal particle *nun*. That is to say, a sentence-initial position<sup>11</sup> is a default grammatical topic position (Lambrecht 1994; Givón 2000a), but a sentence-final position may also be a grammatical topic position, which would call for *nun*-marking (*ka*-marking vs. *nun*-marking:  $\chi^2=14.134$ ,  $P < 0.001$ ). Simply put, a sentence-initial position is regarded as a primary topic position, and the occurrence of *nun*-marking is relatively low for coding a subject in a sentence-initial position (14%). On the other hand, a sentence-final position is not considered a default topic position, though it can also be a topic position, and the occurrence of *nun*-marking is relatively high for coding a sentence-final topic marked subject (29%).

— **OSV and OVS:**

For transitive clauses, the use of *ka*-marking is very common for coding subjects in OSV word order (52%), as is exemplified in (4.23) where the subject, *nay* ‘I’, is coded with *ka*-marking. I attribute this to the assumption that OSV word ordering is viewed as marked, thus a deviation from the norm, which would result in high processing load.

---

<sup>11</sup>Choi (1997) points out that in Korean, topic reading is predominant with a sentence-initial subject.



The processing load invites the use of *ka*-marking to encode these subjects that appear in marked word ordering.

In addition, it may be enough to explicitly mark either a subject only or an object only in order to identify the subject or the object in a sentence. That is to say, the occurrence of *ka*-marked subjects may be relatively high when objects are marked with a zero particle, but the occurrence of zero-marked subjects may be relatively high when objects are marked with the accusative *lul*. In order to see if the choice of subject markings is influenced by object markings, I examined object markings in OSV word ordering as well (see Table 5.8 in Section 5.4.3). In OSV word ordering, objects occur more frequently with zero-marking (50%) than with *lul*-marking (25%). Based on this, it may be said that the high occurrence of *ka*-marked subjects in OSV word ordering is attributed to the low occurrence of *lul*-marked objects appearing in the same clause. However, it still remains as a question to be answered why subjects, but not objects, are coded with an explicit marking. I will further discuss this issue in Section 5.4.3.

In OVS word ordering, the use of *ka*-marking (39%) is more common for coding subjects than the use of zero-marking (28%), as is exemplified in (4.22) where the subject *salamtul* ‘people’ is coded with *ka*-marking. Furthermore, unlike subjects in OSV word ordering, the occurrence of *nun*-marked subjects in OVS word ordering is relatively high. As was proposed above, I relate this to the assumption that a post-predicative subject happens to be a grammatical topic of a given sentence, which is marked by the post-nominal particle *nun*, while subjects occurring in a middle position can hardly be a grammatical topic. Nevertheless, the number of subject tokens in OVS word ordering is too small to make any meaningful generalization.

— **VS<sub>t</sub> and SV<sub>t</sub>:**

Compared to that of subjects in SV<sub>t</sub> word ordering, the occurrence of *ka*-marking is very low (only 14%), and the use of zero-marking is very common (43%) for subjects

appearing in  $VS_t$  word ordering. Furthermore, the occurrence of *nun*-marking is more common for subjects in  $VS_t$  word ordering (21%) than for subjects in  $SV_t$  word ordering (16%). Yet the number of tokens for subjects occurring in  $VS_t$  word order is relatively small, so this distributional tendency may not bear a statistical significance.

Across all word order patterns observed in the present study, the occurrence of *nun*-marking is relatively high for coding subjects that appear in a sentence-initial or post-predicative position, with an exception that the occurrence of *nun*-marked subjects is relatively low in intransitive clauses (14%). On the other hand, *nun*-marking appears not to be preferred for coding subjects that come between an object and its predicate (see Table 4.8). This would be due to the property of grammatical topic; subjects in a sentence-initial or sentence-final position may be a grammatical topic of a given sentence, whereas subjects in a sentence-middle position may not be a grammatical topic of the given sentence.

To summarize, in the present section, subject markings were discussed with relation to word order variations in two respects: processing load and informational prominence. According to the data in Table 4.8, it seems that in conversational Korean, the choice of subject markings is related to word order variations. For example, subjects that appear in a non-canonical, thus marked, word ordering tend to be coded with *ka*-marking (e.g., in OSV word ordering, *ka*-marked subjects take up 52% of the total). Yet the occurrence of *ka*-marking is not always high for all subjects appearing in a marked order variation, which represents processing load. For example, the use of *ka*-marking is less common for post-predicative subjects than it is for pre-predicative subjects (e.g., 34% of *ka*-marked subjects in  $VS_i$  word ordering; 42% of *ka*-marked subjects in  $SV_i$  word ordering). I attributed this pattern to the assumption that although a post-predicative subject is in a marked position, and thus represents a processing load, the post-predicative subject represents attenuated informational prominence, so

the motivation for *ka*-marking is downgraded. Based on this, I argued that both factors of processing load and informational prominence often interplay, as in word order variations, and they both play roles in the selection of subject markings. I also pointed out there is a striking difference between Korean and Japanese in that in conversational Japanese, post-predicative subjects are viewed as thematically unimportant, but in conversational Korean, post-predicative subjects do not necessarily represent thematically unimportant information. In fact, this argument is supported by the findings of the present study, i.e., cataphoric saliency in terms of RP and RP-*f* does not play a role in the selection of subject encoding types (see the discussion in Section 4.5.8). It was also pointed out that the use of *nun*-marking is more common for post-predicative subjects than for pre-predicative subjects, and it is not common for subjects appearing in a sentence-middle position due to the general property of grammatical topic; NPs in a sentence-initial or sentence-final position represent grammatical topic position, whereas NPs in a sentence-middle position do not. Furthermore, the occurrence of *ka*-marked subjects (52%) is relative high when objects of the clause are marked with a zero particle (50%), and I associated this tendency with the assumption it may be enough to mark either subjects only or objects only to identify the subject or object in a clause (see Table 5.8 in Section 5.4.3 for comparison). Lastly, according to Table 4.8, it is more common to postpose subjects than objects: a total of 177 subjects are postposed, while only 15 objects are postposed.

#### 4.5.4 Definiteness

The notion ‘definiteness’ has often been introduced into the discussion of argument as well as particle ellipsis (e.g., Lee 1989; Lee & Thompson 1989; Fujii & Ono 2000; Ono *et al.* 2000; Fry 2003; Shimojo 2005, *inter alia*). For instance, Fry (2003) points out that definiteness is a feature which is important in the direct object marking system of many

languages, including Japanese. More specifically, referents of demonstrative pronouns are accessible to both a speaker and an addressee, and they are most often coded with zero-marking. Lee (1989) argues that as for subject markings in Korean, *ka*-marking is more associated with indefiniteness, whereas *nun*-marking is more with definiteness. Though useful, this bisectional distinction between these two subject markings is not clear-cut because a definite reading can also be derived from *ka*-marked subjects, as was previously pointed out by Kang (1998).

I assume that in conversational Korean, the choice of subject markings is related to definiteness of subject referents. More specifically, I assume that *ka*-marking is not preferred, but zero-marking is instead preferred for encoding subjects with a definite referent due to easy identifiability. In order to better study such a relationship, subject markings are examined with regard to the definiteness of subject referents in a more precise way in this dissertation. For the purpose of this dissertation, all subject NPs are categorized into four different sub-groups based on definiteness of their referents.<sup>12</sup> Based on descriptions of definiteness given by Gundel *et al.* (1993), Lambrecht (1994) and Aissen (2003), definiteness is categorized into four different subtypes: definite, indefinite, indefinite pronoun, indefinite but specific. The definition for each of the four sub-categories is given below along with relevant examples that are utilized in the present study.

- **definite NPs:** have their referents uniquely identifiable to both the speaker and the hearer. Pronouns, personal pronouns, demonstrative pronouns and proper nouns are good examples of definite NPs.

(4.24) *kulem* *na-Ø* *ipwul-Ø* *kacye-ka-l-kkey.* *na-Ø* *echaphi*  
 then 1SG-( ) comforter-( ) bring-go-FUT-SEM 1SG-( ) anyway  
*Thwutoru-Ø* *manna-yatoy-ketun.*  
 Tudor-( ) meet-should-SEM

---

<sup>12</sup>There can be more than these four categories into which subject NPs are categorized.

‘Then, I will bring my comforter. I need to meet Tudor anyway.’

- (4.25) enni-nun eti-Ø sal-a?  
sister-TOP where-( ) live-Q  
‘Where does your sister live?’

Both underlined subjects in (4.24), *na* ‘I’, refer to the speaker, are thus pronouns, and they are easily identifiable to both the speaker and the hearer. The subject in (4.24) is thus viewed as a definite NP. Furthermore, due to its being highly accessible, it is assumed that the subject does not need to be explicitly marked for its identification. Therefore, the use of zero-marking to code *na* ‘I’ in (4.24) is not unexpected. The underlined subject in (4.25), *enni* ‘big-sister’, is counted as a definite NP because both the subject and the hearer know that its referent is the hearer’s sister, and it is coded with *nun*.

- **indefinite NPs:** have their referents identifiable neither to the speaker nor to the hearer. In general, the referents of indefinite NPs are generic/non-specific or new.

- (4.26) salam-tul-i tto nosukholia yok-Ø hay-ss-kwuna?  
people-PL-NOM again North.Korea slander-( ) do-PST-SEM  
‘People did swear about North Korea again, didn’t they?’

The underlined NP in (4.26), *salamtul* ‘people’, does not refer to the specific people, so its referent is identifiable neither to the speaker nor to the hearer. The subject in (4.26) is thus identified as an indefinite NP. Furthermore, the subject in (4.26) is not identifiable, so it is assumed to be explicitly coded with a nominative particle for its identification.

- **indefinite pronoun NPs:** appear in pronominal forms and are referential to

the speaker, but their referents have not been activated for the hearer.<sup>13</sup> Also, indefinite pronoun NPs do not contain any specific information that needs to be recognized.

- (4.27) *mwo-Ø hay-ss-e? keki-Ø ettay-ss-e? mwoya, ceki,*  
 what-( ) do-PST-SEM that.place-( ) how-PST-SEM DP DP  
*posuthen-Ø ka-ss-ta wa-ss-ci?*  
 Boston-( ) go-PST-and come-PST-Q  
 ‘What did you do? How was that place? What you may call it, there..., you’ve been to Boston, haven’t you?’

The underlined subject in (4.27), *keki* ‘that place’, appears in its pronominal form, i.e., demonstrative, and it is referential although the speaker experiences an association problem. That is to say, the speaker is still searching for an appropriate word which can be associated with what is in the speaker’s mind at the time of utterance. Two words that follow, *mwoya* ‘what you may call it’, *ceki* ‘there...’, clearly indicate that the speaker is still in search of a word which can be connected with what is in her mind. The speaker finally finds a word, and she uses the word, *posuthen* ‘Boston’, in the last utterance in (4.27) to associate it with what had been in her mind. Despite the association problem, it is clear that the speaker refers to a specific place whose referent is identifiable to the speaker. Yet its referent had not been activated for the hearer until it is associated with the word, *posuthen* ‘Boston’, in the subsequent utterance, so it is not identifiable to the hearer at this stage. For this reason, the subject in (4.27), *keki* ‘that place’, is classified as an example of an indefinite pronoun NP. Furthermore, unlike indefinite or definite NPs, indefinite pronouns usually do not convey specific

---

<sup>13</sup>Gundel *et al.* (1993) recognize considerable debate concerning the referential status of both definite and indefinite expressions, but argue that indefinite expressions can be used referentially or non-referentially. They (*ibid.* p.5) note that referents of indefinite pronoun phrases with indefinite *this* are likely to be continued in subsequent sentences. Indefinite pronoun *this* can be viewed as an extension of the cataphoric use of the proximal demonstrative. That is to say, it is used in reference to an object that will not be activated for the addressee until the next sentence is processed.

information that is important and needs to be recognized, so they are often coded with zero-marking, as can be seen in (4.27).

- **indefinite but specific NPs:** have their referents identifiable only to the speaker but not to the hearer. Unlike indefinite pronouns, they appear in their noun forms, i.e., full noun.

(4.28) *pap-ilang, mak, ku, [emma-ka haycwu-ten kulen panchan-Ø*  
 rice-with you.know that mom-NOM do-COMP such side.dish-( )  
*pha-nun] tey-ka ccwak issnunkeya.*  
 sell-COMP store-NOM widely exist  
 ‘Along with meals, you know, there were many stores which sell side dishes that (our) moms used to cook (for us).’

Subjects that are categorized as an ‘indefinite but specific NP’ are somewhat similar to those categorized as an ‘indefinite pronoun’, in that referents for both subtypes of subject NPs are identifiable only to the speaker, but not to the hearer. However, in the present study, subjects of ‘indefinite but specific NP’ are distinguished from those of ‘indefinite pronoun’, because unlike subjects of ‘indefinite pronoun’, they appear in a full noun form and contain information that is specific to some degree, though limiting its information to the generic property of the referent. Conversely, subjects of ‘indefinite pronoun’ do not provide the hearer with any information about the referent, not even the generic property of the referent. The underlined subject in (4.28), *tey* ‘a store’, is regarded as an example of an ‘indefinite but specific NP’ in the sense that its referent is newly introduced into the utterance and it is only accessible to the speaker but not to the hearer. In other words, the referent is not identifiable to the hearer although it is to the speaker. Since its referent is identifiable to the speaker only, the subject is assumed to be overtly coded with the nominative *ka* due to processing load in the identification of its referent.

In the present study, it is assumed that the subjects of both a definite NP and an indefinite pronoun NP tend to be coded with zero-marking, while the subjects of both an indefinite NP and an indefinite but specific NP are explicitly marked with *ka*. In order to see if subject markings correlate with definiteness, all subject NPs are examined with regard to definiteness of their referents. The results are summarized in Table 4.9.

Table 4.9: Subject Definiteness and Subject Marking

SUBJ.MARK		NP-Ø	NP-ka	NP-nun	NP-to	NP-etc.	SUB-TOTAL	TOTAL
SUBJ.DEFINITE								
definite	NO.	676	753	464	200	27	1,444	2,120
	%	<b>32</b>	<b>36</b>	<b>22</b>	9	1	68	100
indefinite	NO.	450	747	165	187	28	1,127	1,577
	%	29	<b>47</b>	<b>10</b>	12	2	71	100
indef. pron.	NO.	91	43	15	17	1	76	167
	%	<b>54</b>	<b>26</b>	9	10	1	46	100
indef./specific	NO.	10	21	1	4	0	26	36
	%	28	<b>58</b>	3	11	0	72	100
TOTAL	NO.	1,227	1,564	645	408	56	2,673	3,900
	%	31	40	17	10	1	69	100

According to Table 4.9, it seems that in conversational Korean, definiteness of subject referents plays a role in the selection of subject markings, though the use of zero-marking is generally not common except for subjects that represent an indefinite referent. For instance, the occurrence of *ka*-marking (36%) is relatively low for coding subjects that represent a definite referent. I attribute this to the assumption that definite NPs (e.g., pronouns, personal pronouns, proper nouns, etc.) are easily identifiable to both the speaker and the hearer, thus represent minimal processing load in referent identification, so the use of *ka*-marking is not expected (notice that both underlined subjects in (4.24) are coded with a zero particle). As a matter of fact, this observation is, at least partially, consistent with what has been reported in previous studies of a



similar regard (e.g., Fujii & Ono 2000; Fry 2003), in that an explicit marking (e.g., *ka* for Korean; *ga* for Japanese) is not common for definite NPs. Instead, the occurrence of *nun*-marking is relatively high for coding subjects categorized into ‘definite’ (22%), compared to the occurrence of *nun*-marking for subjects categorized into other subtypes. This tendency supports Lee’s (1989) observation that *nun*-marking is associated with subjects that represent a definite referent.

In Table 4.9, I also find that the occurrence of *ka*-marking is relatively high for coding indefinite subject NPs (47%). I associate this tendency with the assumption that indefinite entities are identifiable neither to the speaker nor to the hearer, thus represent processing load in referent identification, so the occurrence of *ka*-marking is relatively high for indefinite subjects (see the utterance in (4.26)). On the other hand, the occurrence of *nun*-marking is low for indefinite subjects (10%). Simply put, the use of *ka*-marking is more common for indefinite subjects (47%) than for definite subjects, due to processing load in referent identification for indefinite subjects. Contrastively, the use of *nun*-marking is more common for subjects with a definite referent (22%) than for those with an indefinite referent (definite vs. indefinite:  $\chi^2=100.525$ ,  $P < 0.001$ ).

Furthermore, the use of *ka*-marking is high for subjects that represent an indefinite but specific referent (58%). As was stated earlier in the present section, I relate this tendency to the assumption that subjects of an ‘indefinite but specific NP’ are identifiable to the speaker only, thus represent processing load in referent identification, so the occurrence of *ka*-marking is high to alleviate this processing load. For instance, the subject in (4.28), (*emmaka haycwuten kulen panchan phanun*) *tey* ‘a store (that sells side-dishes our moms used to cook)’, is viewed as indefinite but specific, and its referent is identifiable only to the speaker. Thus, processing load is expected in referent identification on the part of the hearer, and it is coded with *ka*-marking to alleviate this processing load, as shown in (4.28).

Lastly, Table 4.9 shows that unlike indefinite subjects, the occurrence of *ka*-marking is low (26%) for coding subjects of an indefinite pronoun NP, and the occurrence of zero-marking (54%) instead is high for coding them (definite vs. indefinite pronoun:  $\chi^2=20.812$ ,  $P < 0.001$ ; indefinite vs. indefinite pronoun:  $\chi^2=45.91$ ,  $P < 0.001$ ). This observation is consistent with what has been claimed in previous studies (e.g., Lee & Thompson 1989; Fujii & Ono 2000), in that indefinite pronouns express non-specific information, and the use of zero-marking is most common for coding them. In an extension of previous studies, I further associate the high occurrence of zero-marking used for indefinite pronoun NPs with the degraded informational prominence encoded in them. That is to say, subjects of ‘indefinite pronouns’ do not contain any information about the referent, not even the generic property of the referent, and thus a low degree of informational prominence is normally given to subjects of ‘indefinite pronouns’. As a result, they are commonly identified with a zero particle, as exemplified in the utterance in (4.27).

According to previous studies (e.g., Lee & Thompson 1989; Fujii & Ono 2000, among others), as well as what is observed in the present study, the relationship between definiteness and subject markings can be summarized as follows: unlike definite subjects, referents of indefinite subjects are not easily identifiable, so the occurrence of *ka*-marking is relatively high for them; referents of indefinite pronoun subjects do not refer to anything specific, so the occurrence of zero-marking is high for coding them.

However, it is paradoxical to claim that referents of indefinite subjects are not identifiable, so the occurrence of *ka*-marking is relatively high for coding them, while referents of indefinite pronoun subjects do not refer to anything specific, thus are not identifiable, so the use of zero-marking is common for coding them. That is to say, indefinite subjects and indefinite pronoun subjects are both not identifiable, and the former is coded with *ka*-marking but the latter is with zero-marking. I believe that the

source of the paradox comes from the overlooked fact that a referent of an indefinite pronoun subject is not only unidentifiable but it also bears low importance. More specifically, indefiniteness encoded in an ‘indefinite pronoun’ subject may represent processing load in referent identification. However, unlike indefinite or definite subjects, indefinite pronoun subjects do not bear any specific information about their referents. Since there is no specific information to be recognized, a low degree of informational prominence is assumed for indefinite pronoun subjects. Based on this, I propose that not only processing load but also informational prominence should be considered when subjects that represent an indefinite pronoun referent are examined for their markings because both factors interplay, as in the definiteness of subject referents.

To summarize the present section, although the occurrence of zero-marking is generally low for coding subjects except indefinite pronoun subjects, it seems that in conversational Korean, the definiteness of subject referents influences the choice of subject markings. Furthermore, what is observed in the present study, at least partially, supports what was claimed in previous studies (e.g., Lee & Thompson 1989; Fujii & Ono 2000; Fry 2003). Subjects whose referents are ‘definite’ are uniquely identifiable, thus are assumed to represent minimal processing load, and the occurrence of *ka*-marking is relatively low (36%) for coding them (see the utterance in (4.24)). Conversely, due to the definiteness, the use of *nun*-marking is more common for definite subjects (22%) than for other subtypes of subjects (see the utterance in (4.25)). Subjects that represent an ‘indefinite’ referent are not identifiable to both the speaker and the hearer, thus high processing load is assumed for referent identification, and the occurrence of *ka*-marking, which I argue alleviates the processing load, is relatively high (47%) for them (see the utterance in (4.26)). Subjects that represent an ‘indefinite but specific’ referent are identifiable only to the speaker, hence processing load is assumed for referent identification, and they are commonly coded with *ka*-marking (58%) (see the

utterance in (4.28)). Lastly, the occurrence of *ka*-marking is low (26%) but instead the occurrence of zero-marking is high (54%) for subjects that represent an ‘indefinite pronoun’ referent. I attributed this tendency to the assumption that due to the lack of informational prominence, indefinite pronoun subjects tend not be recognized explicitly with *ka*-marking, although indefiniteness may impose processing load in referent identification (see the utterance in (4.27)).

#### 4.5.5 Length

It has been pointed out in previous studies (e.g., Tsutsui 1984; Ono *et al.* 2000) that the length of an NP has a correlation with the ellipsis of case particles in Japanese. In colloquial Japanese, monosyllabic NPs are not commonly marked with a zero-particle, and large NPs are commonly marked with an overt particle. More specifically, in Japanese, one-mora or very short NPs are not common, and are often not easily recognizable, especially in fast speech, so they are good candidates for an overt marking (e.g., *ga*-marking), which makes them more recognizable (Ono *et al.* 2000). On the other hand, phonologically larger NPs are more commonly coded with an overt particle than are phonologically shorter NPs (Tsutsui 1984).<sup>14</sup> Based on this tendency observed in spoken Japanese, I make the following assumption; in spoken Japanese, due to its reduced saliency in form, processing load is assumed for monosyllabic NPs, and the use of an explicit particle (e.g., *ga*) is common for coding them to alleviate this processing load. On the other hand, in general, large NPs tend to encode much information, thus are assumed to be high in informational prominence, and the use of an overt particle (e.g., *ga*) is relatively high for coding them.

In order to find out whether the length of subject NPs, in terms of processing load and informational prominence, influences the choice of subject markings in conversa-

---

<sup>14</sup>She used the syllable-length as a way of measuring the length of a given NP.

tional Korean, as was found in colloquial Japanese, I examined subject markings with respect to the length of subject NPs. In previous studies (e.g., Tsutsui 1984; Ono *et al.* 2000), the length of subject NPs was measured based on whether they were monosyllabic or how large they were in terms of syllable-length. In the present study, the length of subject NPs was measured in a similar but somewhat different manner. Just like in previous studies (e.g., Tsutsui 1984; Ono *et al.* 2000), the present study also examined whether or not subject NPs were monosyllabic. However, as a way of measuring how large subject NPs are, I examined how many words were used to modify the subject NPs, instead of using syllable length that was used as a means of length measurement in previous studies (e.g., Tsutsui 1984).

I will first discuss monosyllabic subject NPs with regard to their marking, and later discuss the number of subject modifiers with respect to subject marking. Following are examples of monosyllabic as well as non-monosyllabic subject NPs from the data of this dissertation.

(4.29) (A1) *an kulayto cikum nyuyok-ey iss-e.*  
 not be.so now New.York-LOC be-DCL  
 ‘Speaking of him, (He) is in New York.’

(B1) *ah kulay?*  
 oh be.so  
 ‘Oh, is that so?’

(A2) *yehayng-Ø ha-ko isse.*  
 travel-( ) do-and be  
 ‘(He) is traveling.’

(B2) *honcase?*  
 alone  
 ‘Alone?’

(B3) *kyay-Ø hyuhak-sayng-i-ya?*  
 he-( ) leave.of.absence-student-be-Q  
 ‘Is he a student?’

(4.30) na-nun kuke-Ø kkunh-ess-e.  
 1SG-TOP that-( ) get-PST-SEM  
 ‘I purchased that.’

(4.31) matang-i iss-umyen, mwucoken tongmwul-i wa.  
 front.yard-NOM exist-if no.condition animal-NOM come  
 ‘It is always the case that if there is a front yard, animals come.’

In the utterance labeled as B3 in (4.29), the underlined *kyay* ‘he’, which is marked with a zero particle, is counted as a monosyllabic subject of the clause. The utterance in (4.30) is also viewed as an example of monosyllabic subjects, and the subject *na* ‘I’ that is underlined is marked with *nun*. On the other hand, the utterance in (4.31) is taken as an example of non-monosyllabic subject NPs, in that both the underlined *matang* ‘front yard’ and *tongmwul* ‘animal’ are two-syllable long subjects, and they are both coded with *ka*-marking. Table 4.10 presents the selection of subject markings with relation to their being monosyllabic.

Table 4.10: Monosyllable and Subject Marking

SUBJ.MARK		NP-Ø	NP-ka	NP-nun	NP-to	NP-etc.	SUB-TOTAL	TOTAL
SUBJ.MONO								
¬monosyllabic	NO.	789	1,141	375	299	55	1,870	2,659
	%	<b>30</b>	<b>43</b>	14	11	2	<b>70</b>	100
monosyllabic	NO.	438	423	270	109	1	803	1,241
	%	<b>35</b>	<b>34</b>	22	9	0	<b>65</b>	100
TOTAL	NO.	1,227	1,564	645	408	56	2,673	3,900
	%	31	40	17	10	1	69	100

In Table 4.10, I find that in conversational Korean, the use of the overt marking, either with *ka* or with *nun*, is common for monosyllabic subject NPs, whereas the use of zero-marking is not common for monosyllabic subject NPs (zero-marking vs. overt marking:  $\chi^2=12.399$ ,  $P < 0.001$ ). However, the use of *ka*-marking is not particularly preferred for monosyllabic subject NPs, although it was reported that the use of

*ga*-marking is preferred for monosyllabic subjects in spoken Japanese, as was previously pointed out (Ono *et al.* 2000). In fact, Table 4.10 shows that the occurrence of *ka*-marking is even lower for encoding monosyllabic subjects (34%) than for encoding non-monosyllabic subjects (43%). In short, unlike spoken Japanese, monosyllabicity does not seem to invite *ka*-marking for monosyllabic subject NPs in conversational Korean, at least according to the data presented in Table 4.10. Then, it is natural to wonder why these two languages, which share many linguistic properties, differ in this matter. I suggest that such a difference between these languages may be attributed to the tendency that in Korean, monosyllabic subjects occur quite frequently (see Table 4.10), whereas in Japanese, they do not occur commonly, as was pointed out in Ono *et al.* (2000). In general, entities that commonly occur are more accessible than those that do not, so the occurrence of an overt particle (e.g., *ka* for subjects in Korean; *ga* for subjects in Japanese) is expected to be lower for marking commonly occurring entities, as was in Korean, than for marking uncommonly occurring entities, as reported in Japanese.

One thing that still needs to be pointed out is that (personal) pronouns in Korean (e.g., *na* ‘I’, *ku* ‘he’) are mostly monosyllabic, whereas (personal) pronouns in Japanese (e.g., *watasi* ‘I’, *karae* ‘he’) are not. In general, pronouns are definite, so their referents are more identifiable than that of full nouns. Therefore, less frequent use of *ka*-marking is expected for pronominal, thus definite, subjects, as was exemplified in (4.29) as well as in (4.30). For this reason, I reexamined all tokens of monosyllabic subjects, by categorizing them into two different groups based on their nominal types: monosyllabic pronoun subjects and monosyllabic full-noun subjects. The result is presented in Table 4.11.

According to Table 4.11, the occurrence of *ka*-marking is much lower with monosyllabic pronoun subjects (32%) than with monosyllabic full-noun subjects (47%). Fur-

Table 4.11: Monosyllabic Subject Type and Subject Marking

SUBJ.MARK		NP-Ø	NP-ka	NP-nun	NP-to	NP-etc.	SUB-TOTAL	TOTAL
SUBJ.MONO								
mono.pronoun	NO.	377	336	251	92	1	680	1,057
	%	<b>36</b>	<b>32</b>	<b>24</b>	9	0	64	100
mono.full-noun	NO.	61	87	19	17	0	123	184
	%	<b>33</b>	<b>47</b>	<b>10</b>	9	0	67	100
TOTAL	NO.	438	423	270	109	1	803	1,241
	%	35	34	22	9	0	65	100

thermore, the total number of monosyllabic pronoun subject tokens (1,057, approximately 85% of the entire token) is far more than that of monosyllabic full-nouns (only 184, approximately 15% of the entire token). It is notable that no personal pronouns are monosyllabic in Japanese, while they are mostly monosyllabic in Korean. Therefore, the fact that monosyllabic subjects are uncommon in Japanese further suggests that monosyllabic full-noun subjects are uncommon in Japanese. Based on this, it can further be said that just like in colloquial Japanese, monosyllabic full-noun subjects also occur somewhat infrequently in Korean, and they are more commonly coded with *ka*-marking (47%) rather than zero-marking (33%). On the other hand, in Korean, monosyllabic pronominal subjects not only occur commonly, but they are also easier to identify due to their definiteness, so they are less commonly coded with *ka*-marking than zero-marking, as shown in Table 4.11. In short, monosyllabic full-noun subjects are uncommon in both Korean and Japanese, and they are frequently coded with an overt particle, due to processing load in referent identification (*ka* for Korean; *ga* for Japanese). Conversely, in Korean, monosyllabic pronoun subjects are common as well as definite, thus are highly identifiable, so the overt marking with *ka* is not common for encoding monosyllabic pronoun subjects.



In Table 4.11, I also find that the use of *ka*-marking is less common for monosyllabic pronoun subjects (32%) than for monosyllabic full-noun subjects (47%), and instead the use of zero-marking is more common for monosyllabic pronoun subjects (36%) than for monosyllabic full-noun subjects (33%) (zero-marking vs. *ka*-marking:  $\chi^2=6.666$ ,  $P < 0.01$ ). Instead of zero-marking, the occurrence of *nun*-marking is higher for monosyllabic pronoun subjects than for monosyllabic full-noun subjects (*ka*-marking vs. *nun*-marking:  $\chi^2=23.287$ ,  $P < 0.001$ ).

Based on this, I argue that a simple comparison of encoding types for monosyllabic subjects between these two languages without considering detailed types of subjects can be meaningless. However, it can be claimed, based on the data presented in Table 4.11, that an explicit marking (e.g., *ka* for subjects in Korean; *ga* for subjects in Japanese) is commonly used to encode monosyllabic full-noun subjects in both languages. Despite this similarity, the occurrence of *ka*-marked monosyllabic full-noun subjects in Korean is still lower than that of Japanese because in colloquial Japanese, monosyllabic subjects are much more commonly marked with the nominative *ga* than with a zero particle (Ono *et al.* 2000). This difference may be attributed to other factors such as informational prominence that may facilitate the use of an overt marking (*ka* for subjects in Korean; *ga* for subjects in Japanese). For instance, due to its short length, monosyllabic entities are not informationally loaded, hence less likely to represent informational prominence, than non-monosyllabic entities, so the use of an overt marking (e.g., *ka*-marking) is not common for encoding monosyllabic subjects. Although monosyllabic entities generally represent low informational prominence, they may become informationally prominent when they are modified. Due to informational prominence, the use of overt-marking is assumed to be more common for monosyllabic entities that are modified than for those that are not modified.

To the best of my knowledge, no studies have examined encoding types of mono-

syllabic subjects in Japanese, let alone in Korean, further based on the number of modifying words. Hence, I cannot compare the two languages in this regard, but I can at least discuss whether the encoding types of monosyllabic subjects are influenced by the number of modifying words. For this reason, I categorized all monosyllabic subjects in terms of their number of modifiers, and the results are summarized in Table 4.12.

Table 4.12: Modifiers for Monosyllabic Subjects and Subject Marking

SUBJ.MARK		NP-Ø	NP-ka	NP-nun	NP-to	NP-etc.	SUB-TOTAL	TOTAL
SUBJ.MONO								
0 modifier	NO.	437	418	268	109	1	796	1,233
	%	<b>35</b>	<b>34</b>	22	9	0	65	100
1 modifier	NO.	1	3	2	0	0	5	6
	%	<b>17</b>	<b>50</b>	33	0	0	83	100
2 modifiers	NO.	0	2	0	0	0	2	2
	%	<b>0</b>	<b>100</b>	0	0	0	100	100
TOTAL	NO.	438	423	270	109	1	803	1,241
	%	35	34	22	9	0	65	100

In Table 4.12, I find that the occurrence of *ka*-marked monosyllabic subjects is proportional to their number of modifiers. For instance, the occurrence of *ka*-marked monosyllabic subjects is only 34% when there is no modifier, but it significantly increases up to 50% when there is one modifier. It further increases up to 100% when there are two modifiers. Comparatively, the occurrence of zero-marked monosyllabic subjects is inversely proportional to their number of modifiers. It goes from 35% to 17%, and then to 0% as the number of subject modifiers increases. I attribute this tendency to the increase of informational prominence triggered by the increase of subject modifiers. It should be noted that the number of monosyllabic subjects with modifiers is too small to examine statistical significance.

As noted earlier in the present section, the length of subject NPs was also measured by how large they were, and it was examined with respect to subject encoding types in

the present study. As a means of measuring how large the subject NPs are, I relied on counting the number of subject modifiers, regardless of subject monosyllabicity. That is to say, in the present study, longer subjects refer to subjects with a greater number of modifiers, and shorter subjects refer to those with no or fewer modifiers. Following are examples of subject NPs with and without subject modifiers, in which modified words are underlined, and modifying words are bracketed.

- (4.32) *elkwul-Ø kumpang ppalkaycicanha.*  
 face-( ) quickly get.blushed  
 ‘The face gets blushed quickly.’
- (4.33) [*wuli*] *emma-Ø maynnal ssai-Ø po-nuntey.*  
 1PL mom-( ) every.day cyworld-( ) see-SEM  
 ‘My mom sees CYWORLD every day.’
- (4.34) [*mwun-Ø ye-nun*] *tey-ka epsunikka, mokyoil-un*  
 door-( ) open-REL place-NOM not.exist.because Thursday-TOP  
 ‘Because there is no place that is open on Thursday.’
- (4.35) [*ike-ey tayhan cengpo-lul a-nun*] *ay-ka amwuto*  
 this-at about information-ACC know-REL folks-NOM none  
*eps-nya.*  
 not.exist-SEM  
 ‘There is really no one that has information about this.’
- (4.36) [*kulen sasohan cengpo-tul-ul com allyecwu-nun*] *salam-i*  
 such mere information-PL-ACC just inform-REL person-NOM  
*amwuto eps-nya, ettehkey.*  
 none not.exist how.come  
 ‘How come, there is really no one that informs (me) of such mere information.’

(4.32) is an example of a subject with no modifier because the subject, *elkwul* ‘face’, is not modified. In (4.33), there is one word in brackets, *wuli* ‘my’, that modifies the subject, *emma* ‘mom’, so it is counted as an example of a subject with one modifier. (4.34) is an instance of a subject with two modifiers because the subject, *tey* ‘place’,

is modified by two words in brackets. (4.35) exemplifies a subject with four modifiers, in that the subject, *ay* ‘folks’, is modified by four words in brackets. Lastly, a subject with five modifiers is exemplified in (4.36), and the underlined subject, *salam* ‘person’, is modified by five words that are in brackets.

The present study also examined if and how the length of subject NPs, in terms of the number of subject modifiers, influences the selection of subject markings in conversational Korean. For this reason, all subject NPs were grouped in terms of their number of modifiers, and the results are summarized in Table 4.13.

Table 4.13: Subject Modifiers and Subject Marking

SUBJ.MARK		NP-Ø	NP-ka	NP-nun	NP-to	NP-etc.	SUB-TOTAL	TOTAL
SUBJ.MODIFIER								
zero	NO.	965	1,086	492	257	38	1,873	2,838
	%	<b>34</b>	<b>38</b>	17	9	1	66	100
one	NO.	196	292	115	97	15	519	715
	%	<b>27</b>	<b>41</b>	16	14	2	73	100
two	NO.	36	98	15	29	2	144	180
	%	20	54	8	16	1	80	100
three	NO.	10	33	9	14	0	56	66
	%	<b>15</b>	<b>50</b>	14	21	0	85	100
four or more	NO.	7	24	3	6	0	33	40
	%	<b>18</b>	<b>60</b>	8	15	0	83	101
nominalized	NO.	10	28	7	4	1	40	50
	%	20	56	14	8	2	80	100
that-comp.	NO.	3	3	4	1	0	8	11
	%	27	27	36	9	0	73	100
TOTAL	NO.	1,227	1,564	645	408	56	2,673	3,900
	%	31	40	17	10	1	69	100

In Table 4.13, I find that the occurrence of *ka*-marking is higher for encoding longer subjects than for shorter subjects, though such a tendency is not always consistent (one modifier vs. four or more modifiers:  $\chi^2=7.346$ ,  $P < 0.01$ ).<sup>15</sup> For instance, the

<sup>15</sup>In the present study, longer subjects refer to subjects with more modifiers, and shorter subjects refer to those with no or fewer modifiers.

percentage of *ka*-marked subject NPs is only 38% when there is no subject modifier, and it is 41% when the number of subject modifiers is only one. Yet the percentage of *ka*-marked subjects increases greatly to 50% when there are three subject modifiers, and to 60% when there are four or more subject modifiers. I associate this tendency with informational prominence, based on the assumption that longer entities are informationally more loaded, and are thus informationally more prominent than shorter entities, resulting in the tendency for longer entities to be overtly identified. I also relate this tendency to processing load, in that longer entities are informationally more loaded, hence processing load is assumed to be higher than shorter entities, resulting in the tendency for longer entities to be identified with an explicit particle. For instance, in terms of their number of modifiers, the subjects, *elkwul* ‘face’ in (4.32) and *emma* ‘mom’ in (4.33), are viewed as short(er) NPs, thus would represent a low(er) degree of informational prominence, as well as low processing load, and they are encoded with a zero particle. On the other hand, the subjects, *ay* ‘folks’ in (4.35) and *salam* ‘person’ in (4.36), are regarded as long(er) NPs because of their modifiers, thus a high(er) degree of informational prominence, as well as high processing load, so they are commonly coded with the nominative *ka*.

Based on the observed tendency, I argue that overt marking may be used to denote prominent information, so the occurrence of *ka*-marking is expected to be high for encoding longer subjects than for coding shorter subjects in conversational Korean. In fact, Ono *et al.* (2000) also claims that in spoken Japanese, larger NPs are more likely to be marked with an overt particle than smaller NPs. Mori & Givón (1987) also make a similar claim in this regard, though their way of measuring the length of NPs differs from mine.

To summarize the present section, it was observed that in conversational Korean, the use of *ka*-marking (34%) is not common for coding monosyllabic subjects, though

it was reported to be so in colloquial Japanese (Tsutsui 1984; Ono *et al.* 2000). I attributed this difference to the frequency difference of monosyllabic NPs between these two languages. That is to say, monosyllabic subjects are not common in Japanese, so they are commonly coded with the overt marking with *ga* due to processing load. On the other hand, monosyllabic subjects are common in Korean, so the use of *ka*-marking is not common for coding them. However, when types of monosyllabic subject NPs were also considered as an influencing factor, a similarity was observed between these two languages. For example, the occurrence of *ka*-marking is relatively high (47%) for encoding monosyllabic full-noun subjects in conversational Korean (see Table 4.11), and that of *ga*-marking is high for coding monosyllabic subjects, or one-mora NPs in Mori & Givón's (1987) term, in Japanese. In the present study, encoding types of monosyllabic subjects were also examined with regard to their number of modifiers. As shown in Table 4.12, the percentage of *ka*-marked monosyllabic subjects gradually increases as the number of subject modifiers increases (34% → 50% → 100%). I associated this tendency with an increase of informational prominence incurred by the increase of subject modifiers. Lastly, regardless of whether or not subject NPs are monosyllabic, the occurrence of *ka*-marked subjects proportionally increases as the number of subject modifiers, namely the length of subjects, increases. Again, I attribute this tendency to the assumption that informational prominence, as well as processing load, encoded in larger NPs is higher than that of shorter NPs.

Based on the results discussed thus far, I claim that in conversational Korean, the selection of subject markings is influenced by the length of subject NPs. Simply put, the occurrence of *ka*-marking is relatively high for coding subjects that represent processing load or informational prominence. Korean shares some similarities with Japanese, in that the use of an overt marking (*ka* for Korean; *ga* for Japanese) is common for coding monosyllabic full-noun subjects as well as large subjects in both languages.

### 4.5.6 Repair

As was stated in Section 3.3.1, when a speaker changes what s/he had just said, we say there is an instance of ‘repair’. In general, more attention is paid to reproduced entities than to interrupted entities because the reproduced entities are assumed to represent important information, thus receive attention from the hearer, whereas interrupted entities are assumed not to represent important information, thus can be ignored.<sup>16</sup> Therefore, the entities that are reproduced are identified more explicitly than those that are interrupted. In the present study, when a repair was made during a given utterance, a reproduced entity was only considered for the examination of its marking, and an entity that was interrupted for a repair was ignored. The following is an example which includes subject repair.

- (4.37) *pongkocha thayksi, [pongko]-ka te pissa.*  
van taxi van-NOM more expensive  
‘A van-taxi, a van is more expensive.’

In (3.11), repeated here as (4.37), there is an instance of ‘repair’ because the first element, *pongkocha thayksi* ‘van taxi’, is interrupted, and it is repaired with the second element, *pongko* ‘van’, that is marked with the nominative *ka*. As was repeatedly stated, in the present study, only the second element, namely the reproduced element, was examined with respect to subject markings, and the first element, namely the interrupted element, was ignored. In this dissertation, the selection of subject markings was examined based on whether or not there was a repair made for the subject in a given utterance, but not based on whether it was a reproduced subject or an interrupted subject. For the sake of labeling convenience, subjects appearing in a clause which involves a subject repair are called ‘repaired subjects’, and subjects occurring in a

---

<sup>16</sup>An interrupted entity refers to the entity that is interrupted for a repair (the first element in (4.37)), and a reproduced entity refers to the entity that is reproduced by a repair (the second element in (4.37).)

clause that does not contain a subject repair are called ‘non-repaired subjects’ in this dissertation. The data in Table 4.14 presents the selection of subject markings with respect to whether they are repaired subjects or non-repaired subjects in conversational Korean.

Table 4.14: Subject Repair and Subject Marking

SUBJ.MARK		NP-Ø	NP-ka	NP-nun	NP-to	NP-etc.	SUB-TOTAL	TOTAL
SUBJ.REPAIR								
¬repair	NO.	1,149	1,423	586	371	50	2,430	3,579
	%	<b>32</b>	<b>40</b>	16	10	1	<b>68</b>	100
repair	NO.	78	141	59	37	6	243	321
	%	<b>24</b>	<b>44</b>	18	12	2	<b>76</b>	100
TOTAL	NO.	1,227	1,564	645	408	56	2,673	3,900
	%	31	40	17	10	1	69	100

In Table 4.14, I find that the use of the overt marking, either with *ka* or *nun*, is more common for repaired subjects (62%) than for non-repaired subjects (56%), and conversely, the use of zero-marking is more common for non-repaired subjects (32%) than for repaired subjects (24%) (zero-marking vs. an overt marking:  $\chi^2=7.716$ ,  $P < 0.01$ ). I attribute this tendency to the informational prominence, which is encoded in repaired subjects. That is to say, repaired subjects tend to require attention from the hearer, thus are assumed to be informationally more prominent than non-repaired subjects, so they are likely to be identified with an explicit particle.

I also find that the use of *ka*-marking (44%) is more common for encoding repaired subjects than zero-marking (24%). Furthermore, the use of zero-marking is less common for repaired subjects (24%) than for non-repaired subjects (32%). Based on this tendency, it seems to be clear that due to informational prominence, repaired subjects are more commonly coded with *ka*-marking than with zero-marking, and they are less commonly coded with zero-marking than non-repaired subjects (zero-marking vs. *ka*-



marking:  $\chi^2=6.720$ ,  $P < 0.01$ ). Although the occurrence of *ka*-marking is only slightly higher for repaired subjects than for non-repaired subjects (44% vs. 40%),<sup>17</sup> it is still tenable to claim, based on the result of statistical analysis, that the use of *ka*-marking is more common for encoding repaired subjects than it is for coding non-repaired subjects, and the use of zero marking is less common for repaired subjects than for non-repaired subjects, due to informational prominence encoded in repaired subjects, as exemplified in (4.37).

In the present study, I also examined whether the frequency of repaired subjects may differ based on transitivity of clauses. For this reason, I categorized all tokens of repaired subjects into two subtypes based on the transitivity of clauses: repaired subjects in intransitive clauses and repaired subjects in transitive clauses. The token distribution of repaired subjects with relation to their markings is shown in Table 4.15.

Table 4.15: Subject Repair with Transitivity and Subject Marking

SUBJ.REPAIR		SUBJ.MARK	NP-Ø	NP-ka	NP-nun	NP-to	NP-etc.	SUB-TOTAL	TOTAL
intran.	-repair	NO.	859	1,068	381	284	27	1,760	2,619
		%	33	41	15	11	1	67	100
	repair	NO.	62	116	29	30	2	177	239
		%	<b>26</b>	<b>49</b>	<b>12</b>	13	1	74	100
tran.	-repair	NO.	290	355	205	87	23	670	960
		%	30	37	21	9	2	70	100
	repair	NO.	16	25	30	7	4	66	82
		%	<b>20</b>	<b>30</b>	<b>37</b>	9	5	8	100
TOTAL		NO.	1,227	1,564	645	408	56	2,673	3,900
		%	31	40	17	10	1	69	100

<sup>17</sup>I speculate that this small difference may be due to other factors (e.g., length, animacy, etc.) facilitating the use of *ka*-marking for non-repaired subjects, i.e., subjects not involving a repair.

In Table 4.15, I find that the occurrence of repaired subjects in intransitive clauses (239 tokens, approximately 8% of all subjects in intransitive clauses) is equal to that of repaired subjects in transitive clauses (82%, approximately 8% of all subjects in transitive clauses). This distributional pattern is somewhat unexpected because a ‘repair’ was assumed to occur more frequently with subjects in intransitive clauses than with subjects in transitive clauses. That is to say, it is the general tendency that repairs, as lexical substitutions, most commonly occur with new information because that is where the most planning is required (Jeri Jaeger, personal communication), and new information tends to be represented more commonly by subjects of intransitive clauses than with subjects of transitive clauses. In fact, Muansuwan (2000) observes a similar tendency in her study of lexical SOTs in Thai, and claims that lexical substitutions, i.e., repair, commonly occur with objects where new information is usually encoded. Nevertheless, this general tendency is not observed in conversational Korean, at least according to the data in the present study.

Furthermore, the use of an overt marking, either with *ka* or *nun*, is more common for repaired subjects in transitive clauses (67%) than for repaired subjects in intransitive clauses (61%), while the use of zero-marking is more common for repaired subjects in intransitive clauses (26%) than for repaired subjects in transitive clauses (20%), yet not in a statistically significant way (zero-marking vs. overt marking:  $\chi^2=1.441$ ,  $P=0.23$ ). Also, the use of *ka*-marking is more common for repaired subjects in intransitive clauses (49%) than for repaired subjects in transitive clauses (30%), yet not in a statistically significant way (zero-marking vs. *ka*-marking:  $\chi^2=0.255$ ,  $P=0.6132$ ). These results suggest that in conversational Korean, the encoding types of repaired subjects do not depend upon the transitivity of clauses.

In Table 4.15, I also find that repaired subjects in transitive clauses (37%) are more commonly coded with *nun*-marking than repaired subjects in intransitive clauses (12%),

and conversely, repaired subjects in transitive clauses (30%) are less commonly marked with *ka* than repaired subjects in intransitive clauses (49%) (*ka*-marking vs. *nun*-marking:  $\chi^2=22.881$ ,  $P < 0.001$ ). I relate this tendency with the view that in Korean, *nun*-marking is usually associated with old information and *ka*-marking with new information (Lee 2003). In other words, this skewed tendency supports such a claim that new information is in the subjects of intransitive clauses and old information is in the subjects of transitive clauses.

More importantly, the data in Table 4.15 shows that the use of zero-marking is not common for repaired subjects, but instead the overt marking with *ka* or *nun* is commonly used for coding repaired subjects in both intransitive (zero-marking vs. overt marking in intransitive:  $\chi^2=4.322$ ,  $P < 0.05$ ) and transitive clauses (zero-marking vs. overt marking in transitive:  $\chi^2=3.962$ ,  $P < 0.05$ ). In other words, the use of zero-marking is not common for repaired subjects but the use of an overt marking is common for coding repaired subjects regardless of the transitivity of clauses (zero-marking vs. overt marking:  $\chi^2=6.720$ ,  $P < 0.01$ ). As repeatedly stated, I attribute this tendency to informational prominence encoded in repaired subjects. That is to say, repaired subjects are assumed to be in the focus of attention, hence would be likely to represent informational prominence, and they are commonly identified with an overt marking rather than zero-marking, as exemplified in (4.37).

To summarize the present section, the occurrence of zero-marking is low for coding repaired subjects, compared to that of non-repaired subjects. Contrastively, the occurrence of the overt marking, either with *ka* or *nun*, is higher for coding repaired subjects than for non-repaired subjects (see Table 4.14). I claim that this tendency is due to informational prominence encoded in repaired subjects, which are assumed to be in the focus of attention. Despite the general tendency that repairs occur commonly with new information, in conversational Korean, at least according to my data in Ko-

rean conversation, the percentage of repaired subjects in intransitive clauses is (almost) the same as the percentage of repaired subjects in transitive clauses (see Table 4.15). Furthermore, the occurrence of *nun*-marked subjects is higher in transitive clauses than in intransitive clauses, and conversely, the occurrence of *ka*-marked subjects is higher in intransitive clauses than in transitive clauses (see Table 4.15). I take this tendency as a supporting evidence for the view that *nun*-marking encodes old information and *ka*-marking encodes new information. Most importantly, regardless of the transitivity of clauses, the overt marking, either with *ka* or *nun*, is common for repaired subjects, but the use of zero-marking is not common for repaired subjects, and this tendency is attributed to the assumption that repaired subjects tend to be in the focus of attention, hence represent informational prominence.

Based on these results, I claim that one discourse function of grammatical coding *ka* is certainly to mark informational prominence encoded in subject NPs in conversational Korean.

#### 4.5.7 Anaphoric saliency

As was previously noted in Chapter 2, RD is introduced as one of the quantitative measurements which is extensively discussed in Givón (1983; 1993) and has frequently been utilized in subsequent text analyses in many languages (e.g., Kang 1998; Shimojo 2005). As repeatedly stated, RD is used to indicate the linguistic distance in clausal units, and it is measured by counting clausal units backward to the most recent representation of the co-referential expression which usually includes those of zero-anaphor (Givón 1983). In short, RD is the distance to the most recent co-referential expression. Yet RD is simply a heuristic way of measuring the level, and RD by itself may not be of great cognitive significance (Givón 1993; Shimojo 2005). There can be other factors as well which also influence the level of activation, as was previously noted in Chapter

2. Nevertheless, as Shimojo (2005) points out, in spite of all other potential factors, RD seems to be the only quantifiable anaphoric measurement currently available.

In the meantime, Dryer (1996) points out that the activation status of a given referent can also be measured by examining how frequently the given referent is mentioned within an anaphoric context. For example, the more frequently a given referent is mentioned within an anaphoric context, the more activated it is in one's consciousness, provided that all other things are considered same.

For this reason, in the present study, the anaphoric saliency, that is, the activation status of a given referent, is measured not only by RD (anaphoric recency) but also by AF (anaphoric frequency), as was addressed in Section 2.11.1.

It is generally assumed that processing requires an activation cost, and the cost depends on the activation status of information prior to processing (Dryer 1996). That is to say, information which is activated in one's consciousness may be processed at little or less activation cost, while semi-activated and deactivated information seems to require much or greater activation cost (Dryer 1996; Shimojo 2005). Based on this, I expect RD and AF measurements to be a good means of examining the speaker's choice of subject encoding types in conversational Korean, as was reported to be so in colloquial Japanese (Shimojo 2005). More specifically, with regard to the selection of subject markings, I associate a higher occurrence of zero-marking with a shorter RD and higher AF, and a higher occurrence of *ka*-marking with a greater RD and lower AF.

In spite of the possible correlation between the activation level of a given referent and the choice of subject markings, nearly no attempt, if not absolutely none, has been made to capture a discourse-pragmatic function of subject encoding types (e.g., the usage of *ka/nun*-marking) with respect to the activation level of subject referents in the minds of the speaker and the addressee in actual discourse, as Kang (1998) previously

pointed out. For this reason, the present study attempts to disclose discourse-pragmatic functions of subject encoding types by examining the choice of subject encoding types, including subjects of zero-anaphor, with relation to anaphoric saliency, that is, RD (anaphoric recency) and AF (anaphoric frequency). Table 4.16 and Table 4.18 show distributional patterns of subject encoding types with regard to RD and AF, respectively. In addition, Table 4.17 presents the tokens in terms of the two categories of RD (1-5 vs. 5-NPM) along with the mean RD values, and Table 4.19 presents the tokens in terms of the two categories of AF (0-3 vs. 4-10) along with the mean AF values. Let us first discuss the data in Table 4.16 and Table 4.17.

According to the result shown in Table 4.16 as well as in Table 4.17, RD measurement seems to clearly distinguish the speaker's use of subject encoding types. In other words, in conversational Korean, the choice of subject encoding types can be predicted by the RD measurement.

In Table 4.16, I find that the mean RD for zero-anaphor subjects (4.1 clauses) is much smaller than the one for the other available encoding types (e.g., 10.3 clauses for *ka*-marked subjects). This is a rather expected result, in that entities with a smaller RD are more activated, thus more identifiable, than entities with a greater RD, so they can be processed at little activation cost. Activated entities often appear as a zero-anaphor because they can easily be recovered. This tendency observed in conversational Korean is consistent with Shimojo's (2005) observation in colloquial Japanese. Furthermore, according to Table 4.17, where RD is generalized in terms of the two ranges of RD, in general, subjects of zero anaphor (84% of the total for RD 1-5) are in the anaphorically salient side of the index, whereas *ka*-marked subjects (49% of the total for RD 6-NPM) are in the anaphorically non-salient side of the index (zero anaphor vs. *ka*-marking:  $\chi^2=699.653$ ,  $P < 0.001$ ).

---

<sup>18</sup>NPM stands for no previous mention in the range of 20 preceding clauses, and it is counted as RD 21.

Table 4.16: Subject Anaphoric Recency (RD) and Subject Marking

SUBJ.ET SUBJ.RD		ZERO- ANAPHOR	NP-Ø	NP-ka	NP-mun	NP-to	NP-etc.	SUB- TOTAL	TOTAL
<i>1 cls.</i>	NO.	3,018	449	420	204	90	14	728	4,195
	%	72	11	10	5	2	0	17	100
<i>2 cls.</i>	NO.	787	156	177	94	32	8	311	1,254
	%	63	12	14	7	3	1	25	100
<i>3 cls.</i>	NO.	368	87	94	35	29	4	162	617
	%	60	14	15	6	5	1	26	100
<i>4 cls.</i>	NO.	185	41	70	29	13	1	113	339
	%	55	12	21	9	4	0	33	100
<i>5 cls.</i>	NO.	101	26	35	25	17	1	78	205
	%	49	13	17	12	8	0	38	100
<i>6 cls.</i>	NO.	81	19	25	15	8	1	49	149
	%	54	13	17	10	5	1	33	100
<i>7 cls.</i>	NO.	51	12	16	7	5	1	29	92
	%	55	13	17	8	5	1	32	100
<i>8 cls.</i>	NO.	43	14	20	11	5	0	36	93
	%	46	15	22	12	5	0	39	100
<i>9 cls.</i>	NO.	18	16	10	9	5	1	25	59
	%	31	27	17	15	8	2	42	100
<i>10 cls.</i>	NO.	25	7	10	3	2	0	15	47
	%	53	15	21	6	4	0	32	100
<i>11 cls.</i>	NO.	20	9	9	7	2	0	18	47
	%	43	19	19	15	4	0	38	100
<i>12 cls.</i>	NO.	20	5	19	3	5	0	27	52
	%	38	10	37	6	10	0	52	100
<i>13 cls.</i>	NO.	5	7	6	5	0	0	11	23
	%	22	30	26	22	0	0	48	100
<i>14 cls.</i>	NO.	9	4	0	1	1	0	2	15
	%	60	27	0	7	7	0	13	100
<i>15 cls.</i>	NO.	11	2	3	4	1	0	8	21
	%	52	10	14	19	5	0	38	100
<i>16 cls.</i>	NO.	6	2	1	2	1	0	4	12
	%	50	17	8	17	8	0	33	100
<i>17 cls.</i>	NO.	8	6	4	5	3	0	12	26
	%	31	23	15	19	12	0	46	100
<i>18 cls.</i>	NO.	3	7	6	4	2	0	12	22
	%	14	32	27	18	9	0	55	101
<i>19 cls.</i>	NO.	4	4	0	2	1	1	4	12
	%	33	33	0	17	8	8	33	99
<i>20 cls.</i>	NO.	9	3	6	5	1	0	12	24
	%	38	13	25	21	4	0	50	101
NPM <sup>18</sup>	NO.	577	351	633	175	185	24	1,017	1,945
	%	30	18	33	9	10	1	52	100
TOTAL	NO.	5,349	1,227	1,564	645	408	56	2,673	9,249
	%	58	13	17	7	4	1	29	100
MEAN RD		<b>4.1</b>	<b>8.1</b>	<b>10.3</b>	<b>8.3</b>	11.5	10.6	10.0	6.4

Table 4.17: Subject RD (1-5, 6-NPM) and Subject Marking

SUBJ.ET SUBJ.RD		ZERO- ANAPHOR	NP- $\emptyset$	NP-ka	NP-nun	NP-to	NP-etc.	SUB- TOTAL	TOTAL
<i>1-5</i>	NO.	4,459	759	796	387	181	28	1,392	6,610
	%	<b>84</b>	<b>62</b>	<b>51</b>	<b>60</b>	44	50	52	71
<i>6-NPM</i>	NO.	890	468	768	258	227	28	1,281	2,639
	%	<b>16</b>	<b>38</b>	<b>49</b>	<b>40</b>	56	50	48	29
TOTAL	NO.	5,349	1,227	1,564	645	408	56	2,673	9,249
	%	100	13	17	7	4	1	29	100
MEAN RD		<b>4.1</b>	<b>8.1</b>	<b>10.3</b>	<b>8.3</b>	11.5	10.6	10.0	6.4

More importantly, according to Table 4.16, the mean RD is shown to be highest for *ka*-marked subjects (10.3 clauses), and the mean RD for zero-marked subjects (8.1 clauses) is almost the same as the mean RD for *nun*-marked subjects (8.3 clauses). As a matter of fact, it was initially expected that the mean RD for *ka*-marked subjects would be greatest among other encoding types, at least higher than the one for zero-marked subjects. I based on the assumption that *ka*-marking, which I argue aids the process of referent identification, is the encoding type that fits best for entities with a greater RD. Furthermore, this result is basically consistent with what was observed in colloquial Japanese, though in Japanese, differences in the mean RD among the encoding types was much greater (Shimojo 2005). For instance, the mean RD for *ga*-marked subjects was much greater than the one for zero or *wa*-marked subjects. I attribute this to the different discourse-functional feature encoded in nominative markers in the two languages (i.e., *ka* for Korean; *ga* for Japanese). That is to say, in Korean, the nominative *ka* can encode a (continuing) topic (Choi 1997), whereas in Japanese, the nominative *ga* is not regularly used in the same situation and it cannot be a marker



of the continuation of the same topic. Also, the data in Table 4.17 indicates that *ka*-marked subjects (49% of the total for RD 6-NPM) tend to show up in the anaphorically non-salient side of the index more than zero-marked subjects (38% of the total for RD 6-NPM) do (zero-marking vs. *ka*-marking:  $\chi^2=33.493$ ,  $P < 0.001$ ).

In addition, the mean RD for *nun*-marked subjects (8.3 clauses) is smaller than the one for *ka*-marked subjects (10.3 clauses). This distributional tendency is actually supported by what was suggested by Kang (1998), where *nun*-marking was associated with activated subjects and non-activated predicates, and in contrast, *ka*-marking is with non-activated subjects and with activated predicates. Simply put, in conversational Korean, the RD for *nun*-marked subjects is expected to be smaller than the RD for *ka*-marked subjects, and the result from the present study is similar to what Kang (1998) observed. Kang's (1998) suggestion is also supported by the data in Table 4.17, in that overall, *nun*-marked subjects (60% of the total for RD 1-5) appear in the anaphorically salient side of the index more frequently than *ka*-marked subjects (51% of the total for RD 1-5) (*ka*-marking vs. *nun*-marking:  $\chi^2=15.22$ ,  $P < 0.001$ ).

Lee (1999) also claims that with respect to identifiability of subjects in Korean, *ka*-marking is associated more with indefiniteness, and *nun*-marking is more with definiteness, though he admits that there are cases which do not fit into this bisectonal distinction. Therefore, Lee's (ibid.) claim is, at least partially, supported by the tendency reported in the present section.

The identifiability of a referent also matters greatly in the prediction of subject markings in spoken Japanese. For example, Iwasaki (1987:110) addresses, in his study of spoken Japanese, “[t]he most relevant pragmatic concept which characterizes a *wa*-marked noun phrase is ‘identifiability’, and the referent of such noun phrases must be identifiable before they can take *wa*-marking.” Since *wa*-marked entities are more identifiable than *ga*-marked entities, a smaller RD is more likely for *wa*-marked subjects

than it is for *ga*-marked subjects.

Based on the results, I argue that in conversational Korean, the mean RD for zero anaphor, zero-marked subjects and *nun*-marked subjects is smaller than the mean RD for *ka*-marked subjects because they are generally more identifiable, than *ka*-marked subjects. The overall token distribution found in Table 4.16 and Table 4.17 is consistent enough to show that the index of anaphoric saliency, in terms of RD, is useful to describe the discourse-functional property of subject encoding types: zero anaphor, zero-marking, *ka*-marking, and *nun*-marking. More specifically, the overall choice for subjects in the anaphorically salient side of the index is zero anaphor (84% of the total for RD 1-5) and *ka*-marking for the anaphorically non-salient subjects (49% of the total for RD 6-NPM), and zero-marking (62% of the total for RD 1-5) and *nun*-marking (60% of the total for RD 1-5) are placed between them.

Let us further discuss each of the three encoding types with examples.

(4.38)

A1 *nai-tun yeca-l nai-tun yeca-to manna-ko siph-e?*  
 age-full woman-ACC age-ful woman-also meet-and want-Q  
 ‘Do you want to date a woman who is old enough?’

B1 *ah, na cakkwu yenge naolyekwu kulenta.*  
 ah 1SG repeatedly English be.uttered tend  
 ‘Ah, I feel English is going to be uttered repeatedly without my intention.’

B2 *michikeyssta.*  
 drive.me.crazy  
 ‘It drives me crazy.’

B3 *kulenikka moluntakwu, na-nun.*  
 consequently don’t.know 1SG-TOP  
 ‘As I said, I don’t know (what to do)’

B4 *na ettehkey hayyatoy-l ci molu-l kes kathay.*  
 1SG how do.should-fut comp don’t.know-fut thing seems  
 ‘I feel I would not know what I should do.’

- B5 *cincca manna-myen,*  
 actually meet-if  
 ‘If I actually meet,’
- B6 *cikcang tanikwu,*  
 job work.for  
 ‘(I) work for a company,’
- B7 *han pen-to an haypwa kackwu, kulen kyenghem-ul.*  
 1 NC-even NEG do have such experience-ACC  
 (I) haven’t experienced such a thing.’
- B8 *nay-ka manna-ten yeca-tul-un cenpwu ta haksayng-ikwu.*  
 1SG-NOM meet-comp woman-PL-TOP entirely all student-be  
 ‘Girls I have dated are all students.’
- B9 *cenpwu ta 10 tay-eyse 20 tay chopan.*  
 entirely all 10 NC-from 20 NC early  
 ‘All from teenagers to early twenties.’
- B10 *na-nun kyeysook nulkekanuntey.*  
 1SG-TOP continuously become.old  
 ‘I am continuously getting older.’
- B11 *cham wuskinkekathay.*  
 really funny.seem  
 ‘It seems really funny.’
- B12 *ecey althulo-lang cip-ey ka-myense,*  
 yesterday Althulo-with home-LOC go.while  
 ‘While (I) was going home with Althulo yesterday,’
- B13 *ku yayki hay-ss-nuntey.*  
 that story do-PST-SEM  
 ‘(We) talked about it’
- A2 *oh kulay-ss-kwuna.*  
 oh do.so-PST-SEM  
 ‘Oh, I see.’
- B14 *kuke-l molla-ss-tenkeya, na-nun.*  
 that-ACC unaware-PST-SEM 1SG-top  
 ‘I realized that I wasn’t aware of that.’

- B15 *han pen-to kkaytassci mos hay-ss-ten kulen ke.*  
 1 NC-even realize NEG do-PST-REL such thing  
 ‘Such thing that (I) have never realized’
- A3 *nay nyuyok sa-nun yeca-ay-ka mak sengswukha-ss-ci.*  
 1SG New.York live-REL woman-kid-NOM DP mature-PAST-SEM  
 ‘My girl friend who lived in New York city was very mature.’
- B16 *kyay han sal elita-ko kulayss-na?*  
 3SG 1 NC young-and said-Q  
 ‘Did you say that she was one year younger than (you are)?’
- A4 *ung.*  
 yeah  
 ‘Yeah’
- B17 *kunikka sumwulilkop-i-nikka mwo.*  
 because 27-be-because what  
 ‘because (she) is 27 years old,’
- A5 *cincca nay-ka keyipa kanta kulayssteni,*  
 actually 1SG-NOM gay.bar go said  
 ‘I said I was going to go to a gay bar, and then..’
- A6 *kulen tay way ka-nyakwu.*  
 such place why go-Q  
 ‘What do (you) go to such a place for?’
- A7 *pothong wuli-n hakkyo hwupay-tul-un kitaytoynta, kitaytoynata mak*  
 normally 1PL-TOP school junior-PL-TOP expect expect DP  
*kulenuntey.*  
 say  
 ‘Generally, our school juniors say it must be exciting to go to a gay bar.’
- B18 *waynyahamyen kyay ni yecachinkwu-nun nai-twu isskwu,*  
 that’s.because 3SG 1SG girl.friend-TOP age-also exist.and  
*hankwuksalam-icanha.*  
 Korean-be  
 ‘That’s because your girl friend is old enough and also because she is Korean.’

In the utterance A3 in (4.38), the underlined subject, *yecaay* ‘girl’, is counted as an NPM because its coreferential expression does not appear in the preceding conversation.

Since the referent has not been activated before the utterance, it is assumed that it represents a significant activation cost. Therefore, *ka*-marking best fits the marking of this subject, which is underlined.

Earlier in the present section, it was noted that in Korean, *ka*-marking is normally used to encode subjects with a greater RD, but it can also be used to encode a (continuing) topic, as Choi (1997) pointed out. The following exemplifies an instance in which *ka*-marking is used for coding a (continuing) topic.

(4.39)

- A1 *sihem-un?*  
 exam-TOP  
 ‘How about the exam?’
- B1 *na?*  
 1SG  
 ‘Me?’
- A2 *sihem-un ettehkey pwa-ss-e?*  
 exam-TOP how see-PST-Q  
 ‘How did you do on the exam?’
- B2 *sihem-i, swipki-nun, phyengso-pota swi-woss-nuntey.*  
 exam-NOM easiness-TOP usual-than easy-PST-SEM  
 ‘The exam was easier than the usual, but...’

In the utterance B2 in (4.39), the underlined subject, *sihem* ‘exam’ is counted as a subject with RD1 because its co-referential expression appears in the immediately preceding clause, i.e., the utterance A2. The subject *sihem* ‘exam’ is a topic that continues from the preceding utterance, and it is coded with *ka*-marking. That is to say, the subject in (4.39) is a continuing topic, which is encoded with the nominative *ka*.

Following are examples which show subjects coded with zero-marking and *nun*-marking. Let us first look at an example of zero-marked subject.

(4.40)

A1 *cincca po-l ke-Ø eps-tela.*  
really see-INF thing-( ) not.exist-SEM  
'There was really nothing to see.'

A2 *na-n colla hwuhoy-Ø hay-ss-e,*  
1SG-TOP really regret-( ) do-PST-SEM  
'I greatly regretted.'

A3 *ka-se.*  
go-and  
'When I went there'

A4 *kuntey kuke-nun, ku miniechye catongcha-Ø cincca pissata.*  
yet that-TOP that miniature car-( ) really expensive  
'Yet, that, that miniature car is really expensive.'

B1 *pissa-cyo, kuke-Ø.*  
expensive-SEM that-( )  
'Yes, it is expensive.'

B2 *isipman-won ilehkey hayyo.*  
200,000-NC about cost  
'It costs about \$200.'

A5 *miniechye-Ø* *caksal, caksal-i-ess-e.*  
miniature-( ) great great-be-PST-SEM  
'The miniature was so great.'

In the utterance A5 in (4.40), the underlined subject, *miniechye* 'miniature', is counted as a subject with RD1 because its co-referential expression appears in a clause which immediately precedes, i.e., the utterance B2. The underlined subject in the utterance of (4.40) was mentioned in the immediately preceding utterance, thus can be processed at little activation cost, so the use of zero-marking is not unexpected. Now let us also look at examples of *nun*-marked subject.

(4.41)

- A1 *Tonguni-to na,namcachinkwu mantulki cikcenitentey.*  
 Tongun-also boy.friend make before  
 ‘Tongun is just about to make her boy friend, too.’
- A2 *Tonguni-n ku Taiwan ay mwo com mwo cohta kuletela.*  
 Tongun-top that Taiwan guy dp little dp like said  
 ‘Tongun said she likes that Taiwanese.’

(4.42)

- A1 *meymolielhol-i issta-tentey.*  
 memorial.hall-nom exist-said  
 ‘It is said there is a memorial hall.’
- A2 *mac-e?*  
 right-sem  
 ‘Is it true?’
- B1 *meymolielhol-un kuncheey epstela.*  
 memorial.hall-top nearby not.exist  
 ‘There was no memorial hall.’
- B2 *kuke ta epsecin keskathay.*  
 that all removed seemed  
 ‘That seemed to be removed.’

In the utterance A2 in (4.41), the underlined subject, *Tonguni* ‘Tongun’, is also counted as a subject with RD1 because its co-referential expression appears in the immediately preceding clause, and it is coded with *nun*-marking. In the utterance B1 in (4.42), the subject, *meymolielhol* ‘memorial hall’, is counted as a subject of RD2 because its co-referential expression is in a clause which is two clauses away, and it is marked with *nun*.

As was stated earlier in the section, the degree of activation can also be measured by the anaphoric frequency of a given referent, as well as the anaphoric recency of the referent, though to the best of my knowledge, there have been no previous attempts made to capture a correlation between the choice of subject encoding types and the anaphoric saliency of subject referents. In order to see if the activation level of subject

referents correlates the speaker’s selection of subject encoding types, I also examined the anaphoric frequency of subject referents with respect to the choice of subject markings. The results are presented in Table 4.18 and Table 4.19. Table 4.19 presents the tokens in terms of the two groups of AF (0-3 vs. 4-10) along with the mean AF values.

Table 4.18: Subject Anaphoric Frequency (AF) and Subject Marking

SUBJ.ET SUBJ.AF		ZERO- ANAPHOR	NP-Ø	NP-ka	NP-nun	NP-to	NP-etc.	SUB- TOTAL	TOTAL
<i>0 times</i>	NO.	666	402	687	213	201	25	1,126	2,194
	%	30	18	31	10	9	1	51	99
<i>1 times</i>	NO.	932	306	309	156	97	18	580	1,818
	%	51	17	17	9	5	1	32	100
<i>2 times</i>	NO.	764	155	199	92	38	5	334	1,253
	%	61	12	16	7	3	0	27	100
<i>3 times</i>	NO.	713	124	138	58	28	3	227	1,064
	%	67	12	13	5	3	0	21	100
<i>4 times</i>	NO.	710	95	98	51	22	3	174	979
	%	73	10	10	5	2	0	18	101
<i>5 times</i>	NO.	638	72	74	34	17	0	125	835
	%	76	9	9	4	2	0	15	100
<i>6 times</i>	NO.	484	42	31	23	2	1	57	583
	%	83	7	5	4	0	0	10	100
<i>7 times</i>	NO.	258	20	19	14	3	0	36	314
	%	82	6	6	4	1	0	11	99
<i>8 times</i>	NO.	139	9	8	2	0	1	11	159
	%	87	6	5	1	0	1	7	100
<i>9 times</i>	NO.	35	2	1	2	0	0	3	40
	%	88	5	3	5	0	0	8	101
<i>10 times</i>	NO.	10	0	0	0	0	0	0	10
	%	100	0	0	0	0	0	0	100
TOTAL	NO.	5,349	1,227	1,564	645	408	56	2,673	9,249
	%	58	13	17	7	4	1	29	100
MEAN AF		<b>3.2</b>	<b>1.8</b>	<b>1.5</b>	<b>1.8</b>	1.1	1.1	1.5	2.5

In Table 4.18, I find that the mean AF for zero anaphors (3.2 times) is greater than the mean AF for the other encoding types (e.g., 1.5 times for *ka*-marked subjects). This



Table 4.19: Subject AF (0-3, 4-10) and Subject Marking

SUBJ.ET SUBJ.AF		ZERO- ANAPHOR	NP-Ø	NP-ka	NP-nun	NP-to	NP-etc.	SUB- TOTAL	TOTAL
<i>0-3</i>	NO.	3,075	987	1,333	519	364	51	2,267	6,329
	%	57	80	85	80	89	91	85	68
<i>4-10</i>	NO.	2,274	240	231	126	44	5	406	2,920
	%	43	20	15	20	11	9	15	32
TOTAL	NO.	5,349	1,227	1,564	645	408	56	2,673	9,249
	%	58	13	17	7	4	1	29	100
MEAN AF		<b>3.2</b>	<b>1.8</b>	<b>1.5</b>	<b>1.8</b>	1.1	1.1	1.5	2.5

distributional tendency is expected, based on the assumption that the more frequently entities are given in previous discourse contexts, the more activated the entities are at the time of a given utterance. More specifically, the referents of zero anaphors tend to be frequently given in preceding context. On the other hand, the referents of *ka*-marked subjects are not as frequently given, thus not as easily identifiable. According to the data in Table 4.18, the mean AF for zero anaphors (3.2 times) appears to be distinctively greater than the mean AF for the other encoding types (e.g., *ka*-marking), and the mean AF for *ka*-marked subjects (1.5 times) is slightly smaller than the mean AF for zero-marked subjects (1.8 times) or *nun*-marked subjects (1.8 times).

Furthermore, according to Table 4.19, where AF is generalized in terms of the two ranges of AF (i.e., AF 0-3 vs. AF 4-10), overall, subjects of zero anaphor (43% of the total for AF 4-10) are in the anaphorically salient side of the index, and *ka*-marked subjects (85% of the total for AF 0-3) are in the anaphorically non-salient side of the index (zero anaphor vs. *ka*-marking:  $\chi^2=403.113$ ,  $P < 0.001$ ). Also, zero-marked subjects (20% of the total for AF 4-10) and *nun*-marked subjects (20% of the total for

AF 4-10) is more in the anaphorically salient side of the index than *ka*-marked subjects (15% of the total for AF 4-10) (zero-marking vs. *ka*-marking:  $\chi^2=11.247$ ,  $P < 0.001$ ; *nun*-marking vs. *ka*-marking:  $\chi^2=7.653$ ,  $P < 0.01$ ).

This suggests that the anaphoric frequency (AF) of subject referents distinguishes between the choice of subject encoding types, but not as much as the anaphoric recency (RD) of subject referents does. More specifically, the overall token distribution found in Table 4.18 and Table 4.19 is consistent enough to show that the index of anaphoric saliency, in terms of AF, is useful to describe the discourse property of the four subject encoding types: zero anaphor, zero-marking, *ka*-marking, and *nun*-marking. Overall, the encoding type for subjects in the anaphorically salient side of the index is zero anaphor (43% of the total for AF 4-10) and *ka*-marking for the anaphorically non-salient subjects (85% of the total for AF 1-3), and zero-marking (20% of the total for AF 4-10) and *nun*-marking (20% of the total for AF 4-10) are placed between zero anaphor and *ka*-marking.

To summarize the section, the present study shows, with the results summarized in Table 4.16 and Table 4.17 as well as Table 4.18 and Table 4.19, that in conversational Korean, the anaphoric saliency, either by anaphoric recency (RD) or by anaphoric frequency (AF), of subject referents plays a role in the speaker's selection of subject encoding types, as was reported to be so in studies of conversational Japanese (e.g., Shimajo 2005). More specifically, the mean RD was shown to be smallest for zero anaphor and was to be greatest for *ka*-marked subjects. In general, entities with a smaller RD are assumed to be more activated, thus more identifiable, so they are commonly unexpressed and the use of *ka*-marking is not common for coding them. Furthermore, the mean AF was reported to be greatest for zero anaphor, while it was shown to be smallest for *ka*-marked subjects. Entities that are frequently given in anaphoric contexts are more accessible, hence more identifiable, than entities that appear less frequently in

anaphoric contexts, so it is common that they are not expressed overtly in a clause. On the other hand, entities that appear less frequently in anaphoric contexts are assumed to represent an activation cost, hence processing load, so they are commonly coded with *ka*-marking.

According to the data shown in Table 4.16 as well as Table 4.17, the mean RD for *nun*-marked subjects, as well as the mean RD for zero-marked subjects, is smaller than the mean RD for *ka*-marked subjects. In addition, according to the data shown in Table 4.18 as well as Table 4.19, the mean AF for *nun*-marked subjects, as well as the mean AF for zero-marked subjects, is greater than the mean AF for *ka*-marked subjects. Based on these results, I argued that *ka*-marking tends to be used for anaphorically non-salient subjects that represent a significant activation cost, hence processing load. That is to say, a greater RD or a smaller AF indicates a lower level of activation, and the low level of activation represents a greater processing load. The greater processing load invites the use of *ka*-marking, which I argue alleviates this processing load.

Furthermore, according to the results from the RD and AF measurements in the present study, Korean and Japanese share similarities with regard to the relationship between the anaphoric saliency and the selection of subject encoding types. In other words, the anaphoric saliency of subject referents, in both Korean and Japanese, plays a role in the choice of subject markings, in that the mean RD for *ka*-marked subjects is greater than the mean RD for zero-marked or *nun*-marked subjects in the two languages, though the difference between the mean RD for *ka*-marked subjects and the mean RD for zero-marked or *nun*-marked subjects is much greater in Japanese than it is in Korean. Simply put, the anaphoric saliency distinguishes between the choice of subject encoding types in both languages, but it seems to distinguish more clearly for the subject encoding types in Japanese than it does for the subject encoding types in Korean. I attributed this difference to the discourse-functional differences of subject

marking particles between the two languages, as Choi (1997) previously pointed out. The nominative *ka* in Korean may mark a continuing topic, namely a subject with an activated referent, whereas the nominative *ga* in Japanese does not function as a marker of continuing topic. In short, the usage of the nominative *ka* in Korean is broader than the nominative *ga* in Japanese. Since the primitive function of *ga* in Japanese is to mark unidentifiable referents, the mean RD is expected to be greater for *ga*-marked subjects in Japanese than it is for *ka*-marked subjects in Korean.

#### 4.5.8 Cataphoric saliency

As was previously noted in Chapter 2, RP (*referential persistence*) has been proposed as a means to measure the degree of decay of information in the cataphoric context (Givón 1983). In other words, RP is viewed as a way of measuring importance in the sense that “[m]ore important discourse topics appear more frequently in the register, i.e., they have a higher probability of persisting longer in the register after a relevant measuring point” (Givón 1983:15). Simply put, information which does not persist loses its importance in the subsequent context, while information which persists maintains its importance in the subsequent context.

Although there have been no previous attempts at studying if and how the cataphoric saliency influences the choice of subject encoding types, I assume that in conversational Korean, the cataphoric saliency plays a role in predicting the speaker’s selection of subject markings, as it was reported to be so in colloquial Japanese (Shimojo 2005). More specifically, subjects that are cataphorically salient are assumed to be thematically more prominent, so the use of an overt marking is expected to code them.

In the present study, the cataphoric saliency of subject referents is examined by their RP, which was measured in two different ways: RP in a sense of ‘uninterruption’

and RP in a sense of ‘frequency’. As was noted in Chapter 2, the former way of RP is labeled as RP and the latter type of RP is labeled differently as RP-*f*, for the sake of distinguishing one from the other. Based on what has been discussed, I assume that in conversational Korean, subjects with a greater RP and RP-*f* are expected to be overtly encoded, whereas subjects with a smaller RP and RP-*f* are to be covertly encoded. In order to see if the measurements of RP and RP-*f* distinguish among subject encoding types, I examined the choice of subject encoding types with respect to both RP and RP-*f*. The results are summarized in Table 4.20 and Table 4.21 as well as in Table 4.22 and Table 4.23. Table 4.21 and Table 4.23 also present the tokens in terms of the two categories along with the mean RP and mean RP-*f* values respectively. Let us first discuss the data in Table 4.20 and Table 4.21.

In Table 4.20, I find that the mean RP is generally small for all encoding types. The small mean RP across all encoding types indicates that in conversational Korean, subject referents generally do not persist long, at least without an interruption, regardless of their encoding types, though subjects of zero anaphors (1.1 clauses) persist slightly longer than subjects of the other encoding types (e.g., 0.7 clauses for *ka*-marked subjects). The data in Table 4.21, where RP is generalized in terms of the two ranges of RP, also indicates that overall, subject referents are in the cataphorically non-salient side of the index regardless of their encoding types, although subjects of zero anaphor are slightly less commonly in the cataphorically non-salient side of the index than subjects of the other encoding types (e.g., zero anaphor: 92% of the total for RP 0-3; *ka*-marking: 95% of the total for RP 0-3;  $\chi^2=26.051$ ,  $P < 0.001$ ). This further suggests that in conversational Korean, RP does not seem to distinguish between subject encoding types. This distributional tendency not only differs from what was initially expected but it also differs from what was observed in previous studies of Japanese conversation (e.g., Shimojo 2005), where RP for *ga*-marked subjects is much greater

Table 4.20: Subject Cataphoric Saliency-uninterruption (RP) and Subject Marking

SUBJ.ET SUBJ.RP		ZERO- ANAPHOR	NP-Ø	NP-ka	NP-nun	NP-to	NP-etc.	SUB- TOTAL	TOTAL
<i>0 cl.</i>	NO.	2,752	683	966	336	248	32	1,582	5,017
	%	55	14	19	7	5	1	32	101
<i>1 cl.</i>	NO.	1,248	268	315	150	82	12	559	2,075
	%	60	13	15	7	4	1	27	100
<i>2 cls.</i>	NO.	568	129	158	68	45	8	279	976
	%	58	13	16	7	5	1	29	100
<i>3 cls.</i>	NO.	331	78	54	48	19	3	124	533
	%	62	15	10	9	4	1	23	100
<i>4 cls.</i>	NO.	191	38	36	27	5	0	68	297
	%	64	13	12	9	2	0	23	100
<i>5 cls.</i>	NO.	107	14	20	6	5	0	31	152
	%	70	9	13	4	3	0	20	99
<i>6 cls.</i>	NO.	67	11	4	3	0	1	8	86
	%	78	13	5	3	0	1	9	100
<i>7 cls.</i>	NO.	38	2	5	3	3	0	11	51
	%	75	4	10	6	6	0	22	101
<i>8 cls.</i>	NO.	24	3	1	1	0	0	2	29
	%	83	10	3	3	0	0	7	100
<i>9 cls.</i>	NO.	15	0	4	1	1	0	6	21
	%	71	0	19	5	5	0	29	100
<i>10 cls.</i>	NO.	8	1	1	2	0	0	3	12
	%	67	8	8	17	0	0	25	100
TOTAL	NO.	5,349	1,227	1,564	645	408	56	2,673	9,249
	%	58	13	17	7	4	1	29	100
MEAN RP		<b>1.1</b>	<b>0.9</b>	<b>0.7</b>	<b>1.0</b>	0.7	0.8	0.8	1.0

than RP for zero-marked subjects, that is, *ga*-marked subjects persist much more than zero-marked subjects.

Despite this unexpected tendency, one noteworthy observation is that *ka*-marked subjects (0.7 clauses) persist less than the other encoding types such as *nun*-marked (1.0 clause) or zero-marked subjects (0.9 clauses). Also, according to Table 4.21, *ka*-

Table 4.21: Subject RP (0-3, 4-10) and Subject Marking

SUBJ.ET SUBJ.RP		ZERO- ANAPHOR	NP-Ø	NP-ka	NP-nun	NP-to	NP-etc.	SUB- TOTAL	TOTAL
<i>0-3</i>	NO.	4,899	1,158	1,493	602	394	55	2,544	8,601
	%	<b>92</b>	<b>94</b>	<b>95</b>	<b>93</b>	97	98	95	93
<i>4-10</i>	NO.	450	69	71	43	14	1	129	648
	%	<b>8</b>	<b>6</b>	<b>5</b>	<b>7</b>	3	2	5	7
TOTAL	NO.	5,349	1,227	1,564	645	408	56	2,673	9,249
	%	100	100	100	100	100	100	100	100
MEAN RP		<b>1.1</b>	<b>0.9</b>	<b>0.7</b>	<b>1.0</b>	0.7	0.8	0.8	1.0

marked subjects are slightly more in the cataphorically non-salient side of the index than *nun*-marked subjects (*ka*-marking vs. *nun*-marking:  $\chi^2=4.221$ ,  $P < 0.05$ ). From this observation, I propose that both *ka*-marking and *nun*-marking can mark a continuation of a topic, but the former is simply for local persistence (i.e., a local topic) and the latter is more for global persistence (i.e., a global topic). Furthermore, based on the assumption that important information persists longer than unimportant information, relatively speaking, *nun*-marking encodes cataphorically salient information, whereas *ka*-marking encodes cataphorically non-salient information. Simply put, unlike the nominative *ga* in Japanese, the nominative *ka* in Korean does not seem to mark thematic prominence encoded in referents of subject NPs.

I further attempt to associate this observation, i.e., *ka*-marking for a local topic and *nun*-marking for a global topic, with the notion ‘contrastiveness’. More specifically, I argue that the primitive function of the post-nominal particle *nun* is to mark the contrastiveness, but it can also contribute to the cohesion between utterances, hence naturally a continuation of a topic. In other words, the notion ‘contrastiveness’ should

be extended to the discourse continuity, in that contrastive entities marked by the post-nominal particle *nun* is generally cohesive to another entity, whereas *ka*-marked entities represent disjoint utterances.

To the best of my knowledge, no previous studies have pointed out that *ka*-marking is used for a cataphorically non-salient local topic, and *nun*-marking is used more for coding a cataphorically salient global topic.

As repeatedly mentioned, in the present study, RP-*f* was also used to measure the cataphoric saliency of subject referents. In order to see if RP-*f* distinguishes between subject encoding types more clearly, I also examined the choice of subject encoding type with regard to RP-*f*, and the data in Table 4.22 as well as Table 4.23 show the results.

According to the data in Table 4.22, just as with RP with regard to subject encoding types, there is not much difference in RP-*f* among subject encoding types, which makes RP-*f* not a good means to distinguish between subject encoding types. More importantly, the mean RP-*f* is shown to be smaller for *ka*-marked subjects (2.0 times) than it is to be zero-marked subjects (2.3 times). The data in Table 4.23, where RP-*f* is generalized in terms of the two ranges of RP-*f*, also show that referents of *ka*-marked subjects (79% of the total for RP-*f* 0-3) are more in the cataphorically non-salient side of the index than referents of zero-marked subjects (72% of the total for RP-*f* 0-3;  $\chi^2=15.21$ ,  $P < 0.001$ ) or *nun*-marked subjects (66% of the total for RP-*f* 0-3;  $\chi^2=36.405$ ,  $P < 0.001$ ). In fact, the mean RP-*f* is even smallest for *ka*-marked subjects, among all encoding types. This distributional tendency, with relation to RP-*f*, differs greatly from what was initially expected. That is to say, the mean RP-*f* for *ka*-marked subjects was expected to be greater than the mean RP-*f* for zero-marked subjects because of the high(er) thematic prominence that *ka*-marking encodes. As a matter of fact, in conversational Japanese, the measurement of mean RP-*f* clearly distinguishes between



Table 4.22: Subject Cataphoric Saliency-frequency (RP-*f*) and Subject Marking

SUBJ.ET SUBJ.RP- <i>f</i>		ZERO- ANAPHOR	NP-Ø	NP-ka	NP-nun	NP-to	NP-etc.	SUB- TOTAL	TOTAL
<i>0 times</i>	NO.	1,097	342	519	163	147	21	850	2,289
	%	48	15	23	7	6	1	37	100
<i>1 times</i>	NO.	818	227	295	102	92	10	499	1,544
	%	53	15	19	7	6	1	32	100
<i>2 times</i>	NO.	744	179	230	94	55	12	391	1,314
	%	57	14	18	7	4	1	30	101
<i>3 times</i>	NO.	665	138	185	69	38	5	297	1,100
	%	60	13	17	6	3	0	27	100
<i>4 times</i>	NO.	616	108	121	87	24	2	234	958
	%	64	11	13	9	3	0	24	99
<i>5 times</i>	NO.	549	91	98	47	23	3	171	811
	%	68	11	12	6	3	0	21	100
<i>6 times</i>	NO.	418	86	54	43	13	1	111	615
	%	68	14	9	7	2	0	18	100
<i>7 times</i>	NO.	256	33	36	23	9	2	70	359
	%	71	9	10	6	3	1	19	100
<i>8 times</i>	NO.	125	20	17	9	5	0	31	176
	%	71	11	10	5	3	0	18	100
<i>9 times</i>	NO.	53	2	8	6	2	0	16	71
	%	75	3	11	8	3	0	23	101
<i>10 times</i>	NO.	8	1	1	2	0	0	3	12
	%	67	8	8	17	0	0	25	100
TOTAL	NO.	5,349	1,227	1,564	645	408	56	2,673	9,249
	%	58	13	17	7	4	1	29	100
MEAN RP- <i>f</i>		<b>2.9</b>	<b>2.3</b>	<b>2.0</b>	<b>2.6</b>	1.8	1.6	2.1	2.6

subject encoding types (Shimojo 2005). However, Shimojo's (2005) findings, as to a correlation between RP-*f* and the choice of subject encoding types, are not supported by the data shown in Table 4.22 as well as in Table 4.23.

Despite this surprising tendency, one remarkable observation from Table 4.23, as well as Table 4.22, is that *ka*-marked subjects (2.0 clauses) persist less frequently than the

Table 4.23: Subject RP-*f* (0-3, 4-10) and Subject Marking

SUBJ.ET		ZERO-	NP-Ø	NP-ka	NP-nun	NP-to	NP-etc.	SUB-	TOTAL
SUBJ.RP- <i>f</i>		ANAPHOR						TOTAL	
<i>0-3</i>	NO.	3,324	886	1,229	428	332	48	2,037	6,247
	%	<b>62</b>	<b>72</b>	<b>79</b>	<b>66</b>	81	86	76	68
<i>4-10</i>	NO.	2,025	341	335	217	76	8	636	3,002
	%	<b>38</b>	<b>28</b>	<b>21</b>	<b>34</b>	19	14	24	32
TOTAL	NO.	5,349	1,227	1,564	645	408	56	2,673	9,249
	%	100	100	100	100	100	100	100	100
MEAN RP- <i>f</i>		<b>2.9</b>	<b>2.3</b>	<b>2.0</b>	<b>2.6</b>	1.8	1.6	2.1	2.6

other encoding types such as *nun*-marked subjects (2.6 clauses) (*ka* vs. *nun*:  $\chi^2=36.405$ ,  $P < 0.001$ ). This observation complies with the previous observation of *ka*-marking and *nun*-marking with regard to RP, and in fact, it supports the proposition that both *ka* and *nun* can mark a continuation of a topic, but the former is more for marking local persistence (i.e., a local topic) and the latter is leaning more toward global persistence (i.e., a global topic). Also, since important information tends to persist longer than unimportant information, relatively speaking, *nun*-marking encodes cataphorically salient information, while *ka*-marking encodes cataphorically non-salient information. In short, unlike the nominative *ga* in Japanese, the nominative *ka* in Korean does not seem to mark thematic prominence encoded in referents of subject NPs.

From this observation, I further argue that the primitive discourse function of the post-nominal particle *nun* is to mark the contrastiveness, but it can also contribute to the cohesion between utterances, hence naturally a continuation of a topic. That is to say, the notion ‘contrastiveness’ should be extended to the discourse continuity, since contrastive entities marked by *nun* is usually cohesive to another entity, while

*ka*-marked entities represent disjoint utterances.

In short, unlike Japanese, although cataphoric saliency, in terms of RP-*f*, does not matter much in the speaker's selection of subject encoding types in conversational Korean, it is distinct enough to point out the functional differences between *ka*-marked subjects and *nun*-marked subjects, i.e., *ka* for cataphorically non-salient local persistence (i.e., a local topic); *nun* for cataphorically salient global persistence (i.e., a global topic). This unanticipated tendency makes Korean differ from Japanese, because the cataphoric saliency as defined by the measurements above does not clearly distinguish between subject encoding types in conversational Korean whereas it does in conversational Japanese, as well as because in conversational Korean, the nominative *ka* does not encode thematic prominence, while it does in conversational Japanese.

To summarize the section, information that does not persist loses its importance in the subsequent context, whereas information that persists maintains its importance in the subsequent context. Following Givón (1983) and Shimojo (2005), both RP and RP-*f* are used to measure the cataphoric saliency of subject referents. Based on such an assumption that a greater RP and a greater RP-*f* suggest a greater thematic prominence, the mean RP and mean RP-*f* were expected to be greater for *ka*-marked subjects than for zero-marked subjects. However, the data shown in Table 4.20 and Table 4.21, as well as Table 4.22 and Table 4.23, suggest that in conversational Korean, cataphoric saliency is not a deciding factor for the choice of subject encoding types, but indicates the functional properties of *ka* and *nun*.

Both the mean RP and the mean RP-*f* were somewhat greater for subjects of zero anaphor than they were for other encoding types. More importantly, in conversational Korean, both the mean RP and the mean RP-*f* were smallest for *ka*-marked subjects. Based on these results, I also argued that in conversational Korean, both *ka*-marking and *nun*-marking can mark a continuation of a topic, but the former encodes cataphor-

ically non-salient local persistence (i.e., a local topic) and the latter encodes cataphorically salient global persistence (i.e., a global topic). This is in fact opposite to the tendency Shimojo (2005) observed in his study of Japanese conversation, in that unlike the nominative *ga* in Japanese, the nominative *ka* in Korean does not mark thematic prominence encoded in referents of subject NPs.

Most importantly, in conversational Korean, the measurement of cataphoric saliency, either by RP or by RP-*f*, is not a good means distinguishing between subject encoding types, though it was in conversational Japanese. Nevertheless, it is good enough to point out the functional differences between *ka*-marked subjects and *nun*-marked subjects: *ka* for cataphorically non-salient local persistence (i.e., a local topic); *nun* for cataphorically salient global persistence (i.e., a global topic).

#### 4.5.9 Contrast

The notion of ‘contrast’ has frequently been discussed in previous studies (e.g., Halliday 1967; Chafe 1976; Clancy & Downing 1987; Lambrecht 1994), but the clear-cut definition for the term ‘contrast’ has not necessarily been provided. Lambrecht (1994) argues that this is because almost any element in the clause can inherently be contrastive in a broad sense, mostly due to the presupposition or conversational implicature. Lambrecht (1994) further argues that the term ‘contrastive’ should be defined in a gradient approach rather than in a single-lined term, due to the fact that anything can inherently be contrastive in a certain sense. Clancy & Downing (1987) also argue that the notion of ‘contrast’ needs to be defined more carefully due to varying degrees of contrastivity which is encoded in the elements in the sentence. They attempt to define the notion ‘contrast’ in two different ways, in terms of directiveness; directly contrastive and indirectly contrastive. The present study adopts the former only, i.e., directly contrastive, because the definition for the former seems to be clear-cut, whereas that of the latter

does not seem to be as clear.

The notion ‘directly contrastive’ comprises two subtypes: ‘parallel activities/states’ and ‘action/state reaction’. The first subtype of direct contrastiveness (i.e., parallel activities/states) occurs in the format of “A does/is X, but B does/is Y” (Clancy & Downing 1987). The following exemplifies the contrastiveness in the form of ‘parallel activities/states’.

(4.43) A1: *pam-ey kuntey, neytellantu, [ku ttay-n] yelum-ilase*  
 night-at well Netherlands that time-TOP summer-be  
*nac-i ki-lekaciko sangkwan-Ø epsnuntey,*  
 daytime-NOM long-be matter-( ) not.exist  
 ‘At night, you know, in the Netherlands, at that time, it was summer. So it was okay because the daytime was long.’

A2: *[cikum-un] kyewul-inikka hay-Ø cincca ppalli cicana.*  
 present-TOP winter-be daytime-( ) really early sunset  
 ‘It is winter now, so the sun sets so early.’

In (4.43), *ku ttay* ‘that time’ in the utterance A1 and *cikum* ‘present’ in the utterance A2 represent the contrastive opposition, and both elements in the ‘parallel activities/states’ are encoded with *nun*. More specifically, this instance fits exactly into “A (that time) is X (summer), but B (present) is Y (winter).” Therefore, this utterance is counted as an instance of directly contrastive in the sense of ‘parallel activities/states’.

The second subtype of direct contrastiveness, i.e., an ‘action/state reaction’ type of contrastiveness, typically appears “when the speaker is developing a plot sequence containing successive activities involving different participants” (Clancy & Downing 1987:37). As for the ‘action/state reaction’ type of contrastiveness, Shimojo (2005:176) points out that “[t]he paired elements in ‘action/state reaction’, unlike those in ‘parallel activities/states’, do not appear in parallel; but the first element is presented, and then the second element is (re)introduced into the state of affairs with regard to the first

element.” An example of the direct contrastiveness in the sense of ‘action/state reaction’ is presented in the utterance in (2.5), repeated here as (4.44):

(4.44) A1: [*ne*-Ø    *ppalli*    *chayk*-Ø    *kac-ko*    *wase*    *kongpwu*-Ø    *hay*.]<sub>1st</sub>  
 2SG-( ) quickly book-( ) take-and come.and study-( )  
 ‘You bring your books (to the library) and study now.’

A2: [*hyeng*-Ø    *ecey*    *kongpwu*-Ø    *hay-ss-e*.]<sub>2nd</sub>  
 brother-( ) yesterday study-( ) do-PST-SEM  
 ‘I studied yesterday.’

In (4.44), the speaker, expressed as *hyeng* ‘big brother’ in the utterance A2, points out that the hearer, expressed as *ne* ‘you’ in the utterance A1, would need to bring his books and study right away. In reaction to this, the speaker continues to point out that he studied yesterday. In the ‘action/state reaction’ type of the contrastiveness, only the second element of the contrastive pair is viewed as an instance of direct contrastiveness because the contrastive opposition does not arise until the second element of the pair, as a reaction to the first element, appears in the discourse (Shimojo 2005:177). In (4.44), the second element of the contrastive pair in the second set of brackets appears as a reaction to the first element of the contrastive pair in the first set of brackets. In (4.44), the second element of the contrastive pair is encoded with a zero particle, but the contrastive effect would have been explicit if it had been marked with *nun*, and *ka*-marking would make this utterance sound extremely awkward.

Along with the term ‘contrast’, the notion ‘exhaustive-listing’ (Kuno 1972; 1973), or ‘narrow focus’ (Lambrecht 1994), has also been discussed. Specifically, entities that are viewed as ‘exhaustive or exclusive’ essentially contrast with all other entities, in that only the entity that is singled out exhaustively meets the truth condition represented by the sentence. For this reason, the present study includes all the utterances that contain the exhaustive-listing type of the contrastiveness for the discussion of contrastiveness

with regard to subject as well as object markings. The following is an example of the ‘exhaustive-listing’ type of contrastiveness from the data of the present study.

(4.45) A1: *wuli-ka Kyenghi-Ø anchyehohkwu,*  
 1PL-NOM Kyenghi-( ) sit.and  
 ‘We had Kyenghi sit down and...’

A2: *Kyenghi way ilehkey mak Theyolu-to anikwu,*  
 Kyenghi you.know like.this you.know Teoru-also not.be  
 ‘It was not Teoru.’

A3: *laitu, laitu-Ø hay cwuko maynnal.*  
 ride ride-( ) do give.and every.day  
 ‘(Someone) gives a ride (to Kyenghi) every day.’

A4: *kulen namca-ka issese.*  
 such man-NOM exist  
 ‘There is such a guy.’

A5: *ku salam-uy cengchey-eytayhayse aytul-i mwutnunta.*  
 that person-GEN identity-about they-NOM ask  
 ‘They asked about that guy.’

A6: *nay-ka, [nay-ka] chongtay-Ø meyko,*  
 1SG-NOM 1SG-NOM gun.stock-( ) carry.and  
 ‘It was me (not others) who volunteered, and...’

A7: *kyay-Ø nwukwu-ya?*  
 3RD-( ) who-Q  
 ‘Who’s that guy?’

Before the utterances in (4.45), the speaker was talking to the hearer about how he hung out with his friends several days ago, and the speaker named all of his friends who had hung out with him. While hanging out with his friends, he and his friends happened to talk about *Kyenghi*, one of his friends, who was also there with him. During the conversation, they all became curious about a guy who had been giving a ride to *Kyenghi*, and they started inquiring about the guy. Finally, amongst everyone,

the speaker, but not others, exclusively volunteered and asked *Kyenghi* directly who that guy was, as shown in the utterance labeled as A6 in (4.45).

In the utterance labeled as A6 in (4.45), the speaker, expressed as *nay* ‘I’, was the one and only focus of the sentence, so this sentence, following Kuno (1972; 1973) and Lambrecht (1994), is categorized as an instance that contains an ‘exhaustive-listing’ type of contrastiveness. The following is another example of an ‘exhaustive-listing’ type of contrastiveness.

(4.46) A1: *hakinun yayney-nun sikan-man iss-umyen syophing-Ø hale ka.*  
 as.you.said they-TOP time-only exist-if shopping-( ) to.do go  
 ‘As you said, they go to shop, as long as they have time.’

B1: *satun, an satun.*  
 buy NEG buy  
 ‘Whether they buy or not.’

A2: *kulekwu mwo manhi sacito anh-a.*  
 furthermore what much buy NEG-SEM  
 ‘And, you know, they don’t buy much.’

A3: *sanunke-n ohilye [hankwuk salam-i] hwolssin te manhi sa.*  
 buying-TOP rather Korea people-NOM very more much buy  
 ‘As for buying things, Korean people buy much more.’

A4: *sanunke-n te manhi sanuntey,*  
 buying-TOP more much buy.but  
 ‘As for buying things, (Koreans) buy much more, but’

A5: *kaci-nun ka-nun pintoswu-nun [kyayney-ka] hwolssin manh-ci.*  
 going-TOP go-REL frequency-TOP they-NOM very many-SEM  
 ‘As for how often they go shopping (lit. as for the frequency of going for shopping), they go much more frequently.’

In (4.46), both the speaker and the hearer were talking about Korean and Japanese people’s shopping habits, while comparing one with the other. In the utterance labeled as A3 in (4.46), the nominative-marked subject, *hankwuk salam* ‘Korean’, is the



only entity which allows for the utterance to be interpreted in its intended exclusive manner. That is to say, the utterance labeled as A3 is interpreted as “as for buying things, between Korean and Japanese people, it is Korean, but not Japanese, who buy more,” so it is exclusive, according to Kuno (1972; 1973). Therefore, this utterance is also classified as an ‘exhaustive-listing’ type of contrastiveness. In (4.46), the utterance labeled as A5 is also categorized as an instance of an ‘exhaustive-listing’ type of contrastiveness because the nominative-marked subject in brackets, *kyayney* ‘they (referring to Japanese)’, among Korean and Japanese, is the one and only focus of the sentence. In other words, the utterance labeled as A5 is interpreted as “as for the frequency of shopping, it was Japanese, among Korean and Japanese, who go more frequently,” so it perfectly fits into Kuno’s (1972; 1973) definition of ‘exhaustive-listing’ type of contrastiveness.

More importantly, in (4.46), entities in brackets, *hankwuk salam* ‘Korean people’ in the utterance of A3 and *kyayney* ‘they’ in the utterance of A5 together also exemplify a ‘parallel activities/states’ type of contrastiveness because these paired entities represent a contrastive opposition. More specifically, the two utterances A3 and A5 exactly fit into “A (Korean people) does X (buy more things), but B (Japanese people) does Y (shop more frequently).” Therefore, the utterances of A3 and A5 in (4.46) should also be categorized as instances of the ‘parallel activities/states’ type of direct contrastiveness. Since the utterances of A3 and A5 can be viewed either as a ‘directly contrastive (parallel activities/states)’ (Clancy & Downing 1987) or as an ‘exhaustive-listing’ (Kuno 1972; 1973) type of contrastiveness, the present study classifies these two utterances more precisely into a ‘directly contrastive and exhaustive-listing’ type of contrastiveness, in order to distinguish these instances from instances of either simply ‘directly contrastive’ or ‘exhaustive-listing’ type of contrastiveness.

To briefly summarize what has been discussed, in the present study, the notion ‘con-

trastiveness’ is categorized largely into three different sub-types: ‘directly contrastive’, ‘exhaustive-listing’, as well as ‘directly contrastive and exhaustive-listing’. Following approaches used by Clancy & Downing (1987), the ‘directly contrastive’ type of contrastiveness is classified more precisely, based on its intended description, into two different sub-types: ‘parallel activities/states’ and ‘action/state reaction’.

Based on the indices of contrastiveness defined above, the present study explores the choice of subject markings in conversational Korean with respect to the contrastiveness of subjects, with an assumption that entities that are contrastive in one way or another are overtly identified more commonly than entities that are not, due to informational prominence. Simply put, I assume that an overt marking rather than a zero-marking is preferred for encoding subjects that are contrastive. With this assumption, I will first discuss the choice of subject markings simply based on whether or not subjects are contrastive. Then, I will discuss the choice of subject markings more precisely with regard to each sub-type of contrastiveness defined above. Table 4.24 shows the overall token distribution of subject markings based on a simple distinction of contrastive or non-contrastive.

Table 4.24: Subject Contrast and Subject Marking

SUBJ.MARK SUBJ.CONTRAST		NP-Ø	NP-ka	NP-nun	NP-to	NP-etc.	SUB- TOTAL	TOTAL
¬contrast	NO.	1,187	1,464	483	407	55	2,409	<b>3,596</b>
	%	<b>33</b>	<b>41</b>	<b>13</b>	11	2	67	100
contrast	NO.	40	100	162	1	1	264	<b>304</b>
	%	<b>13</b>	<b>33</b>	<b>53</b>	0	0	87	100
TOTAL	NO.	1,227	1,564	645	408	56	2,673	3,900
	%	31	40	17	10	1	69	100

In Table 4.24, I find that the number of subjects that are viewed as contrastive (304 tokens, approximately 8% of the entire subject tokens) is far smaller than that of

subjects that are viewed as non-contrastive (3,596 tokens, roughly 92% of the entire subject tokens). More importantly, the use of zero-marking is much less common for coding subjects that are contrastive (13%) than it is for coding subjects that are not contrastive (33%). On the other hand, the occurrence of *nun*-marked subjects is much higher when subjects are viewed as contrastive (53%) than when subjects are regarded not as contrastive (13%) (zero-marking vs. *nun*-marking:  $\chi^2=209.795$ ,  $P < 0.001$ ). In fact, *nun*-marked subjects take the largest percentage of the total contrastive subjects. As for the so-called topic marker *nun*, Choi (1996; 1997) claims that the particle *nun* is not a topic marker but instead a contrastive marker, though it has widely been agreed that *nun* functions mainly as marking a (grammatical) topic of the clause (Sohn 1999; Lee 2003). As a matter of fact, Choi (1997) proposes that the topicality of the clause is not encoded by the particle *nun* but by scrambling. Although Choi's (1997) analysis deserves recognition, I view her claim to be untenable, especially because, of all *nun*-marked subjects in the data, only 162 tokens of subjects (approximately 25% of the total) are categorized as contrastive, whereas 483 tokens of subjects (about 75% of the total) are categorized as non-contrastive. In short, Choi's (1997) approach does not explain 483 tokens of *nun*-marked subjects that are viewed as non-contrastive. Nevertheless, it is certain that one discourse function encoded in the particle *nun* is to mark contrastiveness, though it is not its only function.

Shimojo (2005:127) points out that “a sentence containing a *wa*-marked NP is inherently contrastive due to the referent denoted by the NP, which is singled out from the comment provided about the referent as well as other propositions present anaphorically or exophorically.” Shimojo (ibid.) further points out the complementary nature of two encoding types in his study of Japanese conversation; *wa*-marking is associated with direct contrastiveness (82% of the total), and zero-marking is associated with non-contrastiveness (83% of the total).

In Table 4.24, I also find that, although not as many as for *nun*-marked subjects (53%), still a good portion of subjects categorized as contrastive are marked with the nominative *ka* (33%) (*ka*-marking vs. *nun*-marking:  $\chi^2=14.173$ ,  $P < 0.001$ ). This result supports what has been claimed with regard to the discourse function of the nominative *ka* in Korean in previous studies (e.g., Choi 1997; Han 1999; Lee 2003, among others). That is to say, in Korean, the contrastiveness can be marked not only by the particle *nun* but also by the nominative *ka*.

Lastly, according to the data in Table 4.24, zero-marking is not preferred for encoding subjects that are categorized as contrastive, though it occasionally marks the contrastiveness of subject NPs (13%).

Based on what has been discussed thus far, the complementary nature of these three encoding types seems to be obvious. That is to say, the use of *nun*-marking and *ka*-marking is preferred for coding the subjects categorized as contrastive, whereas the use of zero-marking is not preferred for such a purpose. Despite this complementary nature with respect to the contrastiveness marking, the distinction between *ka*-marking and *nun*-marking still remains unclear since the occurrence of *ka*-marking (33%) is also frequent for coding contrastive subjects, though not as frequent as that of *nun*-marking (53%). In order to see if these encoding types mark different senses of contrastiveness, I classified all subjects categorized as contrastive into four different sub-types, as was stated earlier in the present section, and further examined the entire body of contrastive subjects with respect to these four sub-types. The results are summarized in Table 4.25.

In Table 4.25, a clear distinction is observed between *ka*-marking and *nun*-marking as encoders of contrastiveness. In general, the nominative *ka* tends to mark an ‘exhaustive-listing’ type of contrastiveness, while the post-nominal particle *nun* is likely to mark a ‘directly-contrastive’ type of contrastiveness (exhaustive-listing *ka* vs. directly-contrastive *nun*:  $\chi^2=80.75$ ,  $P < 0.001$ ). For example, an ‘exhaustive-listing’ type is encoded most

Table 4.25: Sub-typed Subject Contrast and Subject Marking

SUBJ.MARK		NP-Ø	NP-ka	NP-nun	NP-to	NP-etc.	SUB-TOTAL	TOTAL
SUB-SUBJ.CONTRAST								
¬contrast	NO.	1,187	1,464	483	407	55	2,409	3,506
	%	33	41	13	11	2	67	100
exhaustive	NO.	3	53	0	0	0	53	56
	%	5	<b>95</b>	<b>0</b>	0	0	95	100
parallel	NO.	20	20	70	0	1	91	111
	%	18	<b>18</b>	<b>63</b>	0	1	82	100
reaction	NO.	15	3	92	1	0	96	111
	%	14	<b>3</b>	<b>83</b>	1	86	100	
d.cont.+exhaust	NO.	2	24	0	0	0	24	26
	%	<b>8</b>	<b>92</b>	0	0	0	92	100
TOTAL	NO.	1,227	1,564	645	408	56	2,673	3,900
	%	31	40	17	10	1	69	100

often by the use of *ka*-marking (95%), as exemplified in utterance A6 in (4.45). On the other hand, there is no occurrence of *nun*-marking used for coding this sub-type of contrastiveness. This distributional tendency strongly supports a previous claim, that is, the nominative *ka* marks a subject which denotes an ‘exhaustive-listing’ type (Choi 1997). As a matter of fact, this use of *ka*-marking in Korean is analogous to that of *ga*-marking in Japanese (Kuno 1972; 1973), in that both nominatives, *ka* in Korean and *ga* in Japanese, are used to mark an ‘exhaustive-listing’ type of contrastiveness. Based on this, it can be said that one exclusive discourse-functional feature that both nominative markers, i.e., *ka* in Korean and *ga* in Japanese, have is to mark an ‘exhaustive-listing’ type of contrastiveness.

On the other hand, the use of *nun*-marking constitutes the largest portion of ‘parallel activities/states’ (63%) and ‘action/state reaction’ (83%) types of contrastiveness. Although *nun*-marking is most likely used to encode subjects that are viewed

as the ‘directly-contrastive’ type, zero-marking and *ka*-marking are also used to encode these subjects that are ‘directly contrastive’, as shown in Table 4.25. The use of zero-marking for encoding subjects in a ‘directly-contrastive’ type is exemplified in (4.44), where both subjects in brackets are zero-marked. Lastly, the data in Table 4.25 suggests that *ka*-marking (92%) is the unmarked encoding type for subjects in ‘directly-contrastive+exhaustive-listing’ type of contrastiveness, while *nun*-marking is not an option for coding subjects shown in such a contrastive type.

In general, the use of zero-marking is not preferred for encoding subjects that are regarded as contrastive, regardless of which subtype of contrastiveness it is classified into.

Overall, the observations in the present conversational data point to the complementary nature of subject encoding types: *ka*-marking is associated with the ‘exhaustive-listing’ type of contrastiveness; *nun*-marking is associated with the ‘parallel activities/states’ as well as ‘action/state reaction’ types of contrastiveness; zero-marking is generally not preferred for subjects that are in contrast, regardless of the subtype. The distributional tendency observed in the present study complies with the tendency previously reported in spoken Japanese in the same regard (Shimojo 2005), in that the occurrence of an overt marking is high for contrastive subjects, and conversely, the occurrence of zero-marking is low for coding them.

To summarize the section, contrastiveness appears to strongly correlate with the choice of subject markings in conversational Korean. Overall, zero-marking is not preferred for coding subjects that are in contrast, and an explicit marking, either *ka* or *nun*-marking, is most often used instead to code subjects that are in contrast. More specifically, when subjects are in ‘exhaustive-listing’ or ‘directly contrastive+exhaustive-listing’, *ka*-marking is the default encoding type for these subjects. On the other hand, *nun*-marking is the unmarked encoding type for the subjects that represent ‘directly-

contrastive’ type of contrastiveness, i.e., ‘parallel activities/states’ and ‘action/state reaction’ types of contrastiveness. These observations are consistent with those of the previous studies in both Korean and Japanese (Kuno 1972; 1973; Choi 1997; Han 1999; Lee 2003; Shimojo 2005).

Lastly, I argue that subjects in contrast are informationally more prominent in the sense that in order to correctly process the utterance, in addition to referent identification, the identification of contrastiveness is also necessary. These subjects are thus considered to be informationally loaded, which implies informational prominence, as well as processing load. Therefore, the use of the overt marking with *ka* or *nun* is common to code subjects in contrast, and as expected, the use of zero-marking is not common for subjects that are in contrast due to informational prominence, as well as processing load, encoded in them. It seems that in conversational Korean, the speaker’s selection of one particular encoding type for subjects in opposition to other encoding types is not an arbitrary but a systematic process, particularly when contrastiveness plays a role in the choice of subject encoding types.

#### **4.5.10 Summary of Chapter 4**

In Chapter 4, I have shown that in conversational Korean, the choice of subject markings has a relationship with two discourse related factors: processing load and informational prominence. Furthermore, I have argued that one discourse-pragmatic function of the nominative *ka* in conversational Korean is associated with processing load and informational prominence.

In Section 4.2, I have discussed how subject encoding types are realized in conversational Korean. In the section, I showed that in conversational Korean, at least according to the data of the present study, subjects are most likely unexpressed, that is, subjects appear as zero anaphor. When subjects are explicitly expressed, they tend

to be coded with an overt particle more often than with a zero particle. I pointed out that this tendency is not consistent with that observed in spoken Japanese, in that in Korean, subjects are more likely to be coded with an overt particle (e.g., *ka*), whereas in Japanese, subjects are more likely not to be marked with an overt marker.

In Section 4.3, I have shown that subjects of intransitive clauses tend to be overtly expressed more than subjects of transitive clauses. I also showed that the occurrence of *ka*-marking was higher for coding subjects of intransitive subjects (S) than for coding subjects of transitive clauses (A). I stated that this pattern is consistent with the universal tendency to express at most one argument per clause (Fry 2003), together with the ergative-absolutive pattern that intransitive subjects and transitive objects tend to be overtly expressed (Dixon 1994, Thompson & Hopper 2001).

In Section 4.4, I have demonstrated that subjects either in interrogative or in directive sentences tend to be marked with a zero particle more than subjects in declarative sentences. To explain this tendency, I suggested that subjects appearing in interrogative, as well as those in directive sentences, tend to be presupposed more than subjects in declarative sentences, and they are accessible, thus representing a low degree of processing load, so they are frequently coded with zero-marking.

In Section 4.5, I have discussed the choice of subject markings with relation to the following factors which are related to either processing load or informational prominence or both: negation, animacy, word order, length, repair, anaphoric saliency, cataphoric saliency and contrast.

In Section 4.5.1, it was initially assumed that the negation would generally increase processing load due to a high conceptual complexity. However, according to the data in Table 4.4, the negativity of sentences did not particularly call for the use of *ka*-marking for subjects. I attributed this tendency to the general property encoded in negation, that is, an entity in the focus of negation is normally an already-established topic, so it



can easily be identified, although the negation generally represents processing load due to conceptual complexity. For this reason, the occurrence of *ka*-marking is relatively low for subjects that appear in negative sentences. Contrastively, the use of *nun*-marking is more common for subjects in negative sentences than in affirmative sentences, due to the fact that an entity in the focus of negation usually contrasts with another entity. Also, the use of *to* ‘even’ is more common for subjects in negative sentences than in affirmative sentences.

In Section 4.5.2, I have shown that in conversational Korean, the choice of subject markings was influenced by the animacy of subject NPs, in that the use of *ka*-marking was not common for coding subjects whose referents were animate, though the influence was not always considered significant. It was also stated that this observation was consistent with the previous finding (e.g., Keenan 1976; Comrie 1989; Fry 2003). By examining the choice of subject markings further in several other respects, I discussed the relationship between the choice of subject markings and the animacy of the subject as well as the object. The occurrence of *ka*-marked subjects was low when subject referents were human beings, and even lower when they were interlocutors (i.e., the speaker or the hearer), due to high accessibility, hence low processing load. On the other hand, the occurrence of *ka*-marked subjects was greater when subjects occurred with animate objects, perhaps due to processing load. In short, I argued, in Section 4.5.2, that the use of *ka*-marking was common when processing load was expected, whereas the use of zero-marking was common when processing load was not expected for subjects.

In Section 4.5.3, I have argued that the choice of subject markings is related to word order variations, though the relationship was not always significant. Overall, subjects in non-canonical/marked word orderings tended to be coded with *ka*-marking, due to processing load (e.g., in OSV, 52% of subjects are coded with *ka*-marking). Despite a potential increase in processing load incurred by a marked word ordering, i.e., a

deviation from the norm, post-predicative subjects were not commonly coded with *ka*-marking (e.g., 34% of *ka*-marked subjects in  $VS_i$ ; 42% of *ka*-marked subjects in  $SV_i$ ). This pattern was attributed to the assumption that although a post-predicative subject is in a marked position, hence represents processing load, the post-predicative subject represents attenuated informational prominence, so the motivation for *ka*-marking is downgraded for the post-predicative subject. I further argued that both factors of processing load and informational prominence often interplay, just as in word order variations, and they both play roles in the selection of subject markings. Furthermore, the occurrence of zero-marking was slightly higher for intransitive subjects than for transitive subjects that appear in a post-predicative position. Lastly, unlike subjects appearing in a sentence-initial or sentence-final position, the occurrence of *nun*-marked subjects was low when subjects appear in a sentence-middle position because entities in a sentence-middle position are hardly identified as a (grammatical) topic.

In Section 4.5.4, I have stated that the definiteness of subject referents influences the choice of subject markings. Subjects with a definite referent are generally identifiable, so they tend not to be coded with the nominative *ka* (36%). Due to the definiteness, indicating an already-established topic, the occurrence of *nun*-marking was also reported to be greater for coding definite subjects (22%) than for coding other subtypes of subjects (e.g., 10% for indefinite subjects). Subjects representing an indefinite referent are non-referential, hence not identifiable, to both the speaker and the hearer, so they tend to be coded with the nominative *ka* (47%). Subjects representing an indefinite but specific referent are identifiable only to the speaker, so they are also frequently coded with the nominative *ka* (58%). Lastly, the occurrence of *ka*-marking is low and the occurrence of zero-marking is high for coding subjects of indefinite pronominal NPs because the subjects of indefinite pronoun do not contain any information about the referent, hence would represent low informational prominence. Based on the tendency reported in the

present study, I have claimed that both processing load and informational prominence play roles in the choice of subject markings.

In Section 4.5.5, I have noted that *ka*-marking was not common for coding monosyllabic subjects, though it was reported to be so in colloquial Japanese. I attributed this tendency to the difference in the type of subject NPs: monosyllabic subjects in Korean are mostly personal pronouns, while it is not the case with Japanese. When the types of subject NPs were considered as a factor, a similarity between these two languages was observed. More specifically, monosyllabic full nouns are not common in either languages, and the occurrence of an overt marking was relatively high for encoding monosyllabic full nouns due to processing load. On the other hand, monosyllabic pronouns are not uncommon in Korean, and they are not prominent, so they are not commonly coded with the nominative *ka*. I also examined the choice of subject markings with regard to the number of subject modifiers. The results showed that the percentage of *ka*-marked subjects increases as the number of subject modifiers increases, due to informational prominence. In short, the nominative *ka* is used to code subjects that represent processing load or informational prominence.

In Section 4.5.6, I have shown that in general, the use of zero-marking was not common for coding repaired subjects, which are in the focus of attention, hence represent informational prominence. Furthermore, despite the general tendency of repairs to occur commonly with new information, the percentage of repaired subjects in intransitive clauses is the same as that of transitive clauses. Also, *ka*-marking was reported to be higher for coding repaired subjects of intransitive clauses than for coding those of transitive clauses, and contrastively, *nun*-marking was more common for repaired subjects of transitive clauses than for those of intransitive clauses. I have taken this tendency as a supporting evidence for the argument that *nun* marks old information and *ka* marks new information. Based on this tendency, I claimed that the selection of subject mark-

ings was related to whether or not subjects were repaired, and one discourse-pragmatic function associated with *ka*-marking is to encode informational prominence.

In Section 4.5.7, I have demonstrated, based on the data in Table 4.16 and Table 4.17 as well as Table 4.18 and Table 4.19, that in conversational Korean, the anaphoric saliency, either by recency (RD) or by frequency (AF), of subject referents mattered in the choice of subject encoding types, just like in Japanese where the anaphoric saliency, though measured by recency (RD) only, plays a role in the speaker's selection of subject encoding types (Shimojo 2005). Specifically, the mean RD was smallest for zero anaphor and greatest for *ka*-marked subjects. I attributed this tendency to processing load incurred by activation cost. That is to say, subjects of a small RD remain activated in the addressee's consciousness, and are thus identifiable in the immediately preceding context, so there is no need for them to be overtly expressed in a clause. On the other hand, subjects with a high RD are not as activated, thus are not as identifiable, so they are likely to be coded with *ka*-marking, which I argue functions to alleviate processing load. The mean AF was greatest for subjects of zero anaphor, and it was smallest for *ka*-marked subjects. I also attributed this tendency to processing load. That is to say, subjects with a high(er) AF are more identifiable, and there is less need for them to be overtly expressed in a clause. Conversely, subjects of a small(er) AF are not as activated, thus not as identifiable, so they are commonly coded with *ka*-marking.

With these results mentioned above, I have pointed out that Korean bear cross-linguistic similarities to Japanese in the sense that in the two languages, the anaphoric saliency of subject referents is linked to the use of particular subject encoding types, though the distinction among subject encoding types appeared much greater in Japanese than it appeared in Korean. For example, in both languages, the nominative markers (i.e., *ka* for Korean; *ga* for Japanese) are used to mark subjects that are anaphorically non-salient, but the mean RD for *ga*-marked subjects in Japanese, as reported in Shi-

mojo (2005), is much greater than the mean RD for *ka*-marked subjects in Korean. I attributed this difference to discourse-functional differences of the nominative markers, that is, *ka* in Korean may mark a continuing topic, whereas *ga* in Japanese lacks this property (Choi 1997). Therefore, the mean RD is expected to be greater for *ga*-marked subjects in Japanese than it is for *ka*-marked subjects in Korean.

In Section 4.5.8, I have shown, based on the data in Table 4.20 and Table 4.21 as well as Table 4.22 and Table 4.23, that unlike Japanese, Korean does not exhibit a relationship between the choice of subject encoding types and cataphoric saliency of subject referents. More importantly, both the mean RP and the mean RP-*f* for *ka*-marked subjects appeared to be smaller than the ones for *zero*-marked subjects. In fact, both the mean RP and the mean RP-*f* were smallest for *ka*-marked subjects, among all available encoding types. This tendency is opposite to the tendency that Shimojo (2005) observed in his analysis of Japanese conversation. Furthermore, I proposed that in conversational Korean, both *ka* and *nun* can mark a continuation of a topic, but they differ in that the former tends to encode cataphorically non-salient local persistence (i.e., a local topic) and the latter tends to encode cataphorically salient global persistence (i.e., a global topic). Based on this, I further argued that unlike in Japanese, in conversational Korean, the nominative *ka* does not encode thematic prominence, i.e., cataphoric saliency.

In Section 4.5.9, I have demonstrated that the notion of contrastiveness was a valid means of distinguishing among three different subject encoding types: zero-marking, *ka*-marking and *nun*-marking. More specifically, zero-marking seemed to be a marked/non-canonical encoding type for subjects in contrast, while *ka*-marking seemed to be a default encoding type for subjects shown in ‘exhaustive-listing’ and ‘directly-contrastive+exhaustive-listing’ subtypes of contrastiveness, and *nun*-marking seemed to be an unmarked/canonical encoding type for subjects occurring in a ‘directly-

contrastive' type of contrastiveness, either 'parallel activities/states' or 'action/state reaction' subtypes. In the section, I also pointed out that the observation of the present study is consistent with those of the previous studies in both Korean and Japanese (Kuno 1972; 1973; Choi 1997; Han 1999; Lee 2003; Shimojo 2005). Most importantly, I have claimed, based on the results summarized in Table 4.24 and Table 4.25, that the use of an overt particle, either *ka* or *nun*, is common for coding subjects that appear in a context of contrastiveness. Lastly, I argued, based on these results, that the speaker's selection of one particular encoding type in opposition to other encoding types for contrastive subjects is not an arbitrary but systematic process.

# Chapter 5

## Object Marking

### 5.1 Outline

Chapter 5 presents the detailed discussion of the object encoding types in conversational Korean, from viewpoints of processing load and informational prominence. That is to say, it discusses how a speaker chooses a particular object encoding type over another, largely based on two different factors: processing load and informational prominence. In the present section, the choice of object encoding types is discussed by factors which are related to processing load, informational prominence, or both. Factors that are used for the discussion of object encoding types are: negation, animacy, word order, verb adjacency, definiteness, length, repair, anaphoric saliency, cataphoric saliency, as well as contrast. These factors are related to processing load or informational prominence in one way or another. In addition to these factors, the choice of object markings is discussed with relation to *hata*-verb and split-case marking as well.

## 5.2 Object encoding types

As repeatedly noted, all clausal units in the data were divided into two different subgroups based on transitivity: intransitive clauses and transitive clauses. Unlike subjects, objects appear only in transitive clauses, and therefore only the transitive clausal units were examined with respect to the choice of object encoding types. The number of transitive clauses is 3,692 in total, as was stated in Chapter 3. Just like subjects, objects in Korean may or may not be expressed overtly in a clause, and they may occur with or without an overt particle when they are overtly expressed in a clause. Simply put, objects in Korean are realized largely in three different forms, depending on the situation in which the conversation is carried out; a zero anaphor, a *bare* NP, and a *full* NP with an overt post-nominal marker.

It needs to be pointed out that the selection of a particular NP form for encoding an object is not random, but instead reflects the interaction between the grammatical role and the information status of the NP in a clause, as was noted in Bak (2004). Ko. E (2000), in her study of objects in Korean, also states that the selection of NP forms is based on the information status of the NPs in a clause. In the present study, as I did with subjects, I also examined how object NPs are encoded in conversational Korean, in order to see how the grammatical role and the information status of the NP interact. That is to say, the present study explores the selectional process of object encoding types, and Table 5.1 presents the distributional pattern for object encoding types.

Table 5.1: Object Encoding Types

OBJ.ET	ZERO-ANAPHOR	NP-Ø	NP-lul	NP-nun	NP-to	NP-etc.	SUB-TOTAL <sup>1</sup>	TOTAL
NO.	961	1,688	699	110	200	34	1,043	3,692
%	<b>26</b>	<b>46</b>	19	3	5	1	<b>28</b>	100



In Table 5.1, I find that objects are most likely expressed overtly in a clause (46+28=74%), and when they are explicitly expressed, they are coded more frequently with a zero particle (46%) than with an explicit particle (28%). In short, objects are realized in the following order of occurrence: NP- $\emptyset$  > NP-*lul*<sup>2</sup> > zero anaphor. This distributional pattern observed for object NPs in conversational Korean differs greatly from the distributional pattern observed for subject NPs, in that subjects are most frequently unexpressed, and when they are overtly expressed, they tend to be coded with an overt particle (in short, zero anaphor > NP-*ka* > NP- $\emptyset$ ).<sup>3</sup> I attribute this difference in distributional tendencies to the general characteristics of object NPs. That is to say, unlike subject NPs (of transitive sentences), new information is usually encoded in object NPs, and the new information can only be introduced into the discourse when object NPs are expressed overtly in a clause. Therefore, zero anaphor is not an encoding type that is commonly used for object NPs. Furthermore, object NPs usually denote new information, which may function as a focusing device, so the use of the overt marking is not as high for objects as it is for subjects. Based on this, objects are expected to be realized in the following order of occurrence frequency: NP- $\emptyset$  (46%) > NP-*lul* (28%) > zero anaphor (26%). The tendency for objects observed in the present study of Korean complies with the tendency that was reported in the studies of object encoding types in Japanese (e.g., Fujii & Ono 2000).

Contrastively, the higher occurrence of zero-marked objects reported in Table 5.1 does not comply with Kim's (2006) report on object encoding types in Korean. He (ibid.) reports that objects are coded more frequently with *lul*-marking than with zero-marking, in his study of object markings in Korean, although the occurrence of

---

<sup>1</sup>Recall that the notation 'SUB-TOTAL' indicates the sum of all overt particles used, in other words, the number of overtly marked NPs.

<sup>2</sup>Other particles (e.g., *nun*) can be used to mark objects instead of the particle *lul*, but the particle *lul* represents any use of explicit marking here.

<sup>3</sup>See Section 4.1 for a detailed discussion of subject encoding types.

unexpressed objects is high in both studies. This distributional difference for object markings may be attributed to the difference in the genre of discourse data used in the two studies, as was indirectly suggested by Ko. E's (2000) findings. More specifically, the data used in Kim (2006) is considered to be formal, whereas the data used in the present study is informal, though both sets are from spoken Korean. Simply put, the occurrence of *lul*-marked objects is higher in Kim's (ibid.) study than in the present study, based on the assumption that objects tend to be explicitly coded in formal contexts (Ko. E 2000).

It is notable that the tendency of objects observed in the present study is roughly opposite to that of subjects. That is to say, objects are not likely to appear as a zero anaphor, while subjects of transitive sentences are most likely to occur as a zero anaphor. Furthermore, objects are likely to be coded with zero-marking, whereas subjects tend to be coded with overt marking regardless of the transitivity of the sentences in which they appear. Table 4.2, repeated as Table 5.2, is provided here for the comparison between subject and object encoding types.

Table 5.2: Transitivity and Subject Encoding Types-repeated

SUBJ.ET TRANS.		ZERO- ANAPHOR	NP-Ø	NP-ka	NP-nun	NP-to	NP-etc.	SUB- TOTAL	TOTAL
intran.	NO.	2,700	921	1,183	410	314	29	1,937	5,557
	%	<b>49</b>	<b>17</b>	21	7	6	1	<b>35</b>	101
tran.	NO.	2,649	306	381	235	94	27	736	3,692
	%	<b>72</b>	<b>8</b>	10	6	3	1	<b>20</b>	100
TOTAL	NO.	5,349	1,227	1,564	645	408	56	2,673	9,249
	%	58	13	17	7	4	1	29	100

As repeatedly stated, I attribute a relatively high percentage of overt objects, compared to the percentage of overt subjects, to the general characteristics of object NPs. In other words, objects tend to encode new, thus non-shared/topical, information, so

they are likely to be expressed overtly in a clause.<sup>4</sup> Contrastively, subject NPs tend to denote old information, hence are accessible, and they commonly appear as a zero anaphor in a clause.

It is also notable that according to the data in Table 5.1 and Table 5.2, subjects of transitive clauses, which encode old information, are not likely to be expressed overtly, whereas objects are likely to appear overtly in a clause. Conversely, subjects of intransitive clauses tend to be expressed overtly in a clause. As was pointed out earlier in Chapter 3, these distributional patterns may be attributed to the universal tendency to express at most one argument per clause, which Fry (2003) points out. These distributional patterns may indicate an ergative-absolutive pattern, in that in intransitive clauses, subjects tend to be expressed explicitly, while in transitive clauses, objects tend to be expressed overtly (Dixon 1994). On the other hand, these distributional patterns may simply reflect an observation that referents of subjects of transitive clauses are often humans, perhaps even interlocutors (i.e., the speaker or the hearer), which are easily identifiable, so they are not often expressed overtly in a clause. Conversely, referents of subjects of intransitive clauses are usually inanimate, which are less accessible than human referents, so they tend to appear overtly in a clause, more particularly, they are coded with an explicit marking.

Overall, unlike subjects, objects usually encode new information, so they tend to appear overtly in a clause. However, when objects are overtly expressed, they tend to be coded with a zero particle rather than with an overt particle. One may consider this tendency to be unexpected, based on the assumption that objects usually denote new information, so they are expected to be coded with an overt marking. I attributed this to the assumption that the new information encoded in objects may function as a focusing device.

---

<sup>4</sup>Givón (2001) views the subject as the primary topic, and the object as secondary.

In this chapter, the choice of object encoding types is discussed based on individual factors as well as several factors combined together, which are related to either processing load or informational prominence, or both. I will discuss object markings based on individual factors, and then object markings based on multi-factors when they are relevant.

### 5.3 Sentence types

As was previously noted in Section 4.4, to the best of my knowledge, in Korean, there have been no studies investigating if and how the selection of object encoding types are related to the variation of sentence types. In Japanese, however, there have been several studies (e.g., Makino & Tsutsui 1986; Backhouse 1993) reporting that in Japanese, the use of an overt marking for coding arguments appearing in questions is not preferred, even unnatural (Makino & Tsutsui 1986; Backhouse 1993), though no justification has been provided for such a tendency. Still, to the best of my knowledge, there have been no studies examining object encoding types with regards to directive sentences in either language.

In Section 4.4, I showed that subjects were likely to be zero-marked when they appeared in interrogative or directive sentences. My proposition for such a tendency was that the subjects which are asked about are presupposed, thus activated, so they are identifiable to both the speaker and the hearer. For this reason, they were commonly coded with a zero particle. The subjects in directive sentences tend to be the hearer, thus are identifiable, so they are commonly coded with a zero particle.

In the present section, in order to see if the choice of object encoding types is in correlation with a sentence type, I examine object encoding types based on the sentence types in which objects appear. For this reason, as with subjects, all transitive clausal

units are divided into three different subtypes based on their sentence types: declarative, interrogative, and directive. The following is the table which shows the distributional pattern for object encoding types with relation to the sentence type, in which object NPs occur.

Table 5.3: Sentence Types and Object Marking

OBJ.MARK		NP-Ø	NP-lul	NP-nun	NP-to	NP-etc.	SUB-TOTAL	TOTAL
SEN.TYPES								
declarative	NO.	1,462	642	91	190	30	953	2,415
	%	<b>61</b>	<b>27</b>	4	8	1	39	100
interrogative	NO.	205	55	19	10	4	88	293
	%	<b>70</b>	<b>19</b>	6	3	1	30	100
directive	NO.	21	2	0	0	0	2	23
	%	<b>91</b>	<b>9</b>	0	0	0	9	100
TOTAL	NO.	1,688	699	110	200	34	1,043	2,731
	%	62	26	4	3	1	38	100

In Table 5.3, I find that overall, a zero particle is more commonly used for coding object NPs than an overt particle, across all sentence types. This tendency for object NPs is opposite to the tendency observed for subject NPs, which are more likely to be coded with an overt particle than with a zero particle. More importantly, objects are more likely to be coded with a zero particle when they occur in either interrogative (70%) or directive (91%) sentences than when they appear in declarative sentences (61%). This distributional tendency for object NPs is consistent with the tendency reported for subject NPs (see Table 4.3 in Section 4.4 for a comparison), in that the use of zero-marking is more common for both subjects and objects of interrogative or directive sentences than for those of declarative sentences. Furthermore, it also complies with what was reported in similar studies in colloquial Japanese (e.g., Makino & Tsutsui 1986; Backhouse 1993), in that the use of an overt particle is not preferred for objects in interrogative sentences.

As was stated in Section 4.4, it seems that no justification has been provided for why a zero particle is more commonly used for coding objects of interrogative or directive sentences, in both Korean and Japanese. In the present study, I discuss why it is so. The following (i.e., (5.1) through (5.3)) are examples of objects appearing in each sentence type from the data of the present study.

(5.1)

A1: *na-nun way kulehkey ay-tul-i cal ha-nun kes kath-ci?*  
 1SG-TOP why so guy-PL-NOM well do-COMP that feel-Q  
 ‘Why do I feel that other students are doing great?’

A2: *nay-ka mos hay-se kulen-ka?*  
 1SG-NOM NEG do-because so-Q  
 ‘Is it because I am not doing well?’

A3: *na-nun swuep-ul tul-umyenun,*  
 1SG-TOP class-ACC take-if  
 ‘When I take classes,’

The utterance A3 in (5.1) exemplifies an object shown in a ‘declarative’ sentence type. The underlined object in the utterance A3, *swuep* ‘class’, had not been mentioned before this utterance but is newly introduced into this utterance, and it is marked with the accusative *lul*.

(5.2) *ne-Ø kuke-Ø intheney-s-ulo tul-ess-e?*  
 2SG-( ) that-( ) Internet-by take-PST-Q  
 ‘Did you take it on the Internet?’

The utterance in (5.2) shows an example of an object occurring in an interrogative sentence. In (5.2), the underlined object, *kuke* ‘that’, is what is being asked about, and in fact, it is what had been talked about. That is to say, it has been made accessible to both the speaker and the hearer, so the use of zero-marking is expected

for coding this object that appear in the interrogative sentence. As a matter of fact, the utterance in (5.2) may have sounded a little unnatural if the object had been coded with the accusative *lul*, given that the object had been talked about up to this utterance, hence it is highly accessible. Therefore, this tendency complies with what Backhouse (1993) points out in his study of Japanese, that is, an overt particle after the argument is unnatural in questions. Simply put, both in Korean and Japanese, objects of interrogative sentences are likely to be marked with a zero particle.

(5.3) *ya, ni sa-ka, hayntuwosi-Ø myech-kay.*  
 hey 2SG buy-go handwash-( ) what-NC  
 ‘Hey, why don’t you buy a couple bottles of handwash?’

Lastly, an example of an object of a directive sentence is shown in (5.3). The underlined object, *hayntuwosi* ‘handwash’, had been continuously talked about up to this utterance, and thus it is highly accessible to both the speaker and the hearer. Due to its accessibility, it is identified with the use of a zero particle. Furthermore, it is natural to have objects of directive sentences identifiable to both the speaker and the hearer, in that the hearer should be able to identify the objects of directive sentences in order to carry out the speaker’s request.

To summarize, in the present section, I have examined the choice of object markings with regard to sentence types in which object NPs occur. Just like subject NPs, object NPs are more likely to be coded with a zero particle in interrogative or directive sentences than in declarative sentences. In other words, regardless of whether it is a subject or an object, the use of an overt particle is not preferred for coding arguments that appear either in interrogative or directive sentences. The similar tendency is also observed in spoken Japanese (e.g., Backhouse 1993). I attributed this tendency to the accessibility of the arguments, i.e., subjects and objects. That is, subjects or objects appearing in interrogative or directive sentences are usually accessible to both the speaker

and the hearer, thus represent low processing load, so they tend to be marked with a zero particle.

## 5.4 Factors

Earlier in the present study, I stated that in conversational Korean, the choice of object encoding types would be related to factors such as negation, animacy, word order, definiteness, length, repair, verb adjacency, anaphoric saliency, cataphoric saliency, as well as contrast, which are related to processing load or informational prominence in one way or another. However, the use of these factors is by no means confined to either processing load or informational prominence as separate factors, but it may also be related to both factors concurrently. Throughout this dissertation, it was assumed that a deviation from the norm or the default pattern would indicate (an increase in) processing load, and that extra information to process, i.e., informationally loaded, would represent informational prominence. In the present section, I will discuss how these factors influence the speaker's selection of object markings in conversational Korean, and I will explore the choice of object marking with respect to *hata*-verb and split-case marking as well.

### 5.4.1 Negation

As was noted in the earlier part of this dissertation, there are two major types of negation in Korean (Kim 2000a). One is negation by a morphological negative (*an* 'not'), and the other is negation by a negative auxiliary (*ci anhta* 'be not') (Kim 2000a). The former is often called preverbal or short form negation, and the latter is called post-verbal or long form negation (Sohn 1999; Song 2005). (5.4) and (5.5) are examples of preverbal negation and post-verbal negation, respectively, from the data of the present



study. In addition to these two types of negation, there are other ways of expressing negativity in Korean (Sohn 1999). For instance, lexical items such as adverbs or negative predicates can also encode the negativity of a given utterance. (5.6) and (5.7), which are from the data of the present study, exemplify the sentences of which negative meanings are encoded by lexical items. More specifically, the utterance in (5.6) is negated by the adverb *mos* ‘not’, and the utterance in (5.7) is negated by the negative predicate *moluta* ‘not to know’. (4.7) and (4.8) are repeated as (5.6) and (5.7) here.

(5.4) *kyayney, mwo, hwaktap-ul an cwununtey.*  
 they DP confirmation-ACC NEG give  
 ‘Well, they don’t give an answer’

(5.5) *kulayse na-nun, kulikwu na-nun eyichiothi-lul cohaha-ci*  
 so 1SG-TOP and 1SG-TOP H.O.T-ACC like-COMP  
*anh-ass-ki ttaymwuney.*  
 NEG-PST-NMLZ because  
 ‘So I, and because I didn’t like H.O.T’

(5.6) *ne-hanthey cenhwa-lul mos hayssta.*  
 2SG-DAT phone-ACC not did  
 ‘(I) couldn’t call you,’

(5.7) *ku taumey ttalun ke-n molukeyssko,*  
 that next other thing-TOP not.know  
 ‘And then, (I) don’t know other things,’

As noted earlier in this dissertation, it is assumed that in general, negative constructions are conceptually more complex than their affirmative counterparts. It can further be assumed that due to conceptual complexity, negativity generally increases processing load, which would invite the use of an explicit post-nominal particle for coding object NPs. That is to say, processing load is expected to be higher for objects appearing in negative sentences than for those occurring in their affirmative counterparts, and the use of an overt marking is more common for coding objects of a negative sentence than

it is for that of an affirmative sentence. As a matter of fact, many speech errors occur in utterances involving the negation (Jaeger 2005). Yet, unlike this assumption, it was pointed out in Section 4.5.1 that in conversational Korean, the negativity of a sentence does not call for an overt particle (e.g., *ka*) for marking subject NPs. In other words, the negativity of sentences does not seem to influence the choice of subject encoding types.

In the present section, I explore the choice of object markings with relation to the negative of sentences in which objects occur, based on the assumption that the use of an overt particle (e.g., *lul*) would be more common for objects appearing in negative sentences than it is for those appearing in affirmative counterparts. For this reason, as I did with subject NPs, I divided all clausal units into two different sub-categories based on their polarity: negative and affirmative. Then, I examined object markings with regard to the negativity of the sentence in which objects occur, in order to see if there is a correlation between the selection of object markings and the negativity of the sentence. The distributional patterns of object markings with respect to the negativity of the sentence are summarized in Table 5.4.

Table 5.4: Negation and Object Marking

OBJ.MARK		NP-Ø	NP-lul	NP-nun	NP-to	NP-etc.	SUB-TOTAL	TOTAL
NEGATION								
affirmative	NO.	1,518	628	82	152	34	896	2,414
	%	<b>63</b>	<b>26</b>	<b>3</b>	<b>6</b>	1	<b>37</b>	100
negative	NO.	170	71	28	48	0	147	317
	%	<b>54</b>	<b>22</b>	<b>9</b>	<b>15</b>	0	<b>46</b>	100
TOTAL	NO.	1,688	699	110	200	34	1,043	2,731
	%	62	26	4	7	1	38	100

In Table 5.4, I find that overall, the occurrence of a zero particle is higher for coding object NPs than the occurrence of an overt particle, regardless of the negativity

of the sentences. Yet the use of zero-marking is less common for coding object NPs in negative sentences (54%) than for coding objects in affirmative sentences (63%). Conversely, the use of an overt marking is more frequent with object NPs in negative sentences (46%) than it is with object NPs in affirmative counterparts (37%) (zero-marking vs. overt marking:  $\chi^2=10.168$ ,  $P < 0.01$ ). This tendency suggests that the negativity of sentences influences the speaker's selection of one particular object form over another. The following is another example of an object that occurs in a negative sentence.

(5.8) *kyoswu-ka mwuleponuntey, amwu-to [taytap]-ul mos hanunkeya.*  
 professor-NOM ask none-also answer-ACC not do  
 'The professor asked a question, but no one answered the question.'

The utterance in (5.8) is viewed as a sentence negated by a negative adverb, *mos* 'not', which is underlined, and it is assumed to be conceptually complex because of the negativity encoded in the utterance. Due to conceptual complexity, which represents processing load, the object in (5.8), *taytap* 'answer', is explicitly marked with the accusative *lul*, which I argue would alleviate the processing load.

However, despite the overall tendency that an explicit marking is more commonly used for encoding objects in negative sentences than for those in affirmative counterparts, the percentage of *lul*-marked object NPs is slightly lower in negative sentences (22%) than it is in affirmative counterparts (26%). However, the difference is not statistically significant.

In Table 5.4, I also find that the percentage of *nun*-marked objects is greater in negative sentences than in their affirmative counterparts (zero-marking vs. *nun*-marking:  $\chi^2=24.939$ ,  $P < 0.001$ ; *lul*-marking vs. *nun*-marking:  $\chi^2=20.708$ ,  $P < 0.001$ ). I attribute this to the assumption that an entity in a focus of negation usually contrasts with another entity, so the use of *nun*-marking, which denotes the contrastiveness (Choi 1997),

is more common for objects in a negative sentence (9%) than in an affirmative counterpart (3%). In fact, I argue that the usage of *nun* is what makes negative sentences differ from their affirmative counterparts. Let us see the utterance labeled A3 in (5.9), which exemplifies a *nun*-marked object in a negative sentence.

(5.9)

A1: *mwo-Ø hay-ss-e?*  
 what-( ) do-PST-Q  
 ‘What did you do?’

A2: *kwuntay-eyse kumyen kongpwu-Ø hay-ss-e?*  
 military-LOC then study-( ) do-PST-Q  
 ‘Then, did you study while in military service?’

B1: *kwuntay-eyse mwulken-Ø phal-ass-cyo.*  
 military-LOC things-( ) sell-PST-SEM  
 ‘My duty was to sell things.’

B2: *[ppang]-Ø phal-ass-cyo.*  
 pastry-( ) sell-PST-SEM  
 ‘I was mainly selling pastries.’

A3: *[kongpwu]-nun an hay-ss-ci?*  
 study-TOP NEG do-PST-Q  
 ‘You didn’t study, did you?’

In the utterance A3 of (5.9), the object in brackets, *kongpwu* ‘study’, is in contrast to the bracketed object in the utterance B2, *ppang* ‘pastry’, and it is coded with *nun*, which marks the contrastiveness.

Lastly, it is notable that the occurrence of *to*-marking is higher in negative sentences (15%) than in affirmative sentences (6%) (zero-marking vs. *to*-marking:  $\chi^2=33.968$ ,  $P < 0.001$ ). I attribute this to the fact that the post-nominal particle *to* ‘even’ in Korean

is a negative polarity item (NPI), and it is commonly used to code objects in negative sentences (Sells 2006). Let us see the utterance labeled A2 in (5.10), which exemplifies a *to*-marked object in a negative sentence.

(5.10)

A1: *swul*-Ø      *cohahakin*, *cohahay-ss-nuntey*,  
 alcohol-( ) like              like-PST-though  
 ‘Although (I) used to like an alcohol,’

A2: *a*, [*swul*]-*to*      *mos* *mekkey-ss-e*.  
 ah alcohol-even not drink-PST-SEM  
 ‘Ah, (I) can no longer drink an alcohol.’

In the utterance A2 of (5.10), the object in brackets, *swul* ‘alcohol’, appears in the negative sentence, and it is coded with the particle *to* ‘even’, which often functions as an NPI.

To summarize the present section, regardless of the negativity of sentences, a zero particle is more commonly used for marking objects than an overt particle (see Table 5.4). Nevertheless, the use of zero-marking is less common for objects in negative sentences (54%) than for objects in affirmative sentences (63%). Conversely, the use of an overt-marking is more common for objects in negative sentences (46%) than for objects in affirmative counterparts (37%). In short, objects in negative sentences tend to be coded with the overt marking, whereas objects in affirmative counterparts tend to be coded with zero-marking, due to processing load incurred by the negativity. Yet, unlike an initial expectation, the use of the accusative *lul* is less common for objects in negative sentences than for objects in affirmative sentences. Conversely, the use of *nun*-marking is more common for objects in negative sentences than for objects in affirmative counterparts, and I attributed this to the assumption that an entity in the focus of negation often contrasts with another entity, so it is commonly coded with

*nun*-marking, which denotes contrastiveness (Choi 1997). As a matter of fact, what makes negative sentences differ from their affirmative counterparts is the use of *nun*. Lastly, the use of the particle *to* ‘even’ is common for coding objects that appear in negative sentences.

### 5.4.2 Animacy

As was noted earlier, the notion of ‘animacy’ has often been introduced into the discussion of various issues (e.g., Keenan 1976; Hopper & Thompson 1980; Comrie 1989; Dixon 1994; Givón 2000a; 2000b; Fry 2003, among others). As was noted in Chapter 4, in the present study, the animacy of NPs is determined not based on morphological form (i.e., signifier), but rather on the referents of the NPs in question (i.e., signified).<sup>5</sup> The followings exemplify object NPs with animate and inanimate referents from the data in the present study.

(5.11) *ccacangmyen-un an mek-ko kunyang pap-Ø mek-ess-ta.*  
 blacknoodle-TOP NEG eat-and just rice-( ) eat-PST-DCL  
 ‘I didn’t eat the blacknoodle, but just ate the rice.’

(5.12) *nay-ka mwullon ssakaci-Ø eps-nun ay-lul cohahakin haciman*  
 1SG-NOM of.course manners-( ) lack-REL guy-ACC like though  
 ‘Of course, I like the guy who is ill-mannered, but ...’

(5.13) *chicu-lang chokholeys-ul ilehkey nwatwu-ko cwi-lul*  
 cheese-with chocolate-ACC like.this put-and mouse-ACC  
*phwul-ess-ta.*  
 release-PST-DCL  
 ‘(They) put cheese and chocolate (on the floor), and released the mouse.’

The utterance in (5.11) is an instance showing an object that represents an inanimate referent, in that the underlined objects, *ccacangmyen* ‘black noodle’ and *pap* ‘rice meal’,

<sup>5</sup>See Section 4.5.2 in Chapter 4 for further description of the animacy of NPs provided for the purpose of this dissertation.

are both inanimate. In (5.11), the first object is marked with *nun*, and the second one is coded with a zero particle. The utterance in (5.12) exemplifies an object that represents an animate, more specifically a human, referent because the underlined object *ay* ‘guy’ is a human-being, and it is coded with the accusative *lul*. Lastly, the utterance in (5.13) is also an example of an object that represents an animate, yet not human, referent in that the underlined object *cwi* ‘mouse’ is an animal, and it is overtly coded with the accusative *lul*.

As previously stated, it is a universal tendency that referents of subject NPs are generally animate, and the ellipsis of a subject-marking particle occurs more commonly with subjects that are categorized as animate (Thompson & Hopper 2001; Fry 2003). In Section 4.5.2, it was stated that there was no correlation observed between the choice of subject markings and the simple animacy of subject NPs. However, it was also pointed out that the selection of subject markings correlated with the animacy of subjects and objects when the animacy was categorized in a more precise way.

Unlike referents of subject NPs, referents of objects<sup>6</sup> are not likely to be animate, and the thematic role of objects is usually a patient or theme (Givón 2000a; Saeed 2003; Van Valin 2003). Hence, it is plausible to assume that animate referents may increase processing load, in a sense that an object representing an animate referent is considered a deviation from the norm for an object. Due to processing load, the use of *lul*-marking is expected to be more common for coding objects that represent an animate referent, so as not to be mistaken for subjects. As a matter of fact, it was previously pointed out in Fry (2003) that one discourse function of the object marking particle in Japanese, namely the accusative *o*, is to disambiguate an object from a subject. That is to say, referents of object NPs are usually inanimate, so object NPs representing an animate referent are commonly coded with *o*-marking for the sake of disambiguation.

---

<sup>6</sup>In the present study, an object refers to a direct object.

In the present section, in order to see if there is a correlation between the animacy and the choice of object markings, I examine the speaker’s selection of one particular object form over another with relation to the animacy of the object NPs, as well as that of subject NPs. The distributional pattern for object markings with respect to object animacy is summarized in Table 5.5.

Table 5.5: Object Animacy and Object Marking

ANIMACY	OBJ.MARK	NP-Ø	NP-lul	NP-nun	NP-to	NP-etc.	SUB-TOTAL	TOTAL
	inanimate	NO.	1,576	636	103	185	31	955
	%	<b>62</b>	<b>25</b>	4	7	1	38	100
anim./-human	NO.	2	5	0	8	0	13	15
	%	<b>13</b>	<b>33</b>	0	53	0	87	100
human	NO.	110	58	7	7	3	75	185
	%	<b>59</b>	<b>31</b>	4	4	2	41	100
TOTAL	NO.	1,688	699	110	200	34	1,043	2,731
	%	62	26	4	7	3	38	100

In Table 5.5, I find that overall, the number of object NPs that represent an animate, either human or non-human, referent (200 tokens, approximately 7% of the entire body of object NPs) is far smaller than that of object NPs that represent an inanimate referent (2,531 tokens, approximately 93% of all object NPs). This distributional tendency strongly supports what has been pointed out as to the animacy of object NPs, that is, the referents of object NPs are not likely to be animate (Givón 2000a; Saeed 2003; Van Valin 2003).

More importantly, the data in Table 5.5 shows that the occurrence of *lul*-marking is higher for coding objects that represent an animate referent (33% for objects representing a non-human animate referent; 31% for objects representing a human referent) than for coding those that represent an inanimate referent (25%). Conversely, the use of zero-marking is less common for coding objects that represent a human and non-



human animate referent than it is for coding those that represent an inanimate referent (zero-marking vs. *lul*-marking:  $\chi^2=4.114$ ,  $P < 0.05$ ). I attribute this tendency to the fact that referents of object NPs are not likely to be animate (Hopper & Thompson 1980; Givón 2000a), so objects that represent animate are considered a deviation from the norm, which I argue would indicate processing load. Due to processing load, the use of *lul*-marking is more common for objects that represent an animate referent than for those that represent an inanimate referent, as exemplified in (5.12) and (5.13).

The data in Table 5.5 also shows that the use of zero-marking is much less common for objects that represent a non-human animate referent (13%) than for those that represent an inanimate referent (62%), while the use of *lul*-marking is more common for objects representing a non-human referent (33%) than for those representing an inanimate referent (25%). Yet the number of objects representing a non-human animate referent is too small to examine statistical significance.

Interestingly, the percentage of zero-marking for objects representing a human referent (59%) is only slightly lower than that for objects representing an inanimate referent (62%). Furthermore, the use of *lul*-marking for coding objects that represent a human referent (31%) is not much higher than that for coding objects that represent an inanimate referent (25%) (zero-marking vs. *lul*-marking:  $\chi^2=2.518$ ,  $P=0.1126$ ).

In summary, the use of the accusative *lul* is more common for marking objects that represent a human or non-human animate referent than for marking those that represent an inanimate referent, and conversely, the use of zero-marking is less common for objects representing an animate referent than for those representing an inanimate referent. This tendency, which is statistically significant, complies with what has initially been assumed.

In an attempt to more closely examine the relationship between the animacy of an object NP and its marking, I categorized all object NPs whose referents are human-

beings into two sub-types: object NPs representing an interlocutor referent and object NPs representing a non-interlocutor referent. Then, I examined the choice of object markings based on the two subtypes, and the results are summarized in Table 5.6.

Table 5.6: Object Interlocutor and Object Marking

OBJ.MARK		NP-Ø	NP-lul	NP-nun	NP-to	NP-etc.	SUB-TOTAL	TOTAL
OBJ.INTERLOCUTOR								
-interlocutor	NO.	97	39	6	6	3	54	151
	%	<b>64</b>	<b>26</b>	4	4	2	36	100
interlocutor	NO.	13	19	1	1	0	21	34
	%	<b>38</b>	<b>56</b>	3	3	0	62	100
TOTAL	NO.	110	58	7	7	3	75	185
	%	59	31	4	4	2	41	100

In Table 5.6, I find that object NPs representing an interlocutor (i.e., the speaker or the hearer) referent (34 tokens, roughly 18% of all objects representing a human referent) are much less common than those representing a non-interlocutor human referent (151 tokens, 82% of all objects representing a human referent). Furthermore, the use of *lul*-marking (56%) is more common for objects that represent an interlocutor than the use of zero-marking (38%).

More importantly, the occurrence of *lul*-marking is noticeably higher for coding object NPs representing an interlocutor referent (56%) than it is for coding object NPs representing a non-interlocutor human referent (26%). Conversely, the occurrence of zero-marking is greatly lower for coding object NPs representing an interlocutor referent (38%) than it is for coding object NPs representing a non-interlocutor human referent (64%) (zero-marking vs. *lul*-marking:  $\chi^2=10.8$ ,  $P < 0.001$ ). Simply put, the choice of object markings correlates with whether object referents are an interlocutor or a non-interlocutor referent, though it did not correlate with whether object referents are inanimate or a mere human referent (zero-marking vs. *lul*-marking:  $\chi^2=2.518$ ,

P=0.1126). This tendency may be attributed to the fact that an interlocutor (i.e., the speaker or the hearer) tends not to occur as a referent of an object vis-a-vis as a referent of a subject, as shown in Table 5.6. That is to say, objects representing an interlocutor referent are considered a deviation from the norm, which suggests increased processing load, so the use of *lul*-marking, which I argue would alleviate the processing load, is common for coding them. The following exemplifies an object whose referent is an interlocutor.

(5.14) *caki-ka*    *na-l*            *antey*  
           self-NOM 1SG-ACC know  
           ‘He said he knows me.’

In (5.14), the underlined object NP, *na* ‘me’, is an interlocutor, more specifically the speaker, and it is marked with the accusative *lul*. As repeatedly stated, it is not common that an interlocutor, either the speaker or the hearer, is realized as a referent of an object NP, and the object representing an interlocutor referent is considered a deviation from the norm, which may represent processing load. Due to processing load, it is coded with *lul*-marking, which I argue would alleviate the processing load.

As repeatedly noted, it is a universal tendency that referents of object NPs are generally inanimate, while those of subject NPs are animate (Hopper & Thompson 1980; Givón 2000a). Hence, it is plausible to assume that the choice of object markings may be related to the animacy of subject NPs in the sentence.

More specifically, the use of *lul*-marking may be more common for objects when the referents of subjects are inanimate rather than when they are animate because it is considered a deviation from the norm when subject NPs and object NPs both represent inanimate referents, hence an increase in processing load, resulting in the tendency for objects to be overtly identified. For this reason, I also examined the choice of object markings with regard to the animacy of subject NPs, and the results are presented in

Table 5.7.

Table 5.7: Subject Animacy and Object marking

OBJ.MARK SUBJ.ANIMACY		NP-Ø	NP-lul	NP-nun	NP-to	NP-etc.	SUB- TOTAL	TOTAL
animate	NO.	464	197	28	37	9	271	735
	%	<b>63</b>	<b>27</b>	4	5	1	37	100
inanimate	NO.	33	18	0	5	1	24	57
	%	<b>58</b>	<b>32</b>	0	9	2	42	100
TOTAL	NO.	497	215	28	42	10	295	792
	%	<b>63</b>	<b>27</b>	4	5	1	37	100

In Table 5.7, I find that the use of *lul*-marking is a little more common for object NPs when referents of subject NPs are inanimate (32%) than when they are animate (27%). Contrastively, the use of zero-marking is a little less common for objects when subjects represent inanimate referents (58%) than when they represent animate referents (63%). However, the difference is not statistically significant (zero-marking vs. *lul*-marking:  $\chi^2=0.677$ ,  $P=0.4105$ ).

To summarize this section, the present study examines the choice of object markings with relation to the animacy of object referents as well as subject referents. According to the data in Table 5.5, the choice of object markings correlates with whether objects represent an animate referent or inanimate referent, in that the use of *lul*-marking is more common for objects representing an animate referent than for those representing an inanimate referent. I attribute a higher occurrence of *lul*-marking for objects that represent an animate referent to the assumption that animate objects are considered a deviation from the norm, hence processing load, so the use of *lul*-marking, which would alleviate this processing load, is expected to be common, as exemplified in (5.12) and (5.13). The data in Table 5.5 also shows that the occurrence of zero-marking is low for objects that represent a non-human animate referent (13%), and the occurrence of

*lul*-marking is relatively high for coding them (33%).

In Table 5.6, it is shown that the use of *lul*-marking is more common for object NPs representing an interlocutor referent (56%) than it is for those representing a non-interlocutor human referent (26%), and the use of zero-marking is less common for objects representing an interlocutor referent (38%) than it is for those representing a non-interlocutor human referent (64%). I attribute this tendency to processing load incurred from a deviation from the norm. That is to say, an interlocutor (i.e., the speaker or the hearer) is not likely to occur as a referent of an object vis-a-vis as a referent of a subject, so objects that represent an interlocutor referent are viewed as a deviation from the norm, which may represent processing load. Therefore, the occurrence of *lul*-marking, which would alleviate the processing load, is expected to be high for coding them, as exemplified in (5.14).

Furthermore, according to the data in Table 5.7, the choice of object markings does not correlate with the animacy of subject NPs. In other words, a speaker's selection of one particular object form over another is not influenced by the animacy of subject referents.

Lastly, the distributional pattern observed for the accusative *lul* in Korean is similar to what was reported for the accusative *o* in Japanese (Fry 2003), in that both particles are commonly used for objects that represent an animate referent, thus are considered a deviation from the norm.

In short, it seems that the occurrence of *lul*-marking is (relatively) high for an object when the object is considered a deviation from the norm, hence processing load is assumed. Contrastively, the occurrence of zero-marking seems to be (relatively) high for an object when the object is not viewed as a deviation from the norm. Based on the tendencies reported in this study, I claim that one discourse-pragmatic function of the accusative *lul* in conversational Korean is to alleviate the processing load encoded

in object NPs.

### 5.4.3 Word order

Korean, like Japanese, is known as a predicate-final language, which allows a pre-predicative scrambling (Kuno 1972; Cho 1994; Choi 1996). Yet not all clauses found in the data of this study end with a predicate (e.g., VSO). As was stated earlier in the present study, whether or not post-predicative entities should be viewed as parts of the planned utterance has been a widely discussed issue. For instance, some (e.g., Kuno 1973; 1978) consider only pre-predicative entities relevant for the examination of word order of a sentence, and they preclude any post-predicative entities treating them as ‘after-thoughts’. Contrastively, others (e.g., Shimojo 2005) distinguish instances of postposed arguments from those of simple ‘after-thoughts’, and include the instances of postposed arguments when the word order of a sentence is examined. In fact, Shimojo (2005) claims that arguments are postposed for a discourse function such as defocusing. The latter approach appears to be more convincing, and it is taken in the present study: when the word order of a given sentence is examined, instances of postposed arguments are included, while those of ‘after-thoughts’ are excluded.<sup>7</sup>

The word order in Korean has been claimed to be SOV (Sohn 1999), but such word order is not always observed, particularly in spoken Korean (Ko. S 2000; Kim 2006). As a matter of fact, the data from the present study confirms Kim’s (2006) claim, in that although SOV and OV are the two most commonly observed word orderings in transitive clauses, there are other word orderings reported (e.g., OSV) as well.<sup>8</sup> In subsequent texts are provided examples that show various word orderings from the data

---

<sup>7</sup>In the present study, postposed arguments are distinguished from those that are after-thoughts based on the existence or non-existence of a significant pause between the predicate and the post-predicative entity. See Section 3.3.1 for detail.

<sup>8</sup>In my data, SOV and OV together comprise of approximately 96% of all transitive clauses, excluding clausal units with unexpressed objects (see Table 5.8).

of the present study.<sup>9</sup>

First, following are examples that show the characteristics of a predicate-final language; (5.15) through (5.19) are examples which all end with a predicate.

- (5.15) *na-Ø tto mak kwaynhi capwusim-Ø kacikwu mak*  
1SG-( ) again DP just pride-( ) have DP  
'I took a pride, you know.'

The utterance in (5.15) exemplifies SOV word ordering, and the object, *capwusim* 'pride', that is underlined, is marked with a zero particle.

(5.16)

- A1: *pyello an kellye-ss-e, sikan-Ø*  
much NEG take-PST-SEM time-( )  
'It didn't take long.'
- A2: *iksuphuleysu-Ø tha-ko ileni-kka*  
express-( ) ride-and be.so-because  
'Because I took an express (subway line),'
- A3: *han 30 pwun-to an kellye-ss-e.*  
about 30 minute-even NEG take-PST-SEM  
'It didn't take even 30 minutes.'

The utterance A2 in (5.16) exemplifies an OV word ordering, and the underlined object, *iksuphuleysu* 'express subway line', is coded with a zero particle.

- (5.17) *amwuthun enu kenmwul-uy kipwukum-ul nwuka cwo-ss-nyamyen*  
anyway certain building-GEN donation-ACC who.NOM give-PST-SEM  
'Anyway, if you wonder who gives a donation to a certain building.'

- (5.18) *thekhi-Ø ilehkey ttal-i ilehkey cal hatelakwu*  
turkey-( ) like.this daughter-SM like.this well do  
'his daughter cooked the turkey well, you know.'

---

<sup>9</sup>As repeatedly stated, the present study considers only those clauses containing an overt object for the word order examination of transitive clauses.

- (5.19) yehayng-un [nay]-ka kunyang an hal sayngkakiya.  
 travel-TOP 1SG-NOM just NEG do thought.  
 ‘As for traveling, I am not going to do it’

Lastly, (5.17), (5.18) and (5.19) are examples that show OSV word ordering. The underlined object in (5.18), *kipwukum* ‘donation’, is coded with the accusative *lul*, and the object in (5.17), *thekhi* ‘turkey’, that is underlined, is marked with a zero particle. In (5.19), the underlined object, *yehayng* ‘travel’, is topicalized with *nun*, and precedes the bracketed subject, *nay* ‘I’.

Second, as was stated above, there are also many sentences which do not end with a predicate; (5.20) and (5.25) are examples that do not end with a predicate.

- (5.20) pwutheisse. kulem nwuka salccak tul-e, congi-lul.  
 be.stuck then someone.NOM softly lift-SEM paper-ACC  
 ‘(A mouse) is stuck there. Then someone softly lifts a paper.’

- (5.21) ani kuntey, enni-Ø weynci an cohaha-lkekathay,  
 not well sister-( ) well NEG like-feel  
kopchangcenkol-kathun ke-Ø  
 tripe.stew-like thing-( )  
 ‘Well, you know, you seem not to like something like a tripe stew.’

The utterances in (5.20) and (5.21) exemplify a word ordering of SVO, and both objects are postposed into a position after the predicate. The underlined object in (5.20), *congi* ‘paper’, is coded with the accusative *lul*, while the underlined object in (5.21), *kopchangcenkol-kathun ke* ‘something like tripe’, is marked with a zero particle.

- (5.22) kyepkyepi ip-e, ilehkey yalpun ke?  
 tightly wear-Q like.this thin thing  
 ‘Do (I) need to wear several of clothes that are this thin?’

- (5.23) ettehkey chac-a, kuke-nun?  
 how find-Q that-TOP  
 ‘How do I find that?’



The utterances in (5.22) and (5.23) are viewed as examples of VO word ordering that appear without an overt subject. In (5.22), the object, *ilehkey yalpun ke* ‘clothes like this thin’, that is underlined, is coded with a zero particle, while in (5.23), the underlined object, *kuke* ‘that thing’, is coded with *nun*-marking.

(5.24) *yulichang-eyta ilehkey nakse-lul hay nonta, salam-tul-i.*  
 window-on like.this graffiti-ACC do put person-PL-NOM  
 ‘People graffiti on the window, you know.’

A word ordering of OVS is exemplified in the utterance in (5.24), where the underlined object, *nakse* ‘graffiti’, is coded with the accusative *lul*.

(5.25)

A1: *na-Ø yaysnal-ey ilen ke han pen haypwa-ss-e.*  
 1SG-( ) old.day-in this thing once NC do-PST-SEM  
 ‘A while ago, I had a chance to do something like this.’

B1: *haypwa-ss-e?*  
 do-PST-Q  
 ‘Did you?’

A2: *eh, haypwa-ss-e, na-Ø ilen ke-Ø*  
 yeah do-PST-SEM 1SG-( ) this thing-( )  
 ‘Yeah, I had a chance to do something like this before.’

Lastly, the utterance of A2 in (5.25) exemplifies a VSO word ordering, and the underlined object, *ilen ke* ‘something like this’, is marked with a zero particle.

As was noted earlier in Chapter 4, the present study examines the word order of sentences in two respects: processing load and informational prominence. More specifically, processing load is expected for object NPs when they occur in a non-canonical word order variation, which is regarded as a deviation from the norm. It is further assumed that the processing load would invite the use of *lul*-marking, which

would alleviate the processing load, for coding objects that occur in a non-canonical, thus marked, word ordering. Yanagida (1985) makes a similar claim, based on Givón's (1983) 'degree of predictability', that when there is a potential for increased difficulty (e.g., a non-canonical word order variation), the speaker tends to use a construction type (e.g., arguments with an explicit post-nominal particle) that perceptually demands more attention and eventually results in easier processing. On the other hand, in general, entities are postposed for the sake of defocusing, and postposed entities are not thematically important (Simon 1989; Shimojo 2005). As a matter of fact, Shimojo (2005:203) points out that in colloquial Japanese, zero-marking is the encoding type that fits best for postposed arguments. In this section, I examine if a similar tendency is observed for objects in conversational Korean, although in Section 4.5.3, it was stated that unlike Japanese, in Korean, post-predicative subjects do not necessarily represent unimportant information.

In Japanese, some studies (Shimojo 2005, *inter alia*) have been done discussing a relationship between word orderings and the choice of object markings, yet in Korean, to the best of my knowledge, there have been no studies attempting to examine a speaker's selection of object markings with regard to word order variations. For this reason, with an assumption that the choice of object markings is related to word order variations, I examine the object markings with regard to word order variations. More specifically, processing load is expected for objects that appear in a non-canonical word ordering, and the use of *lul*-marking, which would alleviate the processing load, is expected to be common for coding objects. The use of zero-marking is assumed to be common for coding post-predicative objects, which are considered not to represent informational prominence. The distributional patterns of object encoding types with relation to word order variations are shown in Table 5.8.

— **OV, SOV and OVS:**

Table 5.8: Word Order Variations and Object Marking

OBJ.MARK WORD ORDER		NP-Ø	NP-lul	NP-nun	NP-to	NP-etc.	SUB- TOTAL	TOTAL
OV	NO.	1,179	477	80	158	24	739	1,918
	%	<b>61</b>	25	4	8	1	39	100
SOV	NO.	457	192	20	38	8	258	715
	%	<b>64</b>	27	3	5	1	36	100
OSV	NO.	22	11	6	3	2	22	44
	%	<b>50</b>	25	<b>14</b>	7	5	50	100
VO	NO.	12	7	2	0	0	9	21
	%	57	<b>33</b>	<b>10</b>	0	0	43	100
OVS	NO.	11	6	0	1	0	7	18
	%	<b>61</b>	<b>33</b>	0	6	0	39	100
SVO	NO.	6	6	2	0	0	8	14
	%	<b>43</b>	<b>43</b>	<b>14</b>	0	0	57	100
VSO	NO.	1	0	0	0	0	0	1
	%	100	0	0	0	0	0	100
TOTAL	NO.	1,688	699	110	200	34	1,043	2,731
	%	62	26	4	7	1	38	100

In Table 5.8, I find that most object NPs (2,651 tokens, approximately 97% of the total) immediately precede the predicate in a sentence. That is to say, most objects appear in OV, SOV or OVS word orderings. This strongly suggests that the position immediately preceding the predicate in a sentence is a canonical/unmarked position for object NPs in conversational Korean. More importantly, zero-marking is the most commonly used encoding type for object NPs that occur in such a canonical position in a sentence.<sup>10</sup> In other words, the occurrence of zero-marking is preferred for coding object NPs that appear in a position immediately preceding the predicate (e.g., 64% in SOV). Contrastively, the use of *lul*-marking is generally not common for coding these objects (e.g., 25% in OV). For instance, the object in (5.15), *capwusim* ‘pride’, immediately

<sup>10</sup>The relationship between verb adjacency and object markings is discussed in the following section.

precedes the predicate, *kacita* ‘have’, and it is coded with a zero particle. In (5.16), the object, *iksuhphuleysu* ‘express subway line’, precedes the predicate, *thata* ‘ride’, and it is also marked with a zero particle. I attribute this tendency to the assumption that an entity immediately preceding the predicate in the clause is normally identified as an object, hence minimal processing load in object identification, so the use of zero-marking is common for coding object NPs that immediately precede the predicate.

As stated above, the occurrence of *lul*-marking is relatively low for object NPs immediately preceding the predicate, but interestingly, it is a little more common for object NPs in OVS word ordering (33%) than for object NPs in OV (25%) or SOV (27%) word order variations. For instance, in (5.24), the object *nakse* ‘graffiti’, although it immediately precedes the predicate *hata* ‘do’, occurs in a marked word ordering, i.e., OVS, which may represent processing load, so it is coded with the accusative *lul*. However, the occurrence of zero-marking is nearly equal for object NPs in all variations. In fact, this tendency is not statistically significant (objects in OVS vs. objects in SOV:  $\chi^2=0.259$ ,  $P=0.6111$ ; objects in OVS vs. objects in OV:  $\chi^2=0.345$ ,  $P=0.5569$ ).

— **OSV:**

In Table 5.8, I also find that the occurrence of zero-marking is lower for coding object NPs appearing in OSV (50%), compared to that of other variations of word ordering (e.g., 64% in SOV). For instance, in (5.17), the object, *kipwukum* ‘donation’, is marked with the accusative *lul*. Yet this tendency does not hold statistical significance (zero-marking vs. *lul*-marking:  $\chi^2=0.211$ ,  $P=0.6459$ ).

Interestingly, according to the data in Table 5.8, the occurrence of *lul*-marking appears to be lowest for coding objects in OSV word ordering (25%), among all other word orderings (e.g., 43% for object NPs in SVO). For example, in (5.18), the object, *thekhi* ‘turkey’, is marked with a zero particle. One may find this tendency somewhat

strange, in the sense that due to the processing load incurred by marked word ordering, one would expect the occurrence of *lul*-marking to be relatively high for coding object NPs in OSV word ordering. I attribute the relatively low occurrence of *lul*-marking for objects in OSV word ordering to the relatively high occurrence of *ka*-marking for encoding subject NPs in OSV word ordering (see Table 4.8). Simply put, it may be enough to overtly mark only one of either the subject or object in a sentence for the purpose of argument identification. Still, one may wonder why subject NPs, but not object NPs, tend to be coded with an overt particle in OSV word ordering. I propose that the relatively high occurrence of *nun*-marked object NPs may be a reason for the relatively low occurrence of *lul*-marked objects. More specifically, a sentence-initial position is generally viewed as a primary topic position (Givón 2000a), so the occurrence of *nun*-marking is expected to be higher for object NPs appearing in a sentence-initial position (e.g., 14% in OSV) than for those appearing not in a sentence-initial position (e.g., 3% in SOV) (*lul*-marking vs. *nun*-marking:  $\chi^2=10.457$ ,  $P < 0.01$ ). For instance, the object in (5.19), *yehayng* ‘travel’, which appears in a sentence-initial position, is topicalized with *nun*. Nonetheless, the use of *nun*-marking for object NPs does not distinguish objects from subjects since it may be used to mark subject NPs as well. Therefore, in OSV word ordering, subject NPs are frequently coded with *ka*-marking for the sake of argument identification, while object NPs are coded with *nun*-marking.

— VO:

According to the data in Table 5.8, in VO word ordering, the occurrence of *lul*-marked object NPs is slightly higher (33%), compared to that of other variations of word ordering (e.g., 25% in OV). Conversely, the occurrence of zero-marked object NPs is a little lower (57%), compared to that of other word order variations (e.g., 61% in OV). In Korean, a post-predicative position is not viewed as a canonical or unmarked

position for objects, so processing load is expected. Thus, *lul*-marking is expected to be relatively high and the occurrence of zero-marking is to relatively low. However, this tendency is not statistically significant (zero-marking vs. *lul*-marking for objects in VO and OV:  $\chi^2=0.591$ ,  $P=0.4422$ ).

Earlier in this section, it was noted that in spoken Japanese, zero-marking is the basic encoding type for postposed entities due to the lack of thematic prominence (Shimojo 2005). Yet, according to the data in Table 5.8, unlike Japanese, the zero-marking does not seem to be the basic encoding type for objects in Korean. In short, there is a striking difference between the two languages, in that in Japanese, post-predicative objects represent thematically unimportant information, whereas in Korean, they do not necessarily represent thematically unimportant information.<sup>11</sup>

Lastly, the occurrence of *nun*-marking is slightly higher for object NPs in VO word ordering (10%), compared to those in other word order variations (e.g., 4% in OV). As was noted for subject NPs in Section 4.5.3, I attribute this tendency to the property of grammatical topic; a sentence-final position, as well as a sentence-initial position, is considered a position for the grammatical topic, while a sentence-middle position is rarely a grammatical topic position. Thus, post-predicative objects are commonly coded with *nun*-marking, as was exemplified in (5.23). However, this tendency may not be of significance due to the insufficient number of tokens for *nun*-marked objects in VO word ordering.

— **SVO:**

In SVO word ordering, the occurrence of zero-marking is lower for coding object NPs (only 43%), compared to that of other variations of word ordering (e.g., 61% in OVS).

---

<sup>11</sup>Unlike in Japanese, the cataphoric saliency, either by RP or by RP-*f*, which is used to measure thematic prominence (Shimojo 2005), does not play a role in the choice of argument encoding types in conversational Korean.

Contrastively, the occurrence of *lul*-marking is much higher for coding object NPs in SVO (43%) than that of all other variations of word ordering (e.g., only 33% in OVS). As repeatedly stated, a post-predicative object is viewed as a deviation from the norm, which represents processing load, and the use of *lul*-marking is expected to be common for the post-predicative object, while the occurrence of zero-marking is relatively low for coding such an object. For instance, in (5.20), the object, *congi* ‘paper’, which is shown in a post-predicative position, is coded with the accusative *lul*. Yet this tendency is not statistically significant (zero-marking vs. *lul*-marking:  $\chi^2=0.627$ ,  $P=0.4284$ ).

Lastly, the occurrence of *nun*-marking is relatively high for object NPs in SVO word ordering, compared to that of other word order variations. As was noted before, specifically in Section 4.5.3, I associate this tendency with the property of grammatical topic; a post-predicative entity can be a topic of the sentence, which is marked with NUN in Korean.

— **VSO:**

There is only one instance of VSO word ordering reported in my data, so, due to the insufficient number of tokens, I cannot make any generalization out of this.

To summarize, in the present section, the choice of object encoding types was discussed with respect to word order variations in two different but related respects: processing load and informational prominence. According to the data in Table 5.8, the occurrence of zero-marking is higher for coding objects that immediately precede the predicate than those that do not immediately precede the predicate, due to easiness in object identification. Conversely, the occurrence of *lul*-marking is lower for objects immediately preceding the predicate than for those not immediately preceding the predicate. Yet this tendency does not appear to be statistically significant. Furthermore, the use of *lul*-marking is more common for coding objects that appear in non-canonical

word orderings (e.g., objects in a post-predicative position) than for those in a canonical word ordering (i.e., objects immediately preceding the predicate). However, this tendency is also shown to be not statistically significant. Lastly, the occurrence of *nun*-marking is relatively high for coding objects that appear in a sentence-initial or sentence-final position, whereas the use of *nun*-marking is uncommon for objects in a sentence-middle position. I attribute this tendency to the general property of grammatical topic, that is, entities in a sentence-initial or sentence-final position can be a grammatical topic marked with NUN in Korean.

In short, unlike subjects, in conversational Korean, a speaker's selection of one object encoding type over another does not seem to correlate with word order variations.

#### 5.4.4 Verb adjacency

It has been widely accepted that the word order in Korean is Subject+Object+Verb (SOV), with some variations allowed in spoken Korean (Sohn 1999). Based on this basic word order in Korean, it is assumed that objects are easily recognized when they are close, particularly immediately adjacent, to their predicates. It is further assumed that the easy recognition, hence minimal processing load, would invite the use of zero-marking for objects that are close to their predicates. As a matter of fact, it has also been reported in the previous studies of Japanese (Fry 2003, *inter alia*) that direct objects tend to drop their case marking particles when they are immediately adjacent to their predicates, though no justification for such a tendency has been provided. This tendency observed for direct objects in spoken Japanese has been confirmed by Tsutsui's (1984) and Matsuda's (1996) observations in the same regard. I attribute this tendency to the assumption that objects immediately adjacent to their predicates are easily recognized, thus processing load is assumed to be low, so the use of zero-marking is common for these objects.



I expect the similar, if not same, tendency to be observed for object NPs in conversational Korean. In order to find out whether or not verb adjacency influences the selection of object markings, I examined all object NPs with regard to their distance from their predicates. In the present study, the distance to a predicate is measured by counting the number of words that come between an object NP and its predicate. Following are examples of object NPs with and without words intervening between objects and their predicates. In (5.26) through (5.30), the objects and their predicates are underlined, and intervening words are in brackets.

(5.26) *na-Ø nyuyok-Ø ka-l ttay, pesu-Ø tha-ko ka-ss-e.*  
 1SG-( ) New.York-( ) go-rel time bus-( ) take-and go-PST-sem  
 ‘When I went to New York city, I took a bus to save some money.’

(5.27) *sensayngnim-i an ttaylye, ay-tul-ul.*  
 teacher-NOM NEG beat student-PL-ACC  
 ‘That teacher does not beat students.’

(5.28) *caki pwumonim-i cincca kyay-Ø [nemwu] cohahakwu*  
 self parents-NOM really him-( ) very like.and  
 ‘Her parents really like him a lot.’

(5.29) *ku enni-nun chotunghakkyo ttay mikwuk-ey wakackwu, ku*  
 that sister-TOP elementary.school time USA-LOC come.and that  
*cwupyen-ul [ta cal] anunkeya.*  
 surrounding.area-ACC all well know  
 ‘She came to USA when she was in an elementary school, so she was very familiar with the surrounding area.’

(5.30) *manyakey nay-ka hakwon-ul [keki nyuyok cwungsim-eyta] ete*  
 in.case 1SG-NOM institute-ACC there New.York center-LOC enroll  
*nohko,*  
 place.and  
 ‘In case I enroll in an (English) institute in the center of New York, and ...’

The utterances in (5.26) and (5.27) exemplify instances of no word intervening between the object and its predicate. In (5.26), the object, *pesu* ‘bus’, immediately

precedes the predicate, *thata* ‘ride’, and it is marked with a zero-particle. In (5.27), the object *aytul* ‘students’ appears in a post-predicative position, but it immediately follows the predicate, *ttaylita* ‘beat’. Therefore, this instance is also counted as an object with no intervening word. In (5.27), the object, *aytul* ‘students’, is marked with the accusative *lul*. The utterance in (5.28) exemplifies an object with one intervening word, in that the adverb *nemwu* ‘very’ in brackets intervenes between the object, which is marked with a zero particle, and the predicate. The utterance in (5.29) is an example with two intervening words because two words in brackets, *ta cal* ‘very well’, come between the object and its predicate. In (5.29), the object is marked with the accusative *lul*. Lastly, (5.30) is an instance that shows objects with three words intervening between the object and its predicate, and the object occurs with the particle *lul*. As noted above, the choice of object markings was examined with relation to their distance from the predicates, and the results are summarized in Table 5.9.

Table 5.9: Verb Adjacency and Object Marking

VERB	OBJ.MARK	NP-Ø	NP-lul	NP-nun	NP-to	NP-etc.	SUB-	TOTAL
	ADJACENCY						TOTAL	
zero	NO.	1,364	484	48	120	19	671	2,035
	%	<b>67</b>	<b>24</b>	2	6	1	33	100
one	NO.	253	146	44	64	9	263	516
	%	49	28	9	12	2	51	100
two	NO.	58	56	10	11	3	80	138
	%	<b>42</b>	<b>41</b>	7	8	2	58	100
three	NO.	9	9	7	3	2	21	30
	%	<b>30</b>	<b>30</b>	23	10	7	70	100
four or more	NO.	4	4	1	2	1	8	12
	%	33	33	8	17	8	67	100
TOTAL	NO.	1,688	699	110	200	34	1,043	2,731
	%	62	26	4	7	1	38	100

In Table 5.9, I observe that the choice of object markings is related to the distance from the predicates: The occurrence of *lul*-marked objects increases as the number of

words intervening between an object and its predicate increases, though the increase rate is not always gradual. Likewise, the occurrence of zero-marked objects decreases as the distance from an object to its predicate increases, though the decrease rate is not always gradual.

For instance, according to Table 5.9, the occurrence of *lul*-marked objects constitutes only 24 percent of the total when there is no word intervening between an object and its predicate. Yet it greatly increases to 41 percent when there are two intervening words, and to 30 percent when there are three intervening words. The distributional pattern pertaining to the use of zero-marking for coding objects shows an opposite tendency: When there is no intervening word, the occurrence of zero-marked objects takes up 67 percent of the total. Yet it drops to 58 percent when there are two intervening words (zero-marking vs. *lul*-marking:  $\chi^2=28.308$ ,  $P < 0.001$ ), and to 30 percent when there are three intervening words (zero-marking vs. *lul*-marking:  $\chi^2=5.198$ ,  $P=0.0226$ ). In fact, this tendency observed for objects in conversational Korean supports what has previously been claimed in the studies on particle ellipsis in spoken Japanese (Tsutsui 1984; Matsuda 1996; Fry 2003, *inter alia*), in that the occurrence of zero-marking is high for coding objects that are close to their predicates. The number of objects with four or more is too small to make anything of, so they are not discussed here.

As stated above, I attribute this tendency to the assumption that the closer an object is to its predicate, the easier it can be recognized. More specifically, an object that is closer to its predicate can easily be recognized, thus low processing load is assumed for such an object, so the occurrence of zero-marking is expected to be high for coding the object that is close to its predicate. On the other hand, an object that is in the distance from its predicate can only be recognized with processing load, so the occurrence of *lul*-marking is relatively high for such an object. For instance, the object in (5.26), *pesu* ‘bus’, is immediately adjacent to its predicate, *thako kata* ‘take

and go’, thus can easily be recognized, and due to low processing load, it is marked with a zero particle. The object in (5.30), *hakwon* ‘private academy’, is three words away from its predicate, *ete nohko* ‘find and register’, thus processing load is assumed for the objects, and due to the processing load, it is coded with the accusative *lul*, which I claim alleviates the processing load.

To summarize, just like in colloquial Japanese, in conversational Korean, a speaker’s selection of one particular object form over another seems to correlate with how close or how far the object is away from its predicate. That is to say, according to the data in Table 5.9, the occurrence of *lul*-marked objects gradually increases in proportional to the distance to the verb from the object. The occurrence of zero-marked objects proportionally decreases as the distance from the verb increases. I attribute this tendency to the assumption that in general, processing load would increase as the distance to the verb from an object becomes greater, and the processing load would invite the use of *lul*-marking for the object, which I argue would alleviate this processing load. Furthermore, this tendency observed for objects in conversational Korean complies with the previous finding for the particle ellipsis in spoken Japanese (Tsutsui 1984; Matsuda 1996; Fry 2003, *inter alia*), in that the use of zero-marking is common for objects that are adjacent to their predicates.

#### 5.4.5 Definiteness

As noted in Section 4.5.4, the notion ‘definiteness’ has often been discussed with respect to argument and particle ellipsis (Lee 1989; Lee & Thompson 1989; Fujii & Ono 2000; Ono *et al.* 2000; Fry 2003; Shimojo 2005, *inter alia*). For instance, Fry (2003) points out that definiteness is an important feature in the direct marking system of many languages, including Japanese. More specifically, he (*ibid.*) claims that demonstrative pronouns are highly accessible to both the speaker and the addressee, and they are

most likely marked with a zero particle. Lee (1989) also notes that definiteness is closely related to argument encoding types, particularly to subject encoding types.

I assume that in conversational Korean, the choice of object markings is related to definiteness of object referents. More specifically, I assume that the occurrence of *lul*-marking would be low, but that of zero-marking would be high for coding objects with a definite referent, due to easy identifiability. In order to discuss such a relationship, the choice of object markings is examined with respect to definiteness of object referents in the present section. For its purpose, just like subject NPs, all object NPs are categorized into four different sub-groups based on definiteness of their referents.<sup>12</sup> As was with subject NPs, based on descriptions of definiteness given by Gundel *et al.* (1993), Lambrecht (1994) as well as Aissen (2003), definiteness is classified into four different subtypes: definite, indefinite, indefinite pronoun, indefinite but specific. The definition for each of the four subtypes of definiteness, which is used for object NPs as well as for subject NPs, is provided below along with relevant examples that are from the data of this study.

- **definite NPs:** have their referents uniquely identifiable to both the speaker and the hearer. Pronouns, personal pronouns, demonstrative pronouns and proper nouns are good examples of definite NPs.

(5.31) *pankimwun* *acessi*-Ø *pwa-ss-nya?*  
 Ban.Kimun uncle-( ) see-PST-Q  
 ‘Did you see Mr. Ban Kimun (UN Secretary General)?’

The underlined object in (5.31), *pankimwun accessi* ‘Mr. Ban Kimun’, is a proper noun, referring to the current UN Secretary General, so it is easily identifiable to both the speaker and the hearer. Therefore, the object in (5.31) is categorized as a definite object NP. Furthermore, due to its being highly accessible, thus representing minimal

<sup>12</sup>There may be more than these four sub-groups into which object NPs are categorized.

processing load in referent identification, it is assumed that the object tends not to be explicitly coded with the accusative *lul* for its recognition. The use of zero-marking is instead expected for encoding the object, as shown in (5.31).

- **indefinite NPs:** have their referents identifiable neither to the speaker nor to the hearer. The referents of indefinite NPs are usually generic/non-specific or new.

(5.32) *kyayney-tul-un, ilpon-un, ilehkey picwuelcekin ke-l emcheng*  
 3PL-PL-TOP Japan-sc top like.this visible thing-ACC much  
*culkinapwa.*  
 enjoy  
 ‘They, Japanese seem to like visual things very much.’

The underlined object in (5.32), *picwuelcekin ke* ‘visual things’, does not refer to one specific thing, so its referent is unidentifiable to the speaker as well as to the hearer. For this reason, the object in (5.32) is counted as an instance of an indefinite object NP. The referent of the object in (5.32) is not identifiable, hence processing load is assumed, and the object is expected to be coded with *lul*-marking for its identification, as shown in (5.32).

- **indefinite pronoun NPs:** appear in pronominal forms and are referential to the speaker, but their referents had not been activated for the hearer. Furthermore, indefinite pronoun NPs do not convey specific information that needs to be recognized.

(5.33)

A1: *cihachel-pi-nun elma-ya?*  
 subway-fare-TOP how.much-Q  
 ‘How much was the subway fare?’

B1: *cihachel-pi-nun, na-nun kuke-Ø kkunh-ess-e.*  
 subway-fare-TOP 1SG-TOP that-( ) buy-PST-SEM  
 ‘As for the subway fare, I got that.’

B2: *ppenphaysu-lakwu, 1 il-tongan mwucihakey iyongha-l swu iss-nun*  
 fun.pass-called 1 day-for unlimitedly utilize-COMP can be-REL  
*key issnunkeya.*  
 thing be  
 ‘Being called as Fun Pass, there is a thing that you can utilize unlimitedly for one day.’

The underlined object in the utterance B1 in (5.33), *kuke* ‘that’, appears in its pronominal form, i.e., demonstrative, and it is referential in the sense that the speaker refers to a specific thing whose referent is identifiable to the speaker. Yet the speaker is still searching for a word which can be associated with what is in her mind, and its referent is not identifiable to the hearer until it is associated with the word, *ppenphaysu*, ‘Fun Pass’, in the subsequent utterance B2 in (5.33). For this reason, the object in (5.33), *kuke* ‘that’, is viewed as an example of indefinite pronoun NPs. In addition, unlike indefinite or definite NPs, indefinite pronoun NPs usually do not convey specific information that is important and needs to be recognized, so they are commonly marked with a zero particle, as exemplified in (5.33).

- **indefinite but specific NPs:** have their referents identifiable only to the speaker but not to the hearer. Unlike indefinite pronouns, they appear in their noun forms, i.e., full noun.

(5.34) [*cincca towum an toy-nun*] *yayki-l* *sse nwa-ss-tela.*  
 really help NEG be-REL story-ACC write place-PST-SEM  
 ‘They wrote a note that was not helpful at all.’

The underlined object in (5.34), *yayki* ‘a story’, is viewed as an example of ‘indefinite but specific NP’, in the sense that the object does not refer to a particular story, but it is specific because it typifies a note that was written on the book. Also, its referent

is newly introduced into the utterance and it is only accessible to the speaker but not to the hearer. In other words, although it is to the speaker, the referent is not identifiable to the hearer. Since its referent is identifiable to the speaker only, the object in (5.34) is overtly marked with the accusative *lul* due to processing load in its referent identification. Objects categorized as an ‘indefinite but specific NP’ are somewhat similar to those classified as an ‘indefinite pronoun NP’, in that referents for both subtypes of object NPs are identifiable only to the speaker, but not to the hearer. Yet, in the present study, objects of ‘indefinite but specific NP’ are distinguished from those of ‘indefinite pronoun NP’, because unlike objects of ‘indefinite pronoun NP’, they appear in a full noun form and contain information that is specific to some degree, though limiting its information to the generic property of the referent. Contrastively, objects of ‘indefinite pronoun NP’ do not provide the hearer with any information about the referent, not even its generic property.

In the present study, it is assumed that the choice of object markings is related to definiteness of object NPs. More specifically, as with subjects, it is assumed that the objects of both a ‘definite NP’ and an ‘indefinite pronoun NP’ are likely to be coded with zero-marking due to low processing load and low informational prominence respectively. On the other hand, the objects of an ‘indefinite NP’ and an ‘indefinite but specific NP’ are explicitly marked with the accusative *lul* due to processing load and informational prominence respectively. For this reason, all object NPs are categorized into four different subtypes based on their definiteness, and then they are further examined with regard to their markings. The results are summarized in Table 5.10.

According to the data in Table 5.10, the use of zero-marking is generally more common for objects than the use of *lul*-marking, regardless of definiteness. Yet definiteness of object referents plays a role in a speaker’s selection of object encoding types, though it does in an opposite way to what I had initially expected. For example, the occur-



Table 5.10: Object Definiteness and Object Marking

OBJ.MARK		NP-Ø	NP-lul	NP-nun	NP-to	NP-etc.	SUB-TOTAL	TOTAL
OBJ.DEFINITE								
indefinite	NO.	1,275	518	83	161	24	785	2,061
	%	<b>62</b>	<b>25</b>	4	8	1	38	100
definite	NO.	272	152	26	31	6	215	487
	%	<b>56</b>	<b>31</b>	6	6	1	44	100
indef. pron.	NO.	132	28	1	6	4	39	171
	%	<b>77</b>	<b>16</b>	1	4	2	23	100
indef./specific	NO.	9	1	0	2	0	3	12
	%	<b>75</b>	<b>8</b>	0	17	0	25	100
TOTAL	NO.	1,688	699	110	200	34	1,043	2,731
	%	<b>62</b>	<b>26</b>	4	7	1	38	100

rence of *lul*-marking is high for objects of definite NPs (31%), compared to that of other subtypes of object NPs (e.g., 25% for objects of indefinite NPs) (zero-marking vs. *lul*-marking:  $\chi^2=7.875$ ,  $P < 0.01$ ). One may find this result to be somewhat unexpected, in that referents of definite NPs are identifiable to both the speaker and the hearer, hence represent minimal processing load in referent identification, so the occurrence of *lul*-marking is expected to be relatively low. Yet I argue that this tendency would be attributed to the property of object NPs, that is, referents of object NPs, which typically encode new information, are likely to be indefinite (Fry 2003), while those of subject NPs are normally definite. That is to say, an object NP whose referent is definite may easily be identified due to its accessibility, yet at the same time it is regarded as a deviation from the norm, in the sense that referents of objects are most likely to be indefinite (Fry 2003), and thus processing load is assumed for such an object. Due to processing load, the occurrence of *lul*-marking is expected to be higher for objects of definite NPs (31%) than for those of indefinite NPs (25%). As a matter of fact, the data in Table 5.10 complies with Fry's (2003) statement, in that 2,061 tokens of objects,

roughly 75% of the total, are categorized as objects of indefinite.<sup>13</sup>

Furthermore, indefinite NPs usually denote new information that may function as a focusing device, so the occurrence of *lul*-marking is assumed to be relatively low for coding such indefinite object NPs. That is to say, referents of object NPs are generally indefinite and they represent new information by default, and the new information may function as a focusing device, so the occurrence of *lul*-marking is relatively low for the indefinite object NPs.

In Table 5.10, I also find that the use of zero-marking is more common for coding objects of indefinite pronoun NPs (77%) than for coding objects of definite NPs (56%). Contrastively, the use of *lul*-marking is less common for objects of indefinite pronoun NPs (16%) than for objects of definite NPs (31%) (zero-marking vs. *lul*-marking:  $\chi^2=18.343$ ,  $P < 0.001$ ). Furthermore, the use of zero-marking is more common for objects of indefinite pronoun NPs than for objects of indefinite NPs (62%), and the use of *lul*-marking is less common for objects of indefinite pronoun NPs than for indefinite NPs (25%) (zero-marking vs. *lul*-marking:  $\chi^2=9.462$ ,  $P < 0.01$ ). In fact, this tendency complies with what I had expected. In other words, as noted earlier in the present section, referents of indefinite pronoun NPs are often identifiable to the speaker but not to the hearer, i.e., new to the hearer—default property of objecthood, yet they do not need to be identified. In addition, indefinite pronouns do not convey any specific information, and low informational prominence is assumed to be represented by them. Therefore, they tend not to be explicitly marked with an overt particle. This explains such a tendency that the occurrence of *lul*-marking is relatively low for objects of indefinite pronoun NPs and instead, the occurrence of zero-marking is high for them.

Lastly, what is observed for object NPs differs from what was observed for subject NPs. That is to say, the use of *ka*-marking is less common for definite subjects than

---

<sup>13</sup>On the other hand, a total of 1,577 tokens of subjects have an indefinite referent, which is about 40% of the total.

for their indefinite counterparts, due to easy identifiability, hence minimal processing load in referent identification, whereas *lul*-marking is more common for definite objects than for indefinite counterparts, due to being viewed as a deviation from the norm, hence processing load. This statement may sound paradoxical, in a sense that the use of overt marking is not common for definite subjects due to minimal processing load, while it is common for definite objects due to high processing load. Yet this statement is not paradoxical if a typical property of subject and object NPs is taken into considerations. That is to say, referents of subjects are typically definite, so definite subjects are considered the norm. Definite subjects can be identified with minimal processing load in referent identification. Therefore, the use of *ka*-marking is less common for definite subjects than for their indefinite counterparts. On the other hand, referents of objects are typically indefinite, so definite objects are regarded as a deviation from the norm, though they may be accessible to both the speaker and the hearer. Thus, unlike definite subjects, the occurrence of *lul*-marking is higher for definite objects than for indefinite counterparts.

To summarize this section, in conversational Korean, a speaker's selection of one particular object encoding type over another seems to be related to the definiteness of object referents, and the tendency observed for objects appears to be opposite to the tendency reported for subjects. More specifically, object NPs whose referents are indefinite are considered to be the norm for objects, thus low processing load is assumed for them. Due to minimal processing load, the use of *lul*-marking is less common for indefinite objects than for definite counterparts. It may also be the case that indefinite NPs generally denote new information that may function as a focusing device, so the occurrence of *lul*-marking is relatively low for coding the indefinite object NPs. Contrastively, object NPs whose referents are definite are regarded as a deviation from the norm for objects, thus processing load is expected. Due to the processing load,

the occurrence of zero-marking is relatively low for coding definite objects, and instead the occurrence of *lul*-marking is relatively high for coding them. Lastly, a zero particle is most commonly used for coding objects of indefinite pronoun NPs; conversely, the accusative *lul* is not commonly used for coding them. I attribute this tendency to the fact that in general, indefinite pronouns do not represent informational prominence, and due to the lack of informational prominence, objects of indefinite pronoun tend not to be recognized explicitly with *lul*-marking.

Based on what has been discussed thus far, I claim that the definiteness of object NPs, in the senses of processing load and informational prominence, influences the choice of object markings in conversational Korean.

#### 5.4.6 Length

It has been pointed out that in Japanese, the ellipsis of case particles is related to the length of an NP, i.e., subject or object, in a clause (Tsutsui 1984; Ono *et al.* 2000). More specifically, in spoken Japanese, monosyllabic NPs (or one-mora NPs in Ono *et al.*'s (2000) term) are uncommon, and are often not easily recognized, especially in fast speech, so they tend to be coded with an overt particle, which makes them more recognizable. In other words, processing load is expected for monosyllabic NPs, and the use of an overt particle is common for coding them. On the other hand, phonologically larger NPs are more commonly coded with an overt particle than phonologically shorter NPs (Tsutsui 1984).<sup>14</sup> Based on this tendency observed in spoken Japanese, I make the following assumption; in spoken Japanese, due to its reduced saliency in form, processing load is assumed for monosyllabic NPs, and the use of an explicit particle is common for coding them to alleviate the processing load imposed on monosyllabic NPs. On the other hand, larger NPs generally encode more information than shorter

---

<sup>14</sup>In her study, the syllable-length was used as a way of measuring the length of a given NP.

NPs, and they are assumed to represent informational prominence, which would invite the use of an overt particle for coding them.

In Section 4.5.5, we explored the relationship between the length of subject NPs and the choice of their markings. According to the data in Table 4.10, it appeared that in conversational Korean, the use of *ka*-marking was not common for monosyllabic subjects, though it was reported to be so in colloquial Japanese (Tsutsui 1984; Ono *et al.* 2000). This cross-linguistic difference was due to the frequency as well as type difference of monosyllabic NPs between the two languages. That is to say, in Japanese, monosyllabic subjects are uncommon, and they are also mostly full nouns. On the other hand, in Korean, monosyllabic subjects are not uncommon, and they are mostly personal pronouns. The use of *ga*-marking is common for monosyllabic subjects in Japanese, whereas the use of *ka*-marking is not common for monosyllabic counterparts in Korean. Yet, when types of monosyllabic subjects were also considered as a factor, a similarity was observed between the two languages: the use of nominative marking particles (*ga* in Japanese; *ka* in Korean) is common for monosyllabic full noun subjects in both languages due to processing load (see Table 4.11). In Section 4.5.5, it was also shown that the percentage of *ka*-marked subjects is proportional to the length of subjects, which is measured by the number of their modifiers. For example, the percentage of *ka*-marked monosyllabic subjects gradually increases as the number of subject modifiers increases (34% → 50% → 100%).

In this section, in order to see if a speaker's selection of object encoding types is related to the length of objects, I examined all object NPs with regard to their length. As was with subjects, the length of objects was examined in two different ways: whether or not to be monosyllabic and the number of modifiers. I will first examine monosyllabic objects with relation to their marking, and later examine the number of object modifiers with respect to the choice of object markings. Following are examples of monosyllabic

as well as non-monosyllabic object NPs that are from the data of the present study.

(5.35) mwul-Ø echaphi na-Ø saya-toy-nikka  
 water-( ) anyway 1SG-( ) buy-have.to-because  
 ‘It’s because I have to buy water anyway.’

(5.36) na-l pol ttay, mwonka nwunpich-i com kunyang nwunpich-i  
 1SG-ACC see when something eye.look-NOM well normal eye.look-be  
 anila.  
 not  
 ‘When he looks at me, you know, his eye-look is not like a normal eye-look.’

(5.37) molla. na-nun hangsang naitun salam-ul han pen-to mos  
 don’t.know 1SG-TOP always old person-ACC once NC-even NEG  
 sakwie pwa-ss-e.  
 date-PST-SEM  
 ‘I don’t know. I haven’t dated an old one yet.’

The utterance in (5.35) exemplifies a monosyllabic object NP, and the object, *mwul* ‘water’, which is underlined, is marked with a zero particle. The utterance in (5.36) also exemplifies a monosyllabic object, and the underlined object, *na* ‘me’, is coded with the accusative *lul*. On the other hand, the utterance in (5.37) shows an example of a non-monosyllabic object NP, and the underlined object, *salam* ‘person’, is marked with the accusative *lul*. Table 5.11 shows the distributional pattern of object markings with regard to whether or not object NPs are monosyllabic.

In Table 5.11, I find that in conversational Korean, as was so with subject NPs, the occurrence of overt marking (e.g., *lul*-marking) is not particularly high for coding monosyllabic object NPs (24%, cf., 26% for non-monosyllabic object NPs), whereas in colloquial Japanese, the occurrence of overt marking (e.g., *ga*-marking) was reported to be high for coding monosyllabic NPs (Ono *et al.* 2000).<sup>15</sup> As a matter of fact, the use of

<sup>15</sup>To the best of my knowledge, there are no previous studies done to examine object markings in terms of whether or not objects are monosyllabic in Japanese, let alone in Korean. For this reason, I discuss the choice of object markings in Korean based on the tendency reported for subject NPs in Japanese, assuming that the similar, if not the same, tendency would be observed for object NPs.

Table 5.11: Monosyllabic Objects and Object Marking

OBJ.MARK		NP-Ø	NP-lul	NP-nun	NP-to	NP-etc.	SUB-TOTAL	TOTAL
OBJ.MONOSYLLABLE								
¬monosyllabic	NO.	1,376	589	97	169	32	887	2,263
	%	<b>61</b>	<b>26</b>	4	8	1	39	100
monosyllabic	NO.	312	110	13	31	2	156	468
	%	<b>67</b>	<b>24</b>	3	7	0	33	100
TOTAL	NO.	1,688	699	110	200	34	1,043	2,731
	%	62	26	4	7	1	38	100

zero-marking is more common, though only slightly, for coding monosyllabic object NPs (67%) than it is for coding non-monosyllabic object NPs (61%). Conversely, the use of *lul*-marking is slightly less common for monosyllabic object NPs (24%) than for non-monosyllabic object NPs (26%) (zero-marking vs. *lul*-marking:  $\chi^2=2.562$ ,  $P=0.1094$ ). Simply put, unlike Japanese (Ono *et al.* 2000), in Korean, object NPs' being monosyllabic does not invite the use of *lul*-marking for object NPs.

One may wonder why the two languages, which share many linguistic properties with each other, differ in this regard. As was so with subject NPs, I attribute such a difference to the difference in frequency of monosyllabic object NPs between the two languages. That is to say, in Japanese, monosyllabic object NPs are not common, thus may be viewed as a deviation from the norm, while in Korean, they are common, thus can be regarded as the norm. According to the data in Table 5.11, the number of monosyllabic object NPs is 468 in total, which is approximately 17% of the entire body of object tokens. Although the percentage of monosyllabic objects is much smaller than that of monosyllabic subjects, which constituted roughly 32% of all subject tokens, it is still not considered to be uncommon, at least compared to that of Japanese. As was previously pointed out in Section 4.5.5, frequently occurring entities are generally more

accessible and easier to identify than uncommonly occurring entities, so the occurrence of an overt particle is assumed to be relatively low for marking monosyllabic objects that frequently occur. Since monosyllabic object NPs are common in conversational Korean, hence low processing load in referent identification, the occurrence of *lul*-marking is not expected to be high for coding them, at least not as high as it is for coding those in colloquial Japanese (cf., *o*-marking).

In Chapter 4, I pointed out that in Korean, monosyllabic subjects are mostly (personal) pronouns (e.g., *na* ‘I’, *ku* ‘he’; roughly 85% of the total), while in Japanese, no (personal) pronouns (e.g., *watasi* ‘I’, *karae* ‘he’) are monosyllabic. I assume that in conversational Korean, for the same reason with subject NPs, a speaker’s selection of one object form over another would be related to nominal types of object NPs: monosyllabic pronominal objects or monosyllabic full noun objects. For this reason, just as was with subjects, I examined if the nominal type of object NPs influences the choice of object markings in conversational Korean, and the observed distributional pattern is presented in Table 5.12.

Table 5.12: Monosyllabic Object Type and Object Marking

OBJ.MARK		NP-Ø	NP-lul	NP-nun	NP-to	NP-etc.	SUB-TOTAL	TOTAL
OBJ.MONOSYLLABLE								
mono.pronoun	NO.	87	43	1	4	1	49	136
	%	<b>64</b>	<b>32</b>	1	3	1	36	100
mono.full noun	NO.	225	67	12	27	1	107	332
	%	<b>68</b>	<b>20</b>	4	8	0	32	100
TOTAL	NO.	312	110	13	31	2	156	468
	%	67	24	3	7	0	33	100

In Table 5.12, I find that unlike subject NPs, the number of monosyllabic pronominal object NPs (136 tokens, or 29% of the total) is not only much smaller than that of monosyllabic full noun object NPs (332 tokens, or 71% of the total), but it is also far



lower than that of subject NPs (85%).<sup>16</sup> I associate this distributional tendency to the assumption that in general, personal pronouns are not realized as referents of object NPs, particularly when they are interlocutors, namely the speaker or the hearer. This tendency may also be attributed to the general characteristics of object NPs: direct objects are usually indefinite (Fry 2003). That is to say, referents of objects are usually indefinite, but (personal) pronouns are typically definite, so pronominal object NPs are not common.

Furthermore, according to the data in Table 5.12, the use of zero-marking is more common for object NPs than that of *lul*-marking, regardless of the nominal type of object NPs. I associate this tendency with the fact that object NPs usually denote new information, which may function as a focusing device, so the occurrence of zero-marking is generally high for coding them.

In Table 5.12, I also find that the occurrence of *lul*-marking is higher for coding monosyllabic pronominal objects (32%) than it is for coding monosyllabic full noun objects (20%) (zero-marking vs. *lul*-marking:  $\chi^2=4.791$ ,  $P < 0.05$ ). For example, the object in (5.36), *na* 'me', which is a personal pronoun, is coded with the accusative *lul*. I attribute this to the distributional tendency for object NPs, that is, unlike subjects, monosyllabic full noun objects are common (roughly 71% of the total), whereas monosyllabic personal pronouns tend not to occur as referents of objects. On the other hand, as was noted above, monosyllabic pronominal objects are not common (only 29% of the total), thus may be viewed as a deviation from the norm, which would result in processing load. Due to the processing load, the occurrence of *lul*-marking is expected to be higher for monosyllabic pronominal objects than it is for monosyllabic full noun objects, as can be seen in Table 5.12. Furthermore, this distributional pattern of object NPs is in fact opposite to that of subject NPs (see Table 4.11 for comparison). That

---

<sup>16</sup>See Table 4.11 in Section 4.5.5 of Chapter 4 for comparison.

is to say, as for monosyllabic subject NPs, *ka*-marking is more common for coding full noun subjects than for their pronominal counterparts, whereas as for monosyllabic object NPs, *lul*-marking is more commonly used for coding pronominal objects than their full noun counterparts.

According to the data in Table 5.11 as well as in Table 5.12, unlike in Japanese, the occurrence of *lul*-marking is not particularly high for coding monosyllabic object NPs in conversational Korean. This cross-linguistic difference between the two languages can be attributed to the difference in frequency of monosyllabic object NPs: monosyllabic objects are uncommon in Japanese whereas they are common in Korean. Nevertheless, there may still be other factors that make the two languages differ from each other in this particular matter. For example, monosyllabic objects by themselves may not denote information that is important, although they may convey noteworthy information when they are modified. To the best of my knowledge, no previous studies, either in Korean or in Japanese, have been done to examine monosyllabic object markings with relation to the number of their modifiers. In this section, I examine if the choice of monosyllabic object markings is related to the number of their modifiers. The results are summarized in Table 5.13.

Table 5.13: Monosyllabic Object Modifiers and Object Marking

OBJ.MARK		NP-Ø	NP-lul	NP-nun	NP-to	NP-etc.	SUB-TOTAL	TOTAL
OBJ.MONO.MODIFIERS								
0 modifier	NO.	309	108	13	31	2	154	463
	%	67	23	73	7	0	33	100
1 modifier	NO.	3	2	0	0	0	2	5
	%	60	40	0	0	0	40	100
TOTAL	NO.	312	110	13	31	2	156	468
	%	67	24	3	7	0	33	100

In Table 5.13, I find that the percentage of *lul*-marked object NPs that are monosyl-

labic is proportional to the number of object modifiers (23% with no modifier to 40% with one modifier). As expected, the percentage of zero-marked monosyllabic objects is inversely proportional to the number of object modifiers (67% with no modifier to 60% with one modifier). Yet the number of monosyllabic objects with one modifier is too small for statistically useful comparison.

Based on what has been discussed thus far, I was led to the tentative conclusion that in conversational Korean, the choice of object markings does not correlate with whether or not they are monosyllabic.

As was previously stated earlier in the present section, the length of object NPs was also measured by how large they were, and then it was examined with regard to the choice of object markings. As a means of measuring how large the object NPs are, I relied on counting the number of object modifiers. Following are some examples of object NPs with and without an object modifier, in which modified words are underlined, and modifying words are bracketed.

- (5.38) *waynyamyen nay-ka yosay mak hamyense kunyang kyaysok*  
 because 1SG-NOM these.days hard doing just continuously  
*ku mwonya liphothu-Ø ssumyense*  
 that what report-( ) writing  
 ‘Because, these days, while I was continuously writing a report.’

- (5.39) *nay-ka [cikum pang-ey iss-nun] ke-l ta*  
 1SG-NOM now room-LOC be-REL thing-ACC all  
*chiwoyaha-ki-ttaymwuney, an ka-kwu.*  
 clean.up-NMLZ-because NEG go-and  
 ‘Since I have to clean up things that are now in my room, (I) won’t go and ...’

- (5.40) *kuliko nay-ka tto oppa-nun [hankwuk-eyse manhi tut-ko*  
 and 1SG-NOM also brother-TOP Korea-LOC much take-and  
*o-n] ke-l yekise tto tulumyenun*  
 come-REL thing-ACC here again take.if  
 ‘And, I, and you retake things here again that you had taken in Korea, and then ...’

(5.38) is an example of an object NP with no modifier because the object, *liphothu* ‘report’, is not modified. In (5.39), there are three modifying words in brackets, *cikum pangey issnun* ‘that are now in (my) room’, and thus the object, *ke* ‘things’, is classified into an object NP with three modifiers. Lastly, an object with four modifiers is exemplified in (5.40), in that there are four words in brackets, *hankwukeyse manhi tutko on* ‘that (you) took in Korea’, which modify the object, *ke* ‘thing’.

As was stated above, the present study examined if and how the length of object NPs, in terms of the number of object modifiers, would influence the choice of object markings in conversational Korean, and the results are summarized in Table 5.14.

Table 5.14: Object Modifiers and Object Marking

OBJ.MARK		NP-Ø	NP-lul	NP-nun	NP-to	NP-etc.	SUB-TOTAL	TOTAL
OBJ.MODIFIERS								
zero	NO.	1,274	487	63	132	24	706	1,980
	%	<b>64</b>	<b>25</b>	3	7	1	36	100
one	NO.	298	145	29	48	6	228	526
	%	<b>57</b>	<b>28</b>	6	9	1	43	100
two	NO.	32	34	9	8	3	54	86
	%	<b>37</b>	<b>40</b>	11	9	4	63	100
three	NO.	6	12	4	5	0	21	27
	%	<b>22</b>	<b>44</b>	15	19	0	78	100
four or more	NO.	4	8	0	0	1	9	13
	%	<b>31</b>	<b>62</b>	0	0	8	69	100
that-comp.	NO.	70	10	5	6	0	21	91
	%	77	11	6	7	0	23	100
nominalized	NO.	4	3	0	1	0	4	8
	%	50	38	0	13	0	50	100
TOTAL	NO.	1,688	699	110	200	34	1,043	2,731
	%	62	26	4	7	1	38	100

In Table 5.14, I find that as was with subject NPs, the use of *lul*-marking is more common for longer object NPs than it is for shorter object NPs, though the degree of the increase in frequency is not always gradient. For example, the percentage of *lul*-marked

objects is only 25% when there is no word modifying objects, but it greatly increases to 44% when there are three modifiers, and to 83% when there are five modifiers or more. As expected, the percentage of zero-marked objects is inversely proportional to the number of object modifiers. For instance, the percentage of zero-marked objects is 64% when there is no modifier, yet it drops dramatically to 22% when there are three modifiers (zero modifier vs. three modifiers:  $\chi^2=13.437$ ,  $P < 0.001$ ) and 31% when there are four or more modifiers (zero modifier vs. four or more modifiers:  $\chi^2=9.014$ ,  $P < 0.01$ ). I relate this tendency to informational prominence. That is to say, objects NPs that are modified, thus longer, are considered informationally more prominent than those that are not modified, thus shorter, resulting in the tendency for longer objects to be overtly identified with the accusative *lul*. For instance, the object in (5.38), *liphothu* ‘report’, is viewed as a short NP, hence would represent low informational prominence, so it is marked with a zero particle. Contrastively, the objects, [*cikum pangey issnun*] *ke* ‘things that are now in my room’ in (5.39) and [*hankwukeyse manhi tutko on*] *ke* ‘things that you took in Korean’ in (5.40), are regarded as long NPs, thus would represent high informational prominence, so they are coded with the accusative *lul*. These distributional patterns comply with what was reported in Mori & Givón (1987) and Ono *et al.* (2000), though their way of measuring length differs from mine,<sup>17</sup> in that the larger NPs tend to be marked with an overt particle while the smaller NPs tend to be marked with a zero particle.

Based on the observed tendency, I claim that in conversational Korean, the accusative *lul* can be used to mark informational prominence, which is encoded in object NPs.

To summarize, unlike in colloquial Japanese (Tsutsui 1984), in conversational Korean, the use of zero-marking is common for monosyllabic objects while the use of

---

<sup>17</sup>They measured the length of object NPs in terms of the syllable length.

*lul*-marking is not common for them, though there are some variations in the occurrence of *lul*-marking for the monosyllabic objects observed with regard to their nominal type, as well as the number of their modifiers (refer to Table 5.11, Table 5.12 and Table 5.13). I attribute this tendency to the frequency difference of monosyllabic object NPs between the two languages. That is to say, in Japanese, monosyllabic objects are uncommon, hence a deviation from the norm, and they are commonly coded with an overt particle (e.g., *o*) due to processing load incurred by a deviation from the norm. Conversely, monosyllabic object NPs are common in Korean, thus are not considered a deviation from the norm, and the occurrence of *lul*-marking is relatively low for coding them. The number of pronominal monosyllabic objects is smaller than that of their full noun counterparts (see Table 5.12), and the use of *lul*-marking is more common for pronominal monosyllabic objects (32%) than full noun counterparts (20%). However, this tendency does not appear to be statistically significant. In other words, the choice of object markings does not correlate with whether or not objects are monosyllabic.

The choice of object markings was also examined with relation to the number of words modifying monosyllabic objects. According to Table 5.13, the use of *lul*-marking is more common for monosyllabic objects with a modifier than it is for those without a modifier. This tendency can be attributed to an increase in informational prominence that is proportional to the increase in object modifiers.

Furthermore, the occurrence of *lul*-marked object NPs proportionally increases as the number of object modifiers, namely the length of objects, increases, while the occurrence of zero-marked object NPs is inversely proportional to the number of object modifiers (e.g., 64% for no modifier; 31% for four or more modifiers). In fact, this tendency complies with the previous findings in spoken Japanese (Mori & Givón 1987, *inter alia.*). I attribute this tendency to the assumption that informational prominence is greater in large NPs than in small NPs.

Based on what has been discussed in this section, I conclude that in conversational Korean, a speaker's selection of one particular encoding type for monosyllabic objects over another does not correlate with whether or not the objects are merely monosyllabic, yet it correlates with the type and frequency of monosyllabic object NPs. It also correlates with the number of object modifiers. That is to say, the occurrence of *lul*-marking is relatively high for coding objects that represent processing load or informational prominence, while the occurrence of zero-marking is relatively high for coding objects that represent minimal processing load or low informational prominence. Based on this, I claim that the accusative *lul* in conversational Korean functions as alleviating processing load or marking informational prominence encoded in object NPs.

#### 5.4.7 Repair

As was noted in Section 3.3.1, when a speaker changes what s/he had just said, we say there is an instance of 'repair'. In general, more attention is paid to reproduced entities than to interrupted entities because the reproduced entities are assumed to represent prominent information, thus receive attention from the hearer, while the interrupted entities are assumed not to represent prominent information, thus can be ignored by the hearer. Since the entities that are reproduced represent informational prominence, and they are expected to be identified with an explicit particle. In the present study, when a repair was made during a given utterance, a reproduced entity was only considered for the examination of its marking, and an entity that was interrupted for a repair was ignored. The following are two examples of utterances which include object repair.

- (5.41) *nwun-ul ilehkey [kokay]-lul nophi tul-ci anh-ato*  
 eye-ACC like.this head-ACC high raise-COMP NEG-though  
 'Even though (you) don't raise (your) eye, (your) head high,'

(5.42) *kuntey na-nun icey sapemtay-nikka, kyocik,*  
 yet 1SG-TOP now education.college-be teaching.profession  
*kyocik-to, [kyocik]-ul hay-yatoykwu.*  
 teaching.profession-also teaching.profession-ACC do-must  
 ‘Yet, you know, I go to the education college, so I have to take teacher training courses.’

In (5.41), there is an instance of ‘repair’, in that the first underlined element, *nwun* ‘eye’, is interrupted and reproduced into the second element in brackets, *kokay* ‘head’, which is marked with the accusative *lul*. In this study, only the second element, referred to as a reproduced element in the present study, was examined with regard to its marking, and the first element, namely an interrupted element, was ignored for the examination of object markings. There is also an instance of ‘repair’ in (5.42) because the first two underlined elements, *kyocik*, *kyocikto* ‘teacher training course, teacher training course’, are interrupted, and they are reproduced into the bracketed element, *kyocik* ‘teacher training course’, that is coded with *lul*-marking. As with the selection of subject markings, the choice of object markings was examined based on whether or not there was a repair made for the object in an utterance, but not based on whether it was an interrupted object or a reproduced object. For the sake of labeling convenience, objects appearing in a clause that contains an object repair are called ‘repaired objects’, and objects occurring in a clause that do not include an object repair are labeled as ‘non-repaired objects’ in this dissertation.

In Section 4.5.6, it was shown that the use of zero-marking is less common for repaired subjects than it is for their non-repaired counterparts. Conversely, the use of the overt marking, either with *ka* or with *nun*, is more common for repaired subjects than for non-repaired subjects (see Table 4.14). In this section, I explore the choice of object markings with respect to the object repair, and the results are shown in Table 5.15.



Table 5.15: Object Repair and Object Marking

OBJ.MARK OBJ.REPAIR		NP-Ø	NP-lul	NP-nun	NP-to	NP-etc.	SUB- TOTAL	TOTAL
¬repair	NO.	1,609	638	100	181	31	950	2,559
	%	<b>63</b>	<b>25</b>	4	7	1	<b>37</b>	100
repair	NO.	79	61	10	19	3	93	172
	%	<b>46</b>	<b>36</b>	6	11	2	<b>54</b>	100
TOTAL	NO.	1,688	699	110	200	34	1,043	2,731
	%	62	26	4	7	1	38	100

In Table 5.15, I find that the occurrence of overt marking is high for repaired objects (54%), and the occurrence of zero-marking is high for non-repaired objects (63%) (zero-marking vs. overt marking:  $\chi^2=19.606$ ,  $P < 0.001$ ). I attribute this tendency to the informational prominence encoded in repaired objects, i.e., reproduced objects. That is to say, repaired objects tend to represent informational prominence, so the occurrence of overt marking is high for coding them, and the occurrence of zero-marking is relatively low. Furthermore, the occurrence of *lul*-marking is higher for repaired objects (36%) than it is for non-repaired objects (25%). As expected, the occurrence of zero-marking is lower for repaired objects (46%) than it is for non-repaired objects (63%) (zero-marking vs. *lul*-marking:  $\chi^2=14.661$ ,  $P < 0.001$ ). For instance, the object in (5.41), *kokay* ‘head’, and the object in (5.42), *kyocik* ‘teacher training course’, are both marked with the accusative *lul*.

According to Table 5.15, the selection of object markings seems to correlate with whether or not objects are repaired. In other words, due to informational prominence, the use of zero-marking is not common for coding repaired objects, and instead, the use of overt marking, e.g., *lul*-marking, is common for coding them. Based on this tendency, I argue that the accusative *lul* in Korean functions as marking the informational

prominence that is encoded in repaired object NPs.

In Table 5.15, I also find that the occurrence of repaired objects (172 tokens, 6% of the total) is only slightly lower than that of repaired subjects of transitive clauses (82 tokens, approximately 8% of the total).<sup>18</sup> This distributional pattern does not comply with the general tendency: speech errors occur most frequently with new information where the most speech planning processing is required (Jeri Jaeger, personal communication) and new information is usually encoded in objects.

To summarize this section, the occurrence of *lul*-marking is relatively high for repaired objects, and conversely, the occurrence of zero-marking is high for non-repaired objects. This tendency is attributed to the assumption that repaired objects represent informational prominence, so the use of the overt marking with *lul*, which functions as marking informational prominence encoded in objects, is common for repaired objects. Simply put, the selection of one particular object form over another correlates with whether objects are repaired or not repaired, and more importantly, the accusative *lul* has a function as encoding the informational prominence.

#### 5.4.8 Anaphoric Saliency

As repeatedly stated, RD is introduced as one of the quantitative measurements, which is used to indicate the linguistic distance in clausal units (Givón 1983; 1993), and it is measured by counting clausal units backward to the most recent representation of the co-referential expression that includes those of zero anaphor (Givón 1983). RD is viewed as a heuristic measure for the level of activation of a particular referent in one's consciousness, in that the shorter RD, the more activated a particular referent is in one's consciousness, given that all other things are considered equal (Givón 1983; Shimojo 1995; 2005). It is yet notable that RD is simply a heuristic way of measuring the

---

<sup>18</sup>See Table 4.14 and Table 4.15 for comparison.

level of activation, and RD by itself may not be of great cognitive significance (Givón 1993; Shimojo 2005). Furthermore, there can be other factors, which also influence the level of activation. The activation status of a given referent may also be measured by examining how frequently it is mentioned within an anaphoric context (Dryer 1996). For example, the more frequently a referent is mentioned within an anaphoric context, the more activated it is in one's consciousness, provided that all other things are considered same. Hence, the anaphoric saliency, i.e., the activation status of a referent, is measured not only by its anaphoric recency (RD) but also by its anaphoric frequency (AF), as was previously noted in Section 2.11.1.

In general, processing requires an activation cost, and the cost depends on the activation status of information prior to processing (Dryer 1996). In other words, information that is activated in one's consciousness can be processed at little activation cost, whereas deactivated and semi-activated information seem to require a greater activation cost (Dryer 1996; Shimojo 2005). Based on this, I expect the RD as well as AF measurement to be a good way of examining the speaker's selection of object encoding types in conversational Korean, as was reported to be so in colloquial Japanese (Shimojo 2005). More specifically, with respect to the choice of object markings, I associate a higher occurrence of zero-marking with a shorter RD and higher AF, and a higher occurrence of *lul*-marking with a longer RD and lower AF.

It is notable that in spite of the possible correlation between the activation level of a given referent and the choice of its encoding types, nearly no attempt, if not absolutely none, has been made to capture a discourse-pragmatic function of object encoding types (e.g., the usage of *lul/nun*-marking) with relation to the activation level of the object in the minds of the speaker and the addressee in actual discourse. This study discusses discourse-pragmatic functions of object encoding types by examining the object encoding types with relation to the anaphoric saliency in the sense of RD

and AF of object NPs. Table 5.16 and Table 5.18 show distributional patterns of object encoding types with relation to RD and AF, respectively. Table 5.17 presents the tokens in terms of the two categories of RD (1-5 vs. 6-NPM) along with the mean RD values, and Table 5.19 presents the tokens in terms of the two categories of AF (0-3 vs. 4-10) along with the mean AF values.

As previously stated, it is generally assumed that the smaller the RD is for a given referent, the more activated the referent is in one's consciousness. The more activated the referent is in one's consciousness, the less likely it overtly appears or it is overtly coded with its post-nominal particle. Yet, according to the data in Table 5.16 and Table 5.17, RD does not seem to distinguish between object encoding types except for subjects of zero anaphor, though it was reported to be so for subject NPs. In other words, it seems that except for the subjects of zero anaphor, the RD of object referents does not play a role in the choice of object encoding types. According to Table 5.16, the mean RD for zero anaphor objects (6.4 clauses) is much smaller than that of other available encoding types (e.g., 11.9 clauses for *lul*-marked objects). This supports the assumption that the more activated it is, the less likely it overtly appears in a clause. That is to say, objects with a smaller RD are more activated, thus more identifiable, than objects with a greater RD, so they can be processed at little activation cost. Activated entities commonly appear as a zero anaphor because they can easily be recovered. Table 5.17, where RD is generalized in terms of the two ranges of RD, also indicates that overall, object referents of zero anaphor (73% of the total for RD 1-5) are in the anaphorically salient side of the index, whereas referents of other object encoding types (e.g., *lul*-marked objects: 57% of the total for RD 6-NPM) are in the anaphorically non-salient side of the index (zero anaphor vs. *lul*-marked objects:  $\chi^2=148.783$ ,  $P < 0.001$ ). As a matter of fact, this tendency observed in Korean complies with Shimojo's (2005) observation in colloquial Japanese.

Table 5.16: Object Anaphoric Recency (RD) and Object Marking

OBJ.ET OBJ.RD		ZERO- ANAPHOR	NP-Ø	NP-lul	NP-nun	NP-to	NP-etc.	SUB- TOTAL	TOTAL
1 cls.	NO.	437	410	164	27	37	6	234	1,081
	%	40	38	15	2	3	1	22	100
2 cls.	NO.	136	146	75	18	18	1	112	394
	%	35	37	19	5	5	0	28	100
3 cls.	NO.	68	82	29	2	8	1	40	190
	%	36	43	15	1	4	1	21	100
4 cls.	NO.	42	41	22	7	4	0	33	116
	%	36	35	19	6	3	0	28	100
5 cls.	NO.	16	38	11	1	5	1	18	72
	%	22	53	15	1	7	1	25	100
6 cls.	NO.	15	24	12	0	1	1	14	53
	%	28	45	23	0	2	2	26	99
7 cls.	NO.	4	15	7	0	1	0	8	27
	%	15	56	30	0	4	0	30	101
8 cls.	NO.	7	23	8	0	4	0	12	42
	%	17	55	19	0	10	0	29	101
9 cls.	NO.	3	9	4	0	1	1	6	18
	%	17	50	22	0	6	6	33	100
10 cls.	NO.	3	8	0	0	2	0	2	13
	%	23	62	0	0	15	0	15	100
11 cls.	NO.	2	8	5	0	0	0	5	15
	%	13	53	33	0	0	0	33	99
12 cls.	NO.	5	7	5	0	0	0	5	17
	%	29	41	29	0	0	0	29	99
13 cls.	NO.	1	10	2	1	3	0	6	17
	%	6	59	12	6	18	0	35	100
14 cls.	NO.	2	7	2	0	0	0	2	11
	%	18	64	18	0	0	0	18	100
15 cls.	NO.	2	4	4	1	0	0	5	11
	%	18	36	36	9	0	0	45	99
16 cls.	NO.	0	4	3	0	0	0	3	7
	%	0	57	43	0	0	0	43	100
17 cls.	NO.	1	3	2	0	0	0	2	6
	%	17	50	33	0	0	0	33	100
18 cls.	NO.	1	2	3	0	1	0	4	7
	%	14	29	43	0	14	0	57	100
19 cls.	NO.	1	0	2	0	0	0	2	3
	%	33	0	67	0	0	0	67	100
20 cls.	NO.	0	1	3	0	1	1	5	6
	%	0	17	50	0	17	17	83	100
NPM	NO.	215	846	336	53	113	23	525	1,586
	%	14	53	21	3	7	1	33	100
TOTAL	NO.	961	1,688	699	110	199	35	1,043	3,692
	%	26	46	19	3	5	1	28	100
MEAN RD		<b>6.4</b>	<b>12.0</b>	<b>11.9</b>	<b>11.3</b>	13.4	15.3	12.2	10.6

Table 5.17: Object RD (1-5, 6-NPM) and Object Marking

OBJ.ET OBJ.RD	ZERO- ANAPHOR	NP-Ø	NP-lul	NP-nun	NP-to	NP-etc.	SUB- TOTAL	TOTAL	
<i>1-5</i>	NO.	699	717	301	55	72	9	437	1,853
	%	<b>73</b>	<b>42</b>	<b>43</b>	<b>50</b>	57	26	42	50
<i>6-NPM</i>	NO.	262	971	398	55	127	26	606	1,839
	%	<b>27</b>	<b>58</b>	<b>57</b>	<b>50</b>	43	74	58	50
TOTAL	NO.	961	1,688	699	110	199	35	1,043	3,692
	%	100	100	100	100	100	100	100	100
MEAN RD		<b>6.4</b>	<b>12</b>	<b>11.9</b>	<b>11.3</b>	13.4	15.3	12.2	10.6

In the mean time, the mean RDs for other encoding types differ only slightly among one another (e.g., 12 clauses for zero-marked objects; 11.9 clauses for *lul*-marked objects). In other words, unlike the mean RD for subjects, the mean RD for objects does not distinguish between object encoding types, except for objects of zero anaphor. This result is somewhat unexpected because the mean RD for *lul*-marked objects was expected to be greatest among other encoding types, at least greater than that of zero-marked objects, due to greater activation cost for *lul*-marked objects. Furthermore, this tendency reported in Korean differs from the tendency reported in Japanese, in that the mean RD distinguishes between object encoding types in Japanese, while it does not in Korean. For instance, in Korean, the mean RD for *lul*-marked objects (11.9 clauses) is almost the same as that of *nun*-marked objects (11.3 clauses), whereas in Japanese, the mean RD for *o*-marked objects (14.4 clauses) is much greater than *wa*-marked objects (11.2 clauses) (Shimojo 2005). The data in Table 5.17 also indicates that except for object of zero anaphor, object referents are generally in the anaphorically non-salient side of the index regardless of their encoding types (e.g., *lul*-marked objects: 57% of the

total for RD 6-NPM; *nun*-marked objects: 58% of the total for RD 6-NPM;  $\chi^2=1.857$ ,  $P=0.173$ ). Simply put, according to Table 5.16 and Table 5.17, in conversational Korean, the measurement of RD does not differentiate between object encoding types, though it does between objects of zero anaphor and those of other encoding types.

As repeatedly pointed out, the degree of activation can also be measured by the anaphoric frequency of a referent, though to the best of my knowledge, no studies have examined the speaker's selection of object encoding types with regard to the AF of object referents in conversational Korean. In an attempt to see if the anaphoric saliency, in the sense of anaphoric frequency, plays a role in the speaker's selection of object encoding types, I examine the object encoding types with relation to the measurement of AF. The results are summarized in Table 5.18 and Table 5.19. Table 5.19 presents the tokens based on the two groups of AF (0-3 vs. 4-10) along with the mean AF values.

In Table 5.18, I find that the mean AF for objects of zero anaphor (1.9 times) is greater than the mean AF for the other encoding types (e.g., 0.9 times for *nun*-marked objects). This distributional pattern is expected, given that the more frequently an entity is mentioned in preceding discourse contexts, the more activated the entity is at the time of utterance. In other words, entities that are frequently given in anaphoric discourse contexts, hence are more activated, may be identified at little activation cost, so they tend not to appear overtly in a clause. Table 5.19, where AF is generalized in terms of the two ranges of AF, also suggests, though not as strongly as RD, that object referents of zero anaphor (19% of the total for AF 4-10) appear more frequently in the anaphorically salient side of the index than referents of other object encoding types (e.g., *lul*-marked objects: 7% of the total for AF 4-10;  $\chi^2=44.166$ ,  $P < 0.001$ ).

Yet, except for the mean AF for objects of zero anaphor, there is basically no difference in the mean AF among the other encoding types (e.g., 0.9 times for zero-marked objects; 1 time for *lul*-marked objects). This result does not comply with the

Table 5.18: Object Anaphoric Frequency (AF) and Object Marking

OBJ.ET OBJ.AF		ZERO- ANAPHOR	NP-Ø	NP-lul	NP-num	NP-to	NP-etc.	SUB- TOTAL	TOTAL
0 time	NO.	230	892	366	55	117	24	562	1,684
	%	14	53	22	3	7	1	33	100
1 time	NO.	248	414	151	33	40	3	227	889
	%	28	47	17	4	4	0	26	100
2 times	NO.	184	182	87	10	18	4	119	485
	%	38	38	18	2	4	1	25	101
3 times	NO.	119	108	44	5	13	3	65	292
	%	41	37	15	2	4	1	22	100
4 times	NO.	89	51	23	6	6	0	35	175
	%	51	29	13	3	3	0	20	100
5 times	NO.	47	22	21	1	4	1	27	96
	%	49	23	22	1	4	1	28	100
6 times	NO.	33	12	6	0	1	0	7	52
	%	63	23	12	0	2	0	13	99
7 times	NO.	9	6	1	0	0	0	1	16
	%	56	38	6	0	0	0	6	100
8 times	NO.	2	1	0	0	0	0	0	3
	%	67	33	0	0	0	0	0	100
TOTAL	NO.	961	1,688	699	110	199	35	1,043	3,692
	%	26	46	19	3	5	1	28	100
MEAN AF		<b>1.9</b>	<b>0.9</b>	<b>1.0</b>	<b>0.9</b>	0.8	0.7	0.9	1.2

initial assumption, i.e., the mean AF for *lul*-marked objects, which I assume is less activated, hence less accessible, would be smaller than the mean AF for objects of other encoding types. The data in Table 5.19 also show that except for zero anaphor, in general, object referents of the other encoding types do not commonly occur in the anaphorically salient side of the index (e.g., zero-marked objects: 5% of the total for AF 4-10; *ka*-marked objects: 7% of the total for AF 4-10; zero vs. *ka*:  $\chi^2=3.418$ ,  $P=0.0645$ ). This suggests that the measurement of AF, like that of RD, does not help distinguish among object markings in conversational Korean.



Table 5.19: Object AF (0-3, 4-10) and Object Marking

OBJ.ET OBJ.AF		ZERO- ANAPHOR	NP-Ø	NP-lul	NP-num	NP-to	NP-etc.	SUB- TOTAL	TOTAL
<i>0-3</i>	NO.	781	1,596	648	103	188	34	973	3,350
	%	<b>81</b>	<b>95</b>	<b>93</b>	<b>94</b>	94	97	93	91
<i>4-10</i>	NO.	180	92	51	7	11	1	70	342
	%	<b>19</b>	<b>5</b>	<b>7</b>	<b>6</b>	6	3	7	9
TOTAL	NO.	961	1,688	699	110	199	35	1,043	3,692
	%	100	100	100	100	100	100	100	100
MEAN AF		<b>1.9</b>	<b>0.9</b>	<b>1.0</b>	<b>0.9</b>	0.8	0.7	0.9	1.2

In an additional attempt to study if the choice of object encoding types is related to the anaphoric saliency of objects, either in RD or in AF, I measured RD and AF more precisely in terms of word order variations, and then examined the object encoding types with regard to the specified RD and AF. Yet I did not find any correlation between them.

To summarize this section, the present study shows, with the results in Table 5.16 and Table 5.17 as well as Table 5.18 and Table 5.19, that unlike in colloquial Japanese, in conversational Korean, the anaphoric saliency, either by anaphoric recency (RD) or by anaphoric frequency (AF), of object referents does not play a role in the speaker's selection of object encoding types, except for objects of zero anaphor. More specifically, the mean RD for objects of zero anaphor appeared to be small, compared to other encoding types, and this result is attributed to such a tendency that entities frequently occurring in anaphoric discourse contexts are activated, hence identifiable, so they are commonly realized as a zero anaphor in a clause. On the other hand, there is nearly no difference in the mean RD for other encoding types. In addition, the mean AF for

objects of zero anaphor appeared to be greater than that of other encoding types, yet there is not much difference in the mean AF among other encoding types (e.g., the mean AF for zero-marked objects vs. the mean AF for *lul*-marked objects). Based on these results, I argue that unlike for subjects, the anaphoric saliency does not influence the speaker's choice of object markings, though it distinguishes objects of zero anaphor from objects of other encoding types.

#### 5.4.9 Cataphoric saliency

Given that in general, information that does not persist loses its importance in subsequent context, RP (*referential persistence*) has been proposed as a way of measuring the degree of decay of information in the cataphoric context (Givón 1983). In fact, Givón (1983:15) points out that RP is basically a means to measure importance in the sense that “[m]ore important discourse topics appear more frequently in the register, i.e., they have a higher probability of persisting longer in the register after a relevant measuring point.” With regard to the measurement of RP, Shimojo (2005) argues that the cataphoric saliency should be discussed by measuring how far a given referent persists without an interruption as well as how frequently a given referent appear in the cataphoric context. Following Shimojo's (2005) argument, the cataphoric saliency of an object referent is examined by both its persistence without an interruption and its frequency. In order to distinguish the former from the latter, the former is labeled as RP (i.e., uninterrupted), and the latter is labeled as RP-*f* (i.e., frequency).

In Section 4.5.8, it was noted that in conversational Korean, the cataphoric saliency (RP/RP-*f*) does not play a role in the speaker's selection of subject encoding types, though it does in spoken Japanese (Shimojo 2005). In this section, I examine if the cataphoric saliency influences the choice of object encoding types in conversational Korean, with an assumption that the cataphoric saliency of object referents plays a

role in the speaker’s choice of object encoding types. More specifically, I examine all object tokens with relation to both RP and RP-*f*, and the results are summarized in Table 5.20 and Table 5.21 as well as Table 5.22 and Table 5.23. Table 5.21 and Table 5.23 present the tokens in terms of the two categories along with the mean RP and mean RP-*f* values respectively.

Table 5.20: Object RP and Object Marking

OBJ.ET OBJ.RP		ZERO- ANAPHOR	NP-Ø	NP-lul	NP-nun	NP-to	NP-etc.	SUB- TOTAL	TOTAL
0 cl.	NO.	646	1,190	488	63	141	24	716	2,552
	%	25	47	19	2	6	1	28	100
1 cl.	NO.	202	291	120	22	37	7	186	679
	%	30	43	18	3	5	1	27	100
2 cls.	NO.	71	118	54	15	11	0	80	269
	%	26	44	20	6	4	0	30	100
3 cls.	NO.	21	41	18	7	5	3	33	95
	%	22	43	19	7	5	3	35	100
4 cls.	NO.	12	32	12	1	1	0	14	58
	%	21	55	21	2	2	0	24	100
5 cls.	NO.	6	11	3	0	4	0	7	24
	%	25	46	13	0	17	0	29	100
6 cls.	NO.	2	2	2	2	0	0	4	8
	%	25	25	25	25	0	0	50	100
7 cls.	NO.	0	2	2	0	0	1	3	5
	%	0	40	40	0	0	20	60	100
8 cls.	NO.	1	1	0	0	0	0	0	2
	%	50	50	0	0	0	0	0	100
TOTAL	NO.	961	1,688	699	110	199	35	1,043	3,692
	%	26	46	19	3	5	1	28	100
MEAN RP		<b>0.5</b>	<b>0.5</b>	<b>0.5</b>	<b>0.8</b>	0.5	0.7	0.6	0.5

In Table 5.20, I find that overall, the mean RP was low for all object encoding types. The low mean RP across all encoding types indicates that in conversational Korean, object referents do not persist long, at least without an interruption, regardless of their encoding types. This tendency for object referents is similar to the tendency for

Table 5.21: Object RP (0-3, 4-10) and Object Marking

OBJ.ET OBJ.RP		ZERO- ANAPHOR	NP-Ø	NP-lul	NP-num	NP-to	NP-etc.	SUB- TOTAL	TOTAL
<i>0-3</i>	NO.	940	1,640	680	107	194	34	1,015	3,595
	%	<b>98</b>	<b>97</b>	<b>97</b>	<b>97</b>	97	97	97	97
<i>4-10</i>	NO.	21	48	19	3	5	1	28	97
	%	<b>2</b>	<b>3</b>	<b>3</b>	<b>3</b>	3	3	3	3
TOTAL	NO.	961	1,688	699	110	199	35	1,043	3,692
	%	100	100	100	100	100	100	100	100
MEAN RP		<b>0.5</b>	<b>0.5</b>	<b>0.5</b>	<b>0.8</b>	0.5	0.7	0.6	0.5

subject referents, in that both subject and object referents do not persist long without an interruption, regardless of their encoding types. Table 5.21, where RP is generalized in terms of the two ranges of RP, also indicates that unlike subject referents, overall, object referents are in the cataphorically non-salient side of the index regardless of their encoding types (e.g., zero anaphor: 98% of the total for RP 0-3; *ka*-marked subjects: 97% of the total for RP 0-3;  $\chi^2=0.489$ ,  $P=0.4845$ ). This further suggests that in conversational Korean, there is no correlation between RP and the selection of object encoding types. This distributional pattern not only differs from what was initially expected but it also differs from what was observed in the previous studies of Japanese (e.g., Shimojo 2005), where RP for *o*-marked objects is much greater than RP for zero-marked objects, i.e. *o*-marked objects persist much longer than zero-marked objects do. Simply put, unlike the accusative *o* in Japanese, the accusative *lul* in Korean is not associated with thematic prominence represented by object NPs.

As was stated above, in the present study, the cataphoric saliency was also measured by RP-*f* (frequency). In an attempt to see if RP-*f* distinguishes between object

encoding types more clearly, I also examined object encoding types with relation to RP-*f*, and the data in Table 5.22 and Table 5.23 show the results.

Table 5.22: Object RP-*f* and Object Marking

OBJ.ET		ZERO-ANAPHOR	NP-Ø	NP-lul	NP-nun	NP-to	NP-etc.	SUB-TOTAL	TOTAL
OBJ.RP- <i>f</i>									
0 time	NO.	379	813	298	47	99	18	462	1,654
	%	23	49	18	3	6	1	28	100
1 time	NO.	216	353	156	21	36	7	220	789
	%	27	45	20	3	5	1	28	100
2 times	NO.	160	207	99	15	25	1	140	507
	%	32	41	20	3	5	0	28	101
3 times	NO.	91	133	55	15	17	5	92	316
	%	29	42	17	5	5	2	29	100
4 times	NO.	58	87	39	6	12	1	58	203
	%	29	42	19	3	6	0	29	100
5 times	NO.	33	41	29	2	5	0	36	110
	%	30	38	26	2	5	0	33	101
6 times	NO.	18	23	17	1	2	2	22	63
	%	29	37	27	2	3	3	35	101
7 times	NO.	4	18	6	1	3	1	11	33
	%	12	55	18	3	9	3	33	100
8 times	NO.	2	9	0	2	0	0	2	13
	%	15	69	0	15	0	0	15	99
9 times	NO.	0	3	0	0	0	0	0	3
	%	0	100	0	0	0	0	0	100
10 times	NO.	0	1	0	0	0	0	0	1
	%	0	100	0	0	0	0	0	100
TOTAL	NO.	961	1,688	699	110	199	35	1,043	3,692
	%	26	46	19	3	5	1	28	100
MEAN RP- <i>f</i>		<b>1.4</b>	<b>1.2</b>	<b>1.4</b>	<b>1.4</b>	1.2	1.3	1.4	1.3

According to Table 5.22, just as with RP, there is not much difference in the mean RP-*f* among object encoding types, which makes RP-*f* not a good means to differentiate between object encoding types in conversational Korean. For example, the mean RP-*f* for *lul*-marked objects (1.4 times) is only slightly greater than the mean RP-*f* for zero-marked objects (1.2 times). The data in Table 5.23, where RP-*f* is generalized in

Table 5.23: Object RP-*f* (0-3, 4-10) and Object Marking

OBJ.ET OBJ.RP- <i>f</i>		ZERO- ANAPHOR	NP-Ø	NP-lul	NP-nun	NP-to	NP-etc.	SUB- TOTAL	TOTAL
0-3	NO.	846	1,506	608	98	177	31	914	3,266
	%	<b>88</b>	<b>89</b>	<b>87</b>	<b>89</b>	89	89	88	88
4-10	NO.	115	182	91	12	22	4	129	426
	%	<b>12</b>	<b>11</b>	<b>13</b>	<b>11</b>	11	11	12	12
TOTAL	NO.	961	1,688	699	110	199	35	1,043	3,692
	%	100	100	100	100	100	100	100	100
MEAN RP- <i>f</i>		<b>1.4</b>	<b>1.2</b>	<b>1.4</b>	<b>1.4</b>	1.2	1.3	1.4	1.3

terms of the two ranges of RP-*f*, also suggest that unlike subject referents, overall, referents of object are in the cataphorically non-salient side of the index regardless of their encoding types (e.g., zero-marked subjects: 89% of the total for RP-*f* 0-3; *ka*-marked subjects: 87% of the total for RP-*f* 0-3;  $\chi^2=2.441$ ,  $P=0.1182$ ). This distributional pattern for object encoding types, with relation to RP-*f*, differs from what was initially expected; the mean RP-*f* for *lul*-marked objects was expected to be greater than the mean RP-*f* for zero-marked objects, based on an assumption that the accusative *lul* represents thematic prominence encoded in referents of object NPs. Furthermore, this distributional pattern does not comply with what was reported in conversational Japanese, where the measurement of RP-*f* distinguishes between object encoding types (Shimojo 2005).

In short, I cannot make anything of the data presented in Table 5.22 and Table 5.23. In addition, I do not have a plausible explanation for this cross-linguistic difference between the two languages. I can only suggest, based on the data in Table 5.22 and Table 5.23, that in conversational Korean, the cataphoric saliency does not distinguish

among object encoding types, whereas it does in conversational Japanese. In other words, the cataphoric saliency as defined by the measurements above does not play a role in the speaker's selection of object encoding types in Korean, though it does in Japanese.

As I did with subjects, in an attempt to investigate if the cataphoric saliency of object referents influences the choice of object encoding types in conversational Korean, I measured the cataphoric saliency further in detail with respect to word order variations, and then examined the selection of object encoding types with relation to the specified cataphoric saliency. Yet I did not find any correlation between the choice of object encoding types and the cataphoric saliency.

To summarize the present section, information that does not persist loses its importance in the subsequent context, while information that persists maintains its importance in the subsequent context. Following Givón (1983) and Shimojo (2005), both RP and RP-*f* are used to measure the cataphoric saliency of object referents. Based on the assumption that a greater RP and a greater RP-*f* suggest a greater thematic importance, the mean RP and the mean RP-*f* were expected to be greater for *lul*-marked objects than for zero-marked objects. However, the data in Table 5.22 and Table 5.23 do not suggest that the cataphoric saliency plays a role in the selection of object encoding types, even for objects of zero anaphor. This differs from the tendency that Shimojo (2005) observes in his study of Japanese conversation, where the cataphoric saliency distinguishes between object encoding types. Although I do not know why Korean differs from Japanese in this regard, the findings presented in the present section suggest that unlike Japanese, prominence of information in the sense of the cataphoric saliency is not related to the speaker's selection of object encoding types in conversational Korean.

#### 5.4.10 Contrast

As previously stated in Section 4.5.9, the term ‘contrast’ has frequently been discussed in many previous studies (Halliday 1967; Chafe 1976; Kim 1990; Labmrecht 1994; Choi 1997; Lee 2003, *inter alia*). However, the definition provided for the term ‘contrast’ has not necessarily been clear-cut, mostly because, in a broad sense, almost any element in a clause can inherently be contrastive, as Lambrecht (1994) points out. Clancy & Downing (1987) argue that the notion of ‘contrast’ should be defined carefully due to varying degrees of contrastivity which is encoded in the elements in the sentence. They define the notion ‘contrast’ in two different ways, in terms of directiveness; directly contrastive and indirectly contrastive. The present study adopts the former only, i.e., directly contrastive, because the definition for the former seems to be clear-cut, whereas that of the latter does not seem to be as clear. Along with the term ‘contrast’, the notion of ‘exhaustive-listing’ (Kuno 1972; 1973), also known as ‘narrow focus’ (Lambrecht 1994), has often been discussed because an entity that is referred to as ‘exhaustive-listing’ contrasts with all other entities, in that only the entity that is singled out exhaustively meets the truth condition represented by the sentence. Hence, the present study includes tokens of an exhaustive-listing into the entire body of contrastive elements.

For the purpose of this dissertation, the notion of ‘contrast’ is defined mostly based on Clancy & Downing’s (1987) and Kuno’s (1972; 1973) descriptions given for the term ‘contrast’, as was noted in Section 4.5.9. More specifically, the term ‘contrast’ is categorized largely into two different subtypes based on their contrastivity: directly contrastive (Clancy & Downing 1987) and exhaustive-listing (Kuno 1972; 1973). The subtype of ‘directly contrastive’ is further divided into ‘parallel activities/states’, ‘action/state reaction’, and ‘exhaustive-listing’, and the subtype of ‘exhaustive-listing’ is further categorized into ‘exhaustive-listing’ and ‘directly contrastive & exhaustive-listing’.



As repeatedly stated, the ‘parallel activities/states’ type of directly contrastive most often occurs in the format of “A does/is X, but B does/is Y” (Clancy & Downing 1987). The following is an example of ‘parallel activities/states’ type of contrastiveness.

- (5.43) *ccacangmyen-un an mek-ko, kunayng pap mek-ess-ta,*  
 black.noodle-top NEG eat-and normal meal eat-PST-DCL  
*olaynnaman-ey.*  
 longtime-at  
 ‘(I) didn’t eat a black noodle, but (I) just ate a normal rice meal, (which I haven’t had for a while).’

In (2.4), repeated here as (5.43), the underlined objects, *ccacangmyen* ‘black noodle’ and *kunayng pap* ‘normal rice meal’, represent the contrastive opposition, in the sense that the speaker “doesn’t do X (not eating *ccacangmyen*), but does Y (eating *kunayng pap*).” Therefore, this utterance is counted as an example of directly contrastive in the sense of ‘parallel activities/states’. In (5.43), the first contrastive element is coded with *nun*, while the second one is marked with a zero particle. The following is another instance of ‘parallel activities/states’ subtype of direct contrastiveness.

- (5.44) A1: *toy-l ay an toy-l ay-lul ttak cenghaykacikwu,*  
 be-FUT student NEG be-FUT student-ACC just decide.and  
 ‘He distinguishes students with academic potential from those without academic potential.’

- A2: [*an toy-l ay-nun ayay pocito anh-ko,*  
 NEG be-FUT student-TOP even look.at NEG-and  
 ‘He doesn’t even look at those students without academic potential.’

- A3: [*toy-l ay-tul-un mak ccoakaciko, cincca kongpwu-Ø*  
 be-FUT student-PL-TOP hard force.by really study-( )  
*sikhinunkeya.*  
 let  
 ‘(He) pushes those students with academic potential, and makes them work very hard.’

In (5.44), the speaker was talking about her former high school teacher, who divided his students into two groups based on their academic potential. One group is composed of students who have academic potential, as shown in utterance A2, and the other group is made up of students who do not have academic potential, as seen in utterance A3. More importantly, in (5.44), underlined objects, *an toyl ay* ‘those lacking academic potential’ in utterance A2 and *toyl aytul* ‘those demonstrating academic potential’ in utterance A3 together exemplify a ‘parallel activities/states’ type of contrastiveness because these paired entities represent a contrastive opposition. Specifically, the two utterances of A2 and A3 fit into “(the teacher) does X (not look at students lacking academic potential), but (the teacher) does Y (push students demonstrating academic potential to study hard),” so they are viewed as instances of the ‘parallel activities/states’ type of direct contrastiveness, following Clancy & Downing’s (1987) definition. In (5.44), both contrastive elements, i.e., *an toyl ay* ‘students without academic potential’ in utterance A2 and *toyl aytul* ‘students with academic potential’ in utterance A3, are coded with *nun*.

The ‘action/state reaction’ type of directly contrastive typically appears “when the speaker is developing a plot sequence containing successive activities involving different participants” (Clancy & Downing 1987:37). Shimojo (2005:176) also points out that “[t]he paired elements in ‘action/state reaction’, unlike those in ‘parallel activities/states’, do not appear in parallel; but the first element is presented, and then the second element is (re)introduced into the state of affairs with regard to the first element.” An instance of ‘an action/state reaction’ type of directly contrastiveness is given in the utterance in (5.45).

- (5.45) A1: *mwo-Ø hay-ss-e? kwuntay-eyse kumyen kongpwu-Ø hay-ss-e?*  
 what-( ) do-PST-Q military-LOC then study-( ) do-PST-Q  
 ‘What did you do? Did you study during your military service?’

B1: [*kwuntay-eyse mwulken-Ø phal-ass-cyo. ppang-Ø phal-ss-cyo.*]<sub>1st</sub>  
 military-LOC things-( ) sell-PST-SEM pastry-( ) sell-PST-SEM  
 ‘I used to sell things during my military service. I used to sell pastries.’

A2: [*kongpwu-nun an hay-ss-ci?*]<sub>2nd</sub>  
 study-TOP NEG do-PST-Q  
 ‘You didn’t study, did you?’

In utterance A1 in (5.45), the speaker A asks the speaker B what he did during his military service, and particularly whether or not he studied during his military service. In utterance B1 in (5.45), the speaker B answers the speaker A by stating he used to sell things, particularly pastries, during his military service. In reaction to this, the speaker A again asks the speaker B if he did study during his military service. In the ‘action/state reaction’ type of the contrastiveness, only the second element of the contrastive pair is taken as an instance of direct contrastiveness because the contrastive opposition does not arise until the second element of the pair, as a reaction to the first element, appears in the discourse (Shimojo 2005). In (5.45), the underlined second element of the contrastive pair in the second set of brackets occurs as a reaction to the first element of the contrastive pair in the first set of brackets, and it is coded with *nun*-marking.

Along with the term ‘contrast’, the notion of ‘exhaustive-listing’ (Kuno 1972; 1973), also known as ‘narrow focus’ (Lambrecht 1994), has frequently been discussed in many studies of contrastiveness, because entities that are viewed as ‘exhaustive or exclusive’ essentially contrast with all other entities, in that only the entity that is singled out exhaustively is the focus of the sentence. The following is an example that shows an ‘exhaustive-listing’ type of contrastiveness from the data of this study.

(5.46) *kulayse cikum jeyi-visa-lo-nun an toy. [eyph-visa-lul pat-aya*  
 so now J-VISA-with-TOP NEG do F-VISA-ACC get-should  
*toyketun].*  
 do.SEM

‘So this J-VISA is not useful. It is an F-VISA that (I) must acquire.’

Before the utterances in (5.46), they were talking about the US visa they should acquire in order to continue to study as a regular student at the University at Buffalo, where they are currently registered as an exchange student. After the conversation, they finally learned that among various types of US visas, the F-VISA would be the only visa type to allow them to study at school legally. In other words, the underlined object *F-VISA*, which is in question, is the only focus of the sentence in (5.46). Therefore, this utterance, following Kuno’s (1972; 1973), as well as Lambrecht’s (1994), definition of contrastiveness, is categorized as an example of ‘exhaustive-listing’ type of contrastiveness.

Unlike subject NPs, there is no instance of ‘directly contrastive and exhaustive-listing’ type of contrastiveness reported for object NPs.

In short, the present study categorizes the notion of ‘contrastiveness’ largely into two different subtypes: ‘directly contrastive’ and ‘exhaustive-listing’. Following approaches used by Clancy & Downing (1987), the ‘directly contrastive’ type of contrastiveness is classified more precisely, based on its intended description, into two different sub-types: ‘parallel activities/states’ and ‘action/state reaction’.

Based on the indices of contrastiveness defined above, as with subject NPs, this study discusses the speaker’s selection of object markings in conversational Korean with regard to the contrastiveness of object NPs, with an assumption that due to informational prominence, which are incurred by contrastiveness identification in addition to referent identification, entities which are contrastive in one way or another are explicitly identified with an overt marking more commonly than entities which are not contrastive. Simply put, I assume that the use of overt-marking would be more common for encoding objects which are contrastive than for encoding those which are not, while the use of zero-marking would be less common for coding contrastive objects than for

coding non-contrastive objects. In fact, Shimojo (2005:127) points out the complementary nature of two encoding types in his study of Japanese conversation: *wa*-marking is associated with direct contrastiveness (82% of the total), and zero-marking is associated with non-contrastiveness (83% of the total). With this assumption, I will first discuss the choice of object markings simply based on whether or not objects are contrastive. Then I will discuss the choice of object markings more precisely with regard to each subtype of contrastiveness that is defined above. Table 5.24 shows the overall token distribution of object markings based on a simple distinction of contrastive or non-contrastive.

Table 5.24: Object Contrast and Object Marking

OBJ.MARK OBJ.CONTRAST		NP-Ø	NP-lul	NP-nun	NP-to	NP-etc.	SUB- TOTAL	TOTAL
¬contrast	NO.	1,670	680	90	188	35	993	<b>2,663</b>
	%	<b>63</b>	<b>26</b>	<b>3</b>	7	1	37	100
contrast	NO.	18	19	20	11	0	50	<b>68</b>
	%	<b>26</b>	<b>28</b>	<b>29</b>	16	0	74	100
TOTAL	NO.	1,688	699	110	199	35	1,043	2,731
	%	62	26	4	7	1	38	100

In Table 5.24, I find that the number of objects which are counted as contrastive (68 tokens, approximately 2% of all overt objects) is much smaller than that of objects which are viewed as non-contrastive (2,663 tokens, roughly 98% of all overt objects). I also find that the use of zero-marking (63%) is preferred for coding objects which are not contrastive, whereas the use of overt marking (74%) is preferred for objects which are contrastive. Furthermore, unlike subjects, zero-marked objects (26%), *lul*-marked objects (28%) and *nun*-marked objects (29%) take a roughly equal portion of all objects that are contrastive. More importantly, the occurrence of zero-marking is much lower for coding objects that are contrastive (26%) than it is for coding objects that are not

contrastive (63%). On the other hand, the occurrence of *nun*-marking is much higher for coding objects that are contrastive (29%) than it is for coding objects that are not contrastive (3%) (zero-marking vs. *nun*-marking:  $\chi^2=146.23$ ,  $P < 0.001$ ). This result complies with what was reported in Shimojo (2005:127), in that the use of zero-marking is associated with a non-contrastive element, and the use of *nun*-marking is associated with a contrastive element. Furthermore, this result suggests that the particle *nun* functions as marking the contrastiveness encoded in object NPs.

As for the post-nominal particle *nun* in Korean, Choi (1996; 1997) claims that the particle *nun* is not a topic marker but a contrastive marker, though it has widely been accepted that the particle *nun* functions mainly as marking a (grammatical) topic of the sentence (Sohn 1999; Lee 2003). Choi (1997) further argues that the (grammatical) topic of the sentence is not encoded by the particle *nun* but by scrambling, i.e., the topic position of a sentence. Although Choi's (1997) claim for the particle *nun* deserves recognition, I see her claim to be untenable, especially because out of all *nun*-marked objects in the data, only 20 tokens of objects (roughly 18% of the total) are viewed as contrastive, and 90 tokens of objects (about 82% of the total) are counted as non-contrastive. Simply put, Choi's (1997) analysis does not explain 90 tokens of *nun*-marked objects that are categorized as non-contrastive. Nevertheless, it is certain that one discourse function of the particle *nun* is to mark contrastiveness, though it is not its only function.

As for the particle *wa* in Japanese, which corresponds to the particle *nun* in Korean, Shimojo (2005:127) claims that "a sentence containing a *wa*-marked NP is inherently contrastive due to the referent denoted by the NP, which is singled out from the comment provided about the referent as well as other propositions present anaphorically or exophorically." Shimojo (ibid.) further points to the complementary nature of the two encoding types in his study of Japanese conversation; *wa*-marking is associated

with direct contrastiveness (82% of the total), and zero-marking is associated with non-contrastiveness (83% of the total). This result which is reported in Shimojo (2005) reflects the tokens of direct contrastiveness only, so it cannot be directly compared to the result reported in the present study, which reflects the tokens of exhaustive-listing, as well as the tokens of direct contrastiveness. Still, it can be said that the data in Table 5.24 supports Shimojo's (ibid.) claim, in that the use of zero-marking is certainly most common for non-contrastive objects (63%). On the other hand, the use of zero-marking (26%), *lul*-marking (28%) and *nun*-marking (29%) all seem to be equally common for coding contrastive objects. This suggests that in Korean, *nun*-marking is not the only encoding type available for coding contrastiveness, but the contrastiveness may be encoded by encoding types other than *nun*-marking (e.g., *lul*-marking).

Based on what has been discussed so far, the complementary nature of the three object encoding types appears to be clear. That is to say, the particles *nun* and *lul* tend to mark the contrastiveness encoded in object NPs, whereas a zero particle tends to mark the non-contrastiveness encoded in object NPs. In spite of this complementary nature with regard to the contrastiveness marking, the distinction between *lul*-marking and *nun*-marking still remains unclear since *lul*-marking (28%) is used for coding contrastive objects almost as frequently as *nun*-marking (29%). In an attempt to find out if the two encoding types mark different senses of contrastiveness, I had all contrastive objects categorized as such into four different subtypes, as was stated earlier in the present section, and then examined the choice of object markings with relation to these four subtypes. The results are summarized in Table 5.25.

In Table 5.25, I observe a clear distinction between *lul*-marking and *nun*-marking as encoders of contrastiveness. More specifically, the accusative *lul* is mostly used to encode an 'exhaustive-listing' type of contrastiveness (41%) as exemplified in the utterance in (5.46), whereas no instance of *nun*-marking is reported to encode an 'exhaustive-listing'

Table 5.25: Detailed Object Contrast and Object Marking

OBJ.MARK OBJ.CONTRAST		NP-Ø	NP-lul	NP-nun	NP-to	NP-etc.	SUB- TOTAL	TOTAL
¬contrast	NO.	1,670	680	90	188	35	993	2,663
	%	63	26	3	7	1	37	100
exhaustive	NO.	1	7	0	9	0	16	17
	%	6	<b>41</b>	<b>0</b>	53	0	94	100
parallel	NO.	16	12	15	2	0	29	45
	%	<b>36</b>	<b>27</b>	<b>33</b>	4	0	64	100
reaction	NO.	1	0	5	0	0	5	6
	%	17	<b>0</b>	<b>83</b>	0	0	83	100
TOTAL	NO.	1,688	699	110	199	35	1,043	2,731
	%	62	26	4	7	1	38	100

type of contrastiveness. On the other hand, ‘action/state reaction’ type of direct contrastiveness is most likely coded by *nun*-marking (83%), as shown in utterance A2 in (5.45), and there is no instance of *lul*-marking used for coding such a type of contrastiveness. These results indicate that just like the nominative *ka*, the accusative *lul* exclusively marks an ‘exhaustive-listing’ type of contrastiveness that may be encoded in object NPs, and on the other hand, the particle *nun* exclusively marks an ‘action/state reaction’ type of direct contrastiveness (*lul*-marking vs. *nun*-marking:  $\chi^2=12.0$ ,  $P < 0.001$ ).

In Table 5.25, I also find that unlike subjects, the use of zero-marking (36%) is most common, among other encoding types, for objects that are in a ‘parallel activities/states’ type of direct contrastiveness, and the occurrence of *lul*-marked objects (27%) is a little lower than that of *nun*-marked objects (33%). The use of zero-marking and *nun*-marking for objects in ‘parallel activities/states’ type of contrastiveness is exemplified in (5.43), where the first object is coded with the particle *nun*, and the



second one is marked with a zero particle. The use of only the *nun*-marking for objects in ‘parallel activities/states’ type of contrastiveness is exemplified in (5.44), where both objects are coded with the post-nominal particle *nun*.

Lastly, due to informational prominence of contrasted referents, the occurrence of zero-marking is low for coding objects that are viewed as contrastive, regardless of which subtype of contrastiveness the objects are categorized into, whereas it is high for coding objects that are not contrastive (zero-marking vs. *lul*-marking:  $\chi^2=146.23$ ,  $P < 0.001$ ).

To summarize this section, contrastiveness seems to play a role in the speaker’s selection of object markings in conversational Korean. Overall, according to the data in Table 5.24, the occurrence of zero-marking is high for coding objects that are not viewed as contrastive, whereas it is low for coding objects that are regarded as contrastive. Instead, *lul*-marking and *nun*-marking is most commonly used to encode objects that are contrastive, due to the contrastiveness, which needs to be identified in addition to referent identification, hence informational prominence. Furthermore, the data in Table 5.25 points to the complementary nature of object markings: *lul*-marking is associated with the ‘exhaustive-listing’ type of contrastiveness; *nun*-marking is associated with the ‘action/state reaction’ type of contrastiveness; zero-marking, *lul*-marking and *nun*-marking are almost equally associated with the ‘parallel activities/states’ type of contrastiveness; zero-marking is not commonly used for objects that are in contrast, regardless of their subtype.

Based on the results summarized in Table 5.24 and Table 5.25, I argue that contrastive objects are assumed to be informationally more prominent, and the use of overt marking, either with *lul* or *nun*, is common to mark informational prominence that is encoded in contrastive objects. I further argue that when contrastiveness plays a role in the choice of object markings, the speaker’s selection of one object form over another

is not an arbitrary process but a systematic process.

#### 5.4.11 HA ‘do’ constructions

Fujii & Ono (2000:8-10) point out that in spoken Japanese, direct object NPs most often occur without *o*-marking when they form lexicalized or idiomaticized compound expressions together with their verbs. Lee & Thompson (1989) also report that a similar tendency is observed in spoken Korean. More specifically, Lee & Thompson (ibid.) state that the identity of the nominal which is a part of a “normative activity” expression is not of concern even when the referent in question is specific, and such a nominal is likely to appear without *lul*-marking.

It has been widely accepted that Korean has a productive light verb construction,<sup>19</sup> in which verbal nouns (e.g., Sino-Korean nouns) appear as an object of the verb *hata* ‘do’ (Han 1999; Sohn 1999; Choi 2003).<sup>20</sup> These verbal noun object NPs that are combined with the so-called light verbs predominantly occur without an accusative case marking particle *lul*, though they are occasionally coded with an accusative *lul*.<sup>21</sup> As a matter of fact, Kim (2006) also points out that an object which is viewed as a part of a lexicalized or compound expression is often coded with an accusative *lul*, though it is most commonly coded with a zero particle. For example, the objects *sayngkak* ‘thought’ in (5.47) and *kongpwu* ‘study’ in (5.48) form compound expressions together with a *hata* verb. Yet only the first one occurs without a particle while the second one

---

<sup>19</sup>The notion of ‘light-verb’ was introduced to specifically separate verbs that partially or completely lack their own  $\theta$ -marking capabilities from other verbs that have such capabilities (Grimshaw & Mester 1988).

<sup>20</sup>Japanese also has a light verb, *suru* ‘do’, which corresponds to *hata* ‘do’ in Korean (Grimshaw & Mester 1988)

<sup>21</sup>One may claim that by its definition, the verb, *hata* ‘do’, is no longer viewed as a light verb if a verbal noun, which is combined with the verb, *hata* ‘do’, is overtly coded with the accusative *lul*. The verb *hata* ‘do’ can be labeled as a heavy *hata* ‘do’ when the verbal noun that the verb is combined with is coded with the accusative *lul*, in order to distinguish it from a light *hata* ‘do’. Yet, for the purpose of this dissertation, the verb *hata* ‘do’ is labeled not as a light verb but as a *hata*-verb.

occurs with a particle, that is, the accusative *lul*, just like the object in (5.49), *khayllina ayakho* ‘Kelly or Ayako’, which is also marked with the accusative *lul*.

(5.47) *na-to kulehkey sayngkak-Ø hay-ss-e.*  
 1SG-also so thought-( ) do-PST-SEM  
 ‘I also thought like that.’

(5.48) *tayhakkyo-ttay-pwuthe kongpwu-lul hay-ss-na?*  
 college-time-from study-ACC do-PST-Q  
 ‘I wonder if (he) studied since (he) was a college student?’

(5.49) *kkwum-eyse manyakey mwo khaylli-na ayakho-lul manna-myen,*  
 dream-in in.case DP Kelly-or Ayako-ACC meet-if  
 ‘Well, if I meet Kelly or Ayako in my dream, ...’

In the present section, I explore the choice of object markings with regard to whether or not their verbs are *hata* verb, and the result is summarized in Table 5.26.

Table 5.26: *hata*-V and Object Marking

OBJ.MARK		NP-Ø	NP-lul	NP-nun	NP-to	NP-etc.	SUB-TOTAL	TOTAL
<i>hata</i> -V								
- <i>hata</i>	NO.	896	454	90	149	25	718	1,614
	%	<b>56</b>	<b>28</b>	6	9	2	44	100
<i>hata</i>	NO.	792	245	20	51	9	325	1,117
	%	<b>71</b>	<b>22</b>	2	5	1	29	100
TOTAL	NO.	1,688	699	110	200	34	1,043	2,731
	%	62	26	4	7	1	38	100

In Table 5.26, I find that in conversational Korean, *hata*-verbs occur quite frequently (1,117 tokens, approximately 41% of the total), though they do not appear as frequently as non-*hata*-verbs (1,614 tokens, approximately 59% of the total). The data in Table 5.26 also shows that the occurrence of zero-marking is much higher for coding objects when the objects occur with a *hata* verb (71%) than when they occur with a

non-*hata* verb (56%). In contrast, the occurrence of *lul*-marking is lower for coding objects occurring with a *hata* verb (22%) than for those occurring with a non-*hata* verb (28%) (zero-marking vs. *lul*-marking:  $\chi^2=28.343$ ,  $P < 0.001$ ). More importantly, the use of zero-marking (71%) is more common for coding the objects occurring with a *hata* verb than that of *lul*-marking (22%). In fact, the data in Table 5.26 supports the view that in Korean, verbal nouns that occur with a *hata* verb are predominantly coded with a zero particle. Yet there is still a good portion of objects (22% of all objects) that occurred with a *hata* verb, and it would be interesting to see what linguistic or non-linguistic properties control the use or non-use of *lul*-marking for coding those verbal nouns.

To summarize this section, as previous studies reported, in the present study, the occurrence of zero-marking is reported to be high for coding verbal nouns which are combined with *hata* verbs. However, it still remains as a question when such verbal nouns are overtly marked with its particle and when they are not. For an answer to such a question, I can only propose that the occurrence and non-occurrence of *lul*-marking for verbal nouns may be related to the properties of such verbal nouns. For example, if verbal nouns are abstract or generic, thus represent low informational prominence, they may be likely to occur with a zero particle. If they are specific, thus represent high informational prominence, they may be likely to occur with the accusative *lul*.

#### **5.4.12 Nominative-marked objects: split case marking**

In Korean, direct objects are encoded with *lul*, an accusative marker while subjects are encoded with *ka*, a nominative marker (Sohn 1999). However, there are cases where the nominative *ka* is used to mark the direct objects. This phenomenon is called ‘case

alternation' or 'case shifting'<sup>22</sup> (Han 1999; Um 2003), and it has been at the crux of discussion among Korean as well as Japanese linguists (e.g., Kuno 1973; Yanagida 1985; Kim 1996; Han 1999; 2002; Um 2003; Shimojo 2005, among others).<sup>23</sup> The utterance in (5.50) exemplifies the case alternation in Korean, in that the direct object in brackets, *chopap-ilang sayngsen* 'sushi and fish', is encoded with the nominative *ka*.

(5.50) *Nemwu [chopap-ilang sayngsen]-i nemwu mek-ko siph-unkeyeyyo.*  
 very sushi-and fish-NOM very eat-and want-you.know  
 'I wanted to have sushi and fish so badly.'

In (5.50), the underlined predicate, *mekko siphtha* 'want to eat', is a two-place predicate, which requires two arguments, subject and direct object, in order to be satisfied (Trask 1999). The direct object in brackets is encoded with the nominative *ka*, but it can also be encoded with the accusative *lul* without affecting the acceptability or meaning of the utterance. Um (2002) extensively discusses this particular phenomenon and he also points out, though implicitly, that switching the case markers from *ka* to *lul* and from *lul* to *ka*, depending on the original marking, essentially makes no difference in the meaning of the utterance. He (ibid.) also points out that in Korean, case alternation rarely occurs, and it only occurs in utterances which contain a certain type of auxiliary verb such as *-ko siphtha* 'want', as exemplified in (5.50). He (ibid.) further points out that case alternation is not always allowed even in utterances containing such an auxiliary verb. He (ibid.) argues that there are constraints for allowing case

<sup>22</sup>One may treat this case the same as 'double/multiple nominative construction', in the sense that there are two or more nominative marked elements in a single clause.

<sup>23</sup>In the present study, utterances like *na-nun chakaun namca-ka silh-ta*, '1SG-TOP cold man disliked-DCL' are precluded from the discussion of case alternation, due to the existence of an alternate expression, namely, *na-nun chakaun namca-lul silhehan-ta*, '1SG-TOP cold man dislike-DCL.' Some (e.g., Kim 1990) treat this case as an instance of case alternation because both utterances are translated into 'I dislike a cold man.' They actually view these two examples as the same, though I believe it should not be regarded as the same. It is notable that only a *ka*-marked object is allowed in the first example, whereas only a *lul*-marked object is allowed in the second example. I believe this is because of the change in verb forms (i.e., *silhta* 'be disliked' → *silhehata* 'dislike'). It is thus controversial whether to regard this case as an instance of case alternation, so it is precluded from the discussion in this dissertation.

alternation (i.e., the *ka/lul* alternation) and proposes that the acceptability of the *ka/lul* alternation depends on properties of a main verb with which an auxiliary verb (e.g., *-ko siphtha* ‘want’; *-ki silhta* ‘don’t want’) is combined.<sup>24</sup>

The following are additional instances of direct objects occurring in utterances that include a *-ko siphtha* ‘want’ auxiliary verb. The utterance in (5.51) occurs with the accusative-marked object, *il* ‘job’, whereas the utterance in (5.52) appears with the zero-marked object, *aisukhulim* ‘ice cream’. Interestingly, the case alternation is allowed for both objects in (5.51) and (5.52) without affecting the acceptability and meaning of the utterances.

(5.51) *na-nun nay-ka cincca cohaha-nun [il-ul] chach-ko siph-e.*  
 1SG-TOP 1SG-NOM truly like-REL job-ACC find-and want-SEM  
 ‘I want to find a job that I really like.’

(5.52) *na-Ø [aisukhulim-Ø] mek-ko siph-e.*  
 1SG-( ) ice.cream-( ) eat-and want-sem.  
 ‘I want to eat ice cream.’

As with Korean, there have been studies that explore the case alternation phenomenon in Japanese (Kuno 1973; Yanagida 1985; Shibatani 1990; Shimojo 2005, among others). Kuno (1973) claims that the case alternation, also known as split case marking, is related to the semantic properties of predicates, and he associates the nominative marking with ‘stative’ predicates and the accusative marking with ‘action’ verbs. Furthermore, he (ibid.) points out that the case alternation is most favorably allowed in utterances that include stative derivatives of *-tai* ‘want’ and *-eru* ‘can’. Shimojo (2005:158) views this explanation of the split case marking with the stative/action distinction as unjustifiable because there are no precise stative or active contexts in which a speaker would clearly choose one marking over another.

<sup>24</sup>See Um (2002) for a detailed discussion of constraints for case alternation pertaining to the *-ko siphtha* ‘want’ auxiliary verb.

Yanagida (1985) discusses case alternation from a discourse-pragmatic point of view. She claims that the choice between a nominative marking and an accusative marking reflects the predictability of information (e.g., the nominative *ga* is used for unpredictable information), and is also related to cognitive focusing, in that unpredictable information requires more focusing from interlocutors. Shimojo (2005:51) recognizes Yanagida's (1985) functional approach to case alternation, also known as the split case marking, yet he also criticizes her argument by stating that “[i]t remains unclear whether Yanagida's experimental results reflect the functional properties of the split case marking which she claims.” More interestingly, Shimojo (ibid.) presents an opposite result with respect to the *ga/o* alternation; *o* is a functionally marked encoding type for object NPs that occur in the utterances involving the split case marking in Japanese,<sup>25</sup> whereas Yanagida (1985) claims the opposite result.

Shimojo (2005) claims that the alternation in the split case marking in Japanese is by no means arbitrary but rather it is related with the discourse property. Specifically, accusative-marked NPs exhibit greater cataphoric persistence than nominative-marked NPs, in the sense that the accusative-marked NPs tend to be repeated in subsequent texts, while nominative-marked NPs lack the property of repetition. He (ibid.) further ties repeated information to informational importance, and states that accusative marking is desired for continuing attention from the addressee. In contrast, utterances containing nominative-marked NPs represent a new proposition.

As repeatedly stated above, in both languages, the case alternation in the split case marking seems to be confined to the utterances including certain auxiliary verbs (e.g., *-ko siphta* ‘want’ in Korean; *-tai* ‘want’ in Japanese). For this reason, in the present study, I examine the case alternation only with the utterances that contain either a *-ko siphta* ‘want’ or a *-ki silhta* ‘don't want’ auxiliary verb with respect to the encoding

---

<sup>25</sup>Shimojo (ibid.) labels these object NPs as N2 in order to distinguish them from other direct object tokens.

type of direct object NPs in conversational Korean, and the results are summarized in Table 5.27.

Table 5.27: Case Alternation in Split-case Marking

OBJ.MARK	NP-Ø	NP-lul	NP-ka	TOTAL
NO.	13	2	2	17
%	76	12	12	100

According to the data in Table 5.27, out of the 17 total objects that appear in the utterances containing either a *-ko siphta* ‘want’ auxiliary verb or a *-ki silhta* ‘don’t want’, there are only two objects that are nominative-marked and two objects that are accusative-marked. Contrastively, there are a total of 13 objects (76%) which are zero-marked. This result is consistent with Um’s (2002) observation in the same regard. That is to say, *ka*-marked objects are rarely found in Korean. Unlike Korean, nominative-marked objects appearing with the auxiliary verb *tai* ‘want’ are commonly observed in Japanese (Kuno 1973).

What is also notable in Table 5.27 is that all of the 17 objects essentially allow split-case marking. For instance, a *ka*-marked object can be switched to a *lul*-marked or zero-marked object without affecting the acceptability or meaning of the utterance.<sup>26</sup> However, this does not necessarily indicate that there is absolutely no difference in discourse properties encoded in each encoding type.

As pointed out earlier, there are some studies that have explored the case alternation phenomenon in Korean, particularly focusing on discussing *ka*-marked objects in comparison to *lul*-marked objects (e.g., Han 1999; 2002; Um 2002, *inter alia*). However, to the best of my knowledge, no studies have attempted to investigate discourse properties

<sup>26</sup>Two native speakers of Korean were consulted for the acceptability judgment of the cross-case marking for all of the 17 objects.



encoded in each encoding type for objects in Korean, though some studies have been done exploring such discourse properties in Japanese (Yanagida 1985; Shimojo 2005, *inter alia*). For this reason, the present study attempts to examine discourse properties encoded in each of the three encoding types (i.e., *lul*-marked objects, *ka*-marked objects and zero-marked objects). More specifically, I examined all of the 17 total objects with regard to RD/AF and RP/RP-*f*, in order to find out if information predictability, the discourse property claimed by Yanagida (1985), and cataphoric persistence, the discourse property claimed by Shimojo (2005), play a role in the choice of either encoding type in the alternation. In the present study, I associate predictability of information with processing load in the sense that predictable information is accessible, hence generally easier to process. On the other hand, the cataphoric saliency is associated with thematic prominence because thematically important information tends to appear continuously in subsequent utterances. RD/AF is used to measure the information predictability, and RP/RP-*f* is used to measure the cataphoric saliency. The following table presents each encoding type in split-case marking with respect to its RD/AF and RP/RP-*f*.

Table 5.28: Split-Case Marking and Mean RD;AF;RP;RP-*f*

OBJ.MARK		NP-Ø	NP-lul	NP-ka	MEAN
OBJ.RD;AF;RP;RP- <i>f</i>					VALUES
MEAN RD		9.8	14	11	11.6
MEAN AF		0.8	1	1.5	1.1
MEAN RP		1.2	0	0.5	0.6
MEAN RP- <i>f</i>		2.5	3	2	2.5
TOTAL	NO.	13	2	2	17
	%	76	12	12	100

According to Table 5.28, there is essentially no difference among encoding types with regards to RD and AF or RP and RP-*f*. With the data in Table 5.28, I assume

that unlike Yanagida's (1985) and Shimojo's (2005) claims, neither the information predictability nor the cataphoric saliency seem to play a role in the choice of encoding type in split-case marking in conversational Korean. In other words, processing load and informational prominence in the sense described earlier do not play roles in the alternation in split-case marking. However, the number of the tokens is too small, thus it is not clear whether it may simply be a coincidence, and either predictability of information or the cataphoric saliency actually do play a role in the choice of the encoding type. I propose that the choice of each encoding type reflects a free variation and the token distribution simply reflects the grammatical property exhibited by *ko siphta* that the main verb may incorporate the object argument into. As for the alternation between *lul*-marked objects and *ka*-marked objects, either a *lul*-marked object or a *ka*-marked object can freely be used due to both transitive and intransitive properties encoded in the construction of *-ko siphta* 'want' or a *-ki silhta* 'don't want' verbs. However, this is only a proposition that is based on insufficient data, so further investigation may be necessary to reveal discourse properties encoded in each encoding type.

To summarize the present section, though uncommon, direct objects in Korean may be encoded with the nominative *ka*, and this phenomenon is usually known as 'case alternation', which is also observed in Japanese (Kuno 1973; Yanagida 1985; Shimojo 2005). Interestingly, in both languages, the alternation in split-case marking seems to be limited to utterances that contain certain types of auxiliary verbs (e.g., *-ko siphta* 'want' in Korean; *-tai* 'want' in Japanese). As for the alternation in split-case marking in Japanese, Yanagida (1985) claims predictability of information is responsible for the nominative-marked objects, while Shimojo (2005) argues the cataphoric saliency is a reason for the nominative-marked objects. Contrastively, the present study does not demonstrate either predictability of information or cataphoric saliency to be associated with nominative-marking in this construction of Korean (see Table 5.28). As for the

case alternation phenomenon, I propose that the choice of either encoding type in split-case marking in Korean may be the outcome of a free variation, or insufficient data for a meaningful generalization.

## 5.5 Summary of Chapter 5

In Chapter 5, I have shown that in conversational Korean, the choice of object markings is related to two discourse related factors: processing load and informational prominence. Based on this, I have claimed that one discourse-pragmatic function of the accusative *lul* is certainly to mark processing load and informational prominence.

In Section 5.2, I have presented an overall distributional pattern for object encoding types. In this section, I showed that in conversational Korean, based on the results of the present study, objects were most likely expressed overtly in a clause, and when objects were overtly expressed, they tend to be coded with a zero particle rather than with an overt particle. I attributed this tendency to the general characteristics of object NPs. That is to say, objects usually encode new information, so they tend to appear overtly in a clause. When objects are overtly expressed, they tend to be coded with a zero particle rather than with an overt particle. This tendency is attributed to the assumption that new information encoded in objects may function as a focusing device, so the occurrence of *lul*-marking is not high for objects.

In Section 5.3, I have discussed the choice of object markings with respect to the sentence types in which objects appear. Based on the data in Table 5.3, I stated that similar to subjects, objects of interrogative or directive sentences are more likely to be coded with a zero particle than objects of declarative sentences. I have argued that objects in interrogative or directive sentences are usually accessible to both the speaker and the hearer, thus low processing load, so they tend to be coded with a zero particle.

In Section 5.4, I have discussed the choice of object markings with relation to the following factors that are related to either processing load, informational prominence or both: negation, animacy, word order, definiteness; length, repair, verb adjacency, *hata*-verb, anaphoric saliency, cataphoric saliency and contrast.

In Section 5.4.1, I have stated that the use of zero-marking is less common for coding objects in negative sentences than for objects in affirmative sentences. Likewise, the use of an overt marking is more common for objects in negative sentences than for objects in affirmative sentences. In short, due to processing load, objects in negative sentences tend to be coded with an overt marking, whereas objects in affirmative sentences tend to be coded with zero-marking. Yet the use of *lul*-marking is less common for objects in negative sentences than for those in affirmative sentences, whereas the use of *nun*-marking is more common for objects in negative sentences than for those in affirmative sentences. This tendency is attributed to the assumption that an entity in the focus of negation denotes old, hence identifiable, information and it often contrasts with another entity, so it is commonly coded with *nun*-marking, which encodes contrastiveness (Choi 1997).

In Section 5.4.2, I have shown that the animacy of object NPs influences the choice of object encoding types. More specifically, the use of *lul*-marking is more common for objects representing an animate referent than for those representing an inanimate referent. This tendency is attributed to the assumption that animate objects are considered a deviation from the norm for object NPs, hence processing load, so the use of *lul*-marking, which alleviates processing load, is common. The occurrence of zero-marking is low for objects representing a non-human animate referent, and instead, the occurrence of *lul*-marking is relatively high for coding them. Furthermore, the use of *lul*-marking is more common for object NPs representing an interlocutor referent than it is for those representing a non-interlocutor referent. Likewise, the use of zero-

marking is less common for objects representing an interlocutor referent than it is for those representing a non-interlocutor human referent. This tendency is attributed to processing load incurred from being a deviation from the norm; an interlocutor is not as likely to occur as a referent of an object vis-a-vis as a referent of a subject, so objects representing an interlocutor referent are regarded as a deviation from the norm, which results in processing load. The occurrence of *lul*-marking is high to alleviate the processing load. However, unlike the initial assumption, the animacy of subject NPs does not play a role in the choice of object encoding types.

In Section 5.4.3, I have stated that the speaker's selection of object encoding types does not seem to be related to word order variations. For example, the occurrence of zero-marking is higher for objects that immediately precede the predicate than for those that do not, due to easy identifiability, and the occurrence of *lul*-marking is lower for objects immediately preceding the predicate than for those that do not. Yet this tendency is not statistically significant. Furthermore, the occurrence of *nun*-marking is relatively high for objects that appear in a sentence-initial or sentence-final position. This tendency, which is statistically significant, is attributed the general property of a grammatical topic, i.e., the sentence-initial and sentence-final positions are the topic positions of a sentence.

In Section 5.4.4, I have shown that the choice of object encoding types is related to how close or how far the object is from its predicate. That is to say, the occurrence of *lul*-marked objects gradually increases in proportion to the distance to the verb from the object. Likewise, the occurrence of zero-marked objects gradually decreases as the distance from the verb increases. This tendency is attributed to the assumption that in general, processing load would increase as the distance to the verb becomes greater, and processing load would invite the use of *lul*-marking to alleviate processing load.

In Section 5.4.5, I have noted that the definiteness of object referents seems to play

a role in the selection of object encoding types, and the tendency observed for objects appears to be opposite to that of subjects. Indefinite objects are considered the norm for objects, hence minimal processing load, so the use of *lul*-marking is less common for indefinite objects than for their definite counterparts. Contrastively, definite objects are viewed as a deviation from the norm, hence processing load, so the occurrence of zero-marking is relatively low and the occurrence of *lul*-marking is relatively high for coding them. Furthermore, the use of zero-marking is high for indefinite pronominal objects because indefinite pronominal objects do not represent informational prominence.

In Section 5.4.6, I have shown that unlike Japanese, monosyllabic objects are common in conversational Korean, and they are commonly coded with a zero particle. The use of *lul*-marking is more common for pronominal monosyllabic objects than it is for their full noun counterparts, and this tendency is attributed to processing load; pronominal monosyllabic objects are not as common as full noun monosyllabic objects, so processing load, which would invite the use of overt marking (e.g., *lul*-marking), is assumed. Furthermore, the use of *lul*-marking is more common for monosyllabic objects with a modifier than it is for those with no modifier. This tendency is attributed to informational prominence, which is likely to be encoded in objects that are modified. More importantly, the occurrence of *lul*-marked objects proportionally increases as the number of object modifiers increases, and the occurrence of zero-marked objects is inversely proportional to the number of object modifiers. This tendency is attributed to informational prominence, which is likely to be encoded in modified objects, i.e., large NPs. Based on these tendencies, I argue that the accusative *lul* in conversational Korean functions to alleviate processing load as well as to mark informational prominence.

I have demonstrated that the selection of object markings somewhat correlates with the length of object NPs. That is to say, the larger the object NPs are, the more likely they are coded with *lul*-marking. I attributed this tendency to the assumption

that modified objects were informationally more prominent than unmodified objects. In this section, I also showed that unlike Japanese, an accusative *lul* was not very commonly used for marking monosyllabic objects in conversational Korean.

In Section 5.4.7, I have noted that the occurrence of *lul*-marking is relatively high for repaired objects, and the occurrence of zero-marking is high for non-repaired objects. This tendency is attributed to the assumption that repaired objects represent informational prominence, so the use of *lul*-marking, which functions as marking informational prominence, is expected to be common for repaired objects.

In Section 5.4.8, I have demonstrated that in conversational Korean, unlike for subjects, the anaphoric saliency, either in RD or in AF, does not correlate with the selection of object encoding types, though it distinguishes a zero anaphor from other encoding types for object NPs.

In Section 5.4.9, I have stated that unlike in conversational Japanese, in conversational Korean, the cataphoric saliency does not correlate with the choice of object encoding types.

In Section 5.4.10, I have shown that contrastiveness of object NPs plays a role in the speaker's selection of object markings in conversational Korean. Overall, the use of zero-marking is common for coding objects that are not contrastive, while it is not common for coding those that are contrastive. Instead, *lul*-marking and *nun*-marking is most commonly used to code contrastive objects, due to informational prominence. The data in this section points to the complementary nature of object markings with regard to contrastiveness: *lul*-marking is associated with the 'exhaustive-listing' type of contrastiveness; *nun*-marking is associated with the 'action/state reaction' type of contrastiveness; zero-marking, *lul*-marking and *nun*-marking are almost equally associated with the 'parallel activities/states' type of contrastiveness; zero-marking is generally not preferred for coding contrastive objects, regardless of contrastive type.

In Section 5.4.11, I have discussed that the occurrence of zero-marking is high for coding verbal nouns, which are combined with *hata* verbs. However, it still remains as a question when such verbal nouns are overtly marked with the accusative *lul* and when they are not. Yet at least the greater frequency of the zero-marking for objects with *hata* reflects the grammatical property of the construction. That is to say, the use and non-use of *lul*-marking for verbal nouns may be related to the properties of such verbal nouns. For example, if verbal nouns are abstract or generic, hence low informational prominence, they may be likely to occur with a zero particle. If they are specific, hence high informational prominence, they may be likely to occur with the accusative *lul*.

In Section 5.4.12, though uncommon, in Korean, objects may be coded with the nominative *ka* instead of the accusative *lul*, which is the so-called ‘case alternation’. In general, the alternation in split-case marking is limited to utterances that contain certain types of auxiliary verbs, e.g., *-ko siphta* ‘want’. The present data does not provide a ground for a substantial analysis and I have noted that the choice of either encoding type in split-case marking in Korean may be the outcome of a free variation.



# Chapter 6

## Concluding Remarks

### 6.1 Outline

Chapter 6 concludes this dissertation by briefly recapturing the study from the preceding chapters, as well as by summarizing the findings in this dissertation. Chapter 6 also discusses some issues that remain to be studied in the future.

This dissertation used a total of 9,249 clausal units, which consist of 5,557 intransitive clauses and 3,692 transitive clauses, for the discussion of subject and object encoding types in conversational Korean. In this dissertation, I have investigated discourse-pragmatic perspectives of the subject and object markings in conversational Korean. More specifically, I have examined when the subject and object marking particles, *ka* and *lul* respectively, overtly occur and when they do not in conversational Korean. In this dissertation, I have proposed that the notions of processing load and informational prominence are useful in describing the occurrence and non-occurrence of these particles, and as such, I examined the speaker's selection of subject and object encoding types with relation to several factors that are related to processing load, informational prominence or both.

In Chapters 4 and 5, I have discussed the subject and object encoding types in conversational Korean, more specifically, how a speaker chooses one particular subject and object form over another, largely based on two different but related factors: processing load and informational prominence. For the discussion of subject and object encoding types, I have proposed several individual factors that are related to either processing load, informational prominence, or both, in one way or another: negation, animacy, word order, verb adjacency, definiteness, length, repair, anaphoric saliency, cataphoric saliency, and contrast. The choice of object encoding types is also discussed with relation to the *hata*-verb and the split-case marking.

## 6.2 Subject and object markings

In conversational Korean, at least according to the data of this study based on a total of 9,249 clausal units, subjects are most likely unexpressed, but when they are overtly expressed, they tend to be marked with an overt particle rather than with a zero particle. More specifically, subjects of intransitive clauses tend to be overtly expressed more often than those of transitive clauses, and the use of *ka*-marking was more common for coding subjects of intransitive subjects (S) than for coding subjects of transitive clauses (A). Unlike subjects, objects were most likely to be expressed overtly in a clause, and when overtly expressed, they tend to be marked with a zero particle rather than an overt particle. I have stated that this might be the result of the universal tendency to express at most one argument per clause (Fry 2003), or it might suggest the ergative-absolutive pattern, i.e., subjects tend to be overtly expressed in intransitive constructions, while objects tend to be overtly expressed in transitive constructions (Dixon 1994; Thompson & Hopper 2001). The tendency reported for objects may also be attributed to the general characteristics of object NPs, i.e., objects usually represent

new information, so they tend to appear overtly in a clause, as well as to the assumption that new information encoded in objects itself may function as an attention attractor, so the occurrence of *lul*-marking is not high for objects. That is to say, objects are in a default focus domain of a given clause, therefore they do not need to be explicitly marked with an overt particle.

In general, subject and object referents of interrogative and directive sentences are more accessible/identifiable to both the speaker and the hearer than those of declarative sentences, and hence represent a minimal processing load in referent identification, so they are more commonly coded with a zero particle than those of declarative sentences. More specifically, I have argued that subjects would be likely to be overtly identified with the nominative *ka* when they represent processing load, informational prominence, or both, whereas they would be commonly coded with a zero particle when they represent a minimal processing load or low informational prominence. Likewise, objects are likely to be overtly identified with the accusative *lul* when they represent processing load or informational prominence, and objects are likely to be marked with a zero particle when they represent a minimal processing load or low informational prominence.

As previously stated, two different but related factors, namely processing load and informational processing load, were used for the analysis of subject and object encoding types. It was assumed that a deviation from the norm or from the default pattern would indicate (increased) processing load, and extra information to process, hence informationally loaded, would represent informational prominence. It was also assumed based on many linguistic similarities between Korean and Japanese that tendencies which apply to Japanese particle ellipsis would also be likely to apply to Korean particle ellipsis. Several individual factors that are related to processing load, informational prominence, or both were examined with regard to the choice of subject and object

encoding types, and the results are presented below.

— **Negation:** In general, negative constructions are conceptually more complex than affirmative constructions, so it was assumed that due to the conceptual complexity, negation would impose an increase in processing load on the identification of subject and object. In addition, negative sentences were reported to be less common than affirmative sentences, and hence a deviation from the norm, which would lead to an increase in processing load.

Overall, for both subject and object NPs, a zero particle was less common in negative constructions than for those in their affirmative counterparts. In contrast, an overt marking is more commonly used for subjects and objects in negative sentences than for their affirmative counterparts in order to alleviate the processing load. However, the use of case markers, i.e., *ka* and *lul*, is slightly less common for subjects and objects that occur in negative sentences than those in affirmative sentences. Instead, the occurrence of *nun*-marking is more common for subjects and objects in negative sentences than in affirmative sentences. Negation usually goes with given information, so the particle *nun* is common but *ka* and *lul* are not common for subjects and objects in negative sentences. Furthermore, an entity in the focus of negation often contrasts with another entity, so it is commonly coded with *nun*-marking. It is also notable that the use of the particle *to* ‘even’, often referred to as a negative polarity item (NPI), is also more common for subjects and objects in negative sentences than in affirmative sentences.

— **Animacy:** It is widely accepted that the default animacy of subjects (of transitive clauses) is animate, and that of objects is inanimate. It was assumed that deviation from the default animacy pattern for subject and object NPs would impose an increase in processing load on their identification. The choice of subject and object markings was examined with regard to their animacy in three different ways: human or non-human animate vs. inanimate, interlocutor or non-interlocutor and the animacy of subjects

and objects together.

Subjects and objects roughly reflect the opposite distributional pattern due to the difference in default animacy. More specifically, *ka*-marking is less common for subjects that represent a human referent than for those that represent a non-human animate or inanimate referent because a human referent is viewed as the default animacy for subject NPs, and hence would represent minimal processing load in referent identification. In contrast, *lul*-marking is less common for objects that represent an inanimate referent than for those that represent a human and non-human animate referent, since an inanimate referent is regarded as the default animacy for object NPs, and hence would represent minimal processing load in referent identification. The difference in the occurrence of overt marking became greater when subjects and objects were examined with respect to whether or not they are interlocutors. For subjects, *ka*-marking was less commonly used for subjects that represent an interlocutor referent than for those that represent a non-interlocutor referent. Conversely, *lul*-marking was common for objects that represent an interlocutor referent because an interlocutor referent is not commonly identified as a referent of an object, i.e., a deviation from the norm, and hence represent an increase in processing load in referent identification. The occurrence of overt marking is higher when subjects and objects represent the same animacy than when they do not. More specifically, the use of *ka*-marking is more common when subjects and objects both represent animate referents than when they do not, and the use of *lul*-marking is more common when subjects and objects both represent inanimate referents than when they do not. Both subjects and objects representing the same animacy is a deviation from the norm, hence an increase in processing load, so the use of overt marking (i.e., *ka* for subjects and *lul* for objects) is common.

— **Word order:** The (canonical) word order of Korean is SOV, and non-canonical/marked word order is assumed to impose processing load, due to markedness. It is also notable

that post-predicative entities are generally less important than pre-predicative entities, and hence they would represent low informational prominence.

For both subjects and objects, the use of *ka* and *lul* tends to be commonly used for coding subjects and objects respectively when they represent processing load. More specifically, subjects and objects occurring in a non-canonical/marked position are more commonly coded with an overt particle than those in a canonical/marked position.

For subjects, *ka*-marking is less commonly used when subjects appear in a post-predicative position than in a pre-predicative position, in that they represent low informational prominence. On the other hand, the use of *lul*-marking is more common for objects occurring in a post-predicative position than in a pre-verbal position. This suggests that processing load may play a stronger role than informational prominence in the choice of object markings, whereas informational prominence may influence the choice of subject markings more strongly than processing load. Furthermore, when objects are coded with a zero particle, the occurrence of *ka*-marked subjects is relatively high because at least one argument needs to be overtly identified for the sake of argument identification.

— **Verb adjacency for object markings:** Based on studies on Japanese particle ellipsis, it was assumed that objects tend to drop particles when they are immediately adjacent to the verb, and the greater the distance to the verb from the object, the greater processing load imposed on object referent identification. Just like in colloquial Japanese, the occurrence of *lul*-marked objects is greater when there were words intervening between objects and the verb than when there were no words intervening between them, due to an increased processing load. Likewise, the occurrence of zero-marked objects decreases as the distance from the verb increases.

— **Definiteness:** For the purpose of this study, subjects and objects are categorized into four different subtypes based on their definiteness: definite, indefinite, indefinite

pronoun and indefinite but specific subjects. In general, referents of subjects (of transitive clauses) are definite, and referents of objects are indefinite. Thus, indefinite subjects (of transitive clauses) and definite objects are considered deviations from the norm/default definite pattern, and hence would impose processing load on the identification of subject and object NPs. Furthermore, definite entities are generally easier to identify than indefinite entities. Indefinite pronoun referents do not need to be identified due to the lack of specific information being conveyed, and hence represent low informational prominence.

Subjects and objects roughly reflect the opposite distributional pattern with regard to their markings due to the difference in their default definiteness. More specifically, *ka*-marking is more common for indefinite subjects than for definite subjects, and zero-marking is less common for indefinite subjects than for definite subjects due to processing load imposed on indefinite subjects, which are deviations from the norm for the subject NPs. Conversely, *lul*-marking is more common for definite objects than for indefinite objects due to the processing load imposed on definite objects that are deviations from the norm for object NPs. It is also notable that the difference in the occurrence of *lul*-marking between definite and indefinite objects is smaller than that of *ka*-marking between definite and indefinite subjects, because definite objects are somewhat easier to identify than indefinite objects, though definite objects are viewed as deviations from the norm for object NPs.

Indefinite but specific subjects tend to be coded with *ka*-marking, in the sense that their referents are identifiable only to the speaker, and hence processing load is imposed on referent identification. For both subjects and objects, the occurrence of zero-marking is high when they represent an indefinite pronoun referent.

— **Length:** In the present study, the length of NPs was measured in two different ways, i.e., monosyllabicity and the number of modifiers. In Japanese, monosyllabic

NPs are not common, and are also hard to recognize due to their attenuated form, so they are likely to be overtly identified. Yet, unlike Japanese, in conversational Korean, monosyllabic subjects and objects are not uncommon, and they were more commonly coded with a zero particle than an overt particle. Nevertheless, monosyllabic full-noun NPs are not common in conversational Korean, and hence would represent processing load incurred by the markedness (i.e., a deviation from the norm), so the occurrence of overt marking is relatively high for subjects and objects. The choice of subject and object encoding types was also examined with regard to their modifiers, and for both subjects and objects, the occurrence of overt marking was higher when they are modified than when they are not, in the sense that modified entities are informationally loaded, and hence would represent informational prominence. Likewise, the occurrence of zero-marking was lower when they were modified than when they were not.

In short, in both Japanese and Korean, the use of an overt marking (e.g., *ka* in Korean; *ga* in Japanese) is common for monosyllabic full-noun subjects and objects due to processing load, as well as for large subjects and objects due to informational prominence.

— **Repair:** In this study, repaired entities (i.e., entities involving a repair) are compared to non-repaired entities (i.e., entities not involving a repair), not between an interrupted entity and a reproduced entity. It was assumed that repaired entities would require cognitive attention from the hearer, and hence represent informational prominence, so they would be overtly identified with an explicit particle.

For both subjects and objects, the use of overt marking was more common for repaired subjects and objects than for their non-repaired counterparts, and the use of zero-marking was less common for repaired subjects and objects than for their non-repaired counterparts. Despite the universal tendency for repair or speech error to commonly occur with new information, the percentage of repaired subjects in intransi-



tive clauses in conversational Korean was almost the same as that of transitive clauses. It is also notable that *nun*-marking was less common for repaired subjects in intransitive clauses than for those in transitive clauses, while *ka*-marking was more common for repaired subjects in intransitive clauses than for those in transitive clauses. This is supporting evidence for the view that *nun* encodes old information while *ka* encodes new information.

— **Anaphoric saliency:** In this study, anaphoric saliency was used to measure activation level. In short, the more recently (RD) and frequently (AF) an entity is mentioned in anaphoric context, the entity is assumed to be more activated, and thus more accessible. Therefore, they would represent the minimal processing load in referent identification. In other words, a greater RD or a smaller AF indicates a lower level of activation, and a low level of activation would impose a greater activation cost, and hence indicate the processing load.

For subjects, the mean RD for zero-anaphor was the smallest and the mean RD for *ka*-marked subjects was the greatest. The mean RD for zero-marked subjects was similar to the mean RD for *nun*-marked subjects. Likewise, the mean AF for zero-anaphor was the greatest and the mean AF for *ka*-marked subjects was the smallest. The mean AF for zero-marked subjects was the same as the mean AF for *nun*-marked subjects. This tendency suggests that *ka*-marking tends to be used for anaphorically non-salient subjects which represent a significant activation cost, hence the processing load in referent identification.

It is also notable that in both Korean and Japanese, anaphoric saliency distinguishes between subject encoding types, but more clearly in Japanese than in Korean, due to the discourse-functional differences of subject markers between the two languages, that is, *ka* in Korean may mark given information (or continuing topic) as well as an unidentifiable referent, whereas *ga* in Japanese does not encode given information but

marks an unidentifiable referent. Therefore, it can be said that the usage of *ka* in Korean is broader than the usage of *ga* in Japanese, whose primitive function is to mark an unidentifiable referent.

For objects, the mean RD for zero-anaphor was smallest but there are nearly no differences among the other encoding types. Likewise, the mean AF was greatest for zero-anaphor but the mean AF for the other encoding types was almost the same. This tendency suggests that unlike for subjects, the anaphoric saliency (either by RD or by AF) does not distinguish object encoding types, except for zero-anaphor subjects.

— **Cataphoric saliency:** In this study, cataphoric saliency was used to measure the degree of decay of information in cataphoric context. In short, information that does not persist loses its importance (i.e., informational prominence) in the subsequent contexts, whereas information that persists maintains its importance. For the purpose of the present study, cataphoric saliency was measured in two different ways, that is, uninterrupted persistence (RP) and cataphoric frequency (RP-*f*).

For subjects, the greatest mean RP was reported for zero-anaphor, and the smallest mean RP was reported for *ka*-marked subjects. The mean RP for zero-marked subjects is almost the same as the mean RP for *nun*-marked subjects. Likewise, the greatest mean RP-*f* was observed for zero-anaphor, and the smallest mean RP-*f* was observed for *ka*-marked subjects. The mean RP-*f* for zero-marked subjects and *nun*-marked subjects is placed between that of *ka*-marked subjects and zero-anaphor. Subject referents overall do not persist long, regardless of their encoding types, unlike Japanese, cataphoric saliency either by RP or by RP-*f* does not play a significant role in the choice of subject encoding types. Yet the difference in the mean RP and RP-*f* was good enough to point out the functional differences between *ka* and *nun*. That is to say, *ka*-marking tends to be used for coding cataphorically non-salient information (a local topic or persistence), whereas *nun*-marking to be used for coding cataphorically

salient information (a global topic or persistence). I also pointed out that the primitive function of *nun* is to mark contrastiveness, but it can also contribute to the cohesion between utterances, and hence naturally mark the continuation. This tendency makes Korean differ from Japanese in that in Japanese, *ga* and *wa* both mark cataphorically salient information, and a zero particle marks cataphorically non-salient information.

Just like subjects, object referents do not persist long regardless of their encoding types. Cataphoric saliency either by RP or by RP-*f* does not distinguish between object encoding types, even for zero-anaphor. In other words, unlike Japanese, in conversational Korean, the prominence of information in the sense of cataphoric saliency is not related to the speaker's selection of one object encoding type over another.

— **Contrast:** The definition for the notion of contrast is not clear-cut, hence it is defined more carefully by categorizing it into four different subtypes, that is, directly-contrastive (parallel activities/states; action/state reaction), exhaustive-listing, directly-contrastive & exhaustive-listing. It is generally assumed that entities which are in contrast are informationally more loaded, and hence would represent informational prominence due to the need for contrast identification in addition to referent identification.

For subjects, the complementary nature of the three encoding types was observed, in that *ka*-marking and *nun* are commonly used for coding contrastive subjects whereas zero-marking tends to be used for non-contrastive subjects. More specifically, *ka*-marking is mostly used for coding exhaustive-listing type of contrastiveness, and *nun*-marking is likely to be used for coding directly-contrastive type of contrastiveness.

Just like subjects, *nun*-marking is common for coding contrastive objects, and zero-marking is likely to be used for coding non-contrastive objects. Yet, unlike subjects, *lul*-marking is not particularly high for coding contrastive objects compared to non-contrastive objects. More specifically, *lul*-marking is common for coding the exhaustive-listing type of contrastiveness, whereas *nun*-marking is likely to be used for coding

the action/state reaction type of contrastiveness. Unlike subjects, the three encoding types take a roughly equal portion of all objects of parallel activities/states type of contrastiveness.

Furthermore, I argued that although the primitive function of *nun* is to mark contrastiveness, it is not its only function. Therefore, it would be untenable to claim that *nun* is a contrastive marker, but not a topic marker.

— **Light verbs:** In Korean, as well as in Japanese, direct object NPs often occur without an overt particle when they form lexicalized or idiomaticized compound expressions together with their verbs. The verbal noun object NPs that are combined with the so-called light verb *hata* ‘do’ in Korean predominantly occur with a zero particle but are often overtly coded with *lul* as well. The result of the present study complies with such a strong tendency in the sense that *lul*-marking was uncommon for objects that occur together with *hata* ‘do’ verb, but instead zero-marking was common for them. Yet the question remains of when such verbal nouns are overtly identified with *lul* and when they are not.

— **Nominative-marked objects: split-case marking:** Though uncommon, objects in Korean may be encoded with the nominative *ka*, and nominative-marked objects occur only with certain types of auxiliary verbs, e.g., *-ko siphta* ‘want’. The similar phenomenon is also observed in Japanese. Some argued that predictability of information was responsible for the nominative-marked objects, and some argued that cataphoric saliency was a reason for the nominative-marked objects in Japanese. Yet the present study does not demonstrate either predictability (anaphoric saliency) or cataphoric saliency to be associated with nominative-marking in this construction of Korean. Nevertheless, there were only 17 tokens of nominative-marked objects, so this may be the outcome of free variation, or insufficient data for a meaningful generalization.

According to my data, both the nominative *ka* and the accusative *lul* are allowed for objects due to both the transitive and intransitive properties encoded in the construction of *-ko siphta* ‘want’ or *-ki silhta* ‘don’t want’ verbs. My speculation for this tendency is that the choice between *ka* and *lul* for objects may depend on the main verb that is combined with the object.

### 6.3 Findings of the present study

With a total of 9,249 clausal units, which consist of 5,557 intransitive clauses and 3,692 transitive clauses, I have discussed the selection of subject and object encoding types based on several factors that are related either to processing load, informational prominence or to both, and argued that the speaker’s selection of one subject and object encoding type over another is not an arbitrary but a systematic process. More specifically, I have argued that subjects would be likely to be overtly identified with the nominative *ka* when they represent processing load, informational prominence, or both, whereas they would be commonly coded with a zero particle when they represent a minimal processing load or low informational prominence. Likewise, objects are likely to be overtly identified with the accusative *lul* when they represent processing load or informational prominence, and objects are likely to be marked with a zero particle when they represent a minimal processing load or low informational prominence.

However, the notions of ‘processing load’ and ‘informational prominence’ do not suggest the same thing for subjects and objects, due to the difference of the ‘norm’ for each. That is to say, the functional difference between subjects and objects tends to influence the use and non-use of the overt markers, particularly the case markers *ka* and *lul* in conversational Korean. For instance, subjects tend to occur as zero-anaphor since they tend to encode old information, whereas objects tend to occur overtly in a

clause because they usually represent new information. Furthermore, when subjects appear overtly in a clause, they tend to be identified with an overt marker, e.g., *ka*. In contrast, objects tend to be identified with a zero particle, because the new information that objects represent may function as an attention attractor. In other words, objects are in a default focus domain of a given clause, therefore they tend not to be explicitly marked with an overt particle.

In addition, referents of subjects usually represent old/definite information, so they tend to be identified easily. Due to relative simplicity in referent identification, which implies an attenuated processing load, informational prominence often plays a stronger role than processing load in calling for the use of the nominative *ka* for subjects when subject referents represent old/definite information, which is the norm for the subject NPs. In contrast, referents of objects generally represent new/indefinite information, and the new information per se, i.e., informational prominence, is a default focus domain of a given clause, resulting in the use of a zero particle for coding object NPs. Hence, processing load often plays a stronger role than informational prominence in inviting the use of the accusative *lul* for objects when object referents represent new/indefinite information, which is taken as the norm for object NPs. In short, processing load and informational prominence are interrelated, but informational prominence seems to influence subject encoding types more strongly and processing load tends to influence object encoding types more strongly.

It seems to be clear that the choice between zero-particle and case markers for arguments, i.e., *ka* and *lul*, is decided based on the processing load and informational prominence that subjects and objects may represent. The cataphoric saliency, either by uninterrupted or by frequency, as well as the subcategorization of contrastiveness, points to the functional differences between *ka*-marked subjects and *nun*-marked subjects in the sense that *ka* marks cataphorically non-salient information (a local topic

or persistence) while *nun* marks cataphorically salient information (a global topic or persistence). Korean differs from Japanese in that in Japanese, both *ga* and *nun* mark cataphorically salient information, and a zero particle marks cataphorically non-salient information. Furthermore, it remains an issue to be answered in future studies what distinguishes discourse properties of *nun* from the accusative *lul* in conversational Korean, though the subcategorization of contrastiveness—*lul* for exhaustive-listing type of contrastiveness vs. *nun* for directly-contrastive type of contrastiveness—is one distinguishing factor. Furthermore, the primitive function of the post-nominal particle *nun* is to mark contrastiveness, but it is also used to mark subjects and objects that appear either in a sentence-initial or sentence-final position, which are regarded as a primary and secondary topic position, respectively. However, it is rarely used to code subjects and objects that appear in a sentence-middle position since it is not a topic position. It was also argued in this dissertation that the primitive function of *nun* is to mark contrastiveness, yet it is not its only function.

I have also pointed out the interesting cross-linguistic differences between Korean and Japanese, particularly in terms of case marking. For example, subjects were most likely to occur as zero-anaphor in both Korean and Japanese. However, when they appear overtly in a clause, they tend to be explicitly identified with overt markers (e.g., *ka*) in Korean, whereas they tend to be coded with a zero particle in Japanese. Furthermore, in Japanese, the case markers—*ga* and *o*—encode thematic prominence, but the case markers—*ka* and *lul*—in Korean are not used for such a purpose. The usage of the nominative *ka* in Korean seems to be broader than the nominative *ga* in Japanese in the sense that *ka* can mark given information as well as an unidentifiable referent, but *ga* does not mark given information but marks an unidentifiable referent. Also, in Japanese, the case markers seem to represent discourse prominence, as suggested by cataphoric saliency, rather than to mark deviation from the norm (i.e., markedness).

In contrast, in Korean, case marking seems to be related to prominence in the sense of ‘marked’, hence leaning more toward the processing ground.

Lastly, most studies on case markers in Korean have been done vastly from a structural perspective, so the present study may shed light on the discussion of case markers from an empirical perspective.

## 6.4 Issues for future study

This study examines subject and object markings mostly based on an isolated single factor (e.g., definiteness), but relevant factors combined together seem to better describe the occurrence and non-occurrence of the case markers (e.g., nominal types + monosyllabicity). Therefore, it would be worth furthering research into these cases based on more combined factors (e.g., *hata*-verb + contrast; definiteness + contrast, etc.). It would also be worth while to examine conversation data with respect to two types of a zero particle, i.e., optional and obligatory, in future studies. In other words, it would be worthy of incorporating grammatical properties of the zero particle into discourse observation since the zero particle is an ideal case for the analysis of discourse. The present study examines nominative-marked objects in conversational Korean in order to investigate what is responsible for such nominative marked objects. Yet there were only 17 instances of nominative-marked objects reported in this study, so the nominative-marked objects can be studied further with more data in order to see what enables some objects to be coded with the nominative *ka* in Korean. It would also be interesting to see what linguistic or non-linguistic properties influence the use or non-use of *lul*-marking for objects that are combined with the so-called light verb *hata*-verb. Furthermore, the default properties of intransitive subjects differ from those of transitive subjects, and it would be worth while to investigate the occurrence and non-



occurrence of subject markings with respect to the two different subtypes of subjects. Lastly, there seems to be crucial differences between Japanese and Korean, and those differences can be delineated in future studies.

# Bibliography

- Aissen, Judith. 2003. Differential Object Marking: Iconicity vs. Economy. *Natural Language & Literature Theory* 21: 435-483.
- Backhouse, Anthony. Edgar. 1993. *The Japanese Language: An Introduction*. Oxford University Press. Cambridge, UK.
- Bak, Jae-Hee. 2004. *Optional Case Marking of the Possessor in Korean: A Study of Double Accusative Marking of Possessor/Possessee in Korean Based on Corpus Study and Optimality Computation*. Master's thesis, University of Manitoba.
- Bak, Jae-Hee. 2007. *Double Nominative Constructions in Korean*. Unpublished manuscript. University of Toronto.
- Blake, Barry J. 2001. *Case*. Cambridge, MA: MIT Press.
- Bolinger, Dwight. 1961. Contrastive accent and contrastive stress. *Language* 37. 83-96.
- Chafe, Wallace. 1994. *Discourse, Consciousness, and Time*. Chicago and London: The University of Chicago Press.
- Cho, Jai-Hyoung. 1994. *Scrambling in Korean: crossover, reconstruction and binding theory*. Ph. D. dissertation, University of Connecticut.
- Cho, Sung-Eun. 2003. A Conditioning Factor in Possessor Agreement Constructions. *Japanese/Korean Linguistics* 11, 343-351.
- Choi, Hye-Won. 1996. *Optimizing structure in context: scrambling and information*

- structure*. Ph. D. dissertation, Stanford University.
- Choi, Hye-Won. 1997. Topic and Focus in Korean: The Information Partition by Phrase Structure and Morphology. *Japanese/Korean Linguistics*, Vol 6: 545-561.
- Choi, Hye-Won, and Mitsuaki Shimojo. 2001. The topic marking in Korean: a comparative study with Japanese. *Comparative Korean Studies* 8: 1-17.
- Choi, Incheol, 2003. *Case and Argument Structure in Korean and English*. Ph.D. dissertation. The University of Texas at Austin.
- Choi, Incheol, and Wechsler, Stephen. 2001. Mixed Categories and Argument Transfer in the Korean Light Verb Construction. In *Proceedings of Annual Head-Driven Phrase Structure Grammar Conference*.
- Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Clancy, Patricia, and Pamela Downing. 1987. The Use of 'wa' as a Cohesion Marker in Japanese Oral Narratives. In J. Hinds, S. K. Maynard, and S. Iwasaki (eds), *Perspectives on Topicalization: the Case of Japanese 'Wa'*, 3-56. Amsterdam: John Benjamins.
- Comrie, Bernard. 1989. *Language Universals and Linguistic Typology*. 2nd ed. Chicago, IL: University of Chicago Press.
- Dryer, Matthew S. 1996. Focus, pragmatic presupposition, and activated propositions. *Journal of Pragmatics* 26(4):475-523.
- Um, Jungho. 2003. Case Alternation in '-go siphta' Construction. *Korean Linguistics* 41: 169-195.
- Dixon, Robert M. W. 1994. *Ergativity*. Cambridge, UK: Cambridge University Press.
- Fry, John. 2003. *Ellipsis and wa-marking in Japanese Conversation*. New York and London: Routledge.
- Fujii, Noriko, and Tsuyoshi Ono. 2000. The Occurrence and Non-Occurrence of the

- Japanese Direct Object Marker *o* in Conversation. *Studies in Language* 24(1):1-39.
- Givón, Talmy. 1983. Topic Continuity in Discourse: an Introduction. In T. Givón (Ed.), *Topic Continuity in Discourse: a Quantitative Cross-language Study*, 4-41. Amsterdam: John Benjamins Publishing Company.
- Givón, Talmy. 1988. The Pragmatics of Word-order: Predictability, Importance and Attention. In M. Hammond (ed.), *Studies in Syntactic Typology*, 243-84. Amsterdam: John Benjamins.
- Givón, Talmy. 1993. Coherence in Text, Coherence in Mind. *Pragmatics and Cognition* 1(2):171-227.
- Givón, Talmy. 1995. *Functionalism and grammar*. Amsterdam: John Benjamins Publishing Company.
- Givón, Talmy. 2001a. *Syntax*. Vol. 1. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Givón, Talmy. 2001b. *Syntax*. Vol. 2. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Grimshaw, Jane, and Mester, Armin. 1988. Light Verb and (Theta)-Marking. *Linguistic Inquiry* 19: 205-232.
- Gundel, Jeanette. 1985. 'Shared Knowledge' and Topicality. *Journal of Pragmatics* 9:1, 83-107.
- Gundel, Jeanette, Nancy Headberg, and Ron Zacharski. 1993. Cognitive Status and the Form of Referring Expressions in Discourse. *Language* 69:274-307.
- Han, Chung-Hye. 1998. Asymmetry in the interpretation of *-(n)un* in Korean. *Japanese and Korean Linguistics*. Vol 7. eds., by Noriko Akatsuka, Hajime Hoji, Shoichi Iwasaki, Sung-Oak Sohn, and Susan Strauss. CSLI, Stanford, 1-15.
- Han, Jeonghan. 1999. *Grammatical Coding of Information Structure in Korean: a Role*

- and Reference Grammar (RRG) Account*. Ph. D. dissertation, State University of New York at Buffalo.
- Han, Jeong-han. 2002. Linking Algorithms of KA, LUL, *Language Research* 38(3), 827-850.
- Hasegawa, Yuri. 1993. “hanasi kotoba niokeru ‘muzyosi’ no kinoo” (The Function of the Null Particle in the Spoken Language). *Nihongo kyooiku* 40:158-68.
- Herring, Susan C. 1990. Information structure as a consequence of word order type. In *Proceedings of Berkeley Linguistics Society*, 163-174.
- Hinds, John. 1982. *Ellipsis in Japanese Discourse*. Alberta: Linguistic Research, Inc.
- Hinds, John. 1983. Topic Continuity in Japanese. In T. Givón (Ed.), *Topic Continuity in Discourse: a Quantitative Cross-language Study*, 43-93. Amsterdam: John Benjamins Publishing Company.
- Hong, Ki-Sun. 1991. *Argument Selection and Case Marking in Korean*. Ph. D. dissertation, Stanford University.
- Hong, Minpyo, Kyongjae Park, Inkie Chung, and Ji-young Kim. 1998. Elided Postpositions in Spoken Korean and their Implications on Center Management. *Korean Journal of Cognitive Science* 9:3, 51-61.
- Hopper, Paul J., and Sandra A. Thompson. 1980. Transitivity in Grammar and Discourse. *Language* 56(2):251-299.
- Iwasaki, Shoichi. 1987. Identifiability, scope-setting, and the particle *wa*: a study of Japanese spoken expository discourse. In J. Hinds, S. K. Maynard, and S. Iwasaki (eds), *Perspectives on Topicalization: the Case of Japanese Wa*, 107-141. Amsterdam: John Benjamins.
- Jaeger, Jeri. J. 2005. *Kids’ Slips*. Lawrence Erlbaum Associates, Inc., Publishers: Mahwah, NJ.

- Jun, Jong-Sup. 2003. *Syntactic and Semantic Bases of Case Assignment: A Study of Verbal Nouns, Light Verbs and Dative*. Ph. D. dissertation, Brandeis University.
- Kang, Kyung-Shim. 1998. *The Korean Topic Marker and Activation*. Unpublished manuscript. State University of New York at Buffalo.
- Kang, Young-Sei. 1985. Case Marking in Korean. *Harvard Studies in Korean Linguistics* 1:88-100.
- Keenan, David L. 1976. Toward a universal definition of 'subject'. In C. N. Li (Ed.), *Subject and Topic*, 305-33. New York: Academic Press.
- Kim, Jong-Bok. 2000a. *The Grammar of Negation: A Constraint-Based Approach*. Stanford, California. CSLI publication.
- Kim, Jong-Bok. 2000b. A constraint-based and head-driven analysis of multiple nominative constructions. In D. Flickinger and A. Kathol (Eds.), *Proceedings of the 7th International Conference on Head-Driven Phrase Structure Grammar*, Vol. 7, 166-181, Stanford, California. CSLI publication.
- Kim, Kwang-Sup. 1990. Where do the contrastive and focus readings come from? In *Japanese/Korean Linguistics*, ed. by Hajime Hoji, 395-412.
- Kim, Kyu-Hyun. 1993. Topicality in Korean Conversation: A Conversation Analytic Perspective. *Japanese/Korean Linguistics*, eds. by Patricia M. Clancy, vol 2: 33-54.
- Kim, Sunwoong. 1996. Hankwuke cwukyek mokceke yenkwu 'a study of nominative marked objects in Korean'. *Ene 'Language'*: 21-1 & 2: 241-261.
- Kim, Taeho. 2006. *Zero-marking and -lul marking for direct object in conversational Korean*. Unpublished ms., Paper presented at 36th Annual Meeting of Michigan Linguistic Society, Oakland, Michigan, USA.
- Kim, Young-Joo. 1990. *The Syntax of Semantics of Korean Case: The Interaction Be-*

- tween Lexical and Syntactic Levels of Representation*. Ph. D. dissertation, Harvard University.
- Ko, Eon-Suk. 2000. A Discourse Analysis of the Realization of Object NP Forms in Korean. *Ohak Yonku* 36 (1):43-62.
- Ko, Seok-Ju. 2000. *Hankwuke cosauy yenkwu: '-ka'wa '-lul'ul cwungsimulo* 'Study of Korean Particles: focusing on '-ka' and '-lul'. Ph. D. dissertation, Yonsei University.
- Kuno, Susumu. 1972. Functional sentence perspective: A case study from Japanese and English. *Linguistic Inquiry* 3:269-320.
- Kuno, Susumu. 1973. *The Structure of the Japanese Language*. Cambridge, MA: MIT Press.
- Kuno, Susumu. 1978. Japanese: A Characteristic OV language. In W. P. Lehmann (Ed.), *Syntactic Typology*, chapter 2, 57-138. University of Texas Press.
- Lambrecht, Knud. 1994. *Information structure and sentence form: topic, focus, and the mental representations of discourse referents*. Cambridge University Press.
- Larson, Richard K. 1991. Some issues in verb serialization. In C. Lefebvre (Ed.), *Verb Serialization*. Amsterdam: John Benjamins Publishing Company.
- Lee, Chung-Min. 1989. (In)definite, case markers, classifiers and quantifiers in Korean. In Kuno, Susumu *et al.* ed., *Harvard Studies in Korean Linguistics*. 469-487. Seoul: Hanshin.
- Lee, Chung-Min. 1999. Contrastive Topic: A Locus of the Interface-evidence from Korean and English. In K. Turner *et al.* (eds.), *The Semantics/Pragmatics Interface from Different Points of View (CRiSPI 1)*, Elsevier Science.
- Lee, Chung-Min. 2003. Contrastive Topic and/or Contrastive Focus. *Japanese and Korean Linguistics* 12:352-364.

- Lee, Duck-Young. 2002. The Function of the Zero Particle with Special Reference to Spoken Japanese. *Journal of Pragmatics* 34:645-82.
- Lee, Hyo-Sang, and Sandra A. Thompson. 1989. A Discourse Account of the Korean Accusative Marker. *Studies in Language* 13 (1):105-128.
- Lee, Kiri. 2002. Nominative Case-marker Deletion in Spoken Japanese: an Analysis from the Perspective of Information Structure. *Journal of Pragmatics* 34, 683-709.
- Makino, Seiichi and Michio Tsutsui. 1986. *A Dictionary of Basic Japanese Grammar*. The Japan Times, Tokyo.
- Masunaga, Kiyoko. 1988. Case Deletion and Discourse Context. In W. J. Poser (Ed.), *Papers from the Second International Workshop on Japanese Syntax*, 145-56. Stanford, California: CSLI.
- Matsuda, Kenjiro. 1996. *Variable zero-marking of (O) in Tokyo Japanese*. Ph. D. dissertation, University of Pennsylvania.
- Maynard, Senko K. 1980. *Discourse Functions of the Japanese Theme Marker -wa*. Ph. D. dissertation, Northwestern University.
- Maynard, Senko K. 1981. The Given/New Distinction and the Analysis of Japanese Particles WA and GA. *Papers in Linguistics* 14 (1):109-30.
- Maynard, Senko K. 1989. *Japanese Conversation: Self-contextualization through Structure and Interactional Management*. Ablex., NJ: Norwood.
- Maynard, Senko K. 1993. *Kaiwa bunseki* "Conversation Analysis." Tokyo: Kuroshio.
- Min, Hyun-Sik. 1982. A study of the case in modern Korean. *Korean Studies* 49.
- Mok, Jung-Soo. 1998. Re-examination of Korean case markers and special particles: a typological approach. *Linguistics* 23, 47-78.
- Muansuwan, Nuttanaart. 2000. *Lexical errors in Thai slips of the tongue*. Unpublished manuscript. University at Buffalo (SUNY).



- Nam, Ki-Sim. 1996. *Exploring Korean Grammar 1*. Seoul: Tayhaksa.
- Ono, Tsuyoshi, Sandra A. Thompson, and Ryoko Suzuki. 2000. The pragmatic nature of the so-called subject marker ‘ga’ in Japanese: evidence from conversation. *Discourse Studies* 2(1):55-84.
- Otani, Hiromi. 1995. *Wa to ga to Ø - wa mo ga mo tukaenai bun (Wa, ga, and Ø - Sentences in Which wa or ga Cannot Be Used)*. In T. Miyajima and Y. Nita (eds), *Nihongo ruigi hyoogen no bunpo: tanbun hen (Japanese Grammar of Similar Expressions: the Simple Sentences)*, 287-95. Tokyo: Kuroshio.
- Partee, B. H. 1991. Topic, Focus and Quantification. in *SALT* 1: 159-187.
- Prince, Ellen F. 1981. Toward a Taxonomy of Given-New Information. In P. Cole (Ed.), *Radical Pragmatics*, 223-55. New York, NY: Academic Press.
- Reinhart, T. 1981. Pragmatics and Linguistics: an Analysis of Sentence Topics. *Philosophica* 27, 53-94.
- Rochemont, Michael S. 1986. *Focus in Generative Grammar*. Amsterdam: John Benjamins.
- Rochemont, Michael S., and Peter W. Culicover. 1990. *English Focus Constructions and the Theory of Grammar*. Cambridge: Cambridge University Press.
- Saeed, John I. 2003. *Semantics (Introducing Linguistics)*. 2nd edition. Blackwell.
- Schütze, Carson T. 2001. On Korean “Case Stacking”: The varied functions of the particles ‘ka’ and ‘lul’. *The Linguistic Review* 18:193-232.
- Sells, Peter. 2006. Interactions of Negative Polarity Items in Korean. *Proceedings of the 11th Harvard International Symposium on Korean Linguistics*, Harvard University, 724-737.
- Shibatani, Masayoshi. 1990. *The Language in Japan*. Cambridge University Press.
- Shibatani, Masayoshi. 1991. Grammaticalization of Topic and Subject. In E. C. Trau-

- gott and B. Heine (eds), *Approaches to Grammaticalization, Vol. 2: Focus On Types of Grammatical Markers*, 93-133. Amsterdam: John Benjamins.
- Shimojo, Mitsuaki. 1995. *Focus Structure and Morphosyntax in Japanese: Wa and Ga, and Word Order Flexibility*. Ph. D. dissertation, State University of New York at Buffalo.
- Shimojo, Mitsuaki. 2005. *Argument Encoding in Japanese Conversation*. Palgrave.
- Shimojo, Mitsuaki, and Hye-Won Choi. 2001. On Asymmetry in Topic Marking - The Case of Japanese WA and Korean NUN. In J. Boyle, J-H Lee, and A. Okrent (eds.), *Chicago Linguistic Society 36*, Vol 1: The Main Session, 455-467. Chicago: University of Chicago Press.
- Simon, Mutsuko E. 1989. *An Analysis of the Postposing Construction in Japanese*. Ph. D. dissertation, University of Michigan.
- Sohn, Ho-Min. 1999. *The Korean Language*. Cambridge University Press.
- Song, Hongkyu. 2005. Causatives and resultatives in Korean. Ph. D. dissertation, The University of Wisconsin at Madison
- Suzuki, Satoko. 1995. The Functions of Topic-encoding Zero-marked Phrases: a Study of the Interaction among Topic-encoding expressions in Japanese. *Journal of Pragmatics* 23:607-26.
- Szabolcsi, Anna. 1981. The semantics of topic-focus articulation. In *Formal Methods in the Study of Language*, ed. by T. M. V. Janssen & M. B. J. Stokhoj, 513-540. Amsterdam: Mathematisch Centrum.
- Thompson, Sandra A., and Paul J. Hopper. 2001. Transitivity, clause structure, and argument structure: evidence from conversation. In J. Bybee and P. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure*, 27-60. Amsterdam: John Benjamins Publishing Company.

- Tomlin, Russel. 1995. Focal Attention, Voice, and Word Order: an Experimental, Cross-linguistic Study. In P. Downing and M. Noonan (Eds.), *Word Order in Discourse*, 517-54. Amsterdam: Amsterdam: John Benjamins.
- Trask, Robert Lawrence. 1999. *Key Concepts in Language and Linguistics*. New York, NY: Routledge.
- Tsutsui, Michio. 1984. *Particle Ellipsis in Japanese*. Ph. D. dissertation, University of Illinois, Urbana.
- Um, Jungho. 2003. Case Alternation in ‘-go siphta’ Construction. *Korean Linguistics* 41:169-195.
- Vallduví, Enric. 1993. *Information packaging: A survey*. ms. Centre for Cognitive Science and Human Communication Research Centre, University of Edinburgh.
- Vallduví, Enric, and Maria Vilkuna. 1998. On Rheme and Kontrast. In Peter Culicover and Louise McNally, eds., *The Limits of Syntax, Syntax and Semantics* 29: 79-108. New York: Academic Press.
- Van Valin, Robert D., and Randy J. Lapolla. 1997. *Syntax: structure, meaning, and function*. Cambridge, UK: Cambridge University Press.
- Van Valin, Robert D. 2003. *Role and Reference Grammar: a theory of the syntax-semantics-pragmatics interface*. Cambridge: Cambridge University Press.
- Watanabe, Yasuko. 1989. *The Function of WA and GA in Japanese Discourse*. Ph. D. dissertation,, University of Oregon.
- Wee, Hae-Kyung. 1998. Semantics and Pragmatics of Contrastive Topic in Korean and English. *Japanese/Korean Linguistics*, Vol 7. eds., by Noriko Akatsuka, Hajime Hoji, Shoichi Iwasaki, Sung-Oak Sohn, and Susan Strauss. CSLI, Stanford, 549-65.
- Whitman, John. 1989. Topic, Modality, and IP Structure. *Harvard Studies in Korean Linguistics*, Vol 3, 341-56.

- Yamasaki, Yumi. 2005. *Constituent Ordering in Spoken Japanese*. Ph.D. dissertation, State University of New York at Buffalo.
- Yanagida, Yuko. 1985. Informational predictability and case-marking variation in Japanese. *Text* 5:1-2. 123-145.
- Yang, Dong-Hwi. 1999. Case features and case particles. In S. Bird, A. Carnie, J. D. Haugen, and P. Norquest (Eds.), *Proceedings of the 18th West Coast Conference on Formal Linguistics*, Vol. 18, 626-639, Somerville, MA. Cascadilla Press.
- Yamashita, Hiroko. 2002. Scrambled Sentences in Japanese: Linguistic Properties and Motivations for Production. *Text* 22, 597-633.
- Yngve, Victor H. 1970. On Getting a Word in Edgewise. *Papers from the Sixth Regional Meeting of the Chicago Linguistics Society* 567-78.
- Yoon, James H. 2004. Non-nominative (major) subjects and case stacking in Korean. In P. Bhaskararao and K. V. Subbarao (Eds.), *Non-nominative Subjects*, 265-314. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Yoon, James H. 2007. *Subjecthood and Subject Properties in Multiple Subject Constructions*. Unpublished ms., University of Illinois at Urbana-Champaign.
- Yoon, James H. to appear. The Distribution of Subject Properties in Multiple Subject Constructions. In *Proceedings of the 16th Japanese/Korean Linguistics Conference*. CSLI Publications.