A CORPUS-BASED ANALYSIS OF DISCOURSE ANAPHORA IN ENGLISH AND KOREAN: A NEO-GRICEAN PRAGMATIC APPROACH

by

SOK-HUN KIM

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

May 2013

Acknowledgements

Without constant encouragement and support from faculty members, friends, and family, this dissertation could not have seen the daylight.

First and foremost, a special thanks to my mentor and dissertation supervisor, Dr. Laurel Smith Stvan, whose deep commitment to my research and constant guidance made this journey a rewarding experience of my life. She has helped me to shape and develop this research with insightful comments and warmhearted encouragement. I deeply appreciate her mentorship and unfailing patience for the past years. Additional special thanks go to the other committee members: Dr. Jeffrey Witzel, who always guides me with insightful and critical suggestions not only on research methods but also on statistical analysis, and Dr. Joseph Sabbagh, who always inspires me with thoughtful comments on this research as well as on syntax/semantics.

Beyond the committee, I would like to thank other faculty members: Dr. David J. Silva, who let me begin this valuable journey with fellowship, Dr. Colleen Fitzgerald, who helped me navigate toward successful completion with scholarship, Dr. Jerold Edmondson, Dr. Donald Burquest, Dr. Cynthia Kilpatrick, and Dr. Stephen Parker (GIAL).

I would also like to extend a special note of thanks to my mentor, Prof. Sungbom Lee at Sogang University, Korea, who leads me this great sea of semantics/pragmatics and linguistics, and constantly guides me with a beacon light of his academic enthusiasm. My greatest intellectual debt is to him and I was extremely fortunate to be taught and guided by him during my graduate studies.

I deeply appreciate all the supports from my friends at UT Arlington: JungAe Lee Allman, Joshua Jensen, Vitaly Voinov, Jung Sun Son, Ju-Young Lee, Heilena Sohn, Yilmin Koo, Yi-Ping Wang, and Keunhyung Park. Especially, I would express a special thanks to Yujeong Choi, who encourages and helps me all the time.

I am also deeply grateful to my (senior) friends in Korea: In-Young Kim, Wonbin Lee, Jong Un Park, Seung Ku Park, Inkie Chung, Seungwon Baek, Jung Yeon Kim, Sang Yel Lee, Myeng Hun Lee, and Hyun-Suk Min. Especially, I cannot thank Sang Jo Kim enough for his constant encouragement and care, and I am fortunate to have him as a lighthouse not only during my journey in the U.S. but also in my life.

Last but not least, my dissertation is dedicated to my beloved three ladies: my grandmother, Ye Sun Son, my mother, Jeongsook Yu, and my wife, Bo Young Lee. Their strong belief and unconditional support goes beyond words. SHEs teach me love and sacrifice in my life, and SHEs are my compass at sea of my life. My dissertation is also dedicated to my father, Kyeng Jae Kim, who has been staying in heaven for the last ten years. I would also like thank my parents-in-law, Sang Cheon Lee and Sun Ja Park, for their endless support and care.

April 17, 2013

Abstract

A CORPUS-BASED ANALYSIS OF DISCOURSE ANAPHORA IN ENGLISH AND KOREAN: A NEO-GRICEAN PRAGMATIC APPROACH

Sok-Hun Kim, PhD

The University of Texas at Arlington, 2013

Supervising Professor: Laurel Smith Stvan

This dissertation explores discourse anaphora in English and Korean by using a neo-Gricean pragmatic approach with corpus-based data. Very little study of Korean discourse anaphora has yet taken place at the inter-sentential level, except works looking at zero anaphor and a logophoric reflexive pronoun *caki* 'self'. This research fills this gap by examining two types of discourse anaphora at the discourse level: discourse anaphoric patterns (by order of mention of the referent and by placement of the paragraph) and sentential anaphors.

Two quantitative methods were adopted to verify the distribution and the selection of anaphora: natural data collection and a survey. First, samples of 30,000 running words from newspaper articles (for discourse anaphoric patterns) and the same size of samples from drama scripts (for sentential anaphors) in each language were investigated for each issue. Second, 20 native speakers of English and 20 native speakers

of Korean were recruited to take part in two sets of a threefold acceptability survey for two types of discourse anaphora.

Based on the findings, the distinct characteristics signaling the appropriateness of different anaphors are qualitatively discussed within four relevant theories: topic continuity theory, hierarchy theory, cognitive theory, and principled neo-Gricean theory. First, discourse anaphoric patterns are examined in two respects: by order of mention of the referent and by placement of the paragraph. For both mention types, it is argued that there are general vs. sequential chains of anaphoric patterns: the first type forms a general chain with <full name, single name, pronoun> in both languages, whereas the second type forms a sequential chain with <full name, full name, full name>, notably in Korean. Second, referential properties of sentential anaphora are accounted for in terms of degree of anaphoricity in two languages. Lastly, it is argued that discourse anaphora in both languages can be more comprehensively accounted for through the use of neo-Gricean heuristics.

Table of Contents

List of Illustrations

List of Tables

List of Abbreviations

Nom (NOM)-Nominative
Top (TOP)-Topical Marker
Acc (ACC)-Accusative
Dat (DAT)-Dative
Gen (GEN)-Genitive
LOC -Locative
ABL-Ablative
INST-Instrument
Decl-Declarative
Pres-Present Tense
Past-Past Tense
Rel-Relative Clause
Int-Interjection
FN-Full Name
SN-Surname
Pro-Pronoun
GN-Given Name
(L)NP-(Lexical) Noun Phrase
TT-Title
NN-Nickname
AP-Apposition
AFF-Affiliation
Ø-Zero Anaphor
├ - Entailment
+> - (Conversational) Implicature
~ - Negation
<x, y> - Q-scale
[x, y] - I-scale
{x, y} - M-scale

Chapter 1

Introduction

In this dissertation, I address discourse anaphora in English and Korean in terms of pragmatics. Specifically, I explore different discourse anaphoric patterns and referential properties of sentential anaphora in both languages by analyzing samples of corpus data and survey data. I am interested in anaphora at the inter-sentential level since in natural language (discourse), "anaphora is a specific sinew of textual muscle" (Kittay 1988:206). In fact, an addresser can represent his/her conversational intention in a given discourse (context) with different types of discourse anaphoric forms; the addressee can infer the intention by means of conversational implicature. In this respect, I have noticed that there has been particular discourse anaphora that needs to be explained linguistically such as discourse anaphoric pattern and sentential anaphora.

To create the stepping stone for exploring those discourse anaphoric phenomena in this dissertation, this first chapter presents the relevant problem, the research gap, the research scope, the research questions, the rationale, and the organization of this dissertation.

1.1 The Research Gap

Discourse anaphora, one type of reference, denotes a relation between two linguistic referring expressions in a given discourse, in which an anaphor refers to the same entity (via an antecedent) as another earlier referring expression form. The interpretation of an anaphor is influenced by that of an antecedent (Huang 2007: 245). It is represented by a variety of referring expressions, such as (definite) descriptions with

articles, names, pronouns, reflexives, demonstratives, and gaps (empty categories). It is also represented in the range of intra-sentence, inter-sentence, and inter-turns at speaking in a verbal discourse (Levinson 1983: 86).

The study of discourse anaphora in terms of distribution and selection has been explored in light of four relevant theories: the topic continuity theory (Givón 1983; Brwon 1983; Gundel 1985; Lambrecht 1994); the hierarchy theory (Grimes 1975; Hinds 1978, 1979; Longacre 1979; Mann and Thompson 1987, 1988; Fox 1987); the cognitive theory (Prince 1981; Ariel 1988, 1990, 1994, 2008, 2010; Gundel 1988, 1996, 2003; Gundel, Hedberg, and Zachrski 1988, 1989, 1990, 1993, 2001); the neo-Gricean theory (Grice 1975, 1978; Horn 1984, 1989; Levinson 1987a, 1987b, 1991, 1995, 2000; Huang 1991, 1994, 2000a, 2000b, 2007).

Those previous theories have explored the distribution of discourse anaphora in a variety of languages, including English, German, Spanish, Italian, Irish, Icelandic, Finnish, Russian, Japanese, Chinese, and Korean. Among those languages, the study of discourse anaphora in Korean has been focused on two anaphoric forms: zero anaphora (Ø) in terms of the topicality and a reflexive pronoun *caki* 'self' (corresponding *zibun* 'self' in Japanese; *ziji* 'self' in Chinese) with its variants *casin* and *caki-casin* in terms of logophoricity and long distance anaphora (LDA).

However, it is noteworthy that earlier studies of these two particular items in Korean leave several key areas unexplored. First, these studies have been restricted to the intra-sentential level, not examining the inter-sentential level or discourse level. In fact, Korean discourse anaphora at an inter-sentential level or a discourse level has been little

explored. In addition, as Lee (2010) also notes, little attention has been paid to explore how discourse anaphors are referred and deployed in written discourse such as news articles and TV drama scripts from a contrastive and cross-linguistic perspective.

Second, these studies have been restricted to the components of syntax and semantics, notably classical Chomskyan theory of anaphora, i.e., the binding theory (BT). However, many studies including these four relevant theories I plan to use have verified that the distribution of discourse anaphora in discourse is determined by particular factors, such as topic continuity, hierarchical structure, cognitive condition, and pragmatic heuristics beyond the inter-sentential structure. Third, there is some doubt raised by the fact that the data used in these analyses originated from Korean native speaker's intuition or Korean linguists' intuition about created examples, not from naturally occurring authentic data at the level of discourse. In fact, it seems that the examples related to those two forms are shown in a vacuum. Therefore, the authentic data in dynamically developing context-based discourse can verify the fundamental properties of discourse anaphora.

1.2 Research Scope and Questions

This dissertation focuses on two phenomena of discourse anaphora at a discourse level in English and Korean as follows.

The first aspect focuses on two discourse anaphoric patterns in English and Korean: the general anaphoric pattern (GAP) vs. the sequential anaphoric pattern (SAP). The anaphoric pattern forms a coreferential chain of discourse anaphors <full name,

pronoun/surname, pronoun/surname> [1] , whereas the sequential pattern forms a coreferential chain of discourse anaphors <full name, full name, full name>, notably, this pattern is observed in Korean.

The second point focuses on (different) referential properties of sentential anaphora in English and Korean. In English, *that* and *it* are used to refer to both nominal and sentential referents. In Korean, the literally corresponding items *ce* 'that' and *ku*'it' are used to refer to both nominal referents; however, only *ku* 'it' is used to refer to the sentential referents. That is, a sentential anaphor *ku* 'it' in Korean has dual functions with its two variants: *ku-ken* functioning as *that* and *ku-len* functioning as *it*.

Four relevant studies, i.e., the topic continuity theory, the hierarchy theory, the cognitive theory, and the inferential pragmatic theory, are used to explain the three aspects of discourse anaphora in comparing English with Korean. The strength and weakness in each model are explored and the most comprehensive model is explained through empirical verification.

To verify the distribution and the selection of discourse anaphoric elements, two empirical methods are employed: natural data collection and a survey. First, relevant text samples were collected from newspaper articles in each language and from TV drama scripts such as an American mockumentary comedy *The Office* in English and a serial drama *My Too Perfect Sons* in Korean. Second, a survey with selected news articles from the corpus samples was conducted. 20 native speakers of English and 20 native speakers of Korean were recruited for a threefold survey (i.e., the discourse anaphoric pattern by

[1] The angle brackets < > indicate the ordered sequence through this dissertation.

order of mention, the discourse anaphoric pattern by paragraph, and the sentential anaphors) in each language.

The four relevant models, such as the topic continuity model, hierarchical model, cognitive model, and principled pragmatic model, were tested by the results of the sampled corpus and the survey in terms of predictions of these models in a hypothetico-deductive way. Based on my earlier studies (Kim 2010a, 2010b, 2012), it was hypothesized that each would have some weakness in explaining discourse anaphora in Korean, specifically in (extra-) linguistic respects: socio-cultural points, linguistic points, cognitive processing, and pragmatic heuristics. Thus, a synthetic account was reexamined based on the corpus and the survey results, and a comprehensive model was suggested in terms of pragmatic inferential heuristics from a neo-Gricean perspective.

In order to explore the discourse anaphoric patterns and the properties of sentential anaphors, the following questions are explored:

1. How are discourse anaphora in English and Korean actually distributed and selected at a discourse level in terms of the two linguistic issues stated below?

a. The discourse anaphoric patterns in newspaper articles in two languages.

b. The shared and different referential properties of sentential anaphors: *it* and *that* in English and the corresponding items *ku-ken* ‘that’ and *ku-len* ‘it’ in Korean.

2. What are the distinct characteristics of those particular aspects of discourse anaphora in English and Korean?

3. Which theory is more appropriate and preferable for explaining those aspects of discourse anaphora across two languages?

a. What factors in each theory need to be significantly reinterpreted and modified to explain those aspects?

b. If aspects are left unaccounted for, what other factors are involved in terms of principle-based pragmatic perspective?

1.3 Rationale

In this section, I address the several reasons why I feel that it is important to study referring expressions and why I focus on discourse anaphora through a window of pragmatics, based on key notions presented in Verschueren (1999:55-58).

First, using language necessarily entails making linguistic choices continuously due to linguistic and extra-linguistic reasons. In fact, these linguistic choices can be made at any possible level of linguistic form such as phonetic/phonological, morphological, syntactic, lexical, semantic, and pragmatic. Each point of view in linguistics needs to validly explain these linguistic choices in its own term.

Second, these linguistic choices are made when an addresser produces an utterance and when the addressee understands that utterance, which is significant for the purpose of communication or discourse that they are engaged. In this mutual process, contextual meaning is generated, carried out, and conveyed between interlocutors according to appropriate communicative goal.

Third, unless interlocutors intend to remain silent, interlocutors are under a (semi-) obligatory situation to make a linguistic choice once even a word is uttered. That is, making a linguistic choice is inevitable to interlocutors in a given discourse.

Fourth, as a principle, the linguistic choices carried out by interlocutors are not equivalent in terms of markedness: in between unmarked choice and marked choice. For this reason, making linguistic choices evoke their alternatives, in which conversational implicature is carried out by interlocutors. For instance, interlocutors are under (semi-) compulsory situation to make choice of an appropriate referential form between a variety of referring expressions, specifically, discourse anaphors, in a given discourse. If an addresser chooses one particular discourse anaphoric form with intention between different types of forms, the addressee interprets and understands why that discourse anaphor was chosen between alternatives in a given discourse, which is an inferential process of conversational implicature.

1.4 Organization of the Dissertation

This chapter addresses the research gap, the research scope, the research questions, and the rationale for the research. The remaining five chapters with topics and appendices are organized as follows.

Chapter 2 reviews the theoretical background and scholarly works that are relevant to the two topics in this dissertation. The chapter begins with a discussion of anaphora in syntax vs. discourse anaphora in pragmatics. Then relevant implication of discourse anaphora and the definition are reviewed. The relevant four theoretical models for explaining discourse anaphora in pragmatics are reviewed in terms of the topic

continuity model, the hierarchy model, the cognitive model, and the inferential pragmatic model. Finally, the scholarly works relevant to the two topics in this dissertation are also reviewed.

Chapter 3 describes the twofold methods used in this research: corpus data collection and a survey. First, a detailed overview of the data collection is presented regarding compilation of three parts of the corpus (Corpus I, II, and III), the corpus tool, and the variables used. Second, detailed processes of the survey are presented in terms of subjects, questionnaire (Section A, B, C, and D), variables, and an approval from the Institutional Review Board. Finally, the chapter concludes with a summary of the research methods.

Chapter 4 presents the comparative results and analyses of the sampled corpus in English and Korean. First, the results of the discourse anaphoric patterns by order of mention (Corpus I) are reviewed in terms of the first-, second-, third mention, and case markers. Second, the results of the discourse anaphoric patterns by paragraph (Corpus II) are reviewed in terms of the first-, second-, third mention, and case markers. Third, the results of the distance between discourse anaphors (Corpus I and II) are reviewed. Fourth, the results of sentential anaphors (Corpus III) are reviewed in terms of the givenness of information and the width of reference. Finally, the chapter concludes with a summary of the results and analyses in Corpus I and II, and Corpus III.

Chapter 5 presents the comparative results and analyses of the survey in English and Korean. First, the results of discourse anaphoric patterns by order of mention (Section B) are reviewed in terms of the first-, second, and third mention. Second, the

results of discourse anaphoric patterns by paragraph (Section C) are reviewed in terms of the first-, second, and third paragraph. Third, the results of sentential anaphors (Section D) are reviewed in terms of the givenness of information and the width of reference. Finally, the chapter concludes with a summary of the results and analyses in Section B and Section C, and Section D.

Chapter 6 discusses all of the results and analyses from Chapter 4 and Chapter 5 and concludes this dissertation with the benefits of this research and the future avenues of the research. First, the discussion begins with an argument of two discourse anaphoric patterns (GAP and SAP) in English and Korean. The mention type (order of mention and paragraph) and the distance between discourse anaphors are discussed. The argument of GAP and SAP is discussed in terms of socio-cultural, linguistic, cognitive, and pragmatic points. Second, shared referential properties of sentential anaphors in both languages are discussed regarding the givenness of information and the width of reference. Different properties of sentential anaphors in two languages are also discussed. Third, all discussion and argument are schematized in terms of neo-Gricean heuristics of conversational implicature. Finally, this chapter includes the benefits of this research and the future avenues of the research for investigation with the limitations.

Finally, Appendix A contains the survey questionnaire in English; Appendix B contains the survey questionnaire in Korean. Each survey questionnaire includes four sections: Section A (general questions about subjects); Section B (two sets of discourse anaphoric pattern by order of mention); Section C (two sets of discourse anaphoric pattern by paragraph); Section D (four conversations of sentential anaphors). Appendix C

contains the two approval letters and the informed consent form for subjects approved by the Institutional Review Board of the University of Texas at Arlington.

Chapter 2

Literature Review

The goal of this chapter is to provide an overview of the relevant theoretical issues that play a part in explaining discourse anaphora. Section 2.1 begins with a discussion of anaphora in syntax in terms of Chomsky's binding theory (Section 2.1.1), gives counterexamples to this syntactic view by considering cross-linguistic and pragmatic points (Section 2.1.2), gives an implication of discourse (Section 2.1.3), and gives a definition of discourse anaphora (Section 2.1.4). Section 2.3 through Section 2.5 gives an overview on the four main theoretical models for discourse anaphora in pragmatics: (i) the topic continuity theory (Section 2.3), (ii) the hierarchy theory (Section 2.4), (iii) the cognitive theory (Section 2.5), and (iv) the neo-Gricean theory (Section 2.5). Section 2.6 and Section 2.7 give theoretical backgrounds on the two main subjects of the dissertation: (i) discourse anaphoric pattern (Section 2.6) and (ii) sentential anaphora (Section 2.7). The chapter is wrapped up with a summary of literature review.

2.1 Discourse Anaphora

2.1.1 Anaphora in Syntax

In a syntactic framework, anaphora falls into two main categories: NP- (including N-) anaphora and VP-anaphora[2]. NP-anaphora in syntax has been traditionally discussed in terms of Chomsky's (1981, 1982, 1986, 1995) binding theory (BT) within the principles and parameters theory.

[2] I shall not discuss VP-anaphora throughout the dissertation. VP-anaphora falls into four types in syntactic category: VP-ellipsis (coordinated VP-ellipsis and subordinated VP-ellipsis), gapping, sluicing, and stripping (Huang 2000:3-5). Refer to Huang (2000, 2006).

In NP-anaphora, the anaphor and its antecedent are all NPs, and they are both referring expressions[3]. NP anaphora is encoded by gaps (empty categories), reflexives, pronouns, names, and descriptions. Within a classical Chomskyan theory of anaphora (i.e., the principles-and-parameters theory and its minimalist descendant), NPs are distinguished into two types of abstract binary feature: [±anaphor] and [±pronominal]. First, an anaphor, including a reflexive and a reciprocal, is an NP with the feature [+anaphor] that must be referentially dependent, and that must be bound by an NP in an A[rgument]-position within a certain domain, i.e., the binding domain, which is also referred to as the governing category (GC). Second, a pronominal is an NP with the feature [+pronominal] that may be referentially dependent, but that must be free in its GC. Chomsky (1981, 1982, 1986, 1995) proposes four types of NP in a language in terms of overt and non-overt (i.e., empty) with two independent binary features of anaphor and pronominal, as illustrated in (1) (Haegeman 1994; Huang 1994, 2000, 2007).

(1) Chomsky's typology of NPs

		Overt	Non-overt (Empty)
a.	[+anaphor, -pronominal]	lexical anaphor	NP-trace
b.	[-anaphor, +pronominal]	pronoun	*pro*
c.	[+anaphor, +pronominal]		PRO
d.	[-anaphor, -pronominal]	name	*wh*-trace/variable

[3] An NP-anaphor can refer to a CP (i.e., a clause plus a complementizer) or an IP (i.e., a clause) for its antecedent. In this case, this NP-anaphor indicates the action, event, state, or proposition represented by this CP or IP (Huang 2000:3).

This typology of NPs can be encoded by overt NPs and empty categories, as shown in (2) and (3).

(2) Overt NPs

a. Lexical anaphor

The composers$_1$ admire themselves$_1$/each other$_1$.

b. Pronoun

Poirot$_1$invited him$_{*1/2}$;

Poirot$_1$ says that he$_{1/2}$ is leaving.

c. Names (R[eferential]-expression)

He$_1$ says that Poirot$_{*1/2}$ is leaving.

(3) Empty categories

a. NP-trace

The giant panda$_1$ seems t_1 to live exclusively on bamboo shoots.

` b. *Pro* (Italian, Haegeman 1994)

Gianni$_1$ dice che *pro*$_1$ ha telefonato.

Gianni says that has telephoned

'Gianni says that (he) has telephoned.'

c. PRO

John$_1$ promised PRO$_1$ to compose a light orchestral work for his father.

d. *Wh*-trace/variable

Who$_1$ did Brahms admire t_1?

The three types of NPs listed in (1) and (2), anaphors, pronominals, and r-expressions, are subject to binding theory (BT) A, B, and C, respectively, as shown in (4).

(4) Chomsky's binding theory (BT)

a. An anaphor is bound in a local domain (Principle A).

b. A pronominal is free in a local domain (Principle B).

c. An R-expression is free (Principle C).

The definition of binding is specified in the conditions (5).

(5) α binds β if and only if (iff)

a. α is in an A-position,

b. α c-commands β, and

c. α and β are coindexed.

The c[onstituent]-command is defined in (6).

(6) α c-commands β iif

a. α does not dominate β,

b. β does not dominate α, and

c. the first branching node dominating α also dominates β.

Lastly, the notion of local domain is defined in terms of governing category (GC) or complete functional complex (CFC)[4]. The GC is generally defined in (7) as follows.

(7) α is a GC iff α is the minimal category (i.e., the smallest NP or IP) containing β, a governor of β, and a SUBJECT accessible to β.

The typical examples in English for BT are represented in (8) (Huang 2007:249).

(8) a. Bach_1 adored $\text{himself}_{1/*2}$.

b. Bach_1 adored $\text{him}_{*1/2}$.

[4] According to Chomsky (1986:169), CFC can be defined as 'the smallest maximal category containing all the grammatical functions compatible with its head' of a projection, the predicate that assigns the theta role, the complement that the internal theta role are assigned to, and the subject that the external theta role is assigned to (Haegeman 1994:215; Huang 2000:18).

c. Bach$_1$ adored him$_{*1/2}$.

In (8a), *himself*, the reflexive (i.e., an anaphor) is subject to BT A, by which it must be bound to and coreferential with *Bach*, its local antecedent. In (8b), *him*, the pronoun, is subject to BT B, by which it must be free in its GC and referentially disjoint with *Bach*. In (8c), both Bach$_1$ and Bach$_2$, r-expressions, are subject to BT C, by which Bach$_1$ cannot be coreferential with Bach$_2$ each other.

2.1.2 Discourse Anaphora in Pragmatics

However, Chomsky's binding theory is not cross-linguistically tenable. Let us reconsider the BT A first.

(9) a. John$_1$-i [{caki$_1$-ka aphuta}]-ko malhayssta
John-Nom self-Nom sick-Comp said
'John$_1$ said that self$_1$ is sick.'

b. John$_1$-un [nay-ka caki$_1$-lul cohaha]-nun kes-ul acikto molunta
John-Top I-Nom self-Acc like Comp-Acc yet not know
'John$_1$ still doesn't know that I like self$_1$.'

Kim (1994:11)

In (9), *caki*, the long-distance (LD) reflexive in Korean, have the antecedent *John* outside the local domain, unlike English, where a reflexive must have a strict local antecedent. That is, *caki* takes its antecedent outside the GC, so it is long-distance bound, which runs counter to BT A.

Second, let us reconsider the BT B.

(10) John$_1$-un [{caki$_1$ / ku$_1$}-ka salang-ey ppacyessta]-ko malhayssta

John-Top {self / he}-Nom love-in fell-Comp said
'John$_1$ said that {self$_1$ / he$_1$} was in love.'

Kim (1994:33)

In (10), *ku*, the non-reflexive pronoun in Korean, is bound by the subject NP *John* outside the local domain, where *caki*, the reflexive, is also syntactically allowed. There is no difference in reference. That is, the non-reflexive pronoun is c-commanded by the subject NP, so they must be disjoint with each other according to BT B, by which reflexives and pronominals must be in strict complementary distribution since BT A and B are mirror images of each other. This logophoricity in Korean runs counter to Chomksy's BT B. In addition, this complementarity between reflexives and pronominals (i.e., the complementarity between BT A and B) breaks down even in English as follows.

(11) a. John Kerry$_1$ saw a picture of himself$_1$ / him$_1$ in *The New York Times*.
b. Steve$_1$ looked behind himself$_1$ / him$_1$.
c. [Pavarotti and Domingo]$_1$ adore [each other's]$_1$ / their$_1$ performances.
d. Pavarotti$_1$ said that tenors like himself$_1$ / him$_1$ would not sing operas like that.

(Huang 2007:253)

This anti-complementarity between reflexive and pronominal include 'picture' NP (11a), adjunct PP (11b), possessive NP (11c), and emphatic NP (11d).

Third, let us reconsider the BT C.

(12) (Thai, Lasnik 1989)
a. *Khăw$_1$ chɔ̂p Cɔɔn$_1$
he likes John
'He$_1$ likes John$_1$.'

b. *Khăw$_1$ khít waâ Cɔɔn$_1$ chàlāāt

he thinks that John is smart

'He$_1$ thinks that John$_1$ is smart.'

(13) (Thai, Lasnik 1989)

a. Cɔɔn$_1$ chɔ̂p Cɔɔn$_1$

John likes John

'John$_1$ likes John$_1$.'

b. Cɔɔn$_1$ khít waâ Cɔɔn$_1$ chàlāāt

John thinks that John is smart

'John$_1$ thinks that John$_1$ is smart.'

In (12), the binding of an r-expression by a pronoun is ungrammatical according to BT C. However, in (13), the binding of an r-expression by another coreferential r-expression is tolerant in Thai, which runs counter to BT C. To solve this problem, Lasnik (1989) proposes two distinct sub-binding conditions, as shown in (14).

(14) Lasnik's (1989) BT C

a. C_1: An r-expression is r-expression-free everywhere.

b. C_2: An r-expression is pronoun-free everywhere.

In (14), Lasnik (1989) claims that the parametric BT C_1 explains the examples in (13), while the universal BT C_2 explains the examples in (12).

However, first, the repetition of r-expressions occurs even in English as well at the inter-sentential level, i.e., at the discourse level as follows (Fodor 1975, Evans 1980, Levinson 1987).

(15) a. (Fodor 1975)

Only Churchill$_1$ remembers Churchill$_1$ giving the speech about blood, sweat, toil and tears.

b. (Evans 1980)

I know what John$_1$ and Bill$_1$ have in common. John$_1$ thinks that Bill$_1$ is terrific and Bil1$_1$ thinks that Bill$_1$ is terrific.

c. (Levinson 1987)

Frog$_1$ went Frog$_1$'s house; Toad$_1$ went to Toad$_1$'s house.

In (15), the repetitions of r-expressions support that the BT C_1 cannot be tenable in English[5].

Second, Lasnik's universal BT C_2 can be falsified even in English at the discourse level.

(16) (Evans 1980)

Everyone has finally realized that Oscar$_1$ is incompetent.

Even he$_1$ has finally realized that Oscar$_1$ is incompetent.

In the second sentence of (16), the second r-expression *Oscar* is preceded and c-commanded by the pronoun *he* in the matrix clause of the same sentence, which runs counter to Lasnik's BT C_2.

In short, binding conditions have been discussed at the intra-sentential level in syntax. But even though the BT has been restricted at the intra-sentential level, we have seen that the BT is cross-linguistically problematic since the LD reflexive (9) and logophoricity (10) in Korean, the anti-complementarity between reflexive and

[5] In fact, Huang (1994, 2000) notes that BT C_1 in English is less easily violated than BT C_1 in Thai since English and Thai are essentially different types of language: English is a syntactic language, whereas Thai is a pragmatic language.

pronominal (11) in English, and the repetition of r-expression (13) in Thai run counter to BT A, B, and C, respectively. Moreover, Lasnik's binding conditions C_1 and C_2 are not tenable even at the inter-sentential level, i.e., at the discourse level, in terms of the repetition of r-expressions (15) in English and the precedence and c-command by the pronoun (16) in English, which run counter to C_1 and C_2, respectively.

What ties these together, as Levinson (2000:270) and Huang (2007:258) note, is that anaphoricity, i.e., reference, essentially lies on the systematic use or avoidance of specific lexical items, not on the property of specific lexical items. In other words, anaphora is not a property of specific linguistic expressions with binary features such as anaphor (reflexive and reciprocal), pronoun, and r-expression in syntactic terms, but a property of the use or avoidance of linguistic expressions, i.e., referring expressions, in pragmatic terms. At this point, therefore, it is also important to note that anaphora must be dealt with at the inter-sentential level, i.e., at the discourse level, as well as at the intra-sentential level.

In addition, as Clark and Marshall (1981:35-42) notes, (discourse) anaphora should be analyzed by three levels of heuristics of definite reference in terms of mutual knowledge, e.g., (i) community membership, (ii) physical copresence, and (iii) linguistic copresence. In the same vein, Ariel (1990-5-6) also notes that (discourse) anaphora should be analyzed by a three-way 'geographic' division of context, e.g., (i) encyclopedic knowledge, (ii) physical context, and (iii) linguistic context. With respect to the use of discourse anaphors, both three levels of heuristics and the three-way geographic division of context are correlated to those of the typical use of discourse anaphors. In particular,

(i) the use of names and descriptions refers to the community membership/the encyclopedic knowledge context; (ii) the use of deictics and demonstratives refers to the physical copresence/the physical context; (iii) the use of reflexives and pronouns refers to the linguistic copresence/the linguistic context.

2.1.3 Discourse: Implications for Discourse Anaphora

The first relevant discourse topic to be examined is that of discourse anaphora. I begin by presenting the previous definitions of discourse. A variety of definitions and concepts of discourse provide accounts for discourse anaphora.

According to Brown and Yule (1983:1-3), discourse is considered as 'language in use'. In the transactional view, it is traditionally defined as the transmission of 'content', such as 'factual or propositional information', so it is considered language beyond the clause and the sentence. On the other hand, in the interactional view, it is traditionally defined as the use of language to express 'social relations and personal attitudes'.

However, Brown and Yule (1983), whose concept is used through this paper, pay attention to a concept of discourse that is dependent on language use related to interlocutors' intention or function as well as on the description of linguistic form. This concept has been supplemented by several linguists as follows.

First, Grosz, Pollack, and Sidner (1989) discusses that discourse is formed by the interaction of speaker's intention and addressee's interpretation and understanding of it when they produce utterances. That is, discourse is a mutually constructive and cooperative enterprise.

Second, Schiffrin (1994:20-43) defines discourse as utterances, 'the contextualized sentences'. That is, it is not a collection of decontextualized units of language structure, but rather a collection of contextualized units of language use.

Third, Cornish (1999: 34) defines discourse as 'sequences of utterance and indexical acts' that the speaker and the addressee are participating in while communication unfolds. The sequences having own unified purpose are hierarchically structured and mentally represented.

Fourth, Walker (1998) notes that discourse is constructed both by linguistic and by extralinguistic factors: the former represents the verbal content of information with the elements of the utterance, the ways of structure through the grammatical signs, and the indication of discourse structure by 'cue phrases' such as discourse markers with intonation and gestures, whereas the latter represents features of the context in utterance, interlocutors' generally and particularly mutual knowledge about world and culture, and their personal attitudes and beliefs in their setting.

2.1.4 Definition of Discourse Anaphora

Discourse anaphora is also defined in multiple ways:

First, in pragmatic terms, Lyons (1975, 1977, 1979) and Bühler (1982) point out that anaphora is one type of 'pointing', represented by the discourse context itself, rather than by the speech or utterance context. Lyons (1977:673) explains that "anaphora presupposes that the referent should already have its place in the universe-of-discourse".

Second, as Levinson (1983:85-86) and Huang (2000:1; 2007:245) note, anaphora refers to "a relation between two linguistic elements, wherein the interpretation of one (an

anaphor) is in some way determined by the interpretation of the other (an antecedent)." It is represented by various linguistic pro-forms, such as (definite) descriptions with articles, names, pronouns, reflexives, demonstratives, and gaps (empty categories). It is also represented in the range of intra-sentence, inter-sentence, and inter-turns at speaking in a verbal discourse. As Cornish (1999:1) notes, anaphora is represented at an utterance level, not at a sentence level. In case of discourse anaphora, it is at least represented at the inter-sentence level, i.e., at the discourse level.

Third, Halliday and Hasan (1976:4) represent that anaphora occurs when a writer or a speaker uses a linguistic expression to refer back to a preceding antecedent such as someone or something that is previously identified in discourse. In the meantime, an anaphor act as one of a set of cohesive devices[6], *reference*, pointing back to a previously identified referent in discourse to avoid redundant repetition, forming 'a cohesive chain', such as < A, B, and C > (or 'a coreferential chain'), at which cohesion carries out "where the interpretation of some element is the discourse is dependent on that of another".

On the other hand, in cognitive terms of attention of memory, Ehlich (1982: 325-330) characterizes anaphora as the use of a member (or members) of linguistic devices maintaining the focus of communicators' attention that is already evoked at the point where the anaphor is referred to. That is, anaphora is a linguistic apparatus to focus the addressee's attention on an intended item in a given discourse and allow the addressee to maintain a focused item in subsequent utterances. As Cornish (1999:26) clarifies, under

[6] Halliday and Hasan (1976) suggest five general cohesive devices creating coherence in texts: *reference*, *ellipsis*, *substitution*, *lexical cohesion*, and *conjunction*.

this cognitive 'memory-management' concept, anaphora is defined as a linguistic system that retrieves the salient item or information in a given discourse, in which the linguistic item of information is already established between the speaker and the addressee.

In sum, discourse anaphora, one type of reference, denotes a relation between two linguistic referents in a given discourse, in which an anaphor refers back to the same entity, referred to by an antecedent, in the previous utterance (Huang 2007:245). Yet crucially for this study, it functions as focusing the addressee's attention on an intended referent at some point in memory and maintaining the focus on the item in the subsequent utterances in a given discourse.

2.2 The Topic Continuity Theory

Discourse anaphora has been analyzed by four main approaches in pragmatics, which I will lay out in this section: (i) the topic continuity theory (section 2.2), (ii) the hierarchy theory (section 2.3), (iii) the cognitive theory (section 2.4), and (iv) the neo-Gricean theory (section 2.5).

One of the theoretical models for explaining discourse anaphora is the topic continuity theory. The notion of *topic* is a significant apparatus to introduce a main referent and refer to it in a cohesive and coherent way. Gundel (1985:86) gives the definition of pragmatic topic, described in (17).

(17) Definition of pragmatic topic
An entity, E, is the pragmatic topic of a sentence, S, iff S is intended to increase the addressee's knowledge about, request information about or otherwise get the addressee to act with respect to E.

This definition subsumes Lambrecht's (1994) 'aboutness': the relation between a particular referent and the proposition. It features the topic as an entity for which a proposition in a sentence tells if the sentence implies relevant information about topic and increases the hearer's knowledge about it. In short, topic is what is being talked about in a given discourse.

The topic is hierarchically encoded by different types of anaphoric expressions and determined by topic continuity with a "topic-coding devices scale" in Figure 2.1, put forward by Givón (1983:17).

most continuous/accessible devices topic

zero anaphora (Ø)
unstressed/bound pronouns or grammatical agreement
stressed/independent pronouns
R-dislocated DEF-NPs
neutral-ordered DEF-NPs
L-dislocated DEF-NPs
Y[7]-moved NP's ('contrastive topicalization')
cleft/focus constructions
referential indefinite NPs

most discontinuous/inaccessible topic

Figure 2.1 Topic-coding devices scale

[7] "Y" stands for "Yiddish." The name of "Y-movement" is ascribed to Paul Postal, an American linguist, who considered this term to be a property of Yiddish speakers using contrastive topicalizations.

The topic continuity in Figure 2.1 is fundamentally measured by three parametric factors: "referential distance ('look-back'), potential interference ('ambiguity'), and "persistence ('decay')." The first parameter indicates linear distance, i.e., the number of clause and sentence between an antecedent and its anaphor; the second parameter represents referential interference, i.e., the number of potential interfering referents between two mentions of a referent; the third parameter presents thematic information, i.e., the maintenance or change of protagonists.

Following these parameters, Givón (1983) suggests that the shorter the linear distance between two mentions of a referent, the more continuous a topic; the fewer the competing referents between an antecedent and its anaphor, the more continuous a topic; the more stable thematic status of the protagonists, the more likely that it will be encoded by a minimized anaphoric expression. For this reason, this topic continuity theory is also called the "distance-interference model." In particular, a topic in a given discourse is preferably referred to by a zero anaphor (Ø), whereas a non-topic is preferably referred to by the overt pronoun and noun phrase (NP). In short, a zero anaphor has a preference over other overt referring expressions for a topic in a given discourse.

However, as Kim (2010a, 2012) argues, a topical referent in Korean discourse is actually and repeatedly referred to by a full NP such as a full name rather than by zero anaphors and pronouns even if a zero anaphor is presumably preferred. In this respect, the topic-coding devices scale should be modified to apply for Korean discourse.

2.3 The Hierarchy Theory

Another key role in analysis of discourse anaphora will be played by the hierarchy theory of anaphor. In the hierarchy theory, the most significant factor determining anaphoric selection and distribution is the hierarchical structure of a given discourse. This theory particularizes that initial or non-initial mention for an intended referent at the beginning of a new discourse structure tends to be referred to by a full noun phrase (NP), while the subsequent mentions within the same discourse unit tends to be referred to by a reduced form such as a pronoun or an anaphor.

This hierarchy theory has been discussed by a variety of scholars, such as Fox (1987), Grimes (1975), Hinds (1978, 1979), Longacre (1979), Mann and Thompson (1987, 1988), and Tai (1978). Among these, Fox (1987) is considered the most typical work of hierarchical theory. Analyzing expository written English in terms of rhetorical structure (R-structure)[8], and English conversation in terms of sequential closure or non-closure[9], she argues that there is a strong relationship between hierarchical discourse structure and the use of anaphora,

Fox's (1987:95) main idea about the use of anaphor in (written) discourse is represented in (18).

(18) a. A pronoun is used to refer to a person if there is a previous mention of that person in a proposition that is active or controlling;

[8] Fox (1987:78-79) argues that written texts are organized by hierarchical groups of propositions (i.e., the smallest unit of texts) that have internal R-structures, which contain a "core part (i.e., nucleus) and an ancillary part (i.e., adjunct)": the nucleus identifies the main goals of the writer, while the adjunct represents supplementing information for the nucleus.

[9] The instances of "non-closure" of a sequence in conversation are an adjacency pair and a turn expansion (Fox 1987: 20-21).

b. Otherwise (i.e., all other mentions) a full NP is used.

In (18), *active*[10] "refers to a proposition in an R-structure (either the adjunct or the nucleus) whose R-structure partner (the nucleus or the adjunct) is being produced. That is, the adjunct of an R-structure is active while its nucleus is being produced; similarly, the nucleus of an R-structure is active while its adjunct is being produced. A proposition is *controlling* while its R-structure partner is active" (Fox 1987:95-6). In particular, a pronoun is used to indicate an intended referent in an active or controlling proposition, while a full NP is used to indicate an intended referent other than in active or controlling proposition in a given (written) discourse.

In short, the hierarchy theory argues that the use of anaphoric selection and distribution is determined by the hierarchical structure of a given discourse; i.e., the rhetorical structure (in a written discourse) and the sequential closure or non-closure (in a verbal discourse). In particular, a pronoun indicating a non-closure is used to refer to an intended referent that is mentioned in a rhetorically active or controlling proposition, while a full NP indicating a sequential closure is used to demarcate a new rhetorical unit.

However, as I will show, based on Kim (2010a, 2012) in Korean discourse, a full NP as well as a pronoun is mentioned regardless of the closure of a sequence in discourse. Contrary to Fox's concept, a full NP can be repeatedly used where the mention of the intended referent in a sequence is not closed in Korean discourse.

[10] These terms of *active* and *controlling* are originally taken from Reichman (1981).

2.4 The Cognitive Theory

A third aspect of anaphora resolution is based on the cognitive theory. The basic principle, put forward by Prince (1981), Ariel (1988, 1990, 1994, 2008, 2010), Gundel (1996), and Gundel, Hedberg, and Zacharski (1988, 1989, 1990, 1993), suggests that there is correlation between cognitive status and anaphoric form. That is, the anaphoric distribution in a given discourse is mainly affected by cognitive processes such as activation and attention given to a referent within interlocutors. Three of these theories are discussed in the following sections.

2.4.1 Ariel's Accessibility

Ariel (1988, 1991, 1994, 2008, 2010) suggests that different degrees of accessibility in memory enable the addressee(s) to select an appropriate referring expression from among various possibilities and to decode the intended referent. Thus, an anaphoric procedure in terms of the distribution and the interpretation of anaphoric expressions is a joint achievement between the speaker and the addressee(s), so their roles are important to select and interpret a particular anaphor within a given discourse.

Ariel's accessibility presents different degrees of possibility for the addressee(s) to reach the intended referring expression in accordance with his/her (or their) memory, as shown in the accessibility marking scale in (19), in which the degree of accessibility indicates the scale from the highest to the lowest among different types of referring expressions. The linguistic item in the left in the scale cues relatively higher accessibility, while the one in the right cues relatively lower accessibility. Accessibility implies that a

speaker uses the highest item in the scale among other candidates as a linguistic cue when s/he thinks that the referring expression should be highly accessible to the addressee(s).

(19) The accessibility marking scale:

(← *most continuous/ higher accessible*) zero (Ø) < reflexive < agreement markers < cliticized pronouns < unstressed pronouns < stressed pronouns < stressed pronouns + gesture < proximal demonstrative (+NP) < distal demonstrative (+NP) < proximal demonstrative (+NP) + modifier < distal demonstrative (+NP) + modifier < first name < last name < short definite description < long definite description < full name < full name + modifier (*most discontinuous/lower accessible* →) (Ariel 1994: 30, adapted)

In (19), a zero form (Ø) is the highest in the accessibility marking scale, whereas the form of a full name + modifier is the lowest in terms of attenuation.

In sum, accessibility suggests that an antecedent is coded by a higher accessibility marker when it is a salient, recent, and highly coherent unit, while an antecedent is coded by a lower accessibility marker when it is a non-salient, distant, and less coherent unit. High accessibility is usually translated to coreferentiality, whereas low accessibility is translated to disjointness.

However, as Cornish (1999:8) notes, it is noteworthy that each segmental expression type in Ariel's accessibility marking scale (20) is 'separately' and 'individually' marked for the level of accessibility in memory of the referent that it codes. In other words, each referring expression in Ariel's accessibility hierarchy is not 'inferable' one from another; i.e., each expression type in Ariel's scale is mutually exclusive from one another. To address this 'inferable' or 'implicational' relation in

character between each expression type, Gundel, Hedberg, and Zacharski's Givenness Hierarchy comes into play as a cognitive model.

2.4.2 Gundel, Hedberg, and Zacharski's Givenness Hierarchy

2.4.2.1 The Givenness Hierarchy (GH)

Under this cognitive approach (Gundel, Hedberg, and Zacharski (1988, 1989, 1990, 1993, 2001) (GHZ) and Gundel (1996, 2003)), the givenness hierarchy in (20) represents six statuses of referring expressions in terms of the different memory and attention states that are coded as lexical forms.

(20) The givenness hierarchy (GH)

In focus	> Activated	> Familiar	> Uniquely identifiable	> Referential	> Type identifiable
{it}	{that} {this} {this N}	{that N}	{the N}	{indefinite this N}	{a N}

In this GH, the different linguistic forms represent the different cognitive statuses that assist the addressee in restricting the set of possible referents. Each status is characterized as follows.

First, for an "in-focus" status, an intended referent is at the current center of attention within current short-term memory. The unmarked form is a referring expression *it*. The entity not only represents the topic of the preceding utterance within a discourse, but also illustrates the state of a highly relevant topic.

Second, an "activated" status is the representation either in short-term memory, retrieved from long-term memory, or arisen from the immediate linguistic or

extralinguistic context. Linguistic forms such as *that*, *this*, and *this + N* are felicitous in this status.

Third, a "familiar" status represents that the addressee is able to uniquely identify the intended referent since s/he already has a representation of it in memory. The intended referent is in short-term memory if it has been recently mentioned or perceived, whereas it is in long-term memory if it has not. The appropriate linguistic form is this status is *that N*.

Fourth, a "uniquely identifiable" status shows that the addressee can identify the speaker's intended referent on the basis of the nominal alone. The sufficient linguistic form is the definite article *the* in this status.

Fifth, a "referential" status indicates that the addresser intends to refer to a particular object. The addressee not only can access an appropriate type-representation, but s/he must also retrieve an existing representation of the speaker's intended referent or construct a new representation by the time the sentence has been processed in order to understand the expression of referential status. This status is necessary for the use of all definite expressions.

Sixth, "type identifiable" indicates a cognitive status in which the addressee is able to access a representation of the type of object described by the expression. The sufficient linguistic form is an indefinite article *a* in any nominal expression.

Going back to the first two columns, regarding anaphors *it* and *that* in English, the relevant status is exemplified in (21) (Gundel et al. 1993:279-280).

(21) My neighbor has *a dog*.

a. In-focus: *It* kept me awake.

b. Activated: *That* (/*This*/*This dog*) kept me awake.

In (21a), for an in-focus status, an intended referent is at the current center of attention within current short-term memory. The discourse anaphor *it* refers to a *dog* as being currently focused on. In (21b), retrieved from long-term memory or arisen from the immediate linguistic or extralinguistic context, the discourse anaphor *that* is used appropriately to refer to "the barking of *a dog* only if a dog has actually been barking during that speech event or if the barking has been introduced in the immediate linguistic context" (Gundel et al. 1993:278). The same cognitive concept of in-focus and activated status in English plays out with in Korean, regarding sentential anaphors[11].

2.4.2.2 The GH and Grice's Maxims of Quantity

Within the GH in (20), each cognitive status with its lexical expression type entails that of the item to its right in terms of the implicational relation. Gundel et al. (1993) suggests that the values assigned to given indexical expression types in context are estimated by means of the interactions between the cognitive statuses that are conventionally coded and Grice's two maxims of quantity[12].

First, quantity-1 implicates that the use of an "entailed weaker" expression type conversationally does not obtain an "entailing stronger" expression type. For example, an anaphor *that* is not to be used in a discourse to retrieve referents bearing the in-focus

[11] See fuller application in Section 6.2.2.
[12] See fuller description of Grice's maxims in Section 2.5.1.

status because the use of anaphor *that* requires the activated status in English, like scalar implicature. Thus, using an anaphor *that* for in-focus status, where an anaphor *it* should be sufficiently used, does not observe Grice's quantity-1.

Second, quantity-2 implicates that an "entailed weaker" expression type conversationally can hold the "entailing stronger" status. For instance, the use of anaphor *that* can implicate not only a conventionally coded status, activated, but also the more restrictive cognitive status, in-focus, in the GH. This indicates that an anaphor *it* is to be possible if signaling identifiability for the intended referent is often sufficient, so using an anaphor *that* for the cognitive status in-focus, where an anaphor *it* should be necessarily used, would be possible by observing quantity-2.

Accordingly, based on these implicational characteristics with Grice's quantity-1 and quantity-2, the referring expressions in six cognitive statuses of the GH are fundamentally and pragmatically inferable one from another. That is, each linguistic item in the GH is not mutually exclusive, i.e., inclusive and implicationally related, so each cognitive status with a specific lexical item entails all lower statuses. For example, the expression type {it} not only codes the cognitive status "in-focus", but also implies the next lower statuses such as "activated", "familiar", "uniquely identifiable", "referential", and "type identifiable".

Gundel et al.'s (1993) GH is developed from Prince's (1981:237) familiarity scale in (22), in which Prince does not match each cognitive status with the corresponding form.

(22) The familiarity scale

$$\left\{\begin{array}{c}\text{Evoked}\\ \\ \text{Situationally Evoked}\end{array}\right\} > \text{Unused} > \text{Inferrable} > \begin{array}{c}\text{Containing}\\ \text{Inferrable}\end{array} > \begin{array}{c}\text{Brand}\\ \text{New}\\ \text{Anchored}\end{array} > \begin{array}{c}\text{Brand}\\ \text{New}\end{array}$$

However, it is noteworthy that, unlike the GH, this familiarity scale does not distinguish between "in-focus" and "activated." Instead, the status "evoked" covers both, so we cannot distinguish the distinctive referential characteristics between discourse anaphors *it* and *that*. In other words, since Prince's (1981) familiarity scale is not implicationally related between the statuses, the referring expressions within each status are 'mutually exclusive'. In this case, each status is separately and distinctively marked, like Ariel's accessibility (Gundel et al. 1993:280, Gundel 1996:151; Cornish 1999:8).

2.5 The Neo-Gricean Theory

Finally, the last key aspect to be discussed is the role of neo-Gricean theories[13] of conversational implicature. The basic tenet of the theory, put forward by Grice (1957, 1961, 1969, 1975, 1978, 1989), Horn (1984, 1989), Levinson (1987a, 1987b, 1991, 1995, 2000), Huang (1991, 1994, 2000a, 2000b, 2007) is that anaphoric distribution and selection in a given discourse can be predicted and understood by the inferential interaction between principled pragmatic heuristics.

[13] Grice's original notion has been discussed by a variety of subsequent studies. Among those, there is a unique tradition that has tried to reduce Grice's four conversational maxims into the refined system of pragmatic principles, i.e., neo-Gricean theory, notably including Laurence R. Horn and Stephen C. Levinson.

The basic distributional pattern of discourse anaphora is represented in (23) below (Huang 200b:319).

(23) The basic distributional pattern of anaphora in discourse

a. Establishment of reference tends to be achieved through the use of an elaborated form, notably, a lexical NP.

b. Shift of reference tends to be achieved through the use of an elaborated form, notably, a lexical NP.

c. Maintenance of reference tends to be achieved through the use of an attenuated form, notably, a pronoun or a zero anaphor.

2.5.1 Grice's Cooperative Principle (CP) and Conversational Maxims

The basic idea in neo-Gricean theory originates from Herbert Paul Grice (H. P. Grice). On an account of meaning and communication, Grice proposes two theories: (i) a theory of meaning$_{n[on]n[atural]}$ (Grice 1957, 1969, 1989) and (ii) a theory of conversational implicature (Grice 1961, 1975, 1978, 1989).

In the first theory of meaning$_{nn}$, Grice underlines the relation between natural meaning (i.e., meaning$_{n}$) in the external world, and non-natural and linguistic meaning (i.e., meaning$_{nn}$) of utterances. This meaning$_{nn}$[14] is explained in terms of the speaker's intentions in communication.

In the second theory of conversational implicature, Grice proposes an underlying and fundamental principle in human communication that determines the way of language

[14] <Grice's theory of meaning$_{nn}$>
The speaker (*S*) means$_{nn}$ proposition (*p*) by 'uttering' utterance (*U*) to audience (*A*) if and only if *S* intends:
(i) *A* to think *p*,
(ii) *A* to recognize that *S* intends (i), and
(iii) *A*'s recognition of *S*'s intending (i) to be the primary reason for *A* thinking *p*.

being used in a maximally efficient and effective way to reach a rational interaction. In other words, the human communication has a 'conversational logic' that explains the pragmatic inference in conversation in terms of context, and that it is different from the formal inference based on the logic or the truth conditional semantics. Grice calls this overarching umbrella the *cooperative principle* in human communication, as represented in (24), which presupposes the conversation participants are basically rational people.

(24) The cooperative principle (CP)
Make your conversational contribution such as is required, at the stage at which it occurs by the accepted purpose or direction of the talk exchange in which you are engaged.

The CP explains that conversation participants have to or are supposed to converse with one another appropriately for the current purpose or the direction of the conversation if they are rational people.

In addition, Grice subdivides this CP into the four (nine in detail) maxims of conversation, as represented in (25).

(25) Four conversational maxims[15]
a. *Quality*
Try to make your contribution one that is true.
1. Do not say what you believe to be false.
2. Do not say that for which you lack adequate evidence.

[15] The names of these four maxims are taken from the German philosopher Immanuel Kant (Grice 1989:26).

b. *Quantity*

1. Make your contribution as informative as is required
   (for the current purposes of the exchange).
2. Do not make your contribution more informative that is required.

c. *Relation*

Be relevant.

d. *Manner*

Be perspicuous.

1. Avoid obscurity of expression.
2. Avoid ambiguity.
3. Be brief (avoid unnecessary prolixity).
4. Be orderly.

If the CP and four maxims are observed by interlocutors in conversation, it would be an idealized situation (i.e., conversational implicature$_O$[16]). However, in fact, as Grice (1975) notes, every conversation does not always keep the CP and maxims because it is not considered as an inviolable principle about morality or ethic. Instead, if the maxim is not observed (i.e., it is 'flouted'[17]) by the speaker in conversation, then the conversational implicature can be generated while 'a maxim is being exploited' (i.e., conversational implicature$_F$).

2.5.2 Horn's Bipartite Principle: The Q- and R-principles

Horn (1984, 1989) proposes two fundamental and antithetical principles: Q[uantity]-principle and R[elation]-principle, by reinterpreting and rearranging Grice's

[16] Conversational implicature$_O$ and conversational implicature$_F$ are Yan Huang's terms (Huang 2007:27-29).

[17] 'Flouting' the maxim means that one my 'blatantly' (ostentatiously, intentionally) fails to fulfill the maxim, if one may 'flout' a maxim (Grice 1975:49).

four maxims to reduce redundancy between them. In the meantime, however, he presupposes that the CP and quality maxim are 'primary and essentially unreducible', so they are still being observed.

In fact, the concept for Q- and R-principle is based on Zipf's (1949:21) *auditor's* and *speaker's economy* under the *principle of least effort*: the former indicates "possessing a vocabulary of m different words with one distinct meaning for each word"; the latter means "possessing a vocabulary of one word which will refer to all the m distinct meanings." They are grounded in two opposing forces: the first force with the auditor's economy comes from the *force of diversification* that "increases the diversity of a vocabulary"; the second with the speaker's economy comes from the *force of unification* that "unifies all meanings behind a single word".

Based on this concept, let us consider the Horn's bipartite Q- and R-principle, as represented in Table 2.1.

Table 2.1 Horn's bipartite Q- and R-principle

The Q Principle	The R Principle
Hearer-oriented	Speaker-oriented
Make your contribution sufficient:	Make your contribution necessary:
Say as much as you can (given R)	Say no more than you must (given Q)
Lower-bounding principle, inducing upper-bounding implicata	Upper-bounding principle, inducing lower-bounding implicata
Collects Grice's Quantity 1 maxim and Manner 1 and 2.	Collects Grice's Relation maxim, Quantity 2, and Manner 3 and 4.
Force of diversification (Principle of maximizing contextual effects)	Force of unification (Principle of minimizing processing effort)

As seen in Table 2.1, Q-principle, hearer-oriented and the auditor's economy in Zipfian terms, indicates that the speaker need to 'make your contribution sufficient' in discourse, such that s/he 'say as much as you can'. It collects together Grice's (1975) quantity 1, and manner 1 and 2. On the other hand, R-principle, speaker-oriented and the speaker's economy in Zipfian terms, indicates that the speaker need to 'make your contribution necessary' in discourse, so s/he 'say no more than you must'. It collects together Grice's relation, quantity 2, and manner 3 and 4.

Horn (1989:194) argues that the tension between two pragmatic principles applies to a variety of linguistic phenomena. Moreover, under the name of the 'division of pragmatic labor', Horn (1983:22; 1989:147) implicates that "given two coextensive expressions, the briefer and/or more lexicalized form will tend to become associated through R-based implicature with some unmarked, stereotypical meaning, use, or situation, and the marked, more complex or prolix, less lexicalized expression tends to Q-implicate a marked message, one which the unmarked form could not or would not have conveyed", with respect to the anaphoric distribution.

2.5.3 Levinsion's Tripartite Principle: The Q-, I-, and M-principles

Levinson (1987a, 1987b, 1991, 1995, 2000) proposes three inferential strategies: Q[uantity]-, I[nformativeness]-, and M[anner]-principle, as represented in Table 2.2, in which each of three principles has two poles: (i) a *speaker's maxim* that explains what each principle directs the speaker to say and (ii) a *recipient's corollary* that explains that each principle allows the addressee to infer. These tripartite principles are in fact based

on and reorganized by Grice's CP and four maxims, Zipfian principles, and Horn's Q- and R-principle in terms of conversational implicature.

Table 2.2 Levinson's tripartite Q-, I-, and M-principles

	Speaker's maxim	*Recipient's corollary*
a. The Q-principle (Levinson 2000:76)	Do not provide a statement that is informationally weaker than your knowledge of the world allows, unless providing a stronger statement would contravene the I-principle. Specifically, select the informationally strongest paradigmatic alternate that is consistent with the facts.	Take it that the speaker made the strongest statement consistent with what he knows and therefore that: (i) if the speaker asserted A(*W*), where A is a sentence frame and W and informationally weaker expression than *S*, and the contrastive expression <*S*, *W*> form a Horn-scale (in the prototype case, such that A(*S*) entails A(*W*)), then one can infer that the speaker knows that the stronger statement A(*S*) (with *S* substituted for *W*) would be false (or K~(A(*S*))[18]); (ii) if the speaker asserted A(*W*) and A(*W*) fails to entail an embedded sentence Q, which a stronger statement A(*S*) would entail, and <*S*, *W*> form a contrast set, then one can infer the speaker does not know whether Q obtains or not (i.e., ~K(Q) or equally {P(Q), P~(Q)}[19]).
b. The I-principle (Levinson 2000:114-5)	The maxim of Minimization. "Say as little as necessary"; that is, produce the minimal linguistic information sufficient to achieve your communicational ends (bearing the Q-principle in mind).	The Enrichment Rule. Amplify the informational content of the speaker's utterance, by finding the most *specific* interpretation, up to what you judge to be the speaker's m-intended point, unless the speaker has broken the maxim of Minimization by using a marked or prolix expression.
c. The M-principle (Levinson 2000:136)	Indicate an abnormal, nonstereotypical situation by using marked expressions that contrast with those you would use to describe the corresponding normal, stereotypical situation.	What is said in an abnormal way indicates an abnormal situation, or marked messages indicate marked situations.

[18] This is Gazdar's (1979) notation, in which K~(A(S)) reads as 'the speaker knows that it is not the case that (A(S)).

[19] This is also Gazdar's (1979) notation, in which ~K(Q) or {P(Q), P~(Q)} read as 'it is epistemically possible that Q and epistemically possible that not-Q'.

First, the Q-principle, similar to Horn's Q-principle (i.e., Horn's scalar implicature), implies that the speaker has to make the strongest statement consistent with what s/he knows in discourse. Then the speaker is not supposed to provide a statement that is less informative than your knowledge of the world allows. In hearer's respect, this idea infers that in Horn's scale, as illustrated in (26).

(26) Horn's $Q_{\text{-scalar}}$: <x, y>[20]

a. $x \vdash y; S(x) \vdash S(y)$

b. $y +>Q_{\text{-scalar}} \sim x; S(y) +>Q_{\text{-scalar}} \sim S(x)$

As seen in (26), if the speaker states the strong expression (i.e., x or S(x)), it entails the weak expression (i.e., y or S(y)). On the other hand, if the speaker states the weak expression (i.e., y or S(y)), then it Q-implicates the negation of the statement with the strong expression (i.e., x or S(x)).

Second, the I-principle, mirroring Levinson's Q-principle and similar to Horn's R-principle, implies that the speaker can state the minimal linguistic information necessary to achieve your communication goal, whereas the hearer infer that the use of a linguistic form I-implicates a specific and extended interpretation through the enrichment rule, by which "minimal specification gets maximally informative or stereotypical interpretations", as illustrated in (27).

[20] This < > represents Levinson's (Horn's) Q-scale.

(27) I-scale: [x, y][21]
y +>I x

Third, the M-principle, complementary to I-principle, enables the speaker to choose the unmarked minimal form. Otherwise, one M-implicates the marked situation with disjoint reading rather than coreferential reading. In particular, one infer that in a set of expressions {U, M}[22], the use of a marked form (M) M-implicates the negation of the interpretation associated with the use of an alternative unmarked form (*U*). In fact, Levinson (2000: 135-6) criticizes that "Horn conflates informational minimality and expression brevity within R-principle, and informational maximality with expression markedness within Q-principle". In other words, Horn's scale in terms of semantic entailment fails to draw a distinction between what Levinson calls 'semantic/informational minimization' (i.e., semantically general expressions are preferred to semantically specific ones) and 'formal minimization' (i.e., shorter expressions are preferred to longer ones) because, as Levinson (2000:136) notes, "the use of a minimal expression invokes a maximal interpretation". Then it is argued that Horn's scale does not capture the implicatures induced by the marked or unusual expression, whereby there can be an alternative unmarked expression bearing the same semantic strength; therefore, Levinson postulates additional third heuristic of inference, the M-principle, as illustrated in (28).

[21] This [ ] represents Levinson's I-scale.
[22] The curly brackets { } indicates unordered set in the dissertation.

(28) M-scale: {x, y}[23]

y +>M ~x

Regarding the systematic interaction between these inferential principles, Levinson (2000:157) proposes resolution schema for the interaction between them, as represented in (29).

(29) Levinson's resolution schema

a. Interaction of the Q-, I-, and M-implicature

(i) Genuine Q-implicature from tight contrast sets of equally brief, equally lexicalized linguistic expressions "about" the same semantic relations, take precedence over I-implicatures;

(ii) In all other cases, the I-principle induces stereotypical interpretations, *unless*:

(iii) A marked expression has been used where an unmarked one could have been employed instead, in which case the M-implicature defeats the relevant I-implicature that would have arisen from the unmarked expression.

b. Level of genus: Q > M > I [24]

In (29), the schema orders the priority between pragmatic inferences when there are inconsistent potential implicatures. First, the genuine Q-implicature takes priority over

[23] This { } represents Levinson's M-scale.

[24] The sign of inequality '>' reads as 'defeat inconsistent'.

the I-implicautre. Otherwise, if there is the use of a marked linguistic form, the complementary M-implicature takes priority over I-implicature.

In particular, Levinson derives the scale of minimality in (30) to interpret the relation between anaphoric expressions, such as lexical NP, pronoun, and zero anaphor in terms of coreferentiality, informativeness, and markedness.

(30) The scale of minimality[25]

Zero anaphor	>	Pronoun	>	Lexical NP
← coreferential	……			disjoint inferential →
← informative	……			less informative →
← unmarked	……			marked →

In (30), the scale of minimality indicates that a zero anaphor is more coreferential than a pronoun is, and a pronoun is more coreferential than a lexical NP is, which indicates that the simple form is more informative than the complicated one in the scale. It also indicates the concept of "less is more", that is, "the more minimal the form, the stronger the preference for a coreferential reading."

Moreover, Levinson (2000:285-286) suggests that there is the 'general anaphora pattern'[26] at the discourse level in terms of the inferential interaction between I- and M-principles, as illustrated in (31).

[25] This scale of minimality is also called 'the semantic content hierarchy', in which "the inherent semantic content of a lexical NP tends to be semantically more specific than that of a pronoun, and the inherent semantic content of a pronoun, than that of a zero anaphor" (Huang 1991:310).
[26] This is Levinson's terminology.

(31) The general anaphora pattern

Lexical NP > pronoun > NP-gap

← M-implicates noncoreference

I-implicates coreference →

In (31), the scale indicates that "reduced, semantically general anaphoric expressions (i.e., NP-gaps, zero anaphors) tend to favor locally coreferential interpretations (by the I-principle); full, semantically specific anaphoric expressions (i.e., lexical NPs) tend to favor locally noncoreferential interpretations (by the M-principle)".

2.6 Discourse Anaphoric Patterns: Full Name vs. Single Name

Regarding discourse anaphoric patterning, Mulkern (1996:238-240) argues the relation between anaphoric forms, a full name and a single name, and cognitive statuses as discussed below.

For full names, when a full name refers to an entity, it is not expected by the addressee that s/he has a representation at the beginning part of a discourse. Thus, it seems sufficient that the status related to a proper noun be no higher than uniquely identifiable, similar to definite descriptions. However, it seems necessary that the status related with a proper noun be no lower than uniquely identifiable.

For single names, they are more ambiguous expressions than the full names because they have a broader range of possible referents, so they lead the addressee to have a more restrictive cognitive status, familiar. In other words, a full name only indicates that the referent is uniquely identifiable, whereas a single name indicates that

the writer or the speaker expects the addressee to have a representation of an individual who is already familiar to both the speaker and the addressee in their memory. Therefore, a single name requires that the referent is at least familiar status, similar to a demonstrative determiner *that* in discourse.

In particular, Mulkern concludes that full names are generally used for referents that are uniquely identifiable and familiar, whereas single names are mostly used for referents that are familiar and activated it an appropriate context. It is noteworthy that Mulkern argues that there are no usages of full names that are used as in- focus status, as I show it in Figure 2.2.

In focus	> Activated	> Familiar	> Uniquely Identifiable	> Referential	> Type Identifiable
X[27]	possible	←	Full Names		
X	←	Single Names			

Figure 2.2 The cognitive statuses of full and single names in English in the GH

However, as Kim (2010a, 2012) argues, this GH model does not apply to sequential anaphoric patterning in Korean since a topical referent is repeatedly referred to by a full NP, i.e., a full name through a discourse rather than by its single names and pronouns in the subsequent utterances. Then the full name functions as single names and pronouns indicating their corresponding cognitive statuses.

[27] X slots can be replaced and referred to by the (unstressed) pronoun or zero anaphor (Ø) in some languages like Korean, Japanese, and Chinese (Gundel et al. 1993:283-284).

2.7 Referential Properties of Discourse Anaphors: *that* and *it*

English sentential anaphors *that* and *it* are differentiated by how we understand the context surrounding utterance in discourse regarding the distribution. There are distinctive pragmatic properties between the sentential anaphor *that* and *it* in two respects: prior knowledge and wide reference as follows.

First, Kamio and Thomas (1999:291-295) [28] suggest that a sentential anaphor *it* refers to the referent about which the speaker already has the prior knowledge in one's mind, while *that* refers to the antecedent about which the speaker does not, as shown in (32).

(32) New vs. old Information

A: Guess what! *I just won the lottery*!

B_1: (Yes,) *it*'s amazing! I heard about it on the radio, and I've invited everyone on the block to our house for a party!

(Kamio and Thomas 1999, example (6))

B_2: (Yes,) *that*'s amazing! ?? I heard about it on the radio…..

In ($32B_1$), the use of sentential anaphor *it* indicates that when the speaker B heard the A's utterance, s/he already knew the news, so the following additional statements make B_1's utterance appropriate and coherent in this conversation. On the other hand, as in ($32B_2$), the use of sentential anaphor *that* is felicitous when the speaker B_2 surprised at hearing an

[28] Kamio and Thomas (1999:304) note that the properties of sentential ('propositional', Kamio and Thomas' (1999) term) anaphors *it* and *that* are consistent with those of nominal anaphors *it* and *that*, so discourse anaphors *it* and *that* subsume both nominal and sentential ('propositional') anaphors in this section.

unexpected news, and *that* represents that the speaker B_2 has no idea before hearing A's utterance, so the subsequent utterances become awkward in this conversation.

Second, Kamio and Thomas (1999:295-298) propose that the referential properties of sentential anaphor *that* and *it* are explained by the notion of wide reference, i.e., the breadth of reference. In particular, the sentential anaphors *that* and *it* are differentiated in that *it* refers widely to its referent, while *that* refers narrowly to its referent in (33).

(33) a. *Sonja was born out of wedlock,* but I never revealed *it* to her.
b. *Sonja was born out of wedlock,* but I never revealed *that* to her.

In (33), the semantic property of the verb *reveal* requires that a speaker has some prior knowledge about the direct object, so there seems to be no difference between (33a) and (33b) regarding new vs. old information. However, there is a subtle difference between them in that *it* refers widely to a set of related events and facts in this context in addition to the utterance itself, while *that* refers narrowly to the utterance itself. In particular, as Kamio and Thomas (1999:296) notes, *it* implies that "Sonja was born illegitimately, and the whole story of her mother's disastrous affair with the Prime Minister, the dangerous international intrigue which resulted from it,..," while *that* implies that "the speaker never told Sonja that her parents were unmarried at the time of her birth."

Those referential properties are summarized, as shown in Table 2.3 and Figure 2.3.

Table 2.3 The referential properties of sentential anaphor *it* and *that*

It	*That*
Represent speaker's prior knowledge ("already-learned information")	Need not represent prior knowledge
Refers widely	Points narrowly
Linde (1979) : Information in focus	Linde (1979) : Information out of focus

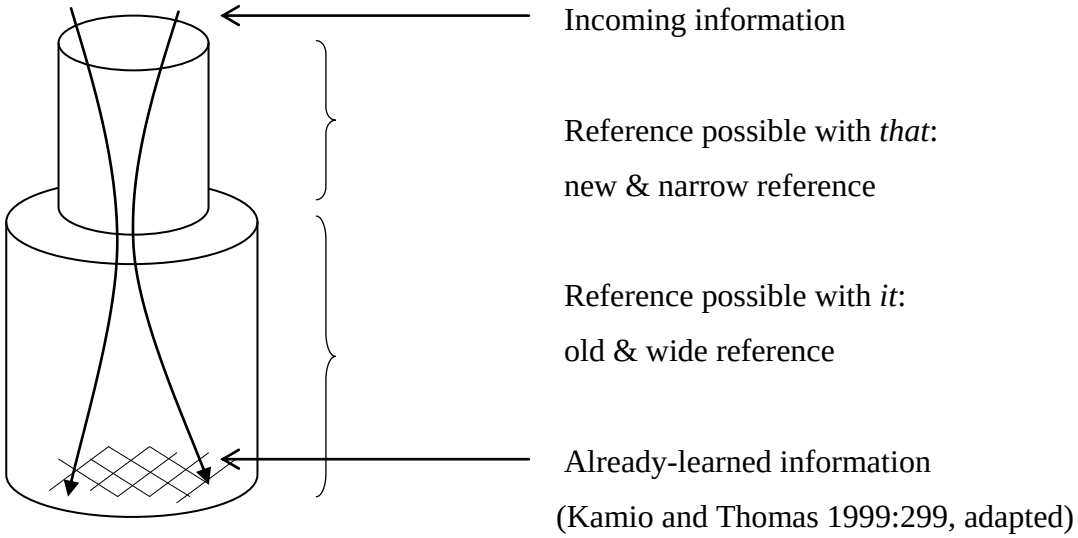


Figure 2.3 The referential properties of sentential anaphor *it* and *that*

As seen in Figure 2.3, accordingly, a sentential anaphor *it* refers widely to its referent combining with background information (i.e., already-learned information), whereas a sentential anaphor *that* refers narrowly to its referent itself (i.e., incoming information). This fact is integrated with that *it* refers to the speaker's prior knowledge, whereas *that*

does not. In sum, a sentential anaphor *it* evokes the context with its referent, whereas a sentential anaphor *that* points to its referent (Kamio and Thomas 1999:298-299).

However, as Kim (2010b) argues, these separate notions for discourse anaphors *it* and *that* in English are not correspondingly represented in Korean discourse in terms of prior knowledge and width of reference. In fact, there is a tendency for *ku* 'it' with its two variants, i.e., *ku-ke(s/n)* 'that' and *ku-len(s/n)* 'it', to have a dual function indicating two referential properties rather than corresponding items *ku* 'it' and *ce* 'that' being used to refer to.

2.8 Chapter Summary

Chapter 2 reviewed the relevant theoretical backgrounds of discourse anaphora in this dissertation. Section 2.1 began with a discussion of anaphora in syntactic approach in terms of Chomsky's binding theory; however, the counterexamples to the binding condition were represented in terms of cross-linguistic and pragmatic points, and the implication of discourse and the definition of discourse anaphora were explained in pragmatic terms. The four main theoretical models for explaining discourse anaphora in pragmatics were explained in the name of (i) the topic continuity theory (Section 2.2), (ii) the hierarchy theory (Section 2.3), (iii) the cognitive theory (Section 2.4), and (iv) the neo-Gricean theory (Section 2.5). Based on these theoretical backgrounds, two main issues in this dissertation were dealt with in terms of referential properties: (i) discourse anaphoric pattern (Section 2.6) and (ii) sentential anaphora (Section 2.7).

Chapter 3

Methodology

3.1 Overview of Methodology

To explore the research questions about the distribution and the selection of discourse anaphora in English and Korean, two methods will be adopted: (1) natural data collection and (2) a survey of native speakers' interpretation of both languages, concerned with two linguistic aspects of discourse anaphora: (a) discourse anaphoric patterns and (b) sentential anaphors.

In the following sections, I identify the types of data collections studied (Section 3.2) and the survey methods used to elicit data from them (Section 3.3). First, information about the data collection is discussed, such as information about the compilation of corpus I, II, and III (Section 3.2.1), the corpus tool (Section 3.2.2), and corpora's independent and dependent variables (Section 3.2.3). Second, information about the survey is discussed, such as information about the subjects (Section 3.3.1), survey questionnaire sections A, B, C, and D (Section 3.3.2), the survey's independent and dependent variables (Section 3.3.3), and the Institutional Review Board and the process of obtaining the approval (Section 3.3.4). The chapter is wrapped up with a summary of how each method correlates to the research questions laid out in Section 1.3 of Chapter 1.

3.2 Data Collection

3.2.1 Corpus Compilation

Text corpora were collected to validate authentic distribution of anaphora in two linguistic aspects: (i) discourse anaphoric patterns and (ii) sentential anaphors. In fact, the data collection is threefold.

3.2.1.1 Corpus I: Discourse Anaphoric Pattern by Order of Mention

First, for discourse anaphoric patterns by order of mention; i.e., for general vs. sequential anaphoric patterning by the first, second, and third mention in one news article, text samples totaling 30,000 running words in each language were collected from news articles mainly about ten public figures to compare and contrast the characteristics of discourse anaphors to them. In fact, in order to make this comparison more focused, the text samples and the news articles in each language were controlled by the same size (i.e., 30,000[29] running words each in total) and by the same number of articles (i.e., 100 articles each in total).

For English, 30,000 word samples were collected from *The New York Times*, *Chicago Tribune, msnbc (news & sports), Los Angeles Times, TIME, The Dallas Morning News, The Washington Post, CNN Sports Illustrated, ESPN, Wikipedia,* and so on. For Korean, 30,000 word samples were collected from *The Hankyoreh*, *The Kyunghyang Shinmun, Sports Khan, The Korea Times, The Korean Herald, Daum,* and *Yahoo Korea, (Sports) Chosun, (Sports) Donga, Munhwa, Yonhap, Newsis, Wikipedia,* and so on.

[29] The sample size of the corpus was determined according to the criterion suggested by Bowker and Pearson (2002:54), in which the initial sample size is suggested more than 25,000 words.

3.2.1.2 Corpus II: Discourse Anaphoric Pattern by Paragraph

Second, for discourse anaphoric patterns by paragraph; i.e., for general vs. sequential anaphoric patterning by the first mention in each paragraph, the same text samples with the first corpus compilation totaling 30,000 running words in each language from news articles were analyzed to compare and contrast the characteristics of discourse anaphors to them.

Information about corpus compilation for discourse anaphoric patterns in each language is summarized in Table 3.1.

Table 3.1 Corpus I and II compilation for discourse anaphoric patterns

Corpus Compilation			English	Korean
Discourse Anaphoric Pattern (general vs. sequential)	Corpus I. By Order of Mention Corpus II. By Paragraph	Word Samples	Samples of 30,000 running words	Samples of 30,000 running words
		Sources	*The New York Times*, *Chicago Tribune*, *msnbc (news & sports)*, *Los Angeles Times*, *TIME*, *The Dallas Morning News*, *The Washington Post*, *CNN Sports Illustrated*, *ESPN*, and *Wikipedia*	*The Hankyoreh*, *The Kyunghyang Shinmun*, *Sports Khan*, *The Korea Times*, *The Korean Herald*, *Daum*, *Yahoo Korea*, *(Sports) Chosun*, *(Sports) Donga*, *Munhwa*, *Yonhap*, *Newsis*, and *Wikipedia*

In particular, in the text corpus compilation for discourse anaphoric patterns both by order of mention and by paragraph, ten public figures were collected from the news articles that mainly describe them in each language in terms of the same size of sampled words and the same number of news articles, as listed in Table 3.2.

Table 3.2 Ten public figures from news articles in each language

English			Korean		
Public Figures	Sampled Words	Number of Articles	Public Figures	Sampled Words	Number of Articles
Derek Jeter	3,015	10	Chan Ho Park	2,887	10
Michael Jordan	2,980	10	Ji-Sung Park	3,946	10
Steve Jobs	3,002	10	Yu-Na Kim	2,432	10
Bill Gates	3,010	10	Tae-Hwan Park	2,548	10
Jay Leno	2,997	10	Cheol- Soo Ahn	3,949	10
Brad Pitt	2,916	10	Jae-Seok Yoo	3,169	10
Morgan Freeman	3,017	10	Suk-Kyu Han	3,406	10
Norah Jones	3,000	10	Sang-Eun Lee	2,357	10
Jonathan Galassi (male author)	2,978	10	In-Ho Choi (male author)	2,785	10
Ann Beattie (female author)	3,085	10	Kyung-Sook Shin (female author)	2,521	10
Total	30,000	100	Total	30,000	100

3.2.1.3 Corpus III: Sentential Anaphors

Third, for referential properties of sentential anaphors, notably *that* and *it* in English; *ku-ken* 'that' and *ku-len* 'it' in Korean, conversation samples with scripts totaling 30,000 running words in each language were investigated from television serial dramas. In the same way with corpus I and II, the script samples in each language were controlled by the same size (i.e., 30,000 running words each) to make this comparison more focused.

For English, script samples totaling 30,000 running words were collected from *The Office* (Season #1 and Season #2), an American mockumentary comedy television series broadcast by National Broadcasting Company (NBC) since March 25, 2005. For Korean, script samples totaling 30,000 running words were collected from *My Too Perfect Sons*, weekend television drama series that were broadcast by Korean Broadcasting System (KBS) from April 11, 2009 to October 11, 2009.

Information about corpus compilation for sentential anaphors in each language is summarized in Table 3.3.

Table 3.3 Corpus III compilation for sentential anaphors in each language

English		Korean	
The Office Season 1&2	Word Samples	*My Too Perfect Sons*	Word Samples
Season 1 Episode #1	4,147	Episode #43	5,201
Season 1 Episode #2	4,161	Episode #46	3,251
Season 1 Episode #3	3,996	Episode #47	3,859
Season 1 Episode #4	4,643	Episode #48	2,507
Season 1 Episode #5	4,722	Episode #50	4,987
Season 1 Episode #6	4,648	Episode #52	4,988
Season 2 Episode #1	3,683	Episode #53	5,207
Total	30,000	Total	30,000

3.2.2 Corpus Tool

For corpus I and II, I manually analyzed the text samples about leading public figures in news articles by comparing the respective frequencies and the respective structural patterns in two languages in terms of the order of mention and by the paragraph

after printing out all news articles. On the one hand, for discourse anaphoric pattern by the order of mention, I manually analyzed the first, second, and the third mentions in all news articles, and the structural patterns in all mentions, and the distance between coreferential mentions in terms of the number of the sentence. In this analysis, the focused anaphoric terms are naming forms such as full names (plus titles and/or affiliation), single names (plus titles), and pronouns (and possibly zero forms in Korean). On the other hand, for discourse anaphoric pattern by the paragraph, I manually analyzed the first mentions in the first three paragraphs in all news articles, and the structural patterns in all mentions, and the distance between coreferential mentions in terms of the number of the sentence. In this analysis, the focused anaphoric terms also are naming forms such as full names (plus titles and/or affiliation), single names (plus titles), and pronouns (and possibly zero forms in Korean).

For corpus III, once word samples for sentential anaphors were collected in each language, the text corpora were saved in MS Word files according to language and the data source. Then they were all converted into the machine readable texts, i.e., the plain text files. Those text files for each sentential anaphor in each language were also opened with AntConc 3.3.5w, i.e., a freeware concordance program, in order to create word lists and to examine word frequency, concordance lines showing concordance, concordance plot, and file view for each sentential anaphoric feature, in which the focused linguistic items are *that* and *it* in English and *ku-ken* ‘that’ and *ku-len* ‘it’ in Korean.

By concordance, concordance plot, and file view in AntConc, I analyzed the each word token in each sentential anaphor, as shown in Figure 3.1 through Figure 3.3.

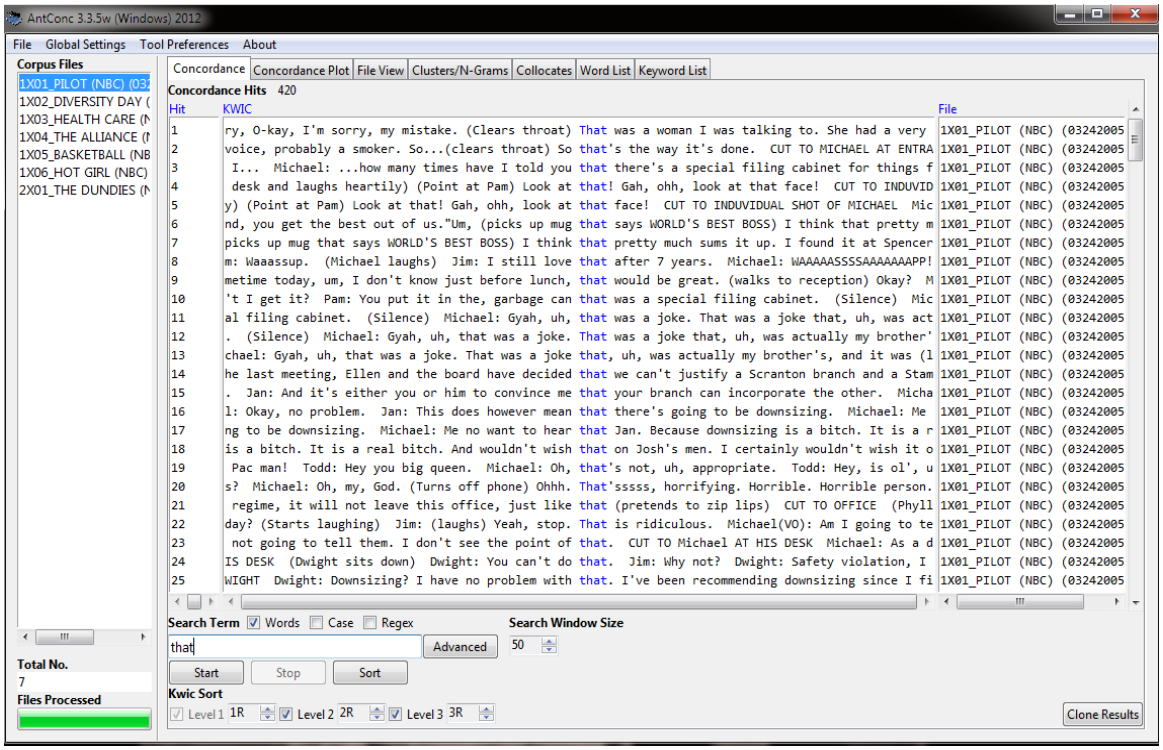


Figure 3.1 The concordance of sentential anaphor *that* in AntConc 3.3.5w

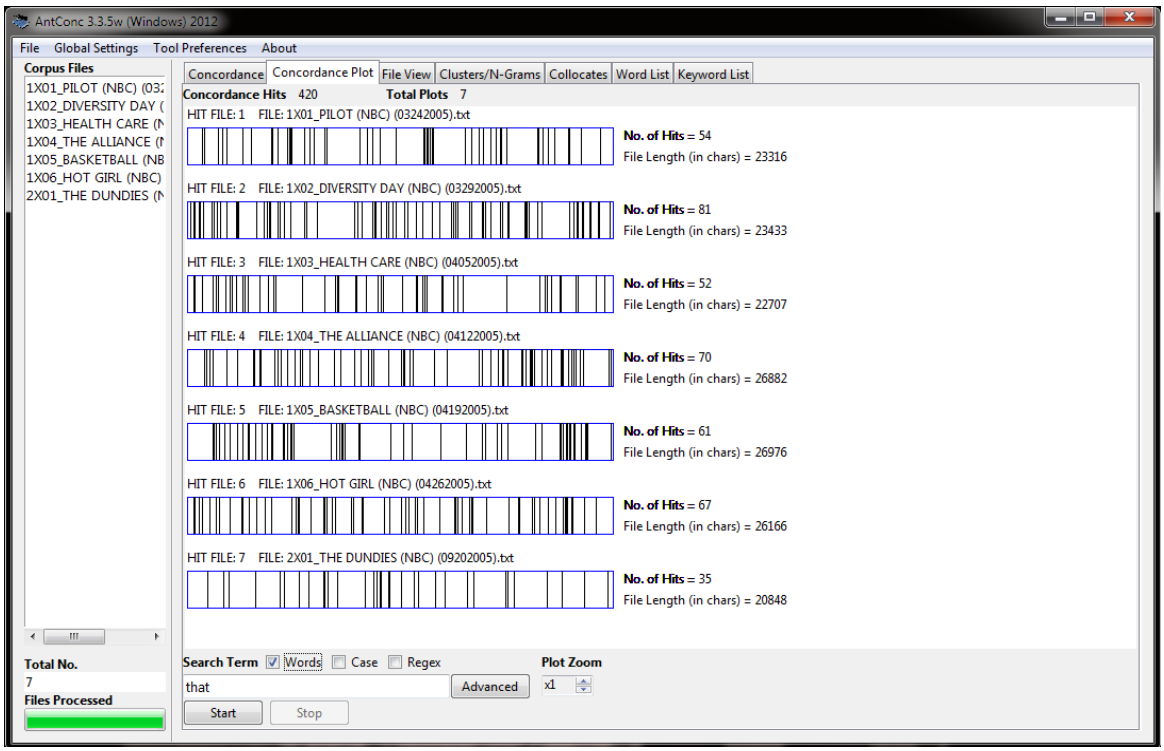


Figure 3.2 The concordance plot of sentential anaphor *that* in AntConc 3.3.5w

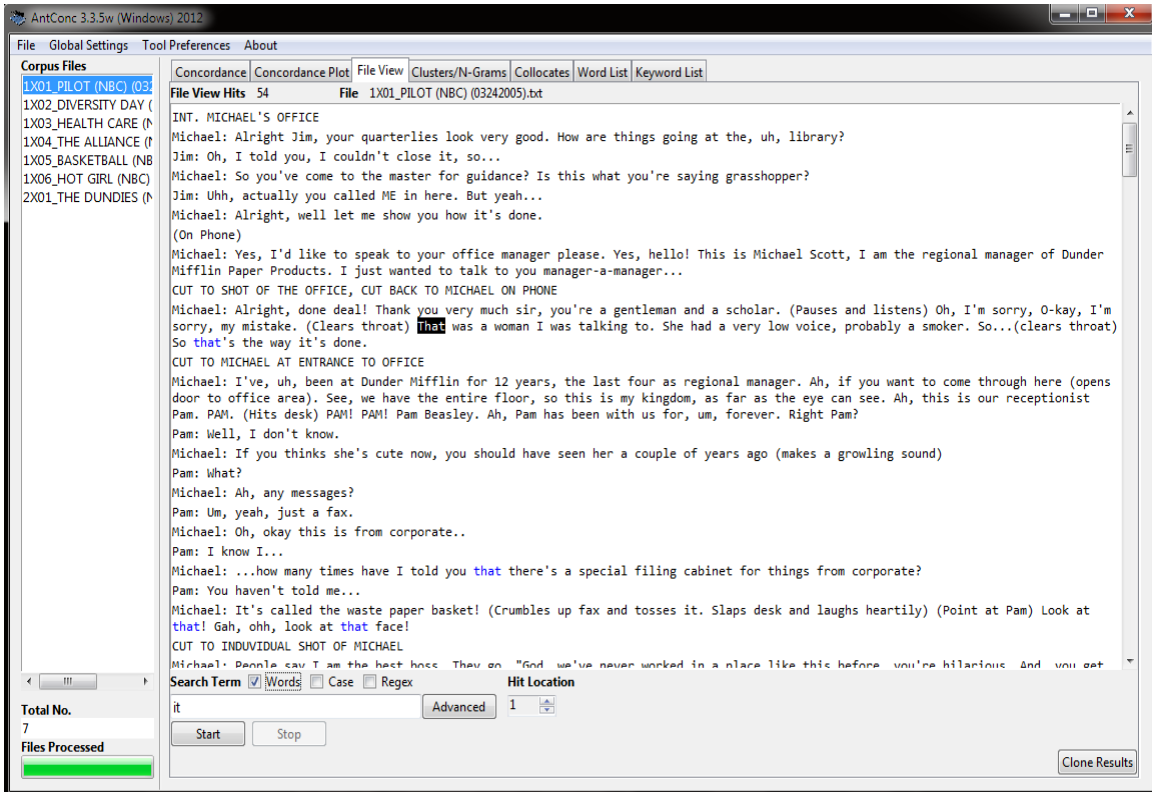


Figure 3.3 The file view of sentential anaphor *that* in AntConc 3.3.5w

For Korean, I changed the Character Encoding Setting in Global Settings ribbon into “korean(euc-kr)” in order to run the plain texts in Korean in AntCont 3.3.5w. Then I opened all seven text files in Korean and analyzed each sentential anaphor, as shown in Figure 3.4 through Figure 3.6.

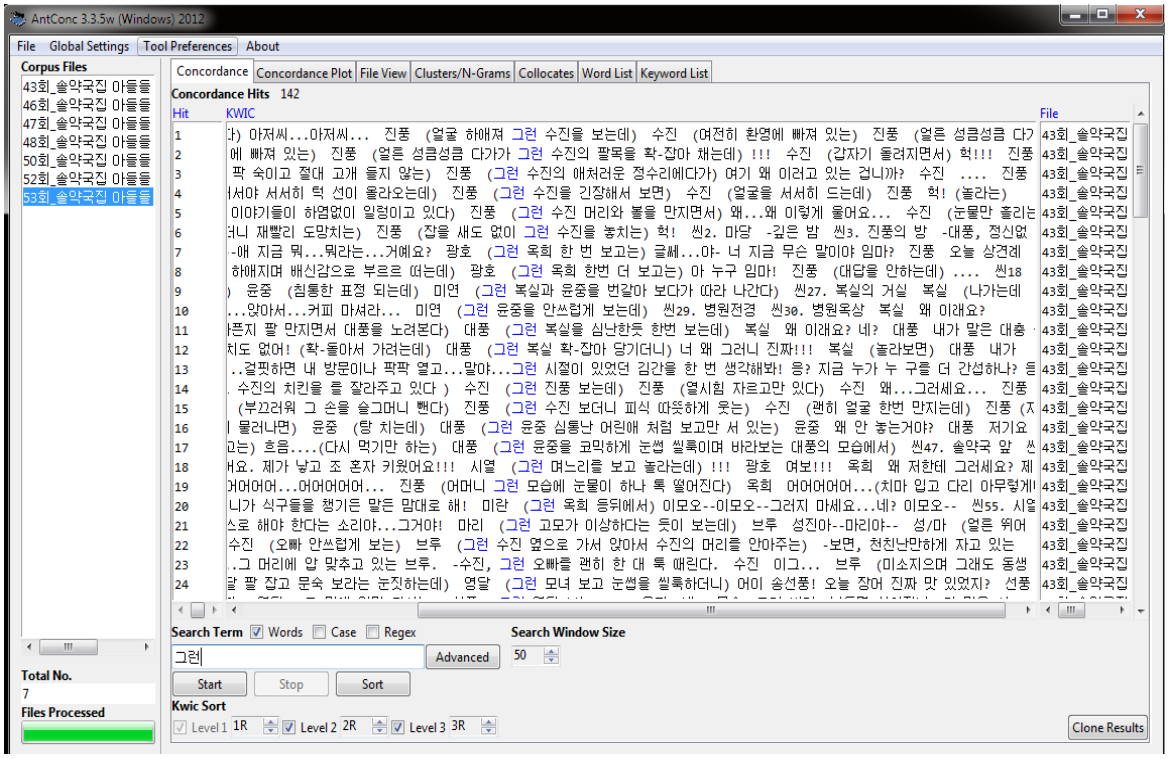


Figure 3.4 The concordance of sentential anaphor *kulen* 'it' in AntConc 3.3.5w

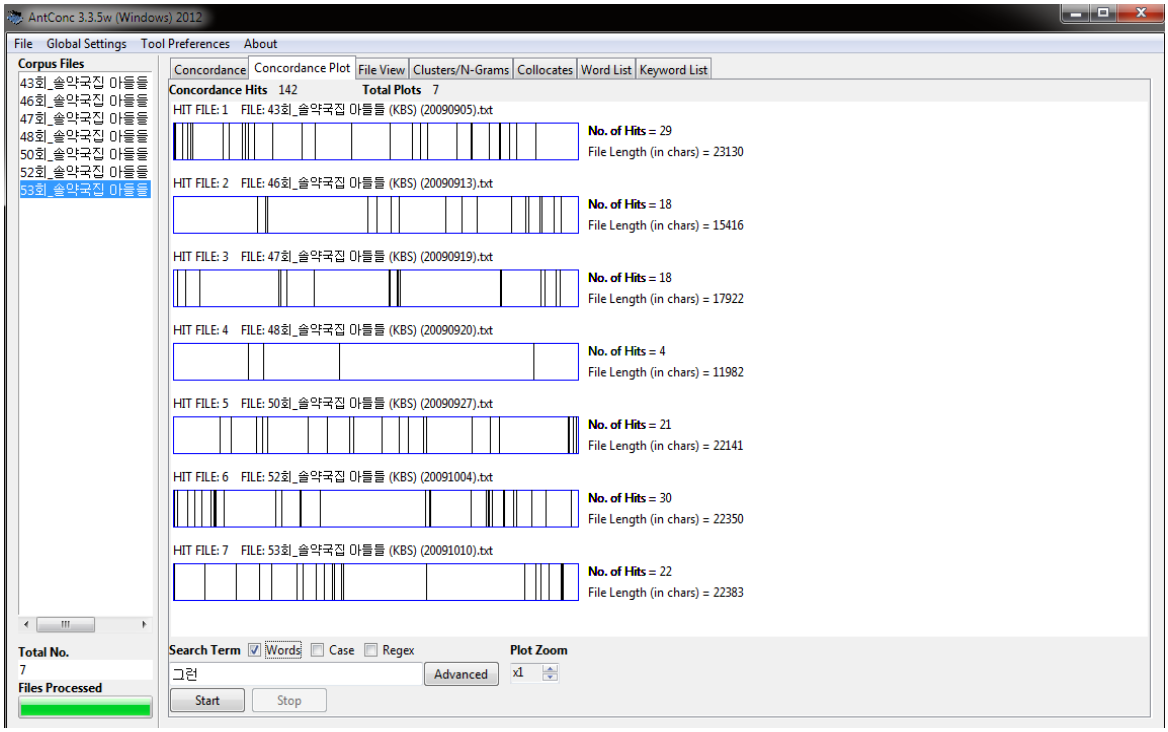


Figure 3.5 The concordance plot of sentential anaphor *kulen* 'it' in AntConc 3.3.5w

Figure 3.6 The file view of sentential anaphor *kulen* 'it' in AntConc 3.3.5w

3.2.3 Variables

The independent variables (IV) and the dependent variables (DV) are differentiated according to the types of discourse anaphoric features: (i) discourse anaphoric patterns by order of mention (Corpus I) and by paragraph (Corpus II) and (ii) referential properties of sentential anaphors (Corpus III).

First, in Corpus I, discourse anaphoric patterns were analyzed by the order of mention (i.e., the first mention, the second mention, and the third mention), which is the independent variable. On the other hand, frequencies of naming forms such as full name, the single name (i.e., surname or given name) (possibly plus titles), and pronoun

(possibly zero forms in Korean) are dependent variables in discourse anaphoric patterns by order of mention.

Second, in Corpus II, discourse anaphoric patterns were analyzed by the paragraph (i.e., the first mention in the first paragraph, the first mention in the second paragraph, and the first mention in the third paragraph), which is the independent variable. On the other hand, frequencies of naming forms such as full name, the single name (i.e., surname or given name) (possibly plus titles), and pronoun (possibly zero forms in Korean) are dependent variables in discourse anaphoric patterns by paragraph.

In Corpus I and II, the distance between discourse anaphors and other retrievals in English and Korean functions as an independent variable. That is, the distance between coreferential expressions and other retrievals except for the first mention in each news article can effect on the distribution of discourse anaphors. This factor is analyzed by the sentence level, i.e., the same sentence, previous sentence, 2 sentences away, 3 sentences away, and more than 4 sentences away.

Third, in Corpus III, sentential anaphors were analyzed by two independent variables: (i) the givenness of information; i.e., new vs. old information and (ii) the width of reference; i.e., narrow vs. wide reference. On the other hand, frequencies of sentential anaphors such as *that* and *it* in English; *ku-ken* 'that' and *ku-len* 'it' in Korean are dependent variables.

For given probabilities between categories, Chi-square tests were conducted, by which statistical significance in any relationship between the variables and each case in terms of the significance between the expected frequency and the observed frequency.

Thus, all frequencies were encoded into the Excel files and then these Excel files were into converted into Comma Separated Values (CSV, comma delimited) files in order to run those in RStudio (v 0.97.248), an open source programming language and software environment for statistical computing and graphics, for obtaining Chi-square test results.

3.3 Survey Design

3.3.1 Subjects

In order to test actual selections of two discourse anaphoric patterns; i.e., discourse anaphoric patterns and sentential anaphors, 20 native speakers of English and 20 native speakers of Korean were actually recruited as subjects to participate in a threefold survey: (i) discourse anaphoric pattern by order of mention, (ii) discourse anaphoric pattern by paragraph, and (iii) sentential anaphors. The survey was actually conducted by me from May 2, 2012 to May 9, 2012 at the University of Texas at Arlington campus.

The participants for the survey were all undergraduate and graduate students at UT Arlington who are aged 19 or over. The participants are all the students majoring in subjects other than linguistics and having no knowledge of linguistics since the main goal of this survey is to see if how subjects (i.e., non-linguists) select discourse anaphoric forms at the expected positions in actual written discourse.

The participants were recruited directly by me by canvassing public areas at UT Arlington campus such as the Central Library, E.H. Hereford University Center, Maverick Activities Center, Hammond Hall, Trimble Hall, English Language Institute

(ELI), Business Building, Chemistry & Physics Building, Engineering Research Building, Architecture Building, and Fine Arts Building.

After being notified that the survey is anonymous and voluntary, and being asked about her/his willingness to participate, each subject was asked to answer two screening questions regarding her/his native language (i.e., English & Korean) and her/his age (i.e., over 18). If s/he was willing to participate in the research survey and met these screening criteria, s/he was presented with the survey questionnaire in her/his own native language.

In fact, when I realized that a subject is aged under 18, the survey was stopped at once. In addition, before an actual survey by the participants, I explained my general plan for this survey with the tile of 'a survey of word choice: naming forms and pronouns', the purpose of the survey, duration, procedures, possible benefits/discomforts, and compensation, according to the Informed Consent Form, approved in March 26, 2012 by Institutional Review Board (IRB) at the University of Texas at Arlington.

After each subject completed the survey questionnaire in each language, s/he was given five dollars in cash by me as a compensation for participating in the survey at the end of the survey. The payment of copies for the questionnaire ($45.00) and the compensation for the subjects ($200.00 for 20 native speakers of English and 20 native speakers of Korean) were reimbursed by the research fund, the Dean's Excellence Award for Graduate Research/Creative Activity by College of Liberal Arts at UT Arlington, which was awarded ($550.00) in April 28, 2011.

3.3.2 Questionnaire

The questionnaire in the research survey was divided into four sections: Section A (General Questionnaire), Section B (Naming Form Type I), Section C (Naming Form Type II), and Section D (Pronoun) in each language, as represented in Table 3.4. The questionnaires that I used for the survey can be found in Appendix A (for native speakers of English) and Appendix B (for native speakers of Korean).

Table 3.4 Information about the survey questionnaire in each language

Section \ Language		English		Korean	
SECTION A		General questions about subjects in English		General questions about subjects in Korean	
SECTION B (by order of mention)		SET A	SET B	SET A	SET B
Degree of famousness[30]	High	Derek Jeter (male)	Steve Jobs (male)	Chan Ho Park (male)	Cheol-Soo Ahn (male)
	Mid-high	Norah Jones (female)	Jay Leno (male)	Sang-Eun Lee (female)	Jae-Seok Yoo (male)
	Mid-low	Jonathan Galassi (male)	Ann Beattie (female)	In-Ho Choi (male)	Kyung-Sook Shin (female)
SECTION C (by paragraph)		SET A	SET B	SET A	SET B
Degree of famousness	High	Derek Jeter (male)	Steve Jobs (male)	Chan Ho Park (male)	Cheol-Soo Ahn (male)
	Mid-high	Norah Jones (female)	Jay Leno (male)	Sang-Eun Lee (female)	Jae-Seok Yoo (male)
	Mid-low	Jonathan Galassi (male)	Ann Beattie (female)	In-Ho Choi (male)	Kyung-Sook Shin (female)
SECTION D (Sentential anaphors)		*that vs. it*		*ku-ken* 'that' vs. *kulen-ken* 'it'	

[30] The degree of famousness was determined in the Google results by hitting each full name in each language: High = -5,000,000 or more results; Mid-high = 1,000,000-5,000,000 results; Mid-low, less than 1,000,000 results.

As shown in Table 3.4, the first section of the questionnaire involved general questions about the subjects in their respective language, such as their native language (one of screening questions), sex, level of education, and age. The second section of the questionnaire involved questions about naming forms by order of mention with two sets according to the degree of famousness (i.e., high, mid-high, and mid-low). The third section of the questionnaire involved questions about naming forms by the first mention in each paragraph with two sets in the same degree of famousness for the same public figures to the second section. The fourth section of the questionnaire involved questions about sentential anaphoric forms.

3.3.2.1 Section A: General Questions

In the first part (Section A) of the survey questionnaire, the general title of the questionnaire 'Word Choice Questionnaire' and five questions were given in each language, as illustrated in Figure 3.7.

Word Choice Questionnaire

SECTION A: General Questionnaire

Date (mm/dd/yr)

		/			/	2	0	1	2

Native Language

English ☐	Korean ☐
Other:	

Sex

Male ☐	Female ☐

Education

High School ☐	Masters (MA) ☐	Other:
Bachelor (BA) ☐	PhD/MD/JD ☐	

Age

~19	☐
20~29	☐
30~39	☐
40~49	☐
50~59	☐
60~	☐

Figure 3.7 The section A of the questionnaire

In Section A, the participants were asked about general questions, such as survey date, native language (one of the screening questions), sex, level of education, and age (the other screening question).

3.3.2.2 Section B: Anaphoric Pattern by Order of Mention

In the second part (Section B: Naming Form [31] Type I) of the survey questionnaire, the questions focus on discourse anaphoric pattern by order of mention in each language. The participants were asked about degree of all possible naming forms (i.e., full name, surname, pronoun, given name, and other) in each question and about level of famousness for each referent mentioned in a paragraph, as illustrated in Figure 3.8.

[31] The title of each section was written in layperson's terms so that the survey participants (i.e., non-linguists) would more easily understand the questionnaire. Thus, discourse anaphoric patterns (by order of mention and by paragraph) and sentential anaphors were written as naming form type (I and II) and pronoun, respectively.

SECTION B: Naming Form Type I

DIRECTION

The following text is taken from a newspaper article.
Please read the paragraph, (i) rate your level of acceptability for **all naming forms (a, b, c, d) in the first three mentions** on a scale of five to one (5-1), (ii) write **other possible naming form(s) (e)** with the level of acceptability, and (iii) answer the question.

SECTION B (1)

Q1>_____ is an American baseball shortstop who has played 17 years in Major League Baseball (MLB) for the New York Yankees. A twelve-time All-Star and five-time World Series champion, **Q2>**_____ has been a central figure of the Yankees during their success of the 1990s and 2000s due to his clubhouse presence, on-field leadership, hitting ability, and baserunning. **Q3>**_____ is the Yankees' all-time career leader in hits (3,088), games played (2,426), stolen bases (329), and at bats (9,868). His accolades include four Silver Slugger Awards and five Gold Glove awards. (*Wikipedia*)

Q1.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Derek Jeter	5	4	3	2	1
b. Jeter	5	4	3	2	1
c. He	5	4	3	2	1
d. Derek	5	4	3	2	1
e. Other:	5	4	3	2	1

Q2.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Derek Jeter	5	4	3	2	1
b. Jeter	5	4	3	2	1
c. He	5	4	3	2	1
d. Derek	5	4	3	2	1
e. Other:	5	4	3	2	1

Q3.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Derek Jeter	5	4	3	2	1
b. Jeter	5	4	3	2	1
c. He	5	4	3	2	1
d. Derek	5	4	3	2	1
e. Other:	5	4	3	2	1

Q4. Please indicate the level of famousness for this person in the United States.

very well-known				very unknown
5	4	3	2	1

Figure 3.8 The section B of the questionnaire

As seen in Figure 3.8, first, the participants read through the direction carefully. Second, in the left-hand of one page, they read through one paragraph (i.e., the first paragraph in an article) in which the first three mentions are blanked like the cloze test format. Third, in the right-hand of one page, they were asked to rate the degree of acceptability according to the format of a typical five-level Likert item (i.e., 5: perfectly acceptable to 1: absolutely unacceptable) for all naming forms (a: full name, b: surname, c: pronoun, and d: given name) in one question (Q1: the first mention, Q2: the second mention, and Q3: the third mention). If possible, they were also asked to write other naming form(s) (e) with the degree of acceptability. Fourth, they were asked to rate the level of famousness for the referent mentioned in the paragraph. In the same way, the participants read through six paragraphs for six referents (i.e., three paragraphs in Set A and three paragraphs in Set B) and answered the same type of questions. In fact, all six paragraphs for all six referents in Section B were selected from the newspaper articles in the text corpora I and II.

3.3.2.3 Section C: Anaphoric Pattern by Paragraph

In the third part (Section C: Naming Form Type II) of the survey questionnaire, the questions focus on discourse anaphoric pattern by paragraph (i.e., the first mention in each paragraph) in each language. The participants were asked about degree of all possible naming forms (i.e., full name, surname, pronoun, given name, and other) in each question and about level of famousness for each referent mentioned in a news article, as illustrated in Figure 3.9.

SECTION C (1)

Paragraph A

Q1>_____ is the shortstop and captain of the New York Yankees. He has been an all-star, a most valuable player and one whose name is spoken in the adoring tones reserved mostly for players with names like Ruth, Gehrig or Mantle. In the long trip from phenom to aging veteran, his skills may have begun to wane, but his popularity has not.

Paragraph B

Q2>_____ has been a team leader who never gets in trouble in an era rife with athlete misbehavior and discontent. He has also been the most marketable athlete in New York — as safe for an advertiser to use as a deceased icon.

Paragraph C

Q3>_____ has played for no other major league team, which puts him in a revered class with seven Yankees Hall of Famers, as well as Thurman Munson and Don Mattingly, his predecessors as team captain. For the current generation of Yankees fans, he is the equivalent of other Yankees-for-life.

(*The New York Times*, *Times Topics*)

Q1.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Derek Jeter	5	4	3	2	1
b. Jeter	5	4	3	2	1
c. He	5	4	3	2	1
d. Derek	5	4	3	2	1
e. Other:	5	4	3	2	1

Q2.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Derek Jeter	5	4	3	2	1
b. Jeter	5	4	3	2	1
c. He	5	4	3	2	1
d. Derek	5	4	3	2	1
e. Other:	5	4	3	2	1

Q3.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Derek Jeter	5	4	3	2	1
b. Jeter	5	4	3	2	1
c. He	5	4	3	2	1
d. Derek	5	4	3	2	1
e. Other:	5	4	3	2	1

Q4. Please indicate the level of famousness for this person in the United States.

very well-known				very unknown
5	4	3	2	1

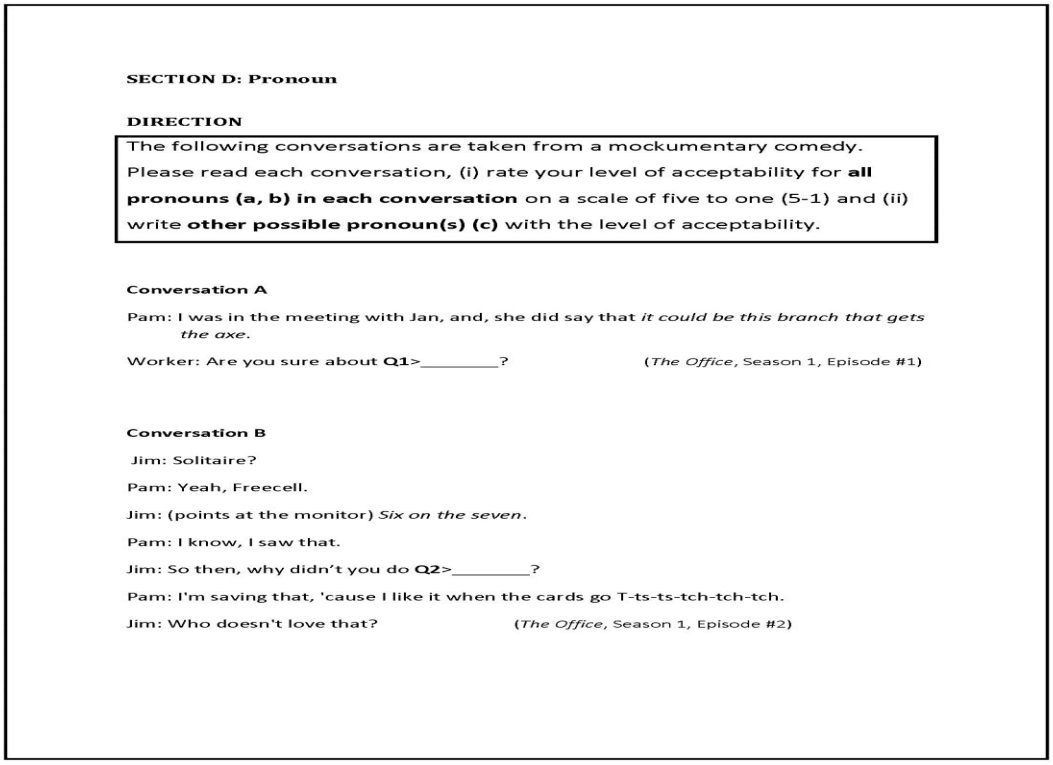
Figure 3.9 The section C of the questionnaire

As seen in Figure 3.9, the participants first read through the directions. Second, in the left-hand page, they read through three paragraphs in order (i.e., paragraph A through paragraph C in an article) in which the first mentions in each paragraph are blanked like the cloze test format. Third, in the right-hand page, they were asked to rate the degree of acceptability according to the format of a typical five-level Likert item (i.e., 5: perfectly acceptable to 1: absolutely unacceptable) for all naming forms (a: full name, b: surname, c: pronoun, and d: given name) in one question (Q1: the first mention, Q2: the second mention, and Q3: the third mention). If relevant, they were also asked to write other naming form(s) (e) with the degree of acceptability. Fourth, they were asked to rate the level of famousness for the referent mentioned in the news article. In the same way, the participants read through six news articles for six referents (i.e., three news articles in Set A and three news articles in Set B) and answered the same type of questions. All six news articles for all six referents in Section C were selected from the newspaper articles in the text corpora I and II.

3.3.2.4 Section D: Sentential Anaphors

In the fourth part of the survey questionnaire (Section D: Pronoun), the questions focus on sentential anaphors in each conversation in terms of two referential properties: the givenness of knowledge (i.e., new vs. old information) and the width of reference (i.e., narrow vs. wide reference), in which Conversation A involves new information; Conversation B involves old information; Conversation C involves narrow reference; Conversation D involves wide reference. The participants were asked about degree of all

possible pro-forms (i.e., *that*, *it*, and other in English; *ku-ken* ‘that’, *ku-len-ken* ‘it’, and other in Korean) in each conversation, as illustrated in Figure 3.10.



SECTION D: Pronoun

DIRECTION

The following conversations are taken from a mockumentary comedy. Please read each conversation, (i) rate your level of acceptability for **all pronouns (a, b) in each conversation** on a scale of five to one (5-1) and (ii) write **other possible pronoun(s) (c)** with the level of acceptability.

Conversation A

Pam: I was in the meeting with Jan, and, she did say that *it could be this branch that gets the axe*.

Worker: Are you sure about **Q1**>_____? (*The Office*, Season 1, Episode #1)

Conversation B

Jim: Solitaire?

Pam: Yeah, Freecell.

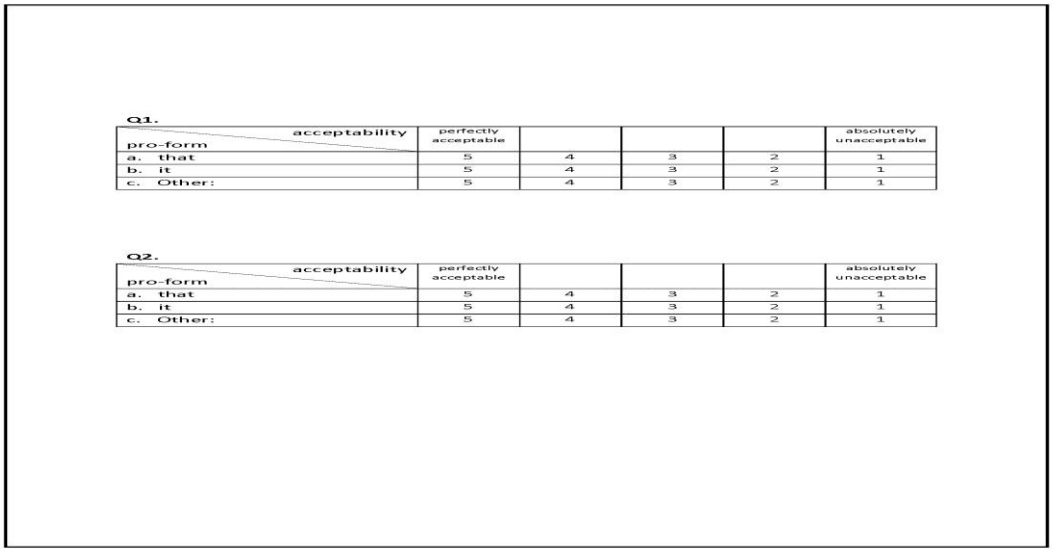
Jim: (points at the monitor) *Six on the seven*.

Pam: I know, I saw that.

Jim: So then, why didn’t you do **Q2**>_____?

Pam: I'm saving that, 'cause I like it when the cards go T-ts-ts-tch-tch-tch.

Jim: Who doesn't love that? (*The Office*, Season 1, Episode #2)



Q1.

acceptability / pro-form	perfectly acceptable				absolutely unacceptable
a. that	5	4	3	2	1
b. it	5	4	3	2	1
c. Other:	5	4	3	2	1

Q2.

acceptability / pro-form	perfectly acceptable				absolutely unacceptable
a. that	5	4	3	2	1
b. it	5	4	3	2	1
c. Other:	5	4	3	2	1

Figure 3.10 The section D of the questionnaire

As seen in Figure 3.10, first, the participants read through the directions. Second, in the left-hand page, they read through conversations in order (i.e., conversation A and B in one page; conversation C and D on the other page) in which the position of a particular sentential anaphor in each conversation is blanked like the cloze test format. Third, in the right-hand page, they were asked to rate the degree of acceptability according to the format of a typical five-level Likert item (i.e., 5: perfectly acceptable to 1: absolutely unacceptable) for two sentential anaphors (a: *that* and b: *it* in English; a: *ku-ken* 'that' and *ku-len-ken* 'it' in Korean). If relevant, they were also asked to write other possible sentential anaphor(s) (e) with the degree of acceptability. In the same way, the participants read through four conversations for two sentential anaphors in each language and answered the same type of questions. All four conversations in Section D were selected from the scripts of the TV drama series from the text corpora III in each language.

3.3.3 Variables

In the research survey, the independent variables (IV) and the dependent variables (DV) are differentiated according to the types of discourse anaphoric features: (i) discourse anaphoric patterns by order of mention and by paragraph and (ii) referential properties of sentential anaphors, and the types of sections in the survey questionnaire.

First, in Section A, native language (i.e., one of screening questions) of the participants functions as an independent variable in the survey, while two discourse anaphoric patterns (i.e., general anaphoric pattern vs. sequential anaphoric pattern) for choosing naming forms with the degree of acceptability in Section B and Section C are

the dependent variables. However, other general questions such as sex, level of education and age were not considered important factors as independent variables in this research survey.

Second, in Section B, the order of mention (i.e., the first mention, the second mention, and the third mention) is the independent variable in each news article, while the degree of acceptability (i.e., mean values) of all possible naming forms (i.e., full name, surname, pronoun, and given name) in each question is the dependent variable. In addition, the degree of famousness (i.e., mean values) by participants in each language is the independent variable, while mean values of all naming forms in each question is the dependent variable.

Third, in Section C, the paragraph (i.e., the first mention[32] in paragraph A, the first mention in paragraph B, and the first mention in paragraph C) is the independent variable in each news article, while the degree of acceptability (i.e., mean values) of all possible naming forms (i.e., full name, surname, pronoun, and given name) in each question is the dependent variable. In addition, the degree of famousness (i.e., mean values) by participants in each language is the independent variable, while mean values of all naming forms in each question is the dependent variable.

Fourth, in Section D, the givenness of information (i.e., new vs. old information) and the width of reference (i.e., narrow vs. wide reference) are the independent variables in the survey, while the degree of acceptability (i.e., mean values) for sentential anaphors

[32] The first mentions in paragraphs were all controlled to be the subject positions in the first sentences without any modifying phrases, relative clauses, or subordinate clauses.

(i.e., *that* and *it* in English; *ku-ken* ‘that’ and *ku-len* ‘it’ in Korean) in each conversation is the dependent variable.

As a note, other possible forms written by the participants in Section B, C, and D were not considered independent variables in the statistical analysis since not all participants wrote the other form(s) in each question and rate ‘0’ even though they wrote the other forms. Instead, other forms filled out by the participants will be descriptively explained by the types of naming forms, by the types of sentential anaphors and by the section.

3.3.4 Institutional Review Board

It was necessary for me to submit a research protocol to the Institutional Review Board (IRB) of the University of Texas at Arlington for approval since this research survey involved the use of human research subjects. The goal of the survey was to explain the natural selection by the participants’ responses in terms of naming forms and sentential pro-forms with confidentiality.

In fact, I intended to conduct the research survey with a quick and written anonymous survey with a set of three news articles for discourse anaphoric pattern by the paragraph (i.e., only Set A in Section C) and four conversations for sentential anaphors (i.e., Section D) in each language. Also, there was no compensation of five-dollar cash when the participants completed the survey questionnaire. In order to comply with the IRB, I submitted a protocol to the IRB of UT Arlington with two forms: IRB Form #1A (Proposal for Research involving Human Subjects Application for Exempt Research) and IRB form #3 (Application for Waiver or Alteration of Informed Consent Requirements).

The protocol, entitled "A Corpus-based Analysis of Discourse Anaphora in English and Korean: A Neo-Gricean Pragmatic Approach (A Survey for Word Choice)", was approved on December 19, 2011, under the protocol number 2012-0250e. The IRB forms that I submitted and the approval letter can be found in Appendix C.

In the meantime, there were some changes in the content and the length of the survey questionnaire and the number of participants in each language. The modified survey included participant compensation of five-dollar cash. Thus, I needed to resubmitted the protocol with modification to the IRB of UT Arlington with the modified IRB form #1A, IRB form #3, and Informed Consent Form that must be attached to the front page of the survey questionnaire. The protocol, the same title as the first submission, was approved on March 20, 2012, under the protocol number 2012-0250, and the Informed Consent Form was approved on March 26, 2012. The revised IRB forms and the informed consent form that I submitted and the approval letter for minor modification can also be found in Appendix C.

3.4 Chapter Summary

The methodology in the dissertation research was two-fold: (i) authentic data collection and (ii) the survey. In order to explore the distribution of discourse anaphors in English and Korean, the data collection was analyzed based on the text corpora from news articles and TV dramas, while in order to the selection of discourse anaphors in two languages, the survey was conducted by 20 native speakers of English and 20 native speakers of Korean. Both data collection and the survey were focused on two linguistic aspects of discourse anaphora: (i) discourse anaphoric pattern and (ii) sentential anaphors.

In the previous sections, I identified the types of data collections studied (Section 3.2) and the survey methods used to elicit data from them (Section 3.3). First, information about the data collection was discussed, such as information about three types of corpus compilation (Section 3.2.1), the corpus tool (Section 3.2.2), and the independent and dependent variables in the data collection (Section 3.2.3). Second, information about the survey was discussed, such as information about the subjects (Section 3.3.1), four sections of the survey questionnaire (Section 3.3.2), the independent and dependent variables (Section 3.3.3), and human subjects and the Institutional Review Board (Section 3.3.4).

In the following chapter, I give a fuller description of the findings of the discourse anaphoric patterns and those of the sentential anaphors in English and Korean.

Chapter 4

Results and Analysis: Corpus

This chapter analyzes the corpus data collected via sampling text from news articles in English and Korean in order to explore the distribution of discourse anaphors in English and Korean. It contains the following sections: Section 4.1 analyzes discourse anaphoric patterns by order of mention (4.1.1 for English; 4.1.2 for Korean) in terms of the first-, the second-, the third mentions, and section summary (4.1.3); Section 4.2 analyzes discourse anaphoric patterns by paragraph (4.2.1 for English; 4.2.2 for Korean) in terms of the first-, the second-, the third paragraph, and section summary (4.2.3); Section 4.3 analyzes the distance between discourse anaphors (4.3.1 for English; 4.3.2 for Korean) in terms of number of sentence; Section 4.4 analyzes sentential anaphors in two referential respects: by new vs. old information (4.4.1) (4.4.1.1 for English; 4.4.1.2 for Korean; 4.4.1.3 for summary) and by narrow vs. wide reference (4.4.2) (4.4.2.1 for English; 4.4.2.2 for Korean; 4.4.2.3 for summary); Section 4.5 summarizes this chapter in terms of two discourse anaphoric patterns (4.5.1) and sentential anaphors (4.5.2) in two languages.

4.1 Corpus I: Discourse Anaphoric Patterns by Order of Mention

4.1.1 By Order of Mention in English

The findings on all structural constructions of discourse anaphors referring to public figures in English news articles are correspondingly shown in Table 4.1 through Table 4.3 and Figure 4.1 through Figure 4.3 in terms of the first-, the second-, and the third mentions.

4.1.1.1 The First Mentions

First, all structural constructions of referring expressions referring to public figures for the first mentions are shown in Table 4.1 and Figure 4.1 in terms of their respective frequencies and examples.

Table 4.1 The types of construction and the relative frequency for the first mentions in English

Type of Construction			Frequency (n (%))	Example
I. Full Name (FN)	a.	FN	77 (77%)[33]	Steve Jobs
	b.	FN + Apposition (AP)	12 (12%)	Brad Pitt, ever the social activist
	c.	(Pseudo-) Title (TT) + FN	8 (8%)	Microsoft founder Bill Gates
	d.	Nickname (NN) + FN	1 (1%)	Late-night TV icon Jay Leno
II. Surname (SN)	a.	SN	0 (0%)	
	b.	TT + SN	0 (0%)	
III. Pronoun (Pro)			2 (2%)	His (Him)
IV. Given Name (GN)			0 (0%)	
V. Lexical Noun Phrase (LNP)			0 (0%)	
VI. Zero (Ø)			0 (0%)	
Total			100 (100%)	

[33] The line in red color represents the most frequent form, while the line in blue color represents the next frequent form in the dissertation.

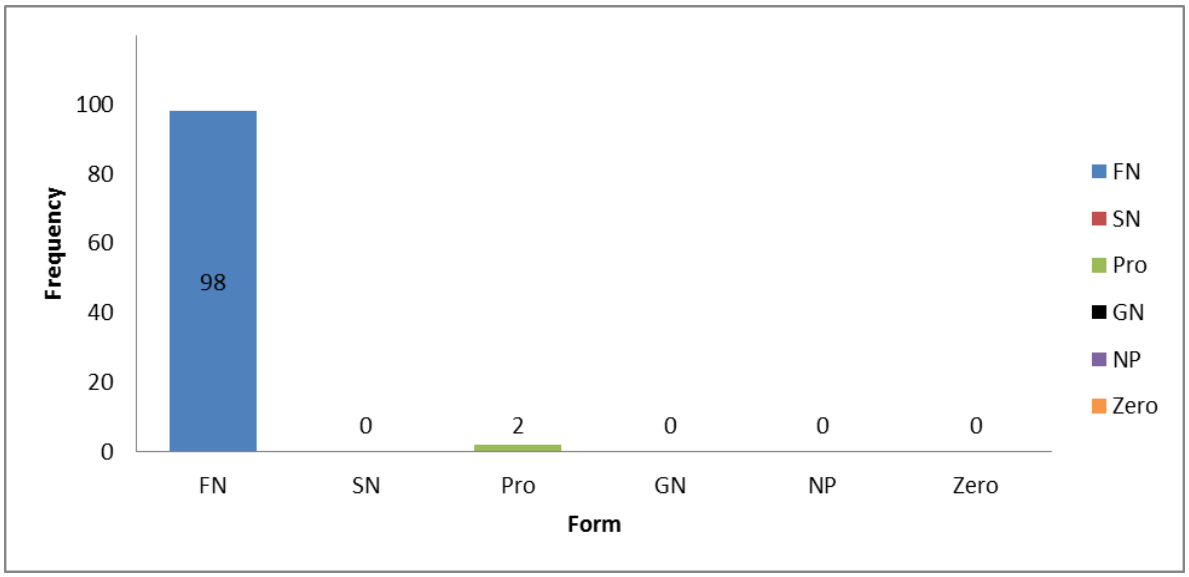


Figure 4.1 The relative frequency of the referring expressions for the first mentions in English

Table 4.1 and Figure 4.1 show that the referring expressions for the first mentions predominantly contain full names (98:98%), classified into four structural constructions in terms of being addressed with apposition, (pseudo-) title, and nickname.

This finding conforms to Ariel (1991, 1994, 2008, 2010) and Lee (2010) in that the first mentions are all referred to by the most discontinuous linguistic form, i.e., full name among other various constructions since the full name (plus modifier) is the lowest accessible marker. Only two occurrences were observed with pronouns (2:2%) for the first mentions, in which pronouns comes before the full names as referring expressions at the beginning of a given discourse. No occurrence was observed with surname, given name, lexical noun phrase, and zero form for the first mention in English.

4.1.1.2 The Second Mentions

Second, all structural constructions of discourse anaphors referring to public figures for the second mentions are shown in Table 4.2 and Figure 4.2 in terms of their respective frequencies and examples.

Table 4.2 The types of construction and the relative frequency for the second mentions in English

Type of Construction			Frequency (n (%))	Example
I. Full Name (FN)	a.	FN	0 (0%)	
	b.	FN + Apposition (AP)	0 (0%)	
	c.	(Pseudo-) Title (TT) + FN	0 (0%)	
	d.	Nickname (NN) + FN	0 (0%)	
II. Surname (SN)	a.	SN	27 (27%)	Jobs
	b.	TT + SN	6 (6%)	Mr. Gates
III. Pronoun (Pro)			64 (64%)	He (his, him)
IV. Given Name (GN)			0 (0%)	
V. Lexical NPs (LNPs)			3 (3%)	the poet
VI. Zero (Ø)			0 (0%)	
Total			100 (100%)	

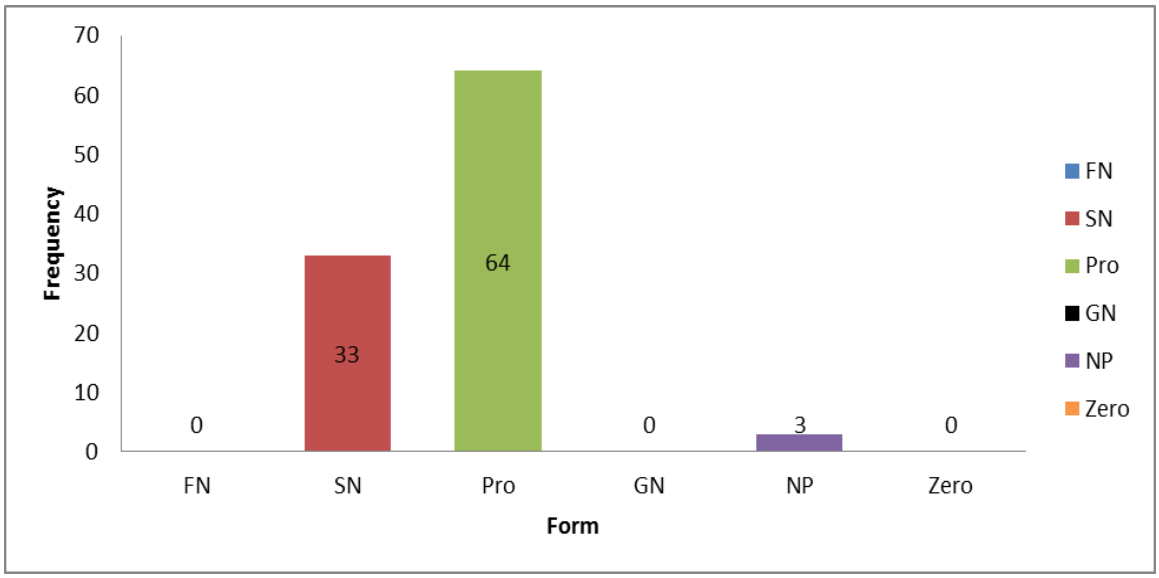


Figure 4.2 The relative frequency of discourse anaphors for the second mentions in English

Table 4.2 and Figure 4.2 show that the second mentions were predominantly retrieved with the pronouns (64:64%). One third of the second mentions were retrieved with the surnames (33:33%), in which 27 (27%) occurrences were retrieved with surname only, while 6 (6%) occurrences were retrieved with surname plus title. Only 3 (3%) occurrences were retrieved with lexical noun phrases for the second mentions. There was no occurrence of full name, given name, zero form for the second mentions in English.

This finding also conforms to Ariel (1991, 1994, 2008, 2010) and Lee (2010) in that the most accessible referent is referred to by the most accessible form, i.e. the pronoun, in English. Moreover, in fact, the pronouns were predominantly retrieved in the same sentence with the antecedents (57.3%).

4.1.1.3 The Third Mentions

Third, all structural constructions of discourse anaphors referring to public figures for the third mentions are shown in Table 4.3 and Figure 4.3 in terms of their respective frequencies and examples.

Table 4.3 The types of construction and the relative frequency for the third mentions in English

Type of Construction			Frequency (n (%))	Example
I. Full Name (FN)	a.	FN	2 (2%)	Derek Jeter
	b.	FN + Apposition (AP)	0 (0%)	
	c.	(Pseudo-) Title (TT) + FN	0 (0%)	
	d.	Nickname (NN) + FN	0 (0%)	
	e.	Abbreviated FN	1 (1%)	MJ
II. Surname (SN)	a.	SN	26 (26%)	
	b.	TT + SN	5 (50%)	
III. Pronoun (Pro)			60 (60%)	He (his, him)
IV. Given Name (GN)			0 (0%)	
V. Lexical NPs (LNPs)			6 (6%)	A phenomenal athlete
VI. Zero (Ø)			0 (0%)	
Total			100 (100%)	

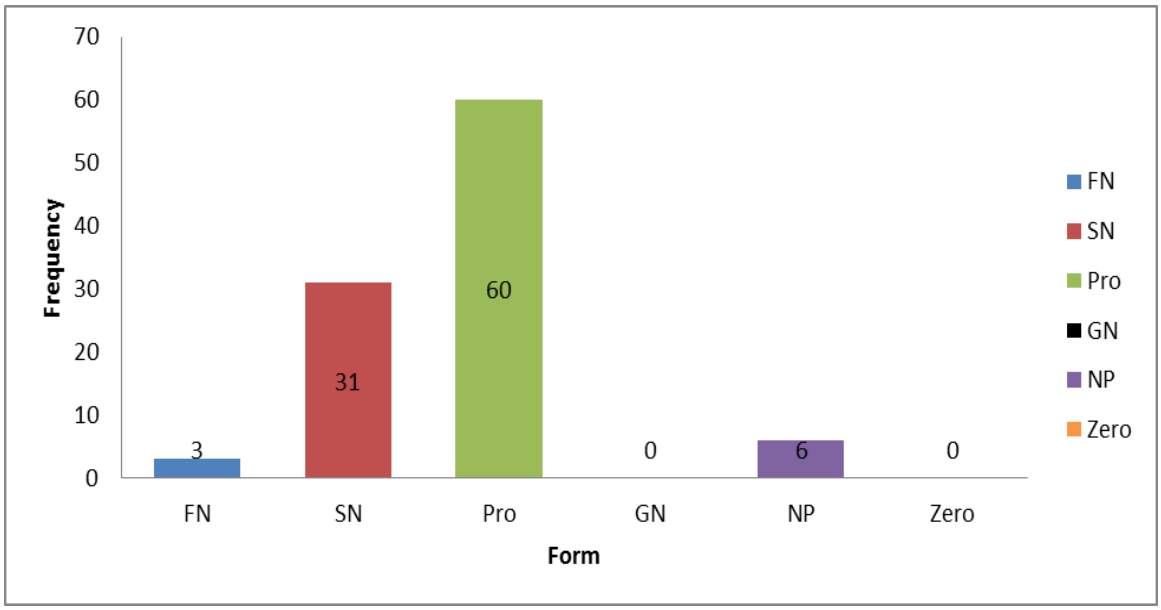


Figure 4.3 The relative frequency of discourse anaphors for the third mentions in English

Table 4.3 and Figure 4.3 show that the third mentions were also predominantly retrieved with the pronouns (60:60%). One third of the second mentions were retrieved with the surnames (31:31%), in which 26 occurrences were retrieved with surname only, while 5 occurrences were retrieved with surname plus title. Only 6 occurrences were retrieved with lexical noun phrases for the third mentions. Only 3 occurrences were retrieved with full names were retrieved for the third mention, in which 2 occurrences were full name only, while 1 occurrence was an abbreviated full name form.

This finding also conforms to Ariel (1991, 1994, 2008, 2010) and Lee (2010) in that the most accessible referent is referred to by the most accessible form, i.e. the pronoun, in English, rather than by other linguistic forms. In fact, the pronouns were predominantly retrieved in the same sentence with the antecedents (57.3%), for which it will be discussed in Section 4.3.1.

4.1.2 By Order of Mention in Korean

The findings on all structural constructions of discourse anaphors referring to public figures in Korean news articles are correspondingly shown in Table 4.4 through Table 4.9 and Figure 4.4 through Figure 4.6 in terms of the first-, the second-, and the third mentions.

4.1.2.1 The First Mentions

First, all structural constructions of referring expressions referring to public figures for the first mentions are shown in Table 4.4 and Figure 4.4 in terms of their respective frequencies and examples.

Table 4.4 The types of construction and the relative frequency for the first mentions in Korean

Type of Construction			Frequency (n (%))	Example
I. Full Name (FN)	a.	(Pseudo-) TT + FN	35 (35%)	paywu Han Seok-Kyu 'actor' (배우 한석규)
	b.	FN	20 (20%)	Kim Yu-Na (김연아)
	c.	NN + FN + AFF	12 (12%)	Korean thukkup Park Chan Ho (Pittsburgh) 'express player' (코리안 특급 박찬호 (피츠버그))
	d.	FN + AFF	12 (12%)	Park Tae-Hwan (Dankuk University) (박태환 (단국대))
	e.	NN + FN	10 (10%)	kwukmin MC Yoo Jae-Seok 'national' (국민 MC 유재석)
	f.	FN + AFF + TT	8 (8%)	Ahn Chel-Soo KAIST sekcwakyoswu 'chair professor' (안철수 KAIST 석좌교수)
	g.	AFF + FN	2 (2%)	mayncheysuthe yunaithitu Park Ji-Sung 'Manchester United' (맨체스터 유나이티드 박지성)
	h.	FN + TT	1 (1%)	Ahn Chel-Soo wencnag 'dean' (안철수 원장)

Table 4.4 – *Continued*

Type of Construction			Frequency (n (%))	Example
II. Surname (SN)	a.	SN	0 (0%)	
	b.	SN + TT	0 (0%)	
III. Pronoun (Pro)			0 (0%)	
IV. Given Name (GN)			0 (0%)	
V. Lexical NPs (LNPs)			0 (0%)	
VI. Zero (Ø)			0 (0%)	
Total			100 (100%)	

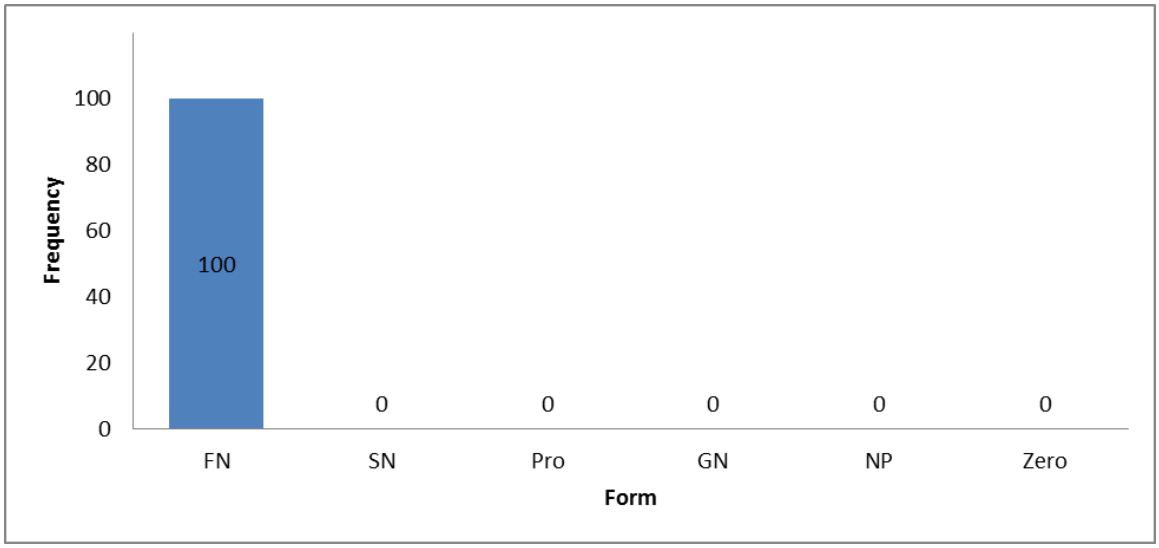


Figure 4.4 The relative frequency of discourse anaphors for the first mentions in Korean

Table 4.4 and Figure 4.4 show that all referring expressions for the first mentions predominantly contain full names (100:100%), classified into eight structural constructions in terms of being addressed with (pseudo-) title, affiliation and nickname. There was no occurrence of full names with apposition, compared with those in English; however, there were occurrence of full names with affiliation, in which all affiliation were all introduced in parenthesis.

This finding conforms to Ariel (1991, 1994, 2008, 2010) and Lee (2010) in that the first mentions are all referred to by the most discontinuous linguistic form, i.e., full name among other various constructions since the full name (plus modifier) is the lowest accessible marker.

In addition, the case markers of the types of construction for the first mentions in Korean were observed, as shown in Table 4.5 below.

Table 4.5 The relative frequency of the case markers for the first mentions in Korean

Frequency & Case marker / Type of Construction			Frequency								Total
			NOM	TOP	ACC	DAT	GEN	LOC	ABL	INST	
I. FN	a.	(Pseudo-) TT + FN	31	1			3				35
	b.	FN	10	2	2	1	5				20
	c.	NN + FN + AFF	11				1				12
	d.	FN + AFF	6	5		1					12
	e.	NN + FN	8	1		1					10
	f.	FN + AFF + TT	7	1							8
	g.	AFF + FN	1	1							2
	h.	FN + TT		1							1
II. SN	a.	SN									
	b.	SN + TT									
III. Pro											
IV. GN											
V. LNPs											
VI. Zero											
Total			74	12	2	3	9				100

Table 4.5 shows that for the first mentions, full names were predominantly introduced with the nominative case markers, i.e., *-i*/*-ka* (74:74%). Next, other cases of full names were introduced with the topical markers, i.e., *-un*/*-nun* (12:12%). Full names were also

introduced with other case markers such as genitive, i.e., *-uy* (9:9%), dative, i.e., *-eykey* (3:3%) and accusative, i.e., *-ul*/*-lul* (2:2%).

4.1.2.2 The Second Mentions

Second, all structural constructions of discourse anaphors referring to public figures for the second mentions are shown in Table 4.6 and Figure 4.5 in terms of their respective frequencies and examples.

Table 4.6 The types of construction and the relative frequency for the second mentions in Korean

Type of Construction			Frequency (n (%))	Example
I. Full Name (FN)	a.	FN	65 (65%)	Yoo Jae-Seok (유재석)
	b.	FN + TT	4 (4%)	Ahn Chel-Soo wencang 'dean' (안철수 원장)
	c.	FN + AFF	1 (1%)	Park Chan Ho (Pittsburgh) (박찬호 (피츠버그))
	d.	NN + FN	1 (1%)	phikye yewang Kim Yu-Na 'Figure Queen' (피겨여왕 김연아)
	e.	TT + FN	1 (1%)	kaswu Lee Sang-Eun 'singer' (가수 이상은)
II. Surname (SN)	a.	SN	0 (0%)	
	b.	SN + TT	14 (14%)	Shin cakka 'writer' (신 작가)
III. Pronoun (Pro)			9 (9%)	ku 's/he' (그)
IV. Given Name (GN)			0 (0%)	
V. Lexical NPs (LNPs)			1 (1%)	cakka 'writer' (작가)
VI. Zero (Ø)			4 (4%)	Ø
Total			100 (100%)	

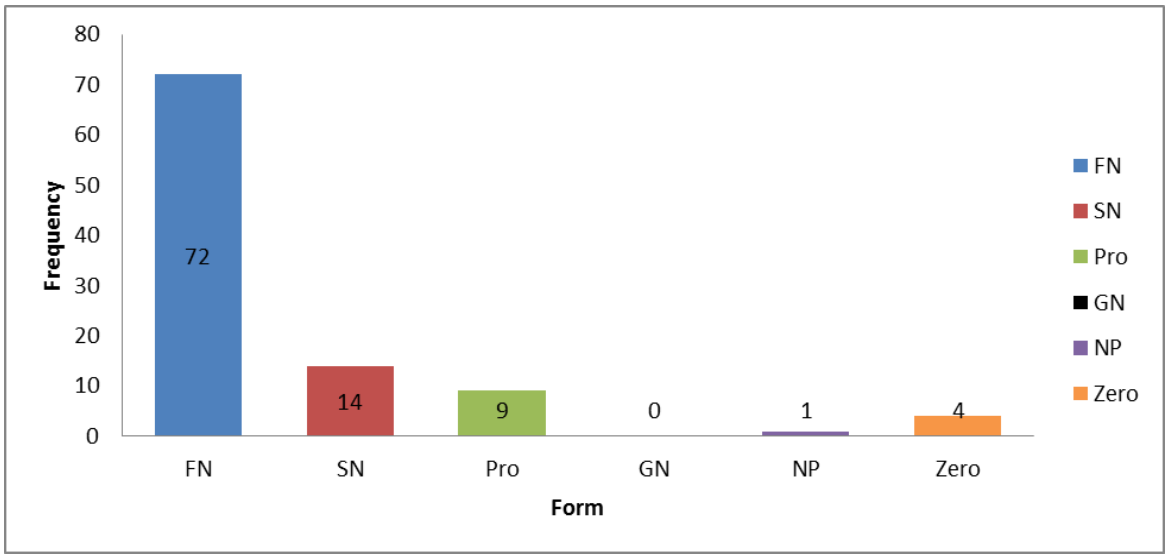


Figure 4.5 The relative frequency of discourse anaphors for the second mentions in Korean

Table 4.6 and Figure 4.5 show that 14 occurrences were retrieved with surnames plus title (14:14%), in which there was no occurrence of surname only. 9 occurrences were retrieved with pronouns (9:9%). 4 occurrences were retrieved with zero anaphors (4:4%). No occurrence with given names was observed.

Like the second mentions in English, these findings seem conform to Ariel (1991, 1994, 2008, 2010) and Lee (2010) in that a surname plus a(n) (abbreviated) title construction, pronoun, and zero anaphors (i.e., relatively higher accessible markers in Ariel's AH) rather than a full name (plus modifier) construction (i.e., a relatively lowest accessible marker) have been used far more frequently for the second retrievals.

However, it is noteworthy that almost of discourse anaphors for the second mentions predominantly contain full names (72:72%), classified into five structural constructions in terms of being addressed with title, affiliation and nickname. This

finding would definitely run counter to Lee's (2010) findings on English/Korean political news articles and Ariel's AH for English mentioned above (Kim 2010a, 2012).

In addition, the case markers of the types of construction for the second mentions in Korean were observed, as shown in Table 4.7 below.

Table 4.7 The relative frequency of the case markers for the second mentions in Korean

Frequency & Case marker / Type of Construction			Frequency								Total
			NOM	TOP	ACC	DAT	GEN	LOC	ABL	INST	
I. FN	a.	FN	6	44	3	1	11				65
	b.	FN + TT		4							4
	c.	FN + AFF	1								1
	d.	NN + FN	1								1
	e.	TT + FN	1								1
II. SN	a.	SN									
	b.	SN + TT		12			2				14
III. Pro			3	2		1	3				9
IV. GN											
V. LNPs				1							1
VI. Zero				4							4
Total			12	67	3	2	16				100

Table 4.7 shows that the second mentions were predominantly retrieved with the topical case markers, *-un*/*-nun* (67:67%), unlike the first mentions. In particular, full names (48:48%); surnames plus titles (12:12%); zero anaphors (4:4%); pronouns (2:2%); lexical NP (1:1%). Next, the second mentions were retrieved with the genitive case markers, *-uy*

(16:16%). Other case markers such as the nominal case, *-i/-ka* (12:12%); the accusative case, *-ul/-lul* (3:3%); the genitive case, *-ekey* (2:2%) were observed.

4.1.2.3 The Third Mentions

Third, all structural constructions of discourse anaphors referring to public figures for the third mentions are shown in Table 4.8 and Figure 4.6 in terms of their respective frequencies and examples.

Table 4.8 The types of construction and the relative frequency for the third mentions in Korean

Type of Construction			Frequency (n (%))	Example
I. Full Name (FN)	a.	FN	71 (71%)	Park Ji-Sung (박지성)
	b.	FN + TT	2 (2%)	Ahn Chel-Soo kyoswu 'professor' (안철수 교수)
	c.	NN + FN	1 (1%)	kwukmin MC Yoo Jae-Seok 'national' (국민 MC 유재석)
	d.	TT + FN	1 (1%)	cakka Lee Sang-Eun 'writer' (작가 이상은)
II. Surname (SN)	a.	SN	0 (0%)	
	b.	SN + TT	15 (15%)	Choi cakka 'writer' (최 작가)
III. Pronoun (Pro)			5 (5%)	ku 's/he' (그)
IV. Given Name (GN)			0 (0%)	
V. Lexical NPs (LNPs)			1 (1%)	cakka 'writer' (작가)
VI. Zero (Ø)			4 (4%)	Ø
Total			100 (100%)	

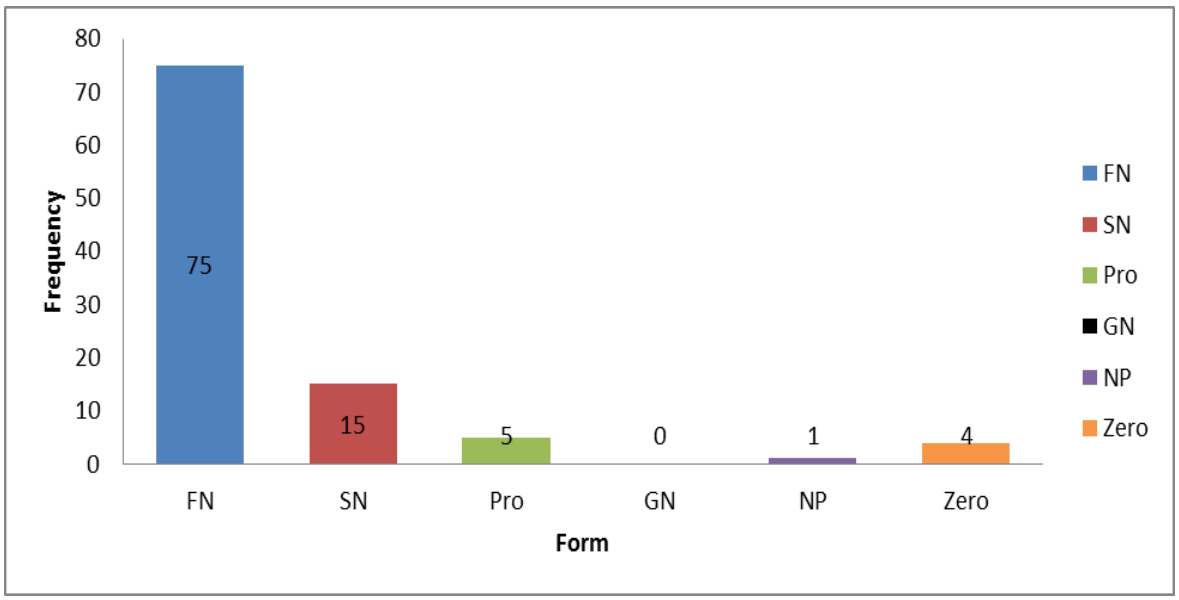


Figure 4.6 The relative frequency of discourse anaphors for the third mentions in Korean

Table 4.8 and Figure 4.6 show that 15 occurrences were retrieved with surnames plus title (15:15%), in which there was no occurrence of surname only. 5 occurrences were retrieved with pronouns (5:5%). 4 occurrences were retrieved with zero anaphors (4:4%). No occurrence with given names was observed.

Like the third mentions in English, these findings conform to Ariel (1991, 1994, 2008, 2010) and Lee (2010) in that a surname plus a(n) (abbreviated) title construction, pronoun, and zero anaphors (i.e., relatively higher accessible markers in Ariel's AH) rather than a full name (plus modifier) construction (i.e., a relatively lowest accessible marker) have been used far more frequently for the second retrievals.

However, like the second mentions in Korean, it is noteworthy that almost of discourse anaphors for the second mentions predominantly contain full names (75:75%), classified into four structural constructions in terms of being addressed with title and

nickname. This finding would definitely run counter to Lee's (2010) findings on English/Korean political news articles and Ariel's AH for English. In fact, in Lee's analysis, more than half of the third retrievals are referred to by pronouns and 40% by surname plus title constructions. According to Ariel's AH, a pronoun (i.e., a relatively higher accessible marker) rather than a surname plus title (i.e., a relatively lower accessible marker) should have been used far more frequently for the third retrievals.

However, surname plus title constructions and even pronouns were relatively lesser used than full name constructions for the third mentions in Korean (Kim 2010a, 2012).

In addition, the case markers of the types of construction for the third mentions in Korean were observed, as shown in Table 4.9 below.

Table 4.9 The relative frequency of the case markers for the third mentions in Korean

Type of Construction \ Frequency & Case marker			Frequency								Total
			NOM	TOP	ACC	DAT	GEN	LOC	ABL	INST	
I. FN	a.	FN	8	41	5	4	13				71
	b.	FN + TT		1			1				2
	c.	NN + FN			1						1
	d.	TT + FN			1						1
II. SN	a.	SN									
	b.	SN + TT	3	10			2				15
III. Pro			1	3	1						5
IV. GN											
V. LNPs				1							1
VI. Zero				4							4
Total			12	60	8	4	16				100

Table 4.9 shows that the third mentions were also predominantly retrieved with the topical case markers, *-un*/*-nun* (60:60%), like the second mentions. In particular, full names (42:42%); surnames plus titles (10:10%); zero anaphors (4:4%); pronouns (3:3%); lexical NP (1:1%). Next, the second mentions were also retrieved with the genitive case markers, *-uy* (16:16%). Other case markers such as the nominal case, *-i*/*-ka* (12:12%); the accusative case, *-ul*/*-lul* (8:8%); with the genitive case, *-ekey* (4:4%) were observed.

4.1.3 Section Summary

All types of discourse anaphors and their relative frequencies in English and Korean Corpus I (i.e., by order of mention) are summarized in Table 4.10 and Figure 4.7.

Table 4.10 Summary of Corpus I for English and Korean

Mention type	Language	Form						Total
		FN	SN	Pro	GN	LP	Zero	
1stM	English	**98**[34]	0	2	0	0	0	100
	Korean	**100**	0	0	0	0	0	100
2ndM	English	0	**33**	**64**	0	3	0	100
	Korean	**72**	**14**	9	0	1	4	100
3rdM	English	3	**31**	**60**	0	6	0	100
	Korean	**75**	**15**	5	0	1	4	100

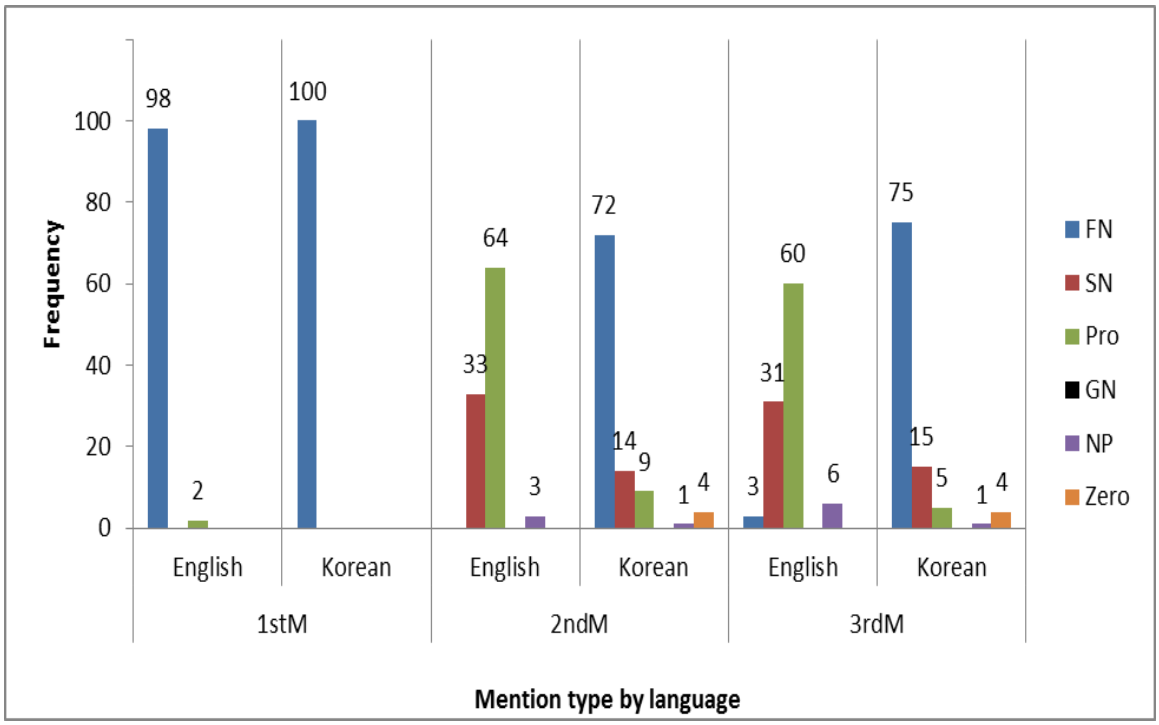


Figure 4.7 Summary of Corpus I for English and Korean

[34] The number in red color represents the most frequent form, while the number in blue color represents the next frequent form in the dissertation.

As seen in Table 4.10 and Figure 4.7, in Corpus I (discourse anaphoric patterns by order of mention), for the first mention, both the English and Korean corpus predominantly contain full names. For the second mention, the English corpus contains three types of constructions such as pronouns, surnames, and lexical NPs in order, while the Korean corpus predominantly contains full name constructions other than discourse anaphoric forms such as surnames, pronouns, zero anaphors, and lexical NPs. For the third mention, the English corpus contains four types of constructions such as pronouns, surnames, lexical NPs, and full names in order, while the Korean corpus predominantly contains full name constructions other than discourse anaphoric forms.

In addition, in order to see if there is a marked difference between English and Korean in terms of the distribution of discourse anaphoric forms depending on the order of mention of the referent (i.e., the first-, the second-, and the third mention), Pearson's Chi-square tests were run for the three discourse anaphoric forms, i.e., FN, SN, and Pro, that occur regularly across mention in two languages, as shown in Table 4.11.

Table 4.11 Summary of Chi-square test results for Corpus I for English and Korean

Mention Type	Language	Pearson's Chi-square test results	
		each mention type vs. three forms in each language	each mention type vs. three forms in both languages
1st M	English	df=1, χ^2=92.16, p<2.2e-16***[35] (significant)	NA
	Korean	NA	
2nd M	English	df=1, χ^2=9.9072, p=.001646** (significant)	df=2, χ^2=121.1115, p<2.2e-16*** (significant)
	Korean	df=2, χ^2=77.4526, p<2.2e-16*** (significant)	
3rd M	English	df=2, χ^2=51.8511, p<5.504e-12*** (significant)	df=2, χ^2=118.5632, p<2.2e-16*** (significant)
	Korean	df=2, χ^2=90.5263, p<2.2e-16*** (significant)	

As seen in Table 4.11, the Pearson's Chi-square tests determined that the three anaphoric forms (i.e., FN, SN, and Pro) in the second (English: df=1, χ^2=9.91, p<.01;Korean: df=2, χ^2=77.45, p<.001) and the third mention (English: df=2, χ^2=51.85, p<.001;Korean: df=2, χ^2=90.53, p<.001) in each language had a statistically significant difference, and that these three forms in the second (df=2, χ^2=121.11, p<.001) and the third mention (df=2, χ^2=118.56, p<.001) in both languages also had a statistically significant difference.

[35] Significance codes: 0;'***' 0.001; '**' 0.01; '*'0.05; '.'0.1'; ' ' 1

Thus, the results of Chi-square tests indicate that there is a marked difference in terms of the distribution of three forms depending on the order of mention of the referent, particularly in the second and the third mention in the English corpus, while the same is not true for the Korean corpus because of the use of FNs regardless of the order of mention.

4.2 Corpus II: Discourse Anaphoric Patterns by Paragraph

4.2.1 By Paragraph in English

The findings on all structural constructions of discourse anaphors referring to public figures in English news articles are correspondingly shown in Table 4.12 through Table 4.14 and Figure 4.8 and Figure 4.9 in terms of (the first-), the second-, and the third paragraphs.

4.2.1.1 The First Paragraphs

First, all structural constructions of referring expressions referring to public figures for the first paragraphs are basically the same as those for the first mentions in terms of their respective frequencies and examples in Table 4.1 and Figure 4.1 in Section 4.1.1.1 since Corpus I and II for English were analyzed from the same sources of news articles.

4.2.1.2 The Second Paragraphs

Second, all structural constructions of referring expressions referring to public figures for the second paragraphs are shown in Table 4.12 and Figure 4.8 in terms of their respective frequencies and examples.

Table 4.12 The types of construction and the relative frequency for the second paragraphs in English

Type of Construction			Frequency (n (%))	Example
I. Full Name (FN)	a.	FN	1 (1%)	Steve Jobs
	b.	FN + Apposition (AP)	0 (0%)	
	c.	(Pseudo-) Title (TT) + FN	0 (0%)	
	d.	Nickname (NN) + FN	0 (0%)	
II. Surname (SN)	a.	SN	68 (68%)	Gates
	b.	TT + SN	12 (12%)	Mr. Jobs
III. Pronoun (Pro)			15 (14%)	His (Him)
IV. Given Name (GN)			0 (0%)	
V. Lexical NPs (LNPs)			4 (4%)	
VI. Zero (Ø)			0 (0%)	
Total			100 (100%)	

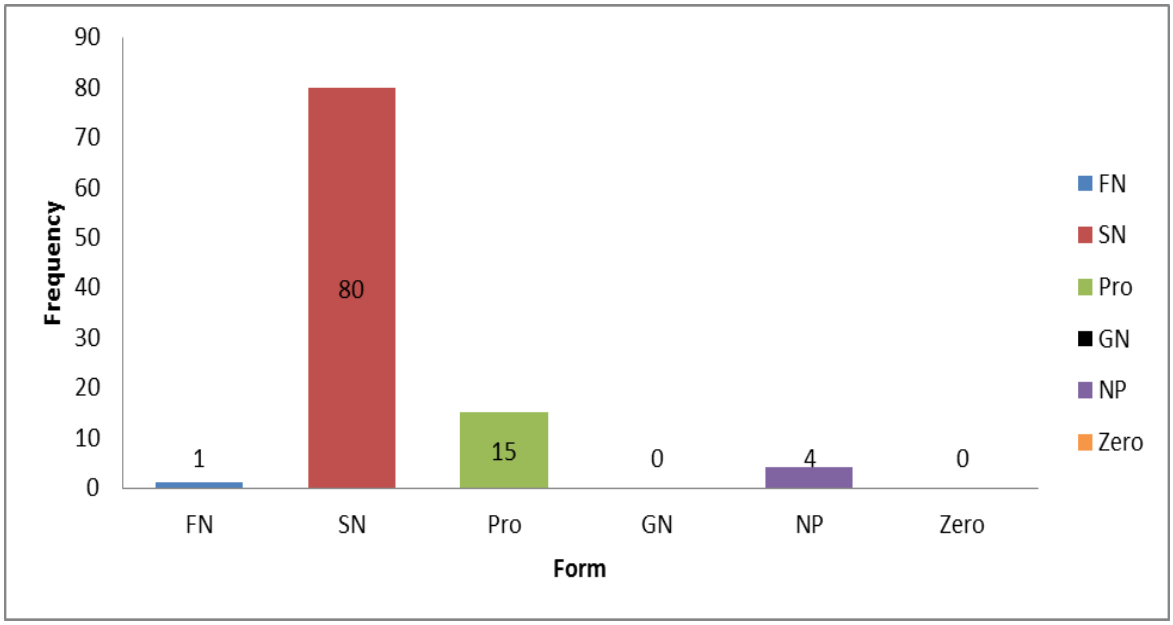


Figure 4.8 The relative frequency of discourse anaphors for the second paragraphs in English

Table 4.12 and Figure 4.8 show that the second paragraphs were predominantly retrieved with surnames (80:80%), in which 68 (68%) occurrences were retrieved with surname only, while 12 (12%) occurrences were retrieved with surname plus title. Next, pronouns were retrieved in 15 (15%) occurrences for the second paragraphs, which is noticeably fewer than pronouns for the second mentions (64:64%) in English. Only 4 (4%) occurrences were retrieved with lexical noun phrases, while only 1 (1%) occurrence was retrieved with full name for the second paragraphs. There was no occurrence of given name and zero anaphor.

This finding conforms to Ariel (1991, 1994, 2008, 2010) and Lee (2010) in that between surname and pronoun, surname (plus title) (i.e., a relatively lower accessible marker in the AH) is more appropriate for the second paragraph than pronoun (i.e., a relatively higher accessible marker) since the concept of paragraph is relatively farther than that of mention in discourse in terms of (intra-) sentential distance. Thus, it is likely that for the second mentions, pronouns are more appropriate than surnames (plus titles), while for the second paragraphs, surnames (plus titles) are more appropriate than pronouns according to the AH.

4.2.1.3 The Third Paragraphs

Third, all structural constructions of referring expressions referring to public figures for the third paragraphs are shown in Table 4.13 and Figure 4.9 in terms of their respective frequencies and examples.

Table 4.13 The types of construction and the relative frequency for the third paragraphs in English

Type of Construction			Frequency (n (%))	Example
I. Full Name (FN)	a.	FN	3 (3%)	Bill Gates
	b.	FN + Apposition (AP)	0 (0%)	
	c.	(Pseudo-) Title (TT) + FN	0 (0%)	
	d.	Nickname (NN) + FN	0 (0%)	
II. Surname (SN)	a.	SN	61 (61%)	Jeter
	b.	TT + SN	15 (15%)	Ms. Jones
III. Pronoun (Pro)			19 (19%)	His (Him)
IV. Given Name (GN)			0 (0%)	
V. Lexical NPs (LNPs)			2 (2%)	
VI. Zero (Ø)			0 (0%)	
Total			100 (100%)	

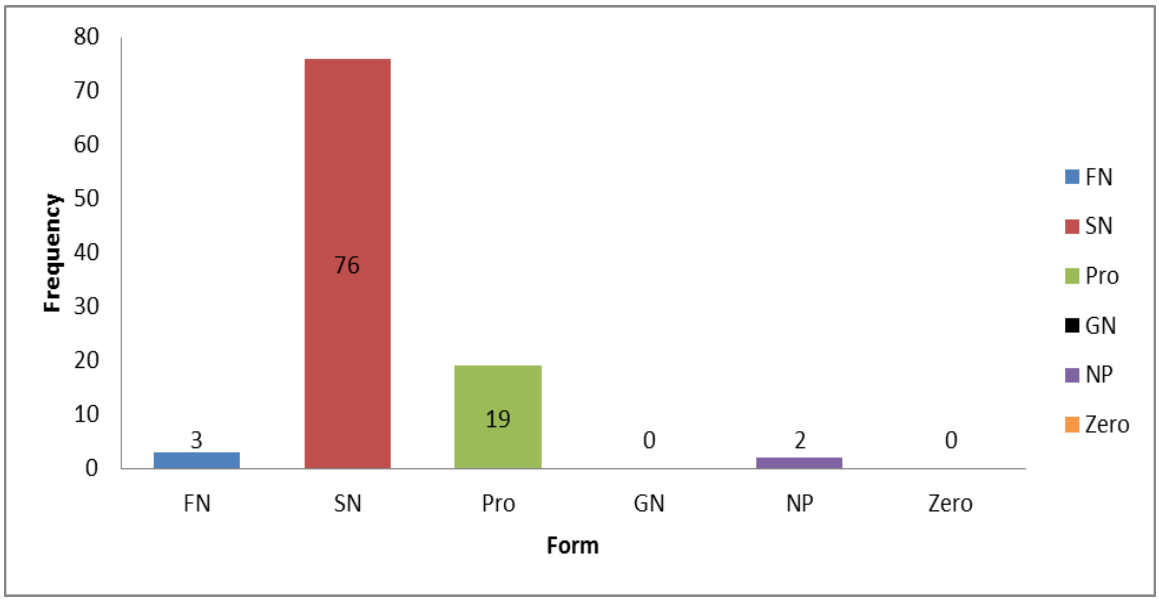


Figure 4.9 The relative frequency of discourse anaphors for the third paragraphs in English

Table 4.13 and Figure 4.9 show that the second paragraphs were predominantly retrieved with the surnames (76:76%), in which 61 (61%) occurrences were retrieved with surname only, while 15 (15%) occurrences were retrieved with surname plus title. Next, pronouns were retrieved in 19 (19%) occurrences for the second paragraphs, which is also noticeably fewer than pronouns for the third mentions (60:60%) in English. Only 3 (3%) occurrences were retrieved with full names, while only 2 (2%) occurrence was retrieved with lexical noun phrases for the second paragraphs. There was no occurrence of given name and zero anaphor.

Like the second paragraphs, this finding also conforms to Ariel (1991, 1994, 2008, 2010) and Lee (2010) in that between surname and pronoun, surname (plus title) (i.e., a relatively lower accessible marker in the AH) is more appropriate for the second paragraph than pronoun (i.e., a relatively higher accessible marker) since the concept of paragraph is relatively farther than that of mention in discourse in terms of (intra-) sentential distance. Thus, it is likely that for the second mentions, pronouns are more appropriate than surnames (plus titles), while for the second paragraphs, surnames (plus titles) are more appropriate than pronouns according to the AH.

4.2.2 By Paragraph in Korean

The findings on all structural constructions of discourse anaphors referring to public figures in English news articles are correspondingly shown in Table 4.14 through Table 4.17 and Figure 4.10 and Figure 4.11 in terms of (the first-), the second-, and the third paragraphs.

4.2.2.1 The First Paragraphs

First, all structural constructions of referring expressions referring to public figures for the first paragraphs are basically the same as those for the first mentions in terms of their respective frequencies, examples, and case markers in Table 4.4, Table 4.5, and Figure 4.4 in Section 4.1.2.1 since Corpus I and II for Korean were analyzed from the same sources of news articles.

4.2.2.2 The Second Paragraphs

Second, all structural constructions of referring expressions referring to public figures for the second paragraphs are shown in Table 4.14 and Figure 4.10 in terms of their respective frequencies and examples.

Table 4.14 The types of construction and the relative frequency for the second paragraphs in Korean

Type of Construction			Frequency (n (%))	Example
I. Full Name (FN)	a.	FN	73 (73%)	Yoo Jae-Seok (유재석)
	b.	FN + TT	5 (5%)	Ahn Chel-Soo wencang 'dean' (안철수 원장)
	c.	NN + FN	2 (2%)	phikye yewang Kim Yu-Na 'Figure Queen' (피겨여왕 김연아)
	d.	FN + AFF	1 (1%)	Park Chan Ho (Pittsburgh) (박찬호 (피츠버그))
	e.	TT + FN	1 (1%)	kaswu Lee Sang-Eun 'singer' (가수 이상은)
II. Surname (SN)	a.	SN	0 (0%)	
	b.	SN + TT	12 (12%)	Shin cakka 'writer' (신 작가)
III. Pronoun (Pro)			5 (5%)	ku 's/he' (그)
IV. Given Name (GN)			0 (0%)	
V. Lexical NPs (LNPs)			1 (1%)	cakka 'writer' (작가)
VI. Zero (Ø)			0 (0%)	
Total			100 (100%)	

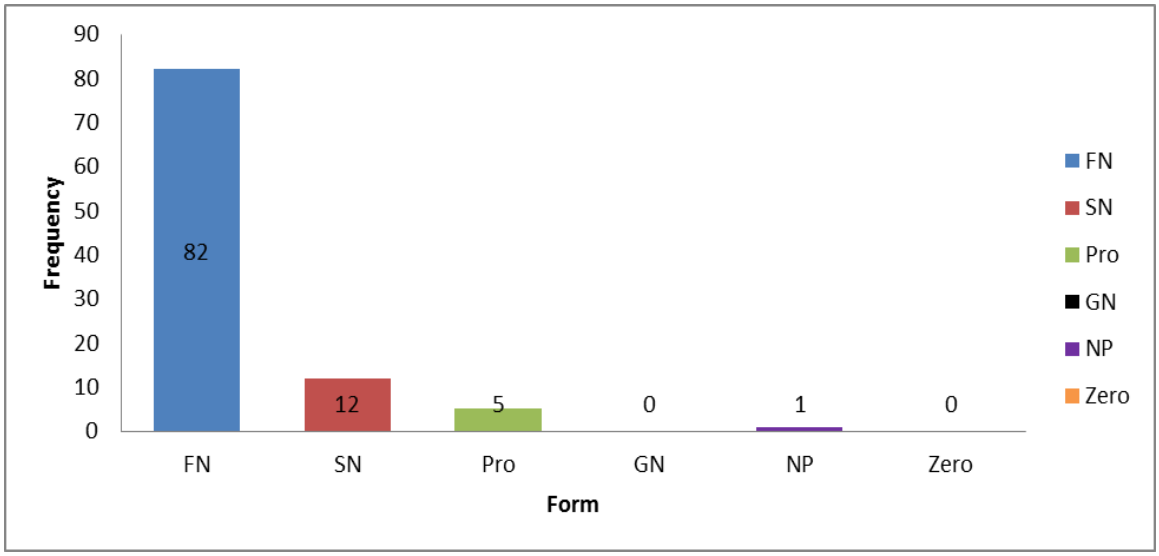


Figure 4.10 The relative frequency of discourse anaphors for the second paragraphs in Korean

Table 4.14 and Figure 4.10 show that 12 occurrences were retrieved with surnames plus title (12:12%), in which there was no occurrence of surname only. 9 occurrences were retrieved with pronouns (5:5%). Only 1 occurrence was retrieved with lexical noun phrase (1:1%). No occurrence with given name and zero anaphor were observed.

Like the second paragraphs in English, this finding seems conform to Ariel (1991, 1994, 2008, 2010) and Lee (2010) in that between surname and pronoun, surname (plus title) (i.e., a relatively lower accessible marker in the AH) is more appropriate for the second paragraph than pronoun (i.e., a relatively higher accessible marker) since the concept of paragraph is relatively farther than that of mention in discourse in terms of (intra-) sentential distance. Thus, it is likely that for the second paragraphs, surnames (plus titles) are more appropriate than pronouns according to the AH.

However, it is noteworthy that like the second mentions in Korean, discourse anaphors for the second paragraphs predominantly contain full names (82:82%), classified into five structural constructions in terms of being addressed with title, affiliation and nickname. This finding would definitely run counter to Lee's (2010) findings on English/Korean political news articles and Ariel's AH for English mentioned above (Kim 2010a, 2012).

In addition, the case markers of the types of construction for the second paragraphs in Korean were observed, as shown in Table 4.15 below.

Table 4.15 The relative frequency of the case markers for the second paragraphs in Korean

Frequency & Case marker / Type of Construction			Frequency								Total
			NOM	TOP	ACC	DAT	GEN	LOC	ABL	INST	
I. FN	a.	FN	7	52	1	2	11				73
	b.	FN + TT		4			1				5
	c.	NN + FN	1		1						2
	d.	FN + AFF	1								1
	e.	TT + FN	1								1
II. SN	a.	SN									
	b.	SN + TT		12							12
III. Pro				4	1						5
IV. GN											
V. LNPs				1							1
VI. Zero											
Total			10	73	3	2	12				100

Table 4.15 shows that like the second mentions, the second paragraphs were predominantly retrieved with the topical case markers, *-un*/*-nun* (73:73%), unlike the first mention/paragraphs. In particular, full names (56:56%); surnames plus titles (12:12%); pronouns (4:4%); lexical NP (1:1%).

Next, the second mentions were retrieved with the genitive case markers, *-uy* (12:12%). Other case markers such as the nominal case, *-i*/*-ka* (10:10%); the accusative case, *-ul*/*-lul* (3:3%); the genitive case, *-ekey* (2:2%) were observed.

4.2.2.3 The Third Paragraphs

Third, all structural constructions of referring expressions referring to public figures for the third paragraphs are shown in Table 4.16 and Figure 4.11 in terms of their respective frequencies and examples.

Table 4.16 The types of construction and the relative frequency for the third paragraphs in Korean

Type of Construction			Frequency (n (%))	Example
I. Full Name (FN)	a.	FN	71 (71%)	Park Ji-Sung (박지성)
	b.	FN + TT	2 (2%)	Ahn Chel-Soo kyoswu 'professor' (안철수 교수)
	c.	NN + FN	1 (1%)	meyttwuki MC Yoo Jae-Seok 'grasshopper' (메뚜기유재석)
II. Surname (SN)	a.	SN	0 (0%)	
	b.	SN + TT	16 (16%)	Choi cakka 'writer' (최 작가)
III. Pronoun (Pro)			8 (8%)	ku 's/he' (그)
IV. Given Name (GN)			0 (0%)	
V. Lexical NPs (LNPs)			2 (2%)	cakka 'writer' (작가)
VI. Zero (Ø)			0 (0%)	
Total			100 (100%)	

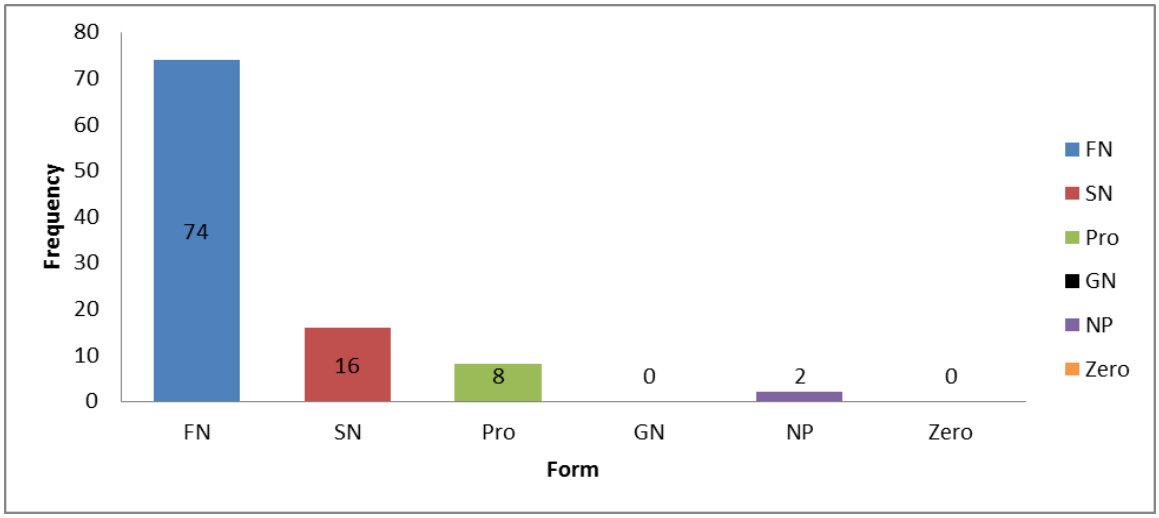


Figure 4.11 The relative frequency of discourse anaphors for the third paragraphs in Korean

Table 4.16 and Figure 4.11 show that 16 occurrences were retrieved with surnames plus title (12:12%), in which there was no occurrence of surname only. 8 occurrences were retrieved with pronouns (8:8%). Only 2 occurrences were retrieved with lexical noun phrase (2:2%). No occurrence with given name and zero anaphor were observed.

Like the third paragraphs in English, this finding seems conform to Ariel (1991, 1994, 2008, 2010) and Lee (2010) in that between surname and pronoun, surname (plus title) (i.e., a relatively lower accessible marker in the AH) is more appropriate for the third paragraphs than pronoun (i.e., a relatively higher accessible marker) since the concept of paragraph is relatively farther than that of mention in discourse in terms of (intra-) sentential distance. Thus, it is likely that for the second paragraphs, surnames (plus titles) are more appropriate than pronouns according to the AH.

However, it is noteworthy that like the second/third mentions and the second paragraphs in Korean, almost discourse anaphors for the third paragraphs predominantly

contain full names (74:74%), classified into five structural constructions in terms of being addressed with title, affiliation and nickname. This finding would definitely run counter to Lee's (2010) findings on English/Korean political news articles and Ariel's AH for English mentioned above (Kim 2010a, 2012).

In addition, the case markers of the types of construction for the third paragraphs in Korean were observed, as shown in Table 4.17 below.

Table 4.17 The relative frequency of the case markers for the second paragraphs in Korean

Frequency & Case marker / Type of Construction			Frequency								Total
			NOM	TOP	ACC	DAT	GEN	LOC	ABL	INST	
I. FN	a.	FN	5	47	3	3	13				71
	b.	FN + TT		1			1				2
	c.	NN + FN	1								1
II. SN	a.	SN									
	b.	SN + TT	4	10			2				16
III. Pro			2	5		1					8
IV. GN											
V. LNPs				1			1				2
VI. Zero											
Total			12	64	3	4	17				100

Table 4.17 shows that like the second/third mentions and the second paragraphs, the third paragraphs were predominantly retrieved with the topical case markers, *-un*/*-nun* (64:64%), unlike the first mention/paragraphs. In particular, full names (48:48%);

surnames plus titles (10:10%); pronouns (5:5%); lexical NP (1:1%). Next, the second mentions were retrieved with the genitive case markers, *-uy* (17:17%). Other case markers such as the nominal case, *-i*/*-ka* (12:12%); the accusative case, *-ul*/*-lul* (3:3%); the genitive case, *-ekey* (4:4%) were observed.

4.2.3 Section Summary

All types of discourse anaphors and their relative frequencies of English and Korean Corpus II (i.e., by paragraph) are summarized in Table 4.18 and Figure 4.12.

Table 4.18 Summary of Corpus II for English and Korean

Mention Type	Language	Form						Total
		FN	SN	Pro	GN	LP	Zero	
1stP	English	**98**[36]	0	2	0	0	0	100
	Korean	**100**	0	0	0	0	0	100
2ndP	English	1	**80**	**15**	0	4	0	100
	Korean	**82**	**12**	5	0	1	0	100
3rdP	English	3	**76**	**19**	0	2	0	100
	Korean	**74**	**16**	8	0	2	0	100

[36] The number in red color represents the most frequent form, while the number in blue color represents the next frequent form in the dissertation.

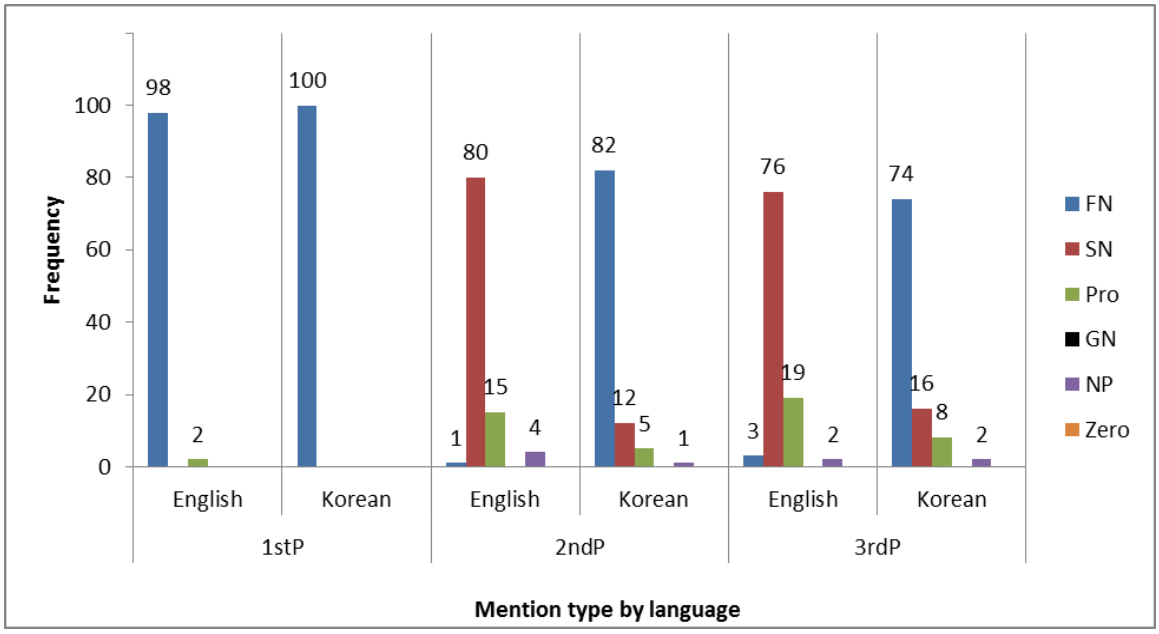


Figure 4.12 Summary of Corpus II for English and Korean

As seen in Table 4.18 and Figure 4.12, in Corpus II (discourse anaphoric patterns by paragraph), for the first paragraph, both the English and Korean corpus predominantly contain full names. For the second paragraph, the English corpus contains three types of constructions such as surnames, pronouns, and lexical NPs in order, while the Korean corpus predominantly contains full name constructions other than discourse anaphoric forms such as surnames, pronouns, zero anaphors, and lexical NPs. For the third paragraph, the English corpus contains four types of constructions such as surnames, pronouns, full names, and lexical NPs in order, while the Korean corpus predominantly contains full name constructions other than discourse anaphoric forms.

In addition, in order to see if there is a marked difference between English and Korean in terms of the distribution of discourse anaphoric forms depending on the

placement of the paragraph (i.e., the first-, the second-, and the third paragraph), Pearson's Chi-square tests were run for the three discourse anaphoric forms, i.e., FN, SN, and Pro, that occur regularly across mention in two languages, as shown in Table 4.19.

Table 4.19 Summary of Chi-square test results for Corpus II for English and Korean

Mention Type	Language	Pearson's Chi-square test results	
		each mention type vs. three forms in each language	each mention type vs. three forms in both languages
1stP	English	df=1, x^2=92.16, p<2.2e-16***[37] (significant)	NA
	Korean	NA	
2ndP	English	df=2, x^2=111.0625, p<2.2e-16*** (significant)	df=2, x^2=134.2947, p<2.2e-16*** (significant)
	Korean	df=2, x^2=109.8788, p<2.2e-16*** (significant)	
3rdP	English	df=2, x^2=90.1429, p<2.2e-16*** (significant)	df=2, x^2=109.0794, p<2.2e-16*** (significant)
	Korean	df=2, x^2=79.4286, p<2.2e-16*** (significant)	

As seen in Table 4.19, the Pearson's Chi-square tests determined that the three anaphoric forms (i.e., FN, SN, and Pro) in the second (English: df=2, x^2=111.06, p<.001; Korean: df=2, x^2=109.88, p<.001) and the third paragraph (English: df=2, x^2=90.14, p<.001; Korean: df=2, x^2=79.43, p<.001) in each language had a statistically significant

[37] Significance codes: 0;'***' 0.001; '**' 0.01; '*'0.05; '.'0.1'; ' ' 1

difference, and that these three forms in the second (df=2, x^2=134.29, p<.001) and the third paragraph (df=2, x^2=109.08, p<.001) in both languages also had a statistically significant difference.

Thus, the results of Chi-square tests indicate that there is a marked difference in terms of the distribution of the three forms depending on the placement of the paragraph, particularly in the second and the third paragraph in the English corpus, while the same is not true for the Korean corpus because of the use of FNs regardless of placement of the paragraph.

4.3 Corpus I and II: Distance between Discourse Anaphors

The distance between each referring expression and its nearest coreferential expression, i.e., the distance between discourse anaphors, was examined in order to see the correspondence between referring expressions and their pragmatic/cognitive statuses as accessibility markers, based on Ariel's AH and Lee (2010) analysis. Except for all first mentions in all news articles in both languages, the distances between all discourse anaphors and their nearest previous coreferential expressions (i.e., the distance between all retrievals) were examined in terms of the sentential level. The analysis results are summarized in Table 4.20 and Figure 4.13 for English and in Table 4.21 and Figure 4.14 for Korean.

4.3.1 Distance between Discourse Anaphors in English

First, the distances between discourse anaphors, i.e., other retrievals and their nearest coreferential items in the English news articles are shown in Table 4.20 and Figure 4.13 in terms of the sentence level.

Table 4.20 Distance between discourse anaphors in English

Frequency & Distance / DA Form	Same sentence	Previous sentence	2 Sentences away	3 Sentences away	More than 4 sentences away	Total
1. FN	5 (38.4%) (1: FN-FN) (4: Pro-FN)	3 (23.1%)	1 (7.7%)	3 (23.1%)	1 (7.7%)	13 (100%)
2. SN	32 (6.7%)	**363**[38] **(75.4%)**	**66** (13.7%)	11 (2.3%)	9 (1.9%)	481 (100%)
3. Pro	**451** **(57.3%)** (33: FN-Pro)	**332** (42.2%)	3 (0.4%)	1 (0.1%)		787 (100%)
4. GN		1 (100%)				1 (100%)
5. LNPs	3 (17.6%) (2: SN-NP) (1: Pro-NP)	14 (82.4%)				17 (100%)
6. Zero						0

[38] The number in red color represents the most frequent position, while the number in blue color represents the next frequent position.

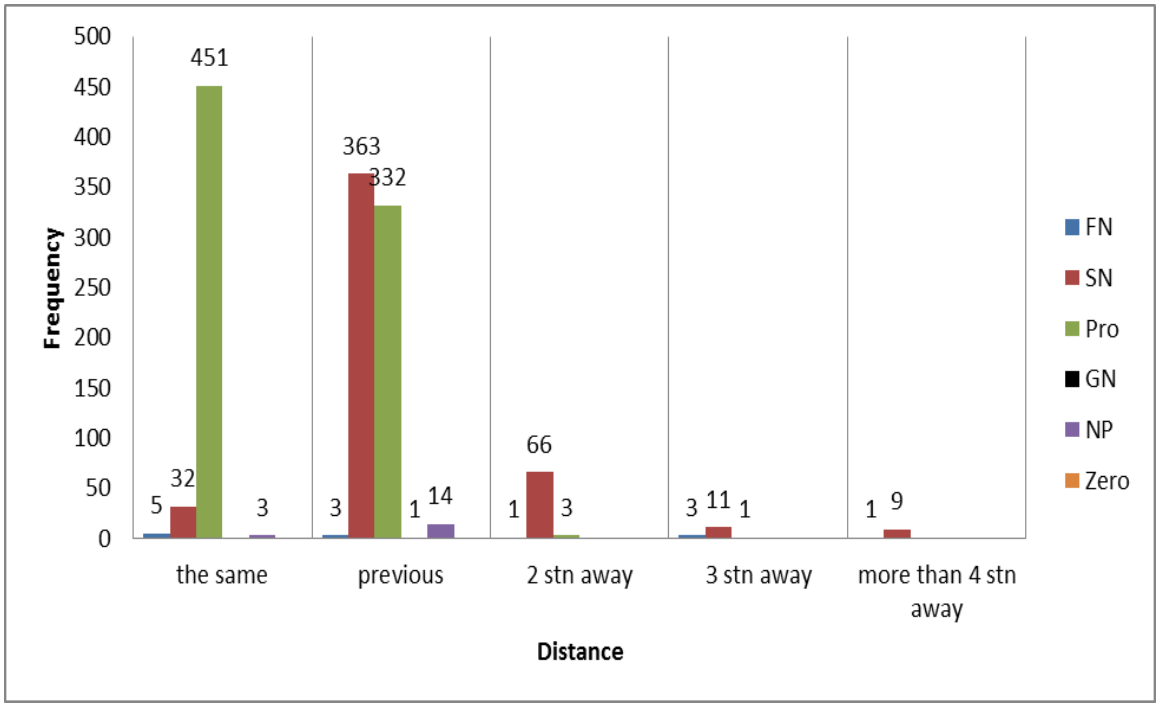


Figure 4.13 Distance between discourse anaphors in English

As seen in Table 4.20 and Figure 4.13, first, in terms of retrievals and the distance, full names were retrieved with previous pronouns (4 occurrences) and another full name (1 occurrence) in the same sentences (5:38.4%); in the previous sentences (3:23.1%); in 2 sentences away (1:7.7%); in 3 sentences away (3:23.1%); in more than 4 sentences away (1:7.7%). Next, surname constructions were retrieved very frequently in the previous sentences (363:75.4%) and also retrieved in 2 sentences away (66:13.7%); in the same sentences (32:6.7%); in 3 sentences away (11:2.3%); in more than 4 sentences away (9:1.9%). Then, pronouns were retrieved very frequently both in the same sentences (451:57.3%) including retrievals for previous full names (33 occurrences) and in the previous sentences (332:42.2%), and also retrieved in 2 sentences away (3:0.4%); in 3

sentences away (1:0.1%); however, no occurrence was observed in more than 4 sentences away. Lexical noun phrases were retrieved in the previous sentences (14:82.4%) and in the same sentences (3:17.6%); however, no occurrence was observed in other cases. Only one occurrence of given name, for example, Michael for Michael Jordan, was observed in the previous sentence. No retrieval with zero anaphor was observed in English articles.

In addition, in terms of distribution, full names and surnames show the widest distribution from the same sentence to more than 4 sentences away, but surnames were far more frequent than full names in terms of frequency. Compared with surnames, pronouns show relatively narrow distribution, very frequently retrieved both in the same sentences and in the previous sentence, although they were scarcely retrieved in the 2 sentences away and in the 3 sentences away. Lexical noun phrases with few frequencies also show narrow distribution, rarely retrieved in the same sentences and in the previous sentences. These findings conform to Ariel's AH and Lee's (2010)[39] in that surname constructions, frequently retrieved in the previous sentences, show the widest distribution from the same sentence to more than 4 sentences away, while pronouns were far frequently retrieved in the same sentences first, and then in the previous sentences.

4.3.2 Distance between Discourse Anaphors in Korean

Second, the distances between discourse anaphors, i.e., other retrievals and their nearest coreferential items in Korean news articles are shown in Table 4.21 and Figure 4.14 in terms of the sentence level.

[39] In fact, in Lee's (2010:2513-2516) analysis, no full name construction was retrieved in the same sentence in English and Korean political news articles.

Table 4.21 Distance between discourse anaphors in Korean

Frequency & Distance / DA Form	Same sentence	Previous sentence	2 Sentences away	3 Sentences away	More than 4 sentences away	Total
1. FN	37 (5.4%) (33: FN-FN) (4: Pro-FN)	**497** **(72.6%)**	**86** (12.5%)	45 (6.6%)	20 (2.9%)	685 (100%)
2. SN	10 (7%) (1:FN-SN) (9:SN-SN)	**95** **(65.5%)**	**26** (17.9%)	8 (5.5%)	6 (4.1%)	145 (100%)
3. Pro	**28** (20.9%) (22: FN-Pro) (1:SN-Pro) (4:Pro-Pro) (1:Pro-LNP)	**97** **(72.4%)**	9 (6.7%)			134 (100%)
4. GN						0
5. LNPs	2 (10.5%)	14 (73.7%)	3 (15.8%)			19 (100%)
6. Zero		**117** **(100%)**				117 (100%)

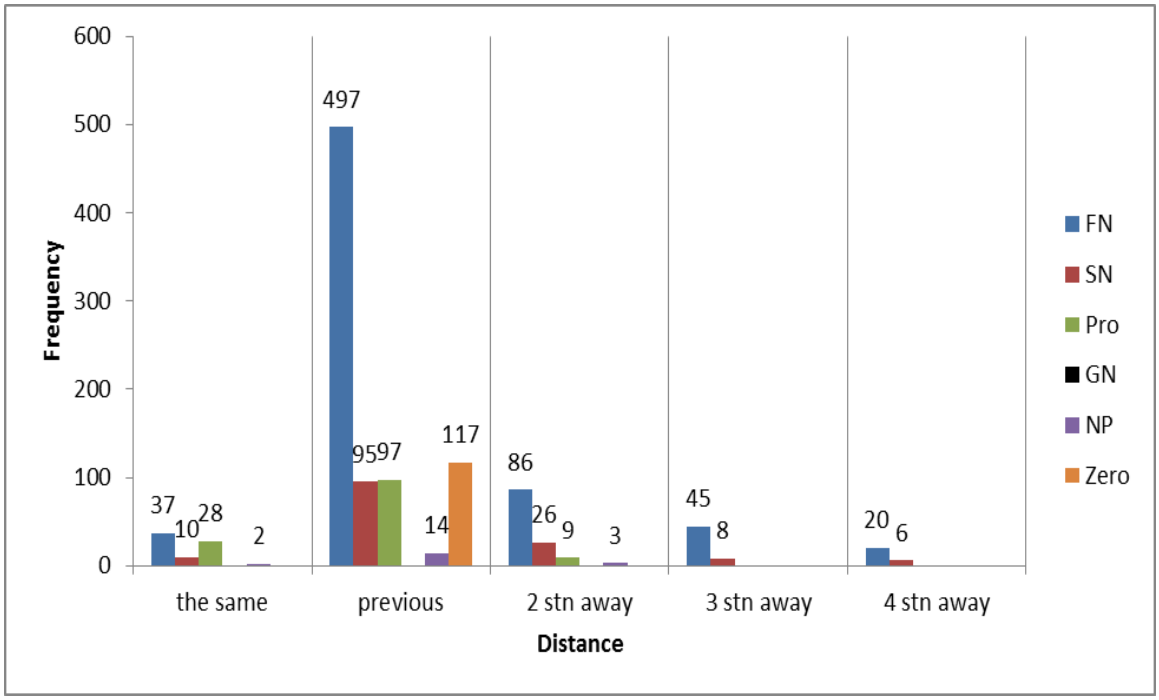


Figure 4.14 Distance between discourse anaphors in Korean

As seen in Table 4.21 and Figure 4.14, first, in terms of retrievals and the distance, it is noteworthy that full names were retrieved very frequently in the previous sentences (497:72.6%); in the 2 sentences away (86:12.5%); in the 3 sentences away (45:6.6%); in the same sentences (37:5.4%) with another full names (33 occurrences) and previous pronouns (4 occurrences); in more than 4 sentences away (20:2.9%). Next, surname constructions were retrieved frequently in the previous sentences (95:65.5%) and also retrieved in 2 sentences away (26:17.6%); in the same sentences (10:7%); in 3 sentences away (8:5.5%); in more than 4 sentences away (6:4.1%). Then, pronouns were also retrieved frequently both in the previous sentences (97:72.4%), which is different from those in English, and they retrieved in the same sentences (28:20.9%) and in the 2

sentences away (9:6.7%). Note that the frequencies of surnames and pronouns were relatively fewer than those in English; instead, the frequencies of full names were definitely very frequent in Korean. Also note that zero anaphors were retrieved frequently in the previous sentences (117:100%) in Korean, frequently functioning as surnames and pronouns in English. Lexical noun phrases were retrieved in the previous sentences (14:73.7%); in 3 sentences away (3:15.8%); in the same sentences (2:10.5%). No retrieval with given name was observed in Korean articles.

In addition, in terms of distribution, full names and surnames show the widest distribution from the same sentence to more than 4 sentences away like the distribution in English. However, it is noteworthy that full names (685 occurrences) were far more frequently retrieved than surnames (145 occurrences) in terms of frequency. Both surnames and pronouns also show similar distribution as those in English; in fact, surnames (95:65.5%) and pronouns (97:72.4%) were retrieved frequently in the previous sentences; however, the frequencies were relatively fewer than those in English. Instead, zero anaphors in Korean, frequently retrieved in the previous sentences, show narrow distribution in replacement of surnames and pronouns in English. Lexical noun phrases with few frequencies relatively wider distribution those in English in the previous sentence, in 2 sentences away and in the same sentences. Note that these findings would run counter to Ariel's AH and Lee's (2010), in which surname constructions, frequently retrieved in the previous sentences, show the widest distribution from the same sentence to more than 4 sentences away, while pronouns were far frequently retrieved in the same sentences first, and then in the previous sentences.

4.4 Corpus III: Sentential Anaphors

With the results by AntConc 3.3.5w corpus tool, I analyzed sentential anaphors *that* and *it* in English, and *ku-ken* 'that' and *ku-len* 'it' in Korean in terms of the givenness of information (4.4.1) (i.e., new vs. old information) for English (4.4.1.1) and Korean (4.4.1.2) and the width of reference (4.4.2) (i.e., narrow vs. wide reference) for English (4.4.2.1) and Korean (4.4.2.2).

4.4.1 New vs. Old Information

4.4.1.1 In English

First, the findings on all sentential anaphors in English TV drama scripts are shown in Table 4.22 and Table 4.23 and in terms of the givenness (i.e., new vs. old information).

For English, I gained the result of nominal vs. sentential anaphor *that* and *it* by hitting *that* and *it* in AntConc 3.3.5w oncordance, as shown in Table 4.22.

Table 4.22 The relative frequency of Corpus III for English

Frequency & type / Form	Frequency			Total
	Nominal	**Sentential**	Other	
that	145 (34.5%)	**68 (16.2%)**	207 (49.3%)	420 (100%)
it	271 (49.5%)	**32 (5.8%)**	245 (44.7%)	548 (100%)

As seen in Table 4.22, I gained 145 hits (34.5%) of the nominal anaphor *that* and 68 hits (16.2%) of the sentential anaphor *that*. On the other hand, I gained 271 hits (49.5%) of the nominal anaphor *it* and 32 hits (5.8%) of the sentential anaphor *it*. Out of these

results, I analyzed the sentential anaphors by the givenness of information (i.e., new vs. old information) and by the width of reference (i.e., narrow vs. wide reference) with concordances, concordance plots and file views in AntConc 3.3.5w.

The relative frequencies of the sentential anaphor *that* for new information and those of the sentential anaphor *it* for old information are shown in Table 4.23.

Table 4.23 The relative frequency of new (*that*) vs. old (*it*) information in English

Frequency & forms / Condition		Frequency (n (%))	
		English sentential anaphors (total n of each form)	
		that (68)[40]	*it* (32)
Givenness	New	41 (60.3%)	0
	Old	0	23 (71.9%)

As seen in Table 4.23, I gained 41 occurrences (60.3% out of 68 hits) of the sentential anaphor *that* for new information, while I gained 23 occurrences (71.9% out of 32 hits) of the sentential anaphor *it* for old information in English. There was no occurrence of *that* for old information and of *it* for new information.

In a Pearson's Chi-square test, the given probabilities for the relative frequencies between two categorical variables: each condition in the givenness and the form of the sentential anaphor (i.e., new/old and *that/it*) were statistically significant (df=3, x^2=74.13, p<.001***).

[40] The rest of the total number of *that*, 21, is the relative frequency for narrow reference of *that*, while the rest of the total number of *it*, 9, is the relative frequency for wide reference of *it*. See Section 4.4.2.

4.4.1.2 In Korean

Second, the findings on all sentential anaphors in Korean TV drama scripts are shown in Table 4.24 and Table 4.25 and in terms of the givenness (i.e., new vs. old information).

For Korean, I gained the result of nominal vs. sentential anaphor *ku-ken* 'that' and *ku-len* 'it' by hitting *ku-ken* and *ku-len*[41] in AntConc 3.3.5w oncordance, as shown in Table 4.24.

Table 4.24 The relative frequency of Corpus III for Korean

Frequency & type / Form	Frequency			Total
	Nominal	**Sentential**	Other	
ku-ken 'that'	24 (40.7%)	**33 (55.9%)**	2 (3.4%)	59 (100%)
ku-len 'it'	4 (2.8%)	**28 (19.7%)**	110 (77.5%)	142 (100%)

As seen in Table 4.24, I gained 24 hits (40.7%) of the nominal anaphor *ku-ken* 'that' and 33 hits (55.9%) of the sentential anaphor *ku-ken* 'that'. On the other hand, I gained 4 hits (2.8%) of the nominal anaphor *ku-len* 'it' and 28 hits (19.7%) of the sentential anaphor *ku-len* 'it'.

Out of these results, I analyzed the sentential anaphors by the givenness of information (i.e., new vs. old information) and by the width of reference (i.e., narrow vs. wide reference) with concordances, concordance plots and file views in AntConc 3.3.5w.

[41] In fact, the representative base form of the sentential anaphor *ku-ken* 'that' includes the hitting of *ku-ke*, *ku-kes*, and *ku-ken* 'that', i.e., *ku-ke(s/n)*, while that of *ku-len* 'it' includes the hitting of *ku-len*, *ku-len-ke*, *ku-len-kes*, and *ku-len-ken* 'it', i.e., *ku-len(ke/kes/ken)* in AntConc.

The relative frequencies of the sentential anaphor *ku-ken* 'that' for new information and those of the sentential anaphor *ku-len* 'it' for old information are shown in Table 4.25.

Table 4.25 The relative frequency of new (*ku-ken*) vs. old (*ku-len*) information in Korean

Condition \ Frequency & forms		Frequency (n (%))	
		Korean sentential anaphors (total n of each form)	
		ku-ken'that' (33)	*ku-len*'it' (28)
Givenness	New	12 (36.4%)	0
	Old	0	11 (39.3%)

As seen in Table 4.25, I gained 12 occurrences (36.4% out of 33 hits) of the sentential anaphor *kuken* 'that' for new information, while I gained 11 occurrences (39.3% out of 28 hits) of the sentential anaphor *kulen* 'it' for old information in Korean. There was no occurrence of *kuken* 'that' for old information and of *kulen* 'it' for new information.

In a Pearson's Chi-square test, the given probabilities for the relative frequencies between two categorical variables: each condition in the givenness and the form of the sentential anaphor (i.e., new/old and *kuken/kulen*) were statistically significant (df=3, χ^2=23.09, p<.001***).

4.4.1.3 Section Summary

The relative frequencies of the sentential anaphors in English and Korean in terms of the givenness of information (i.e., new vs. old information) and the relevant Chi-square test results are summarized in Table 4.26, Figure 4.15, and Table 4.27.

Table 4.26 The summary of Corpus III (the givenness of information) in English and Korean

Condition \ Frequency & forms		Frequency (n (%))			
		Sentential anaphors (total n of each form)			
		English		Korean	
		that (68)	*it* (32)	*ku-ken*'that' (33)	*ku-len*'it' (28)
Givenness	New	41 (60.3%)	0	12 (36.4%)	0
	Old	0	23 (71.9%)	0	11 (39.3%)

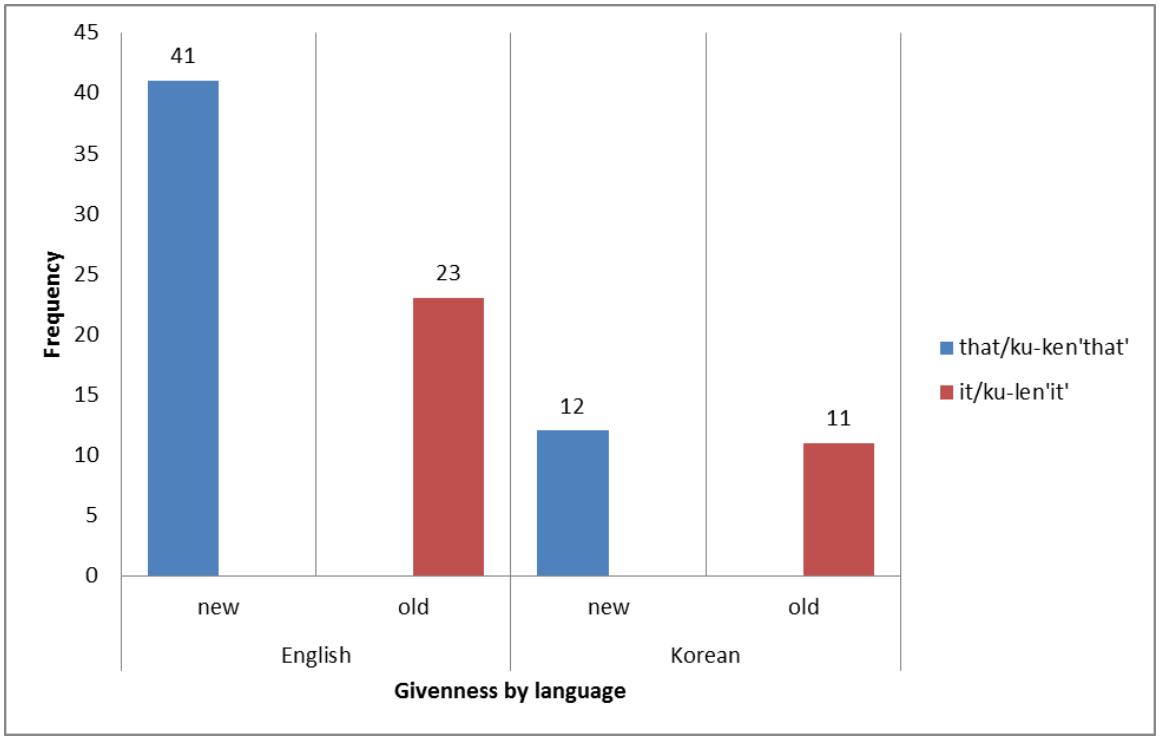


Figure 4.15 The summary of Corpus III (the givenness of information) in English and Korean

As seen in Table 4.26 and Figure 4.15, for Corpus III (the givenness of information), new information is referred to by *that* in English and *ku-ken* 'that' in Korean, whereas old information is referred to by *it* in English and *ku-len* 'it' in Korean.

Table 4.27 The summary of the Chi-square test results of Corpus III (the givenness of information) in English and Korean

Language & chi-square test / Condition	Language	Pearson's Chi-square test results
		each condition vs. all forms in each language
Givenness (new vs. old)	English	df=3, χ^2=74.13, p<.001*** (significant)
	Korean	df=3, χ^2=23.09, p<.001*** (significant)

In the Chi-square test for Corpus III, the given probabilities for the frequencies between two categorical variables (i.e., the givenness of information and the discourse anaphoric forms) were all statistically significant, which indicates that there is a marked difference between English and Korean in terms of the givenness of information.

4.4.2 Narrow vs. Wide Reference

4.4.2.1 In English

First, the findings on all sentential anaphors in English TV drama scripts are shown in Table 4.33 in terms of the width of reference (i.e., narrow vs. wide reference).

The relative frequencies of the sentential anaphor *that* for narrow reference and those of the sentential anaphor *it* for wide reference are shown in Table 4.28.

Table 4.28 The relative frequency of narrow (*that*) vs. wide (*it*) reference in English

Condition \ Frequency & forms		Frequency (n (%))	
		English sentential anaphors (total n of each form)	
		that (68)[42]	*it* (32)
Width	Narrow	27 (39.7%)	0
	Wide	0	9 (28.1%)

As seen in Table 4.28, I gained 27 occurrences (39.7% out of 68 hits) of the sentential anaphor *that* for narrow reference, while I gained 9 occurrences (28.1% out of 32 hits) of the sentential anaphor *it* for wide reference in English. There was no occurrence of *that* for wide reference and of *it* for narrow reference.

In a Pearson's Chi-square test, the given probabilities for the relative frequencies between two categorical variables: each condition in the width and the form of the sentential anaphor (i.e., narrow/wide and *that/it*) were statistically significant (df=3, x^2=54, p<.001***).

4.4.2.2 In Korean

Second, the findings on all sentential anaphors in Korean TV drama scripts are shown in Table 4.33 in terms of the width of reference (i.e., narrow vs. wide reference).

The relative frequencies of the sentential anaphor *ku-ken* 'that' for narrow reference and those of the sentential anaphor *ku-len* 'it' for wide reference are shown in Table 4.29.

[42] The rest of the total number of *that*, 41, is the relative frequency for new information of *that*, while the rest of the total number of *it*, 23, is the relative frequency for old information of *it*. See Section 4.4.1.

Table 4.29 The relative frequency of narrow (*ku-ken*) vs. wide (*ku-len*) reference in Korean

Condition \ Frequency & forms		Frequency (n (%))	
		Korean sentential anaphors (total n of each form)	
		ku-ken 'that' (33)	*ku-len* 'it' (28)
Width	Narrow	21 (63.6%)	0
	Wide	0	17 (60.7%)

As seen in Table 4.29, I gained 21 occurrences (63.6% out of 68 hits) of the sentential anaphor *ku-ken* 'that' for narrow reference, while I gained 17 occurrences (60.7% out of 28 hits) of the sentential anaphor *ku-len* 'it' for wide reference in Korean. There was no occurrence of *ku-ken* 'that' for wide reference and of *ku-len* 'it' for narrow reference.

In a Pearson's Chi-square test, the given probabilities for the relative frequencies between two categorical variables: each condition in the width and the form of the sentential anaphor (i.e., narrow/wide and *kuken/kulen*) were statistically significant (df=3, x^2=38.84, p<.001***).

4.4.2.3 Section Summary

The relative frequencies of the sentential anaphors in English and Korean in terms of the width of reference (i.e., narrow vs. wide reference) relevant Chi-square test results are summarized in Table 4.30, Figure 4.16, and Table 4.31.

Table 4.30 The summary of Corpus III (the width of reference) in English and Korean

Condition \ Frequency & forms		Frequency (n (%))			
		Sentential anaphors (total n of each form)			
		English		Korean	
		that (68)	*it* (32)	*ku-ken*'that' (33)	*ku-len*'it' (28)
Width	Narrow	27 (39.7%)	0	21 (63.6%)	0
	Wide	0	9 (28.1%)	0	17 (60.7%)

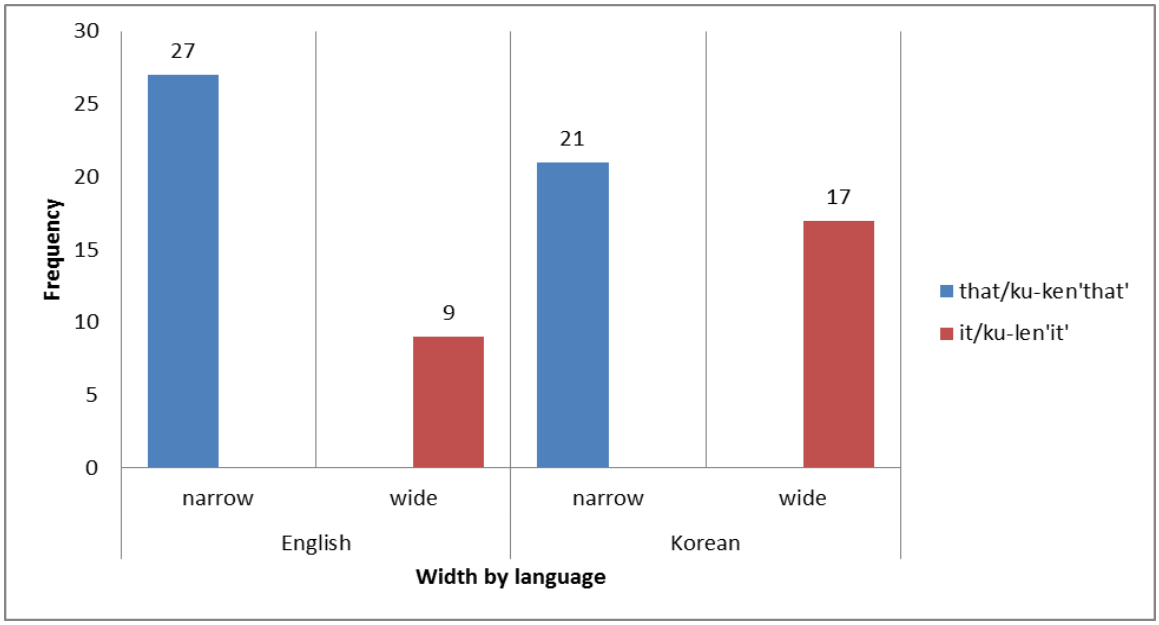


Figure 4.16 The summary of Corpus III (the width of reference) in English and Korean

As seen in Table 4.30 and Figure 4.16, for Corpus III (the width of reference), narrow reference is referred to by *that* in English and *ku-ken* 'that' in Korean, whereas wide reference is referred to by *it* in English and *ku-len* 'it' in Korean.

Table 4.31 The summary of the Chi-square test results of Corpus III (the width of reference) in English and Korean

Language & chi-square test / Condition	Language	Pearson's Chi-square test results
		each condition vs. all forms in each language
Width (narrow vs. wide)	English	df=3, x^2=54, p<.001*** (significant)
	Korean	df=3, x^2=38.84, p<.001*** (significant)

In the Chi-square test for Corpus III, the given probabilities for the frequencies between two categorical variables (i.e., the width of reference and the discourse anaphoric forms) were all statistically significant, which indicates that there is a marked difference between English and Korean in terms of the width of reference.

4.5 Chapter Summary

4.5.1 Corpus I and II: Discourse Anaphoric Patterns in English and Korean

The results of Corpus I (by order of mention) and Corpus II (by paragraph) for discourse anaphoric patterns in English and Korean analyzed in Section 4.1 and Section 4.2 are summarized by mention types, languages, anaphoric forms, frequencies, and Chi-square test results in Table 4.32 and Table 4.33.

Table 4.32 Summary of Corpus I and II for English and Korean

Mention type	Language	Form						Total
		FN	SN	Pro	GN	LP	Zero	
1stM	English	**98**	0	2	0	0	0	100
	Korean	**100**	0	0	0	0	0	100
2ndM	English	0	**33**	**64**	0	3	0	100
	Korean	**72**	**14**	9	0	1	4	100
3rdM	English	3	**31**	**60**	0	6	0	100
	Korean	**75**	**15**	5	0	1	4	100
1stP	English	**98**	0	2	0	0	0	100
	Korean	**100**	0	0	0	0	0	100
2ndP	English	1	**80**	**15**	0	4	0	100
	Korean	**82**	**12**	5	0	1	0	100
3rdP	English	3	**76**	**19**	0	2	0	100
	Korean	**74**	**16**	8	0	2	0	100

As seen in Table 4.32, in Corpus I (discourse anaphoric patterns by order of mention), for the first mention, both the English and Korean corpus predominantly contain full names. For the second mention, the English corpus mainly contains pronouns and surnames, while the Korean corpus predominantly contains full names. For the third mention, the

English corpus mainly contains pronouns and surnames, while the Korean corpus predominantly contains full names.

In Corpus II (discourse anaphoric patterns by paragraph), for the first paragraph, both the English and Korean corpus predominantly contain full names. For the second paragraph, the English corpus mainly contains surnames and pronouns, while the Korean corpus predominantly contains full names. For the third paragraph, the English corpus mainly contains surnames and pronouns, while the Korean corpus predominantly contains full names.

Table 4.33 Summary of Chi-square test results for Corpus I and II for English and Korean

Mention Type	Language	Pearson's Chi-square test results	
		each mention type vs. three forms in each language	each mention type vs. three forms in both languages
1st M	English	df=1, χ^2=92.16, p<.001*** (significant)	NA
	Korean	NA	
2nd M	English	df=1, χ^2=9.91, p=.01** (significant)	df=2, χ^2=121.11, p<.001*** (significant)
	Korean	df=2, χ^2=77.45, p<.001*** (significant)	
3rd M	English	df=2, χ^2=51.85, p<.001*** (significant)	df=2, χ^2=118.56, p<.001*** (significant)
	Korean	df=2, χ^2=90.53, p<.001*** (significant)	

Table 4.33 – *Continued*

Mention Type	Language	Pearson's Chi-square test results	
		each mention type vs. three forms in each language	each mention type vs. three forms in both languages
1stP	English	df=1, χ^2=92.16, p<.001*** (significant)	NA
	Korean	NA	
2ndP	English	df=2, χ^2=111.06, p<.001*** (significant)	df=2, χ^2=134.29, p<.001*** (significant)
	Korean	df=2, χ^2=109.88, p<.001*** (significant)	
3rdP	English	df=2, χ^2=90.14, p<.001*** (significant)	df=2, χ^2=109.08, p<.001*** (significant)
	Korean	df=2, χ^2=79.43, p<.001*** (significant)	

In the Pearson's Chi-square test, the given probabilities for the frequencies between two categorical variables in Corpus I (i.e., the order of mention and the three discourse anaphoric forms, i.e., FN, SN, and Pro) and Corpus II (i.e., the order of paragraph and the three discourse anaphoric forms, i.e., FN, SN, and Pro) were all statistically significant except for the first mention/paragraph in the Korean corpus.

Thus, the results of Chi-square tests indicate that there is a marked difference in terms of the distribution of three forms depending on the order of mention and on the paragraph, particularly on the second and the third mention/paragraph in the English

corpus, while the same is not true for the Korean corpus because of the use of FNs regardless of order of mention of the referent and placement of the paragraph.

4.5.2 Corpus III: Sentential Anaphors in English and Korean

The results of Corpus III for sentential anaphors in English and Korean analyzed in Section 4.4 are summarized by referential properties, languages, anaphoric forms, frequencies, and Chi-square test results in Table 4.34 and Table 4.35.

Table 4.34 The summary of Corpus III for English and Korean

Frequency & forms / Condition		Frequency (n (%))			
		English		Korean	
		that	*it*	*ku-ken* 'that'	*ku-len* 'it'
Givenness	New	41 (60.3%)	0	12 (36.4%)	0
	Old	0	23 (71.9%)	0	11 (39.3%)
Width	Narrow	27 (39.7%)	0	21 (63.6%)	0
	Wide	0	9 (28.1%)	0	17 (60.7%)
Total		68 (100%)	32 (100%)	33 (100%)	28 (100%)

As seen in Table 4.34, for Corpus III, new information is referred to by *that* in English and *ku-ken* 'that' in Korean, whereas old information is referred to by *it* in English and *ku-len* 'it' in Korean. On the other hand, narrow reference is referred to by *that* in English and *ku-ken* 'that' in Korean, whereas wide reference is referred to by *it* in English and *ku-len* 'it' in Korean.

Table 4.35 The summary of Chi-square test results for Corpus III for English and Korean

Language & chi-square test / Condition	Language	Pearson's Chi-square test results
		each condition vs. all forms in each language
Givenness (new vs. old)	English	df=3, χ^2=74.13, p<.001*** (significant)
	Korean	df=3, χ^2=23.09, p<.001*** (significant)
Width (narrow vs. wide)	English	df=3, χ^2=54, p<.001*** (significant)
	Korean	df=3, χ^2=38.84, p<.001*** (significant)

In the Pearson's Chi-square test, the given probabilities for the frequencies between two categorical variables in the first part of Corpus III (i.e., the givenness of information and the discourse anaphoric forms) and in the second part of Corpus III (i.e., the width of reference and the discourse anaphoric forms) were all statistically significant, which indicates that there is a marked difference between English and Korean in terms of the givenness and the width.

Now let us consider the findings of the survey results in English and Korean in the following chapter.

Chapter 5

Results and Analysis: Survey

This chapter analyzes the survey data collected and encoded via the survey questionnaire in order to explore the natural selection of discourse anaphors by language users. The samples in the questionnaire were selected from news articles in English and Korean. The survey results coded every single observation in MS Excels Files sorted in terms of discourse anaphors by the order of mention (Section B), discourse anaphors by the paragraph (Section C), and the properties of sentential anaphors (Section D), and the converted CSV files were run in RStudio for the mean values and the statistical analysis. This chapter lays out the results of the surveys in the following sections: Section 5.1 analyzes discourse anaphoric patterns by order of mention (5.1.1 for English; 5.1.2 for Korean) in terms of the first-, the second-, the third mentions, the interactions between variables (5.1.3), and section summary (5.1.4); Section 5.2 analyzes discourse anaphoric patterns by paragraph (5.2.1 for English; 5.2.2 for Korean) in terms of the first-, the second-, the third paragraph, the interactions between variables (5.2.3), and section summary (5.2.4); Section 5.3 analyzes sentential anaphors in two referential respects: by new vs. old information (5.3.1) (5.3.1.1 for English; 5.3.1.2 for Korean; 5.3.1.3 for summary), by narrow vs. wide reference (5.3.2) (5.3.2.1 for English; 5.32.2 for Korean; 5.3.2.3 for summary), and the interactions between variables (5.3.3); Section 5.4 summarizes this chapter in terms of two discourse anaphoric patterns (5.4.1) and sentential anaphors (5.4.2) in two languages.

5.1 Section B: Discourse Anaphoric Patterns by Order of Mention

5.1.1 By Order of Mention in English

The findings on mean values of all degrees of participants' acceptability for all types of discourse anaphors referring to public figures in English news articles are correspondingly shown in Table 5.1 through Table 5.3 and Figure 5.1 through Figure 5.3 in terms of the first-, the second-, and the third mentions.

5.1.1.1 The First Mentions

First, the mean values of all degrees of participant acceptability for all types of referring expressions referring to public figures in the first mentions are shown in Table 5.1 and Figure 5.1.

Table 5.1 The mean values of participant acceptability for the first mentions in English

Mean by discourse anaphoric form / Factors			Mean			
			Form			
Condition	Language	Famousness (Section)	a_FN	b_SN	c_Pro	d_GN
Q1_1st M	English	high (B1 & B4)	**5.00**[43]	2.9	1.675	2.125
		mid-high (B2 & B5)	**4.95**	2.725	1.675	2.05
		mid-low (B3 & B6)	**5.00**	2.475	1.45	1.85

[43] The number in red color represents the highest mean value, while the number in blue color represents the next one in the dissertation.

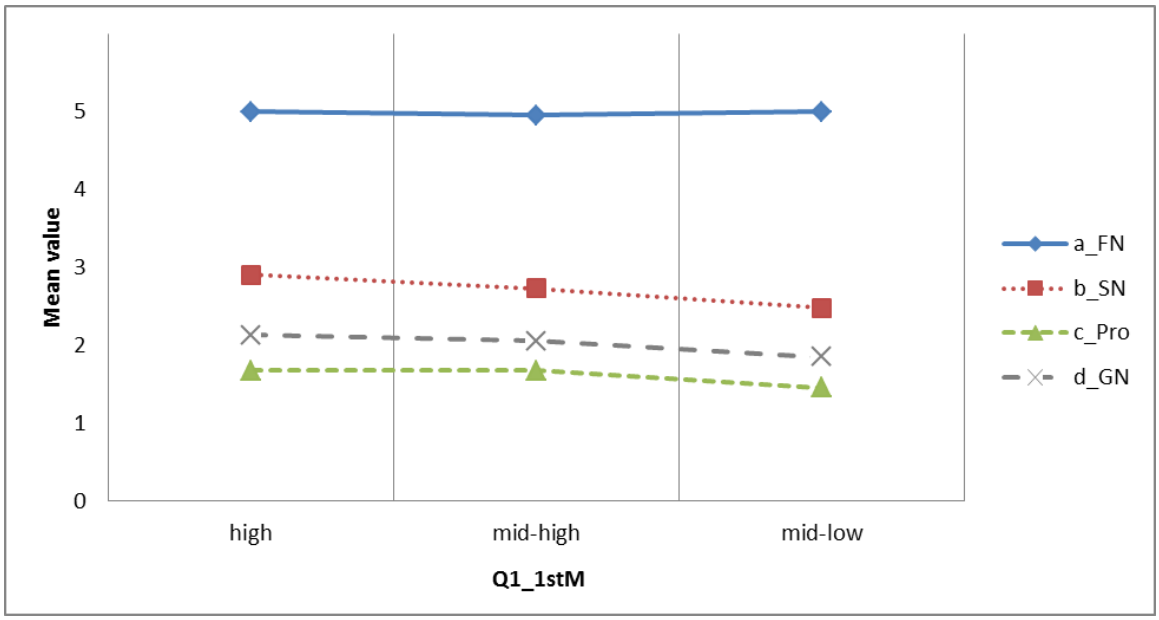


Figure 5.1 The mean values of participant acceptability for the first mentions in English

The red column in Table 5.1 and the blue line in Figure 5.1 show that regardless of degree of famousness, the degree of acceptability for full names in the first mentions (high, M=5; mid-high, M=4.9; mid-low, M=5) are predominantly higher than that for other forms of referring expressions such as surnames (high, M=2.9; mid-high, M=2.725; mid-low, M=2.475), pronouns (high, M=1.675; mid-high, M=1.675; mid-low, M=1.45) and given names (high, M=2.125; mid-high, M=2.05; mid-low, M=1.85).

This finding conforms to Ariel's (1991, 1994, 2008, 2010) AH and Lee (2010) in that the first mentions are all referred to by the most discontinuous linguistic form, among other various constructions, where full name is the highest and the full name (plus modifier) is the lowest accessible marker.

For the first mentions, naming forms other than those suggested in the survey questionnaire were observed in terms of the structural constructions (frequency) and the examples, as shown in (34).

(34) a. TT + SN (45) Mr. Jeter
b. NP (4) the athlete, the singer
c. TT + FN (2) Ms. Ann Beattie

As seen in (34), surnames with titles (45) were mainly written in by the participants; but definite noun phrases (4) and full names with titles (2) were also observed.

5.1.1.2 The Second Mentions

Second, the mean values of all degrees of participant acceptability for all types of discourse anaphors referring back to public figures in the second mentions are shown in Table 5.2 and Figure 5.2.

Table 5.2 The mean values of participant acceptability for the second mentions in English

Mean by discourse anaphoric form / Factors			Mean			
			Form			
Condition	Language	Famousness (Section)	a_FN	b_SN	c_Pro	d_GN
Q2_2nd M	English	high (B1 & B4)	2.775	**4.6**	**4.475**	3.7
		mid-high (B2 & B5)	2.925	**4.275**	**4.35**	3.725
		mid-low (B3 & B6)	2.9	**4.525**	**4.225**	2.925

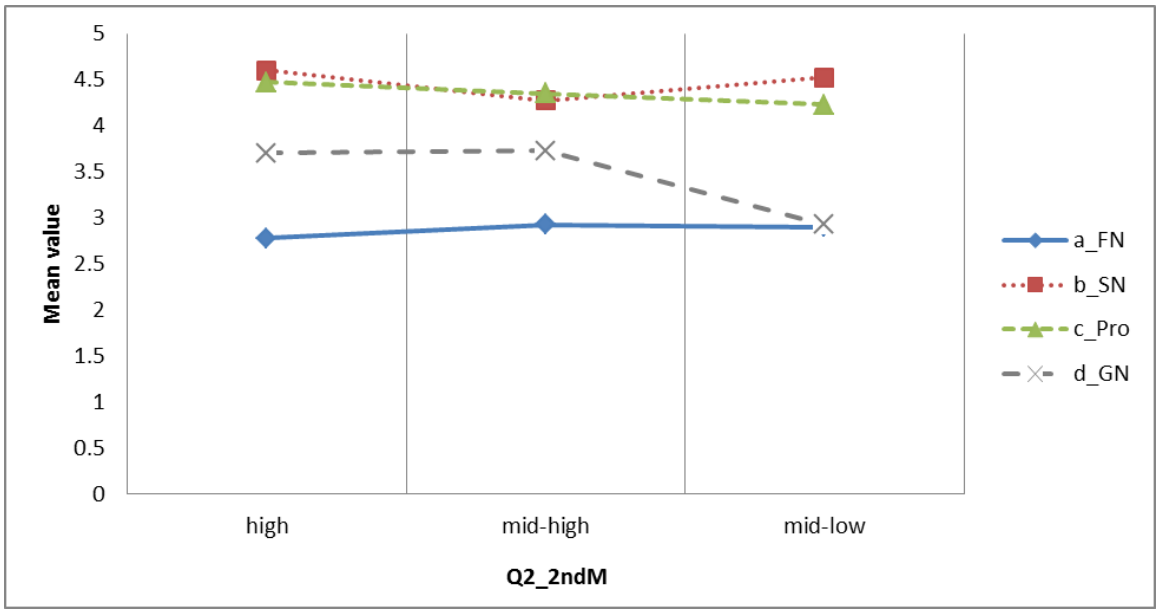


Figure 5.2 The mean values of participant acceptability for the second mentions in English

Table 5.2 and Figure 5.2 show that the degree of acceptability for surnames (high, M=4.6; mid-high, M=4.275; mid-low, M=4.525) and pronouns (high, M=4.475; mid-high, M=4.35; mid-low, M=4.225) in the second mentions are relatively higher than that for other forms of referring expressions such as given names (high, M=3.7; mid-high, M=3.925; mid-low, M=2.925) and full names (high, M=2.775; mid-high, M=2.925; mid-low, M=2.9).

This finding also conforms to Ariel's AH and Lee (2010) in that the most accessible referent is referred to by the most accessible forms such as pronouns and surnames in English. In fact, pronouns were predominantly retrieved for the second mentions in Corpus I (Section 4.1.1) in terms of the distribution and the acceptability rate between surnames and pronouns for the second mentions are very close; however, that of surnames is relatively higher than that of pronouns in the survey. In addition, the

acceptability rate of full names for the second mentions is predominantly lower than that of full names in the first mentions.

For the second mentions, naming forms other than those suggested in the survey questionnaire were observed in terms of the structural constructions (frequency) and the examples, as shown in (35).

(35) a. TT + SN (41) Mr. Jobs
b. NP (8) the athlete, the singer, the comedian

As seen in (35), surnames with titles (41) were mainly written in by the participants; definite noun phrases (8) were also observed.

5.1.1.3 The Third Mentions

Third, the mean values of all degrees of participant acceptability for all types of referring expressions referring back to public figures in the third mentions are shown in Table 5.3 and Figure 5.3.

Table 5.3 The mean values of participant acceptability for the third mentions in English

Mean by discourse anaphoric form / Factors			Mean			
			Form			
Condition	Language	Famousness (Section)	a_FN	b_SN	c_Pro	d_GN
Q3_3rd M	English	high (B1 & B4)	2.85	**4.35**	**4.25**	3.75
		mid-high (B2 & B5)	3.075	**4.475**	**4.35**	3.75
		mid-low (B3 & B6)	2.65	**4.3**	**4.3**	3.35

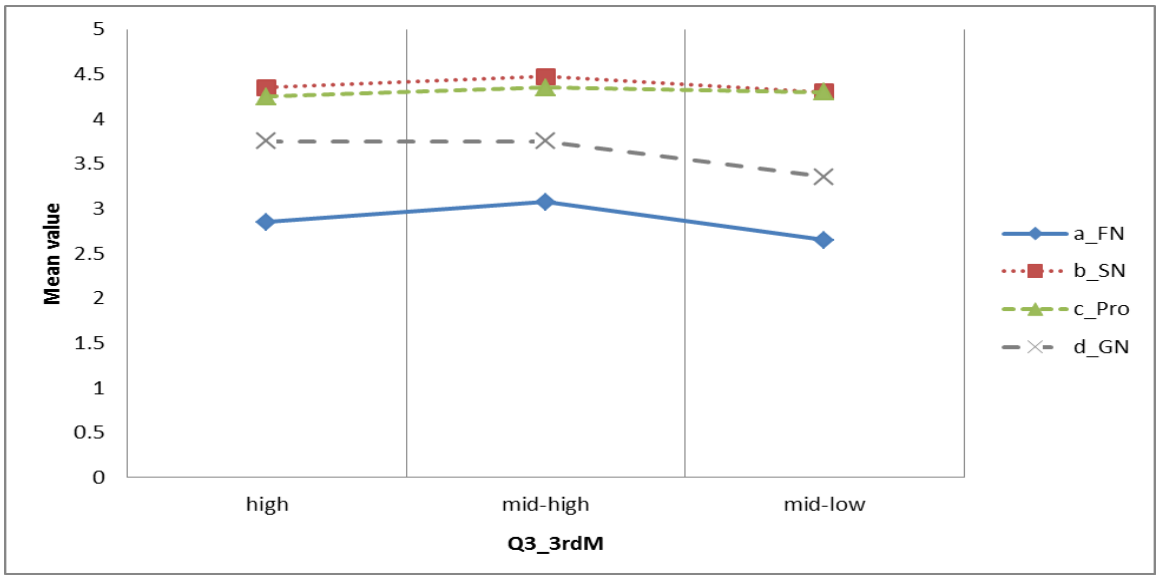


Figure 5.3 The mean values of participant acceptability for the third mentions in English

As seen in Table 5.3 and Figure 5.3, the mean values of surnames and pronouns for the third mentions shows similar mean values in the second mentions in terms of degree of acceptability. In particular, the degree of acceptability for surnames (high, M=4.35; mid-high, M=4.475; mid-low, M=4.3) and pronouns (high, M=4.25; mid-high, M=4.35; mid-low, M=4.3) in the third mentions are relatively higher than that for other forms of referring expressions such as given names (high, M=3.75; mid-high, M=3.75; mid-low, M=3.35) and full names (high, M=2.85; mid-high, M=3.075; mid-low, M=2.65).

This finding also conforms to Ariel's AH and Lee (2010) in that the most accessible referent is referred to by the most accessible forms such as pronouns and surnames in English. Like in the second mentions in the survey, pronouns were predominantly retrieved for the second mentions in Corpus I (Section 4.1.1) in terms of the distribution and the acceptability rate between surnames and pronouns for the second mentions are very close; in fact, that of surnames is relatively higher than that of

pronouns in the survey. Again, the acceptability rate of full names for the third mentions is predominantly lower than that of full names in the first mentions.

For the third mentions, naming forms other than those suggested in the survey questionnaire were observed in terms of the structural constructions (frequency) and the examples, as shown in (36).

(36) a. TT + SN (40) Ms. Jones, Ms. Beattie
b. NP (6) the CEO, the innovator, the comedian

As seen in (36), surnames with titles (40) were mainly written in by the participants; definite noun phrases (6) were also observed.

5.1.2 By Order of Mention in Korean

The findings on mean values of all degrees of participants' acceptability for all types of discourse anaphors referring to public figures in Korean news articles are correspondingly shown in Table 5.4 through Table 5.6 and Figure 5.4 through Figure 5.6 in terms of the first-, the second-, and the third mentions.

5.1.2.1 The First Mentions

First, the mean values of all degrees of participant acceptability for all types of referring expressions referring to public figures in the first mentions are shown in Table 5.4 and Figure 5.4.

Table 5.4 The mean values of participant acceptability for the first mentions in Korean

Mean by discourse anaphoric form / Factors			Mean			
			Form			
Condition	Language	Famousness (Section)	a_FN	b_SN	c_Pro	d_GN
Q1_1st M	Korean	high (B1 & B4)	**4.95**	1.55	1.8	1.475
		mid-high (B2 & B5)	**4.95**	1.4	1.65	1.575
		mid-low (B3 & B6)	**4.95**	1.45	1.45	1.525

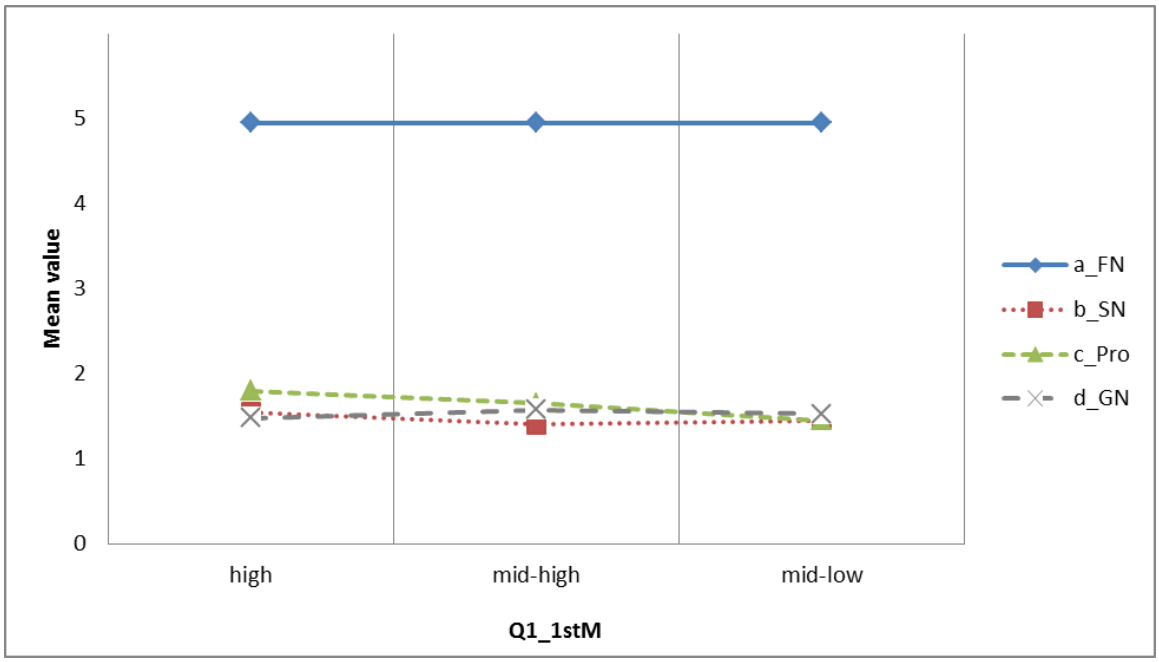


Figure 5.4 The mean values of participant acceptability for the first mentions in Korean

The red column in Table 5.4 and the blue line in Figure 5.4 show that like the mean values for the first mentions in English, the degree of acceptability for full names in the first mentions (high, M=4.95; mid-high, M=4.95; mid-low, M=4.95), regardless of degree of famousness, are predominantly higher than that for other forms of referring expressions such as surnames (high, M=1.55; mid-high, M=1.4; mid-low, M=1.45),

pronouns (high, M=1.8; mid-high, M=1.65; mid-low, M=1.45) and given names (high, M=1.475; mid-high, M=1.575; mid-low, M=1.525). In fact, the gap of the mean values between full names and other forms in Korean is relatively wider than those in English.

This finding conforms to Ariel's AH and Lee (2010) in that the first mentions are all referred to by the most discontinuous linguistic form, i.e., full name among other various constructions since the full name (plus modifier) is the lowest accessible marker.

For the first mentions, naming forms other than those suggested in the survey questionnaire were observed in terms of the structural constructions (frequency) and the examples, as shown in (37).

(37) a. TT + FN (+ TT) (21) Park Chan Ho senswu 'player'
b. SN + TT (4) Choi ssi 'Mr.'
c. NN + FN (2) Korean thukkup Park Chan Ho 'express'

As seen in (37), it is noteworthy that full names with titles (21) were mainly written in by the participants; surnames with titles (4) and full names with nicknames (2) were also observed.

5.1.2.2 The Second Mentions

Second, the mean values of all degrees of participant acceptability for all types of discourse anaphors referring back to public figures in the second mentions are shown in Table 5.5 and Figure 5.5.

Table 5.5 The mean values of participant acceptability for the second mentions in Korean

Factors / Mean by discourse anaphoric form			Mean			
			Form			
Condition	Language	Famousness (Section)	a_FN	b_SN	c_Pro	d_GN
Q2_2nd M	Korean	high (B1 & B4)	**4.325**	2.025	**4.425**	1.525
		mid-high (B2 & B5)	**4.575**	1.875	**3.875**	1.775
		mid-low (B3 & B6)	**4.3**	1.8	**3.775**	1.45

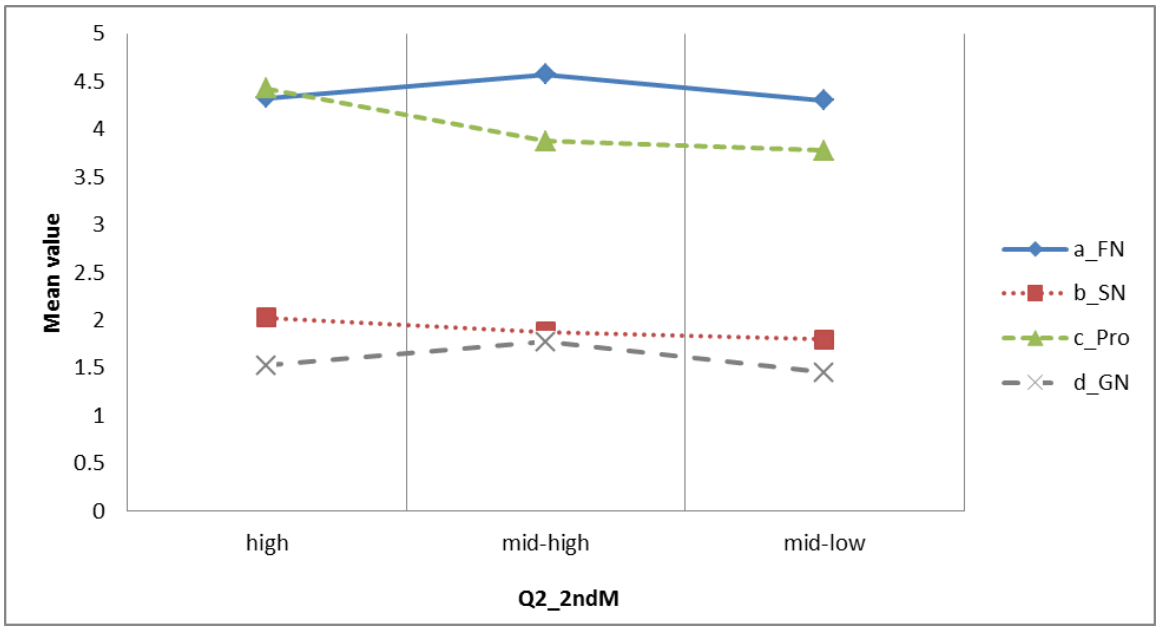


Figure 5.5 The mean values of participant acceptability for the second mentions in Korea

Table 5.5 and Figure 5.5 show that like the second mentions in Corpus I for Korean (Section 4.1.2.2), the degree of participants' acceptability for full names (high, M=4.325; mid-high, M=4.575; mid-low, M=4.3) in the second mentions is predominantly higher than other forms of discourse anaphors, regardless of the degree of famousness. However, unlike Corpus I in terms of the distribution, it is noteworthy that the degree of

acceptability for pronouns (high, M=4.425; mid-high, M=3.876; mid-low, M=3.375) in the second mentions is also relatively higher than other naming forms such as surnames (high, M=2.025; mid-high, M=1.875; mid-low, M=1.8) and given names (high, M=1.525; mid-high, M=1.775; mid-low, M=1.45) in terms of the selection.

This finding would definitely run counter to Lee's (2010) findings on English/Korean political news articles and Ariel's AH for English, in which the second mentions are referred to by more accessible linguistic markers such as pronouns and surnames, not by less accessible markers such as full names (cf. Kim 2010a, 2012).

For the second mentions, naming forms other than those suggested in the survey questionnaire were observed in terms of the structural constructions (frequency) and the examples, as shown in (38).

(38) a. SN + TT (18) Choi cakka 'writer'
b. FN + TT (10) Park Chan Ho senswu 'player'
c. this + TT (4) i cakka 'this writer'

As seen in (38), surnames with titles (18) were mainly written in by the participants; full names with titles (10) and 'this' plus titles (4) were also observed.

5.1.2.3 The Third Mentions

Third, the mean values of all degrees of participant acceptability for all types of discourse anaphors referring back to public figures in the third mentions are shown in Table 5.6 and Figure 5.6.

Table 5.6 The mean values of participant acceptability for the third mentions in Korean

Factors \ Mean by discourse anaphoric form			Mean			
			Form			
Condition	Language	Famousness (Section)	a_FN	b_SN	c_Pro	d_GN
Q3_3rd M	Korean	high (B1 & B4)	**4.375**	1.975	**4.325**	1.575
		mid-high (B2 & B5)	**4.475**	1.650	**3.95**	1.75
		mid-low (B3 & B6)	**4.4**	1.675	**3.825**	1.4

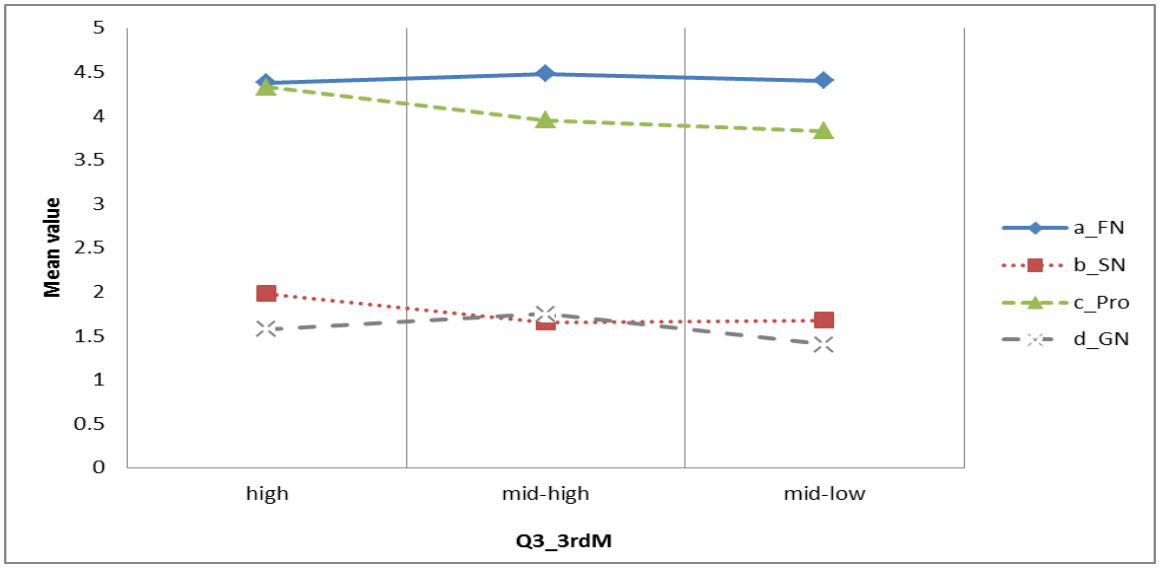


Figure 5.6 The mean values of participant acceptability for the third mentions in Korean

Table 5.6 and Figure 5.6 show that like the second/third mentions in Corpus I for Korean (Section 4.1.2.3) and the second mentions in the survey, the degree of participants' acceptability for full names (high, M=4.375; mid-high, M=4.475; mid-low, M=4.4) in the third mentions is predominantly higher than other forms of discourse anaphors, regardless of the degree of famousness. However, unlike Corpus I in terms of the distribution, it is

also noteworthy that the degree of acceptability for pronouns (high, M=4.325; mid-high, M=3.95; mid-low, M=3.825) in the second mentions is also relatively higher than other naming forms such as surnames (high, M=1.975; mid-high, M=1.650; mid-low, M=1.675) and given names (high, M=1.575; mid-high, M=1.75; mid-low, M=1.4) in terms of the selection.

This finding would also definitely run counter to Lee's (2010) findings on English/Korean political news articles and Ariel's AH for English mentioned in the previous section (cf. Kim 2010a, 2012).

For the third mentions, naming forms other than those suggested in the survey questionnaire were observed in terms of the structural constructions (frequency) and the examples, as shown in (39).

(39) a. FN + TT (15) Park Chan Ho senswu 'player'
b. SN + TT (12) Shin cakka 'writer'
c. this + TT (4) i cakka 'this writer'

As seen in (39), it is noteworthy that full names with titles (15) were mainly written in by the participants; surnames with titles (12) and 'this' plus titles (4) were also observed.

5.1.3 Interactions between Variables

Now let us consider the interactions between the independent variables (IVs) and the dependent variable (DV) in Section B. The results of ANOVA test are shown in Table 5.7.

Table 5.7 The results of ANOVA test for Section B

Response: rating ~	Sum Sq	Df	F value	Pr(>F)
condition Q1_1st Mention Q2_2nd Mention Q3_3rd Mention	389.5	3	121.6765	< 2.2e-16***[44]
famousness high mid-high mid-low	63.7	2	29.8723	< 1.420e-13***
language English Korean	279.6	1	262.0495	< 2.2e-16***
naming form a_Full Name (FN) b_Surname (SN) c_Pronoun (Pro) d_Given Name (GN)	1266.6	3	395.7049	< 2.2e-16***
condition:famousness	118.8	6	18.5570	< 2.2e-16***
condition:language	38.9	3	12.1634	< 6.537e-08***
famousness:language	3.3	2	1.5512	0.21216
condition:naming form	1243.2	6	194.1978	< 2.2e-16***
famousness:naming form	15.4	6	2.3991	< 0.02576*
language:naming form	1058.0	3	330.5161	< 2.2e-16***
condition:famousness:language	14.0	6	2.1943	< 0.04077*
condition:famousness:naming form	7.3	12	0.5690	0.86858
condition:language:naming form	240.0	6	37.4863	< 2.2e-16***
famousness:language:naming form	9.5	6	1.4864	0.17862
condition:famousness:language:naming form	6.9	12	0.5362	0.89241
Residuals	3245.7	3042		

As seen in Table 5.7, concerning each single-factor repeated-measure ANOVA for Section B, there are significant main effects for rating (i.e., degree of acceptability). Condition (the first-, the second-, the third mention) as the independent variable (IV) show a significant main effect on rating as the dependent variable (DV), F(3, 3042) =

[44] Significance codes: 0;'***' 0.001; '**' 0.01; '*'0.05; '.'0.1'; ' ' 1

121.68, p <.001***. Famousness (high, mid-high, mid-low) as the IV show a significant main effect on rating as the DV, F(2, 3042) = 29.87, p<.001***. Language (English, Korean) as the IV show a significant main effect on rating as the DV, F(1, 3042) = 262.05, p<.001***. Naming form (full name, surname, pronoun, given name) as the IV show a significant main effect on rating as the DV, F(3, 3042) = 395.7, p<.001***.

The results of two-factor repeated measure ANOVA in Section B are as follows. Condition and famousness as IVs show a significant main effect on rating as the DV, and significant interaction between two IVs for rating, F(6, 3042) = 18.56, p<.001***. Condition and language as IVs show a significant main effect on rating as the DV, and significant interaction between two IVs for rating, F(3, 3042) = 12.16, p<.001***. Condition and naming form as IVs show a significant main effect on rating as the DV, and significant interaction between two IVs for rating, F(6, 3042) = 194.20, p<.001***. Famousness and naming form as IVs show a significant main effect on rating as the DV, and significant interaction between two IVs for rating, F(6, 3042) = 2.40, p<.05*. Language and naming form as IVs show a significant main effect on rating as the DV, and significant interaction between two IVs for rating, F(3, 3042) = 330.52, p<.001***.

The results of more than two-factor repeated measure ANOVA in Section B are as follows. Condition, famousness, and language as IVs show a significant main effect on rating as the DV, and significant interaction between three IVs for rating, F(6, 3042) = 2.19, p<.05*. Condition, language, and naming form as IVs show a significant main effect on rating as the DV, and significant interaction between three IVs for rating, F(6, 3042) = 37.49, p<.001***.

In addition, as seen in more than two-factor repeated measure ANOVA, condition other than factors might predict the effects on rating. Thus, this ANOVA needs to be broken down into subanalysis ANOVA by condition, i.e., the first-, the second-, and the third mention, as shown in Table 5.8.

Table 5.8 The results of subanalysis ANOVA test by condition for Section B

Response: rating for 1st M ~	Sum Sq	Df	F value	Pr(>F)
famousness	4.40	2	2.7859	0.06218
language	42.50	1	53.7984	4.809e-13***
naming form	1814.98	3	765.7536	<2.2e-16***
famousness:language	0.70	2	0.4443	0.64140
famousness:naming form	2.52	6	0.5322	0.78405
language:naming form	63.11	3	26.6276	<2.2e-16***
famousness:language:naming form	1.86	6	0.3916	0.88470
Residuals	739.50	936		
Response: rating for 2nd M ~	Sum Sq	Df	F value	Pr(>F)
famousness	10.34	2	4.2945	0.01391*
language	156.01	1	129.5953	<2e-16***
naming form	360.06	3	99.6997	<2e-16***
famousness:language	0.02	2	0.0078	0.99224
famousness:naming form	16.24	6	2.2478	0.03685*
language:naming form	595.73	3	164.9550	<2e-16***
famousness:language:naming form	9.06	6	1.2538	0.27634
Residuals	1126.77	936		
Response: rating for 3rd M ~	Sum Sq	Df	F value	Pr(>F)
famousness	8.14	2	3.4236	0.0330*
language	169.18	1	142.3136	<2e-16***
naming form	334.81	3	93.8829	<2e-16***
famousness:language	1.94	2	0.8158	0.4426
famousness:naming form	3.89	6	0.5447	0.7743
language:naming form	639.10	3	179.2079	<2e-16***
famousness:language:naming form	5.47	6	0.7667	0.5962
Residuals	1112.68	936		

As seen in Table 5.8, the results of subanalysis ANOVA by condition for Section B are as follows. For a factor, language and naming form as IVs show a statistically significant main effect on rating for each order of mention as the DV. For more than two factors, it is worth noting that there is a statistically significant interaction between language and naming form as IVs on rating for each order of mention as DVs, i.e., for the first mention, $F(3, 936) = 26.63$, $p<.001$***; for the second mention, $F(3, 936) = 164{,}96$, $p<.001$***; for the third mention, $F(3, 936) = 179.21$, $p<.001$***. In addition, famousness as an IV shows a statistically significant effect on rating for the second and the third mention as DVs.

5.1.4 Section Summary

All mean values of degrees of participants' acceptability for all types of discourse anaphors in Section B (i.e., by order of mention) by native speakers of English and Korean are summarized in Table 5.9 and Figure 5.7.

Table 5.9 Summary of Section B for English and Korean

Factors \ Mean by discourse anaphoric form			Mean			
			Form			
Condition	Language	Famousness (Section)	a_FN	b_SN	c_Pro	d_GN
Q1_1st M	English	high (B1 & B4)	**5.00**[45]	2.9	1.675	2.125
		mid-high (B2 & B5)	**4.95**	2.725	1.675	2.05
		mid-low (B3 & B6)	**5.00**	2.475	1.45	1.85
	Korean	high (B1 & B4)	**4.95**	1.55	1.8	1.475
		mid-high (B2 & B5)	**4.95**	1.4	1.65	1.575
		mid-low (B3 & B6)	**4.95**	1.45	1.45	1.525
Q2_2nd M	English	high (B1 & B4)	2.775	**4.6**	**4.475**	3.7
		mid-high (B2 & B5)	2.925	**4.275**	**4.35**	3.725
		mid-low (B3 & B6)	2.9	**4.525**	**4.225**	2.925
	Korean	high (B1 & B4)	**4.325**	2.025	**4.425**	1.525
		mid-high (B2 & B5)	**4.575**	1.875	**3.875**	1.775
		mid-low (B3 & B6)	**4.3**	1.8	**3.775**	1.45
Q3_3rd M	English	high (B1 & B4)	2.85	**4.35**	**4.25**	3.75
		mid-high (B2 & B5)	3.075	**4.475**	**4.35**	3.75
		mid-low (B3 & B6)	2.65	**4.3**	**4.3**	3.35
	Korean	high (B1 & B4)	**4.375**	1.975	**4.325**	1.575
		mid-high (B2 & B5)	**4.475**	1.650	**3.95**	1.75
		mid-low (B3 & B6)	**4.4**	1.675	**3.825**	1.4

[45] The number in red color represents the highest mean value, while the number in blue color represents the next one in the dissertation.

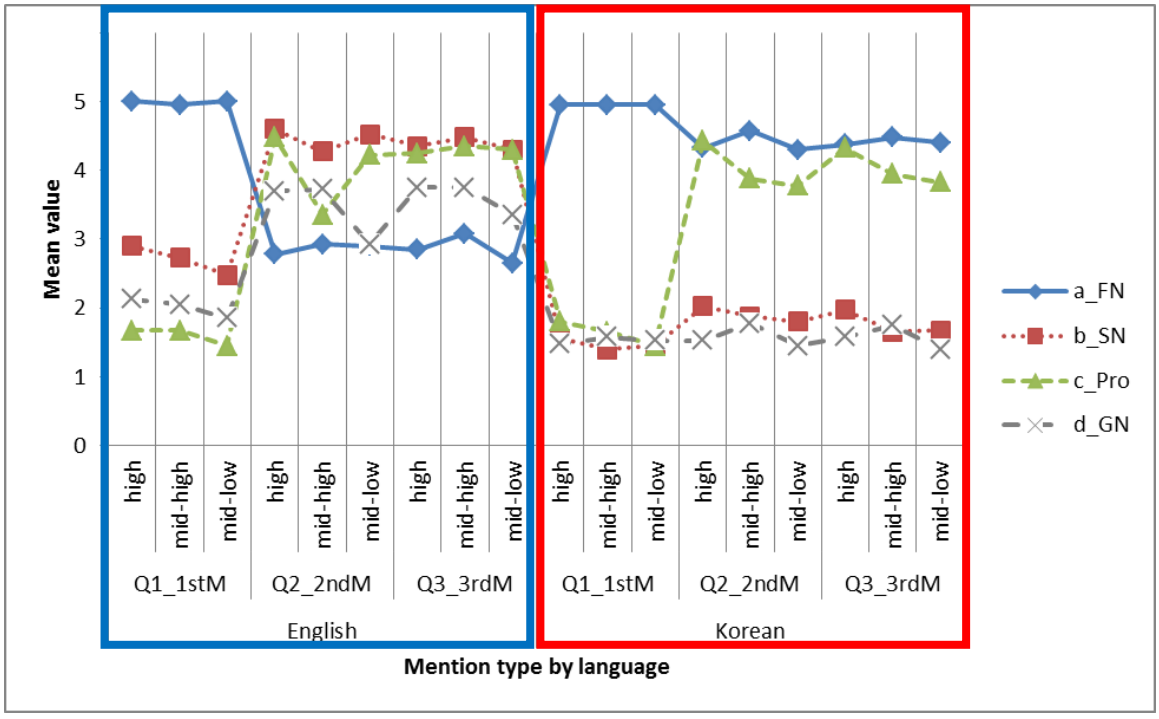


Figure 5.7 Summary of Section B for English and Korean

As seen in Table 5.9 and Figure 5.7, in Section B (discourse anaphoric patterns by order of mention), for the first mention, both English (in blue) and Korean (in red) survey show that the mean values of the full names are the highest other than naming forms in terms of degree of acceptability. For the second mention, the English survey shows that surnames are the most preferred form and surnames are also acceptable, while the Korean survey shows that full names are the most preferred form and pronouns are also acceptable. For the third mention, the English survey shows that surnames are the most preferred form and surnames are also acceptable, while the Korean survey shows that full names are the most preferred form and pronouns are also acceptable, like the second mention.

5.2 Section C: Discourse Anaphoric Patterns by Paragraph

5.2.1 By Paragraph in English

The findings on mean values of all degrees of participants' acceptability for all types of discourse anaphors referring to public figures in English news articles are correspondingly shown in Table 5.10 through Table 5.12 and Figure 5.8 through Figure 5.10 in terms of the first-, the second-, and the third paragraphs.

5.2.1.1 The First Paragraphs

First, the mean values of all degrees of participant acceptability for all types of referring expressions referring to public figures in the first paragraphs are shown in Table 5.10 and Figure 5.8.

Table 5.10 The mean values of participant acceptability for the first paragraphs in English

Mean by discourse anaphoric form / Factors			Mean			
			Form			
Condition	Language	Famousness (Section)	a_FN	b_SN	c_Pro	d_GN
Q1_1st P	English	high (C1 & C4)	**5.00**	2.875	1.5	1.95
		mid-high (C2 & C5)	**4.95**	2.45	1.525	1.9
		mid-low (C3 & C6)	**5.00**	2.35	1.425	1.725

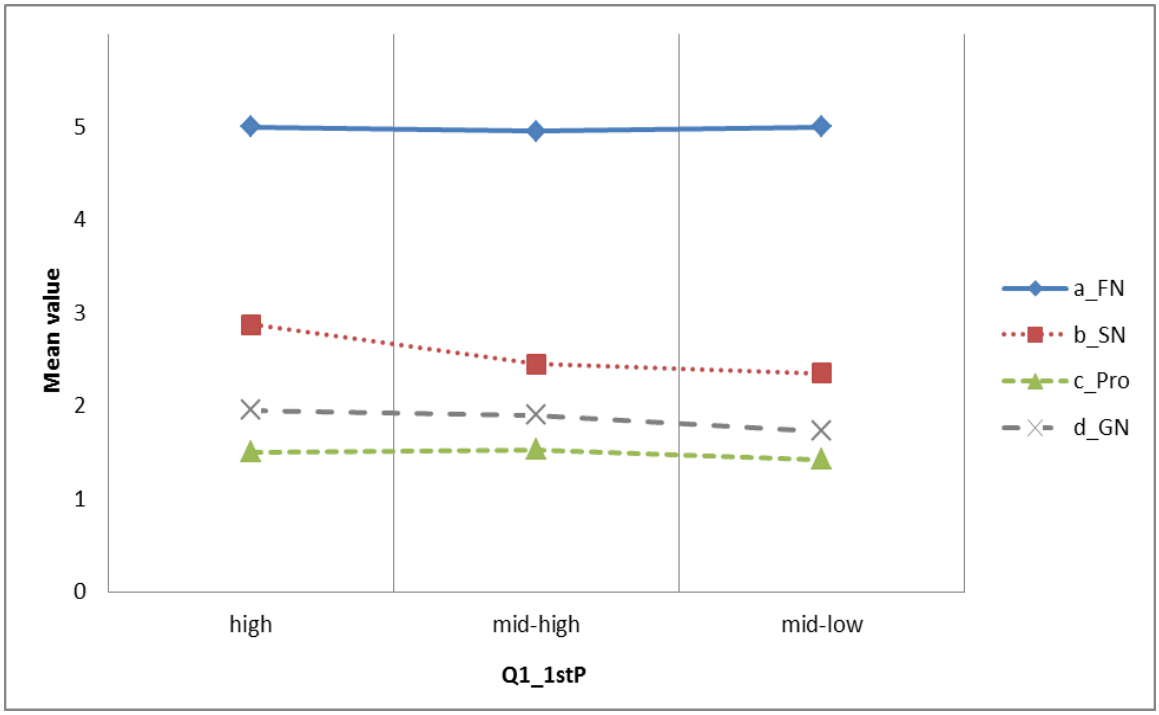


Figure 5.8 The mean values of participant acceptability for the first paragraphs in English

The red column in Table 5.10 and the blue line in Figure 5.8 show that like the first mentions in Section B, the degree of acceptability for full names in the first mentions (high, M=5; mid-high, M=4.95; mid-low, M=5), regardless of degree of famousness, are predominantly higher than that for other forms of referring expressions such as surnames (high, M=2.875; mid-high, M=2.45; mid-low, M=2.35), pronouns (high, M=1.5; mid-high, M=1.525; mid-low, M=1.425) and given names (high, M=1.95; mid-high, M=1.9; mid-low, M=1.725).

This finding conforms to Ariel's (1991, 1994, 2008, 2010) AH and Lee (2010) in that the first mentions are all referred to by the most discontinuous linguistic form, i.e., full name among other various constructions since the full name (plus modifier) is the lowest accessible marker.

For the first paragraphs, naming forms other than those suggested in the survey questionnaire were observed in terms of the structural constructions (frequency) and the examples, as shown in (40).

(40) a. TT + SN (41) Mr. Leno
b. NP (6) the comedian, the writer
c. TT + FN (5) Mr. Derek Jeter, Mr. Steve Jobs

As seen in (40), surnames with titles (41) were mainly written in by the participants; definite noun phrases (6) and full names with titles (5) were also observed.

5.2.2.2 The Second Paragraphs

Second, the mean values of all degrees of participant acceptability for all types of discourse anaphors referring back to public figures in the second paragraphs are shown in Table 5.11 and Figure 5.9.

Table 5.11 The mean values of participant acceptability for the second paragraphs in English

Mean by discourse anaphoric form / Factors			Mean			
			Form			
Condition	Language	Famousness (Section)	a_FN	b_SN	c_Pro	d_GN
Q2_2nd P	English	high (C1 & C4)	3.225	**4.75**	**3.625**	3.425
		mid-high (C2 & C5)	3.225	**4.35**	**3.9**	3.05
		mid-low (C3 & C6)	3.175	**4.45**	**3.7**	2.875

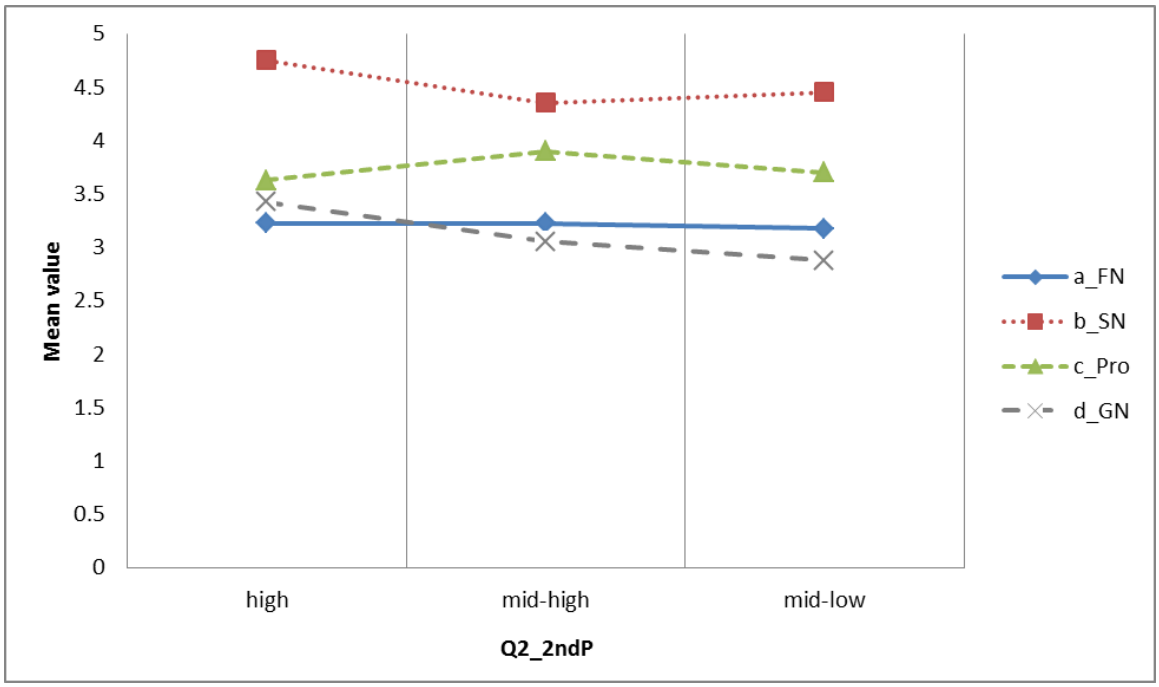


Figure 5.9 The mean values of participant acceptability for the second paragraphs in English

Table 5.12 and Figure 5.9 show that unlike the second mentions in Section B, the degree of acceptability for surnames (high, M=4.75; mid-high, M=4.35; mid-low, M=4.45) in the second paragraphs are relatively higher than and that for pronouns (high, M=3.625; mid-high, M=3.9; mid-low, M=3.7), which is the similar distribution for the second paragraphs in Corpus II (Section 4.2.1.2). However, the mean values of pronouns are still higher than other discourse anaphoric forms such as full names (high, M=3.225; mid-high, M=3.225; mid-low, M=3.175) and given names (high, M=3.425; mid-high, M=3.05; mid-low, M=2.875). In addition, the mean values of full names in the second paragraphs are relatively higher than those of full names in the second mentions in Section B. Like the second mentions in Section B, this finding also conforms to Ariel's

AH and Lee (2010) in that the most accessible referent is referred to by the most accessible forms such as pronouns and surnames in English.

For the second paragraphs, naming forms other than those suggested in the survey questionnaire were observed in terms of the structural constructions (frequency) and the examples, as shown in (41).

(41) a. TT + SN (41) Mr. Galassi
b. NP (9) the comedian, the writer

As seen in (41), surnames with titles (41) were mainly written in by the participants; definite noun phrases (9) were also observed.

5.2.3.3 The Third Paragraphs

Third, the mean values of all degrees of participant acceptability for all types of referring expressions referring back to public figures in the third paragraphs are shown in Table 5.12 and Figure 5.10.

Table 5.12 The mean values of participant acceptability for the third paragraphs in English

Mean by discourse anaphoric form / Factors			Mean			
			Form			
Condition	Language	Famousness (Section)	a_FN	b_SN	c_Pro	d_GN
Q3_3rd P	English	high (C1 & C4)	3.2	**4.6**	**4.1**	3.4
		mid-high (C2 & C5)	3.5	**4.3**	**3.95**	3.425
		mid-low (C3 & C6)	3.125	**4.6**	**3.6**	3.025

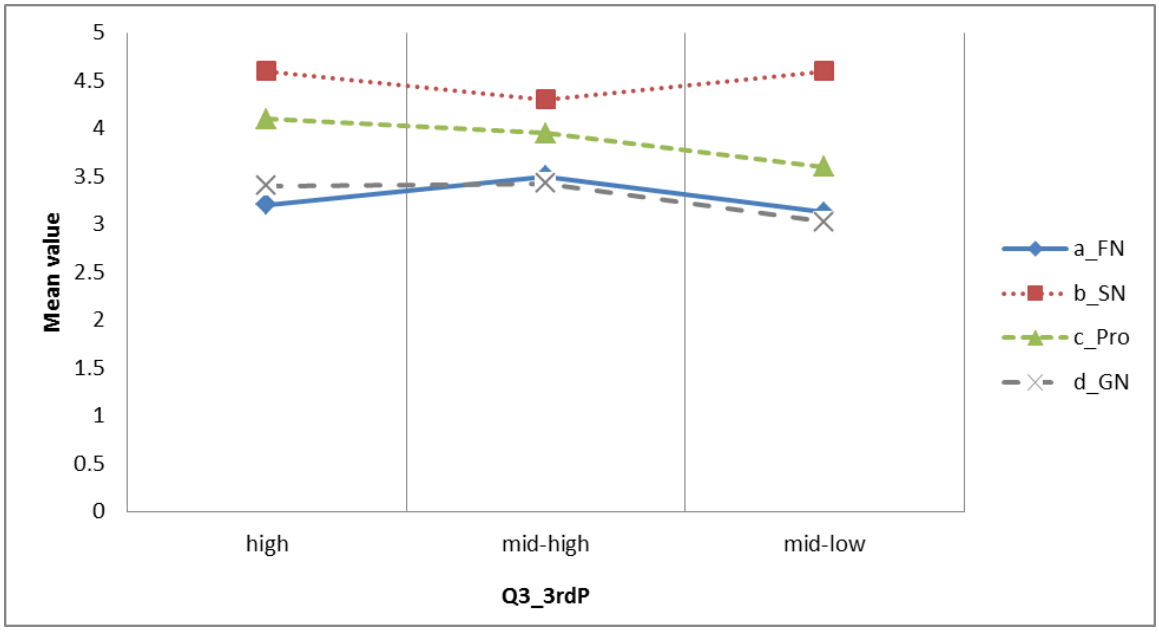


Figure 5.10 The mean values of participant acceptability for the third paragraphs in English

Table 5.12 and Figure 5.10 show that like the second paragraphs, the degree of acceptability for surnames (high, M=4.6; mid-high, M=4.3; mid-low, M=4.6) in the third paragraphs are relatively higher than and that for pronouns (high, M=4.1; mid-high, M=3.95; mid-low, M=3.6), which is also the similar distribution for the third paragraphs in Corpus II (Section 4.2.1.3). However, the mean values of pronouns are still higher than other discourse anaphoric forms such as full names (high, M=3.2; mid-high, M=3.5; mid-low, M=3.125) and given names (high, M=3.4; mid-high, M=3.425; mid-low, M=3.025). In addition, the mean values of full names in the third paragraphs are also relatively higher than those of full names in the third mentions in Section B. Like the third mentions in Section B and the second paragraphs in Section C, this finding also conforms

to Ariel's AH and Lee (2010) in that the most accessible referent is referred to by the most accessible forms such as pronouns and surnames in English.

For the third paragraphs, naming forms other than those suggested in the survey questionnaire were observed in terms of the structural constructions (frequency) and the examples, as shown in (42).

(42) a. TT + SN (43) Mr. Galassi, Ms. Beattie
b. NP (5) the singer, the comedian

As seen in (42), surnames with titles (43) were mainly written in by the participants; definite noun phrases (5) were also observed.

5.2.2 By Paragraph in Korean

The findings on mean values of all degrees of participants' acceptability for all types of discourse anaphors referring to public figures in Korean news articles are correspondingly shown in Table 5.13 through Table 5.15 and Figure 5.11 through Figure 5.13 in terms of the first-, the second-, and the third paragraphs.

5.2.2.1 The First Paragraphs

First, the mean values of all degrees of participant acceptability for all types of referring expressions referring to public figures in the first paragraphs are shown in Table 5.13 and Figure 5.11.

Table 5.13 The mean values of participant acceptability for the first paragraphs in Korean

Factors / Mean by discourse anaphoric form			Mean			
			Form			
Condition	Language	Famousness (Section)	a_FN	b_SN	c_Pro	d_GN
Q1_1st P	Korean	high (C1 & C4)	**4.975**	1.4	1.475	1.35
		mid-high (C2 & C5)	**4.975**	1.375	1.35	1.375
		mid-low (C3 & C6)	**5**	1.425	1.3	1.45

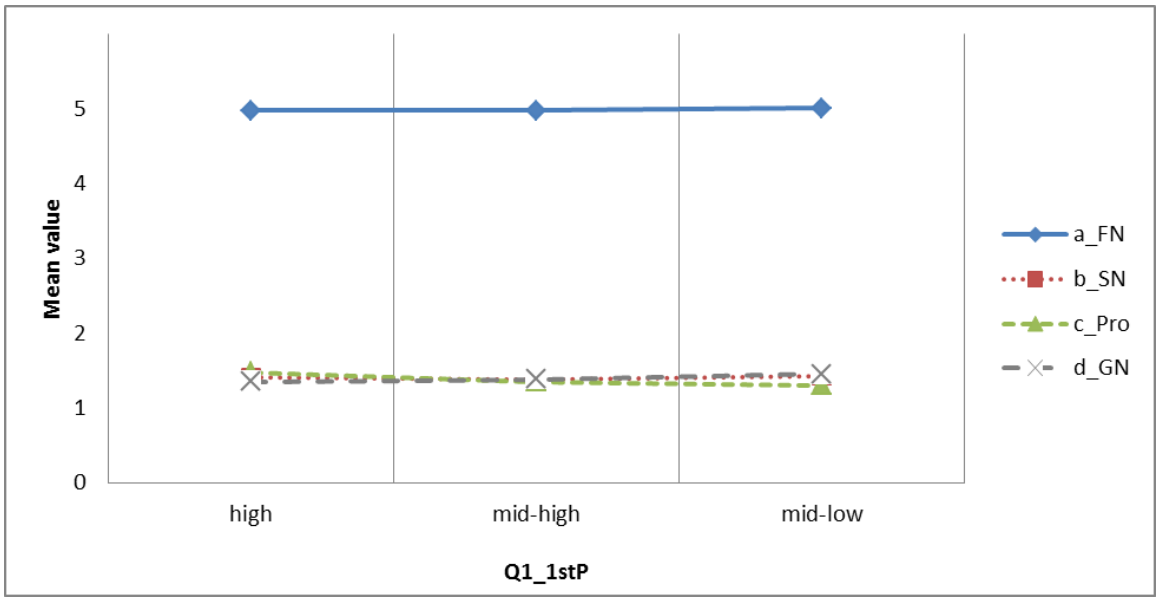


Figure 5.11 The mean values of participant acceptability for the first paragraphs in Korean

The red column in Table 5.13 and the blue line in Figure 5.11 show that like the mean values for the first mentions in English, the degree of acceptability for full names in the first mentions (high, M=4.975; mid-high, M=4.975; mid-low, M=5), regardless of degree of famousness, are predominantly higher than that for other forms of referring expressions such as surnames (high, M=1.4; mid-high, M=1.375; mid-low, M=1.425),

pronouns (high, M=1.475; mid-high, M=1.35; mid-low, M=1.3) and given names (high, M=1.35; mid-high, M=1.375; mid-low, M=1.45). In fact, the gap of the mean values between full names and other forms in Korean is also relatively wider than those in English.

This finding conforms to Ariel's AH and Lee (2010) in that the first mentions are all referred to by the most discontinuous linguistic form, i.e., full name among other various constructions since the full name (plus modifier) is the lowest accessible marker.

For the first paragraphs, naming forms other than those suggested in the survey questionnaire were observed in terms of the structural constructions (frequency) and the examples, as shown in (43).

(43) a. TT + FN (20) cakka Shin Kyung Sook 'writer'
b. SN + TT (6) Ahn ssi 'Mr.'
c. NN + FN (2) Korean thukkup Park Chan Ho 'express'

As seen in (43), it is noteworthy that full names with titles (20) were mainly written in by the participants; surnames with titles (6) and full names with nicknames (2) were also observed.

5.2.2.2 The Second Paragraphs

Second, the mean values of all degrees of participant acceptability for all types of discourse anaphors referring back to public figures in the second paragraphs are shown in Table 5.14 and Figure 5.12.

Table 5.14 The mean values of participant acceptability for the second paragraphs in Korean

Factors / Mean by discourse anaphoric form			Mean			
			Form			
Condition	Language	Famousness (Section)	a_FN	b_SN	c_Pro	d_GN
Q2_2nd P	Korean	high (C1 & C4)	**4.575**	1.9	**4.25**	1.325
		mid-high (C2 & C5)	**4.55**	1.825	**3.9**	1.6
		mid-low (C3 & C6)	**4.35**	1.85	**3.875**	1.325

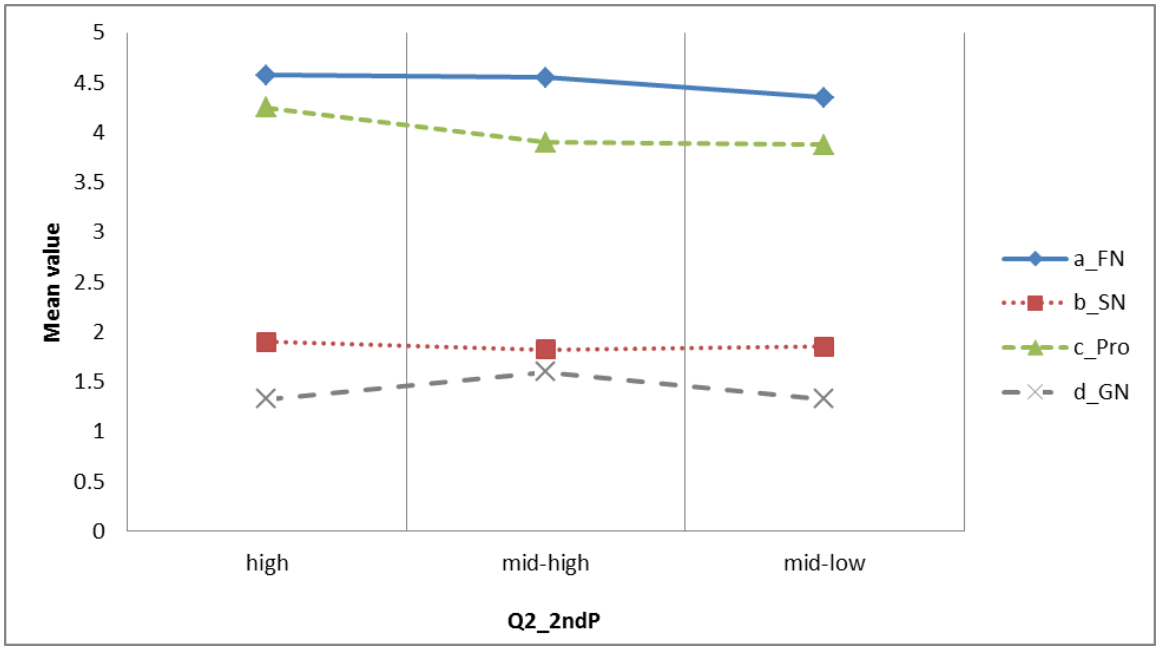


Figure 5.12 The mean values of participant acceptability for the second paragraphs in Korean

Table 5.14 and Figure 5.12 show that like the second mentions in Corpus II for Korean (Section 4.2.2.2), the degree of participants' acceptability for full names (high, M=4.575; mid-high, M=4.55; mid-low, M=4.35) in the second paragraphs is predominantly higher than other forms of discourse anaphors, regardless of the degree of famousness.

However, unlike Corpus II in terms of the distribution, it is noteworthy that the degree of acceptability for pronouns (high, M=4.25; mid-high, M=3.9; mid-low, M=3.875) in the second paragraphs is also relatively higher than other naming forms such as surnames (high, M=1.9; mid-high, M=1.825; mid-low, M=1.85) and given names (high, M=1.325; mid-high, M=1.6; mid-low, M=1.325) in terms of the selection.

This finding would definitely run counter to Lee's (2010) findings on English/Korean political news articles and Ariel's AH for English, in which the subsequent mentions are referred to by the more accessible linguistic markers such as pronouns and surnames, not by the less accessible markers such as full names (Kim 2010a, 2012).

For the second paragraphs, naming forms other than those suggested in the survey questionnaire were observed in terms of the structural constructions (frequency) and the examples, as shown in (44).

(44) a. SN + TT (19) Choi cakka 'writer'
b. FN + TT (13) soselka Choi In Ho 'novelist'
c. this + TT (4) i cakka 'this writer'

As seen in (44), like the second mentions, surnames with titles (19) were mainly written in; full names with titles (13) and 'this' plus titles (4) were also observed.

5.2.2.3 The Third Paragraphs

Third, the mean values of all degrees of participant acceptability for all types of discourse anaphors referring back to public figures in the third paragraphs are shown in Table 5.15 and Figure 5.13.

Table 5.15 The mean values of participant acceptability for the third paragraphs in Korean

Factors / Mean by discourse anaphoric form			Mean			
			Form			
Condition	Language	Famousness (Section)	a_FN	b_SN	c_Pro	d_GN
Q3_3rd P	Korean	high (C1 & C4)	**4.5**	1.85	**4.175**	1.45
		mid-high (C2 & C5)	**4.5**	1.875	**3.725**	1.65
		mid-low (C3 & C6)	**4.45**	1.775	**3.85**	1.325

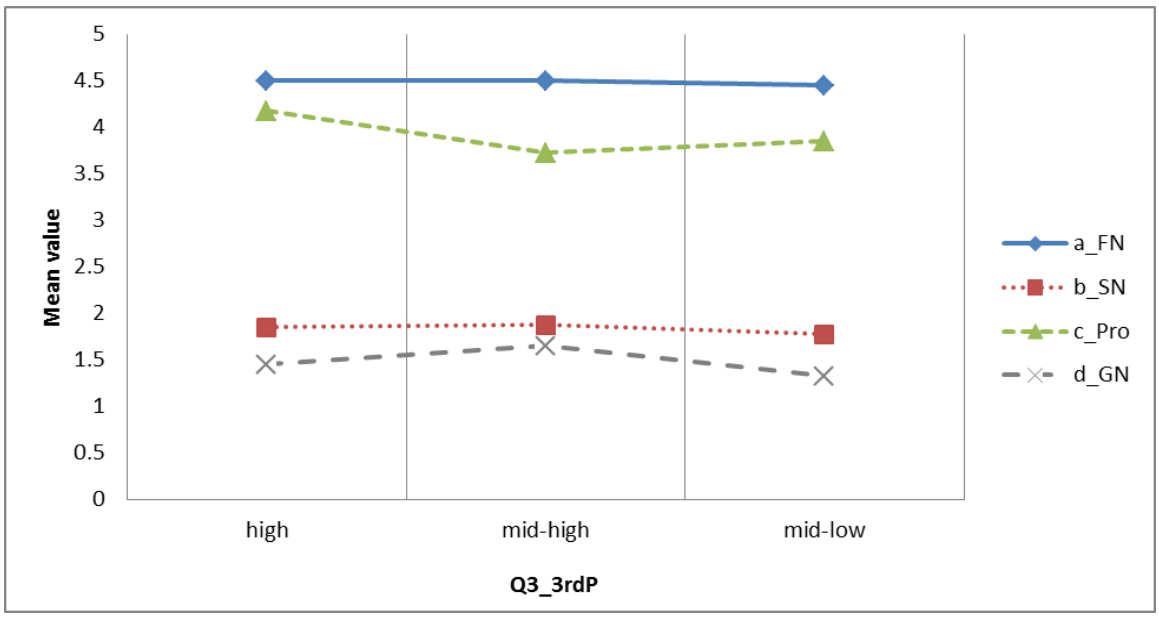


Figure 5.13 The mean values of participant acceptability for the third paragraphs in Korean

Table 5.15 and Figure 5.13 show that like the second/third mentions in Corpus II for Korean (Section 4.2.2.3) and the second mentions/paragraphs in the survey, the degree of participants' acceptability for full names (high, M=4.5; mid-high, M=4.5; mid-low, M=4.45) in the third paragraphs is predominantly higher than other forms of discourse

anaphors, regardless of the degree of famousness. However, unlike Corpus II in terms of the distribution, it is also noteworthy that the degree of acceptability for pronouns (high, M=4.175; mid-high, M=3.725; mid-low, M=3.85) in the third paragraphs is also relatively higher than other naming forms such as surnames (high, M=1.85; mid-high, M=1.765; mid-low, M=1.775) and given names (high, M=1.45; mid-high, M=1.65; mid-low, M=1.325) in terms of the selection.

This finding would also definitely run counter to Lee's (2010) findings on English/Korean political news articles and Ariel's AH for English mentioned in the previous section (Kim 2010a, 2012).

For the third paragraphs, naming forms other than those suggested in the survey questionnaire were observed in terms of the structural constructions (frequency) and the examples, as shown in (45).

(45) a. FN + TT (17) Park Chan Ho senswu 'player'
b. SN + TT (14) Choi cakka 'writer'
c. this + TT (4) i cakka 'this writer'

As seen in (45), like the third mentions, full names with titles (17) were mainly written in by the participants; surnames with titles (14) and 'this' plus titles (4) were also observed.

5.2.3 Interactions between Variables

Now let us consider the interactions between the independent variables (IVs) and the dependent variable (DV) in Section C. The results of ANOVA test are shown in Table 5.16.

Table 5.16 The results of ANOVA test for Section C

Response: rating ~	Sum Sq	Df	F value	Pr(>F)
condition Q1_1st Paragraph Q2_2nd Paragraph Q3_3rd Paragraph	441.1	3	137.6227	< 2.2e-16***[46]
famousness high mid-high mid-low	46.8	2	21.7567	< 4.151e-10***
language English Korean	243.2	1	226.0499	< 2.2e-16***
naming form a_Full Name (FN) b_Surname (SN) c_Pronoun (Pro) d_Given Name (GN)	1731.9	3	536.7050	< 2.2e-16***
condition:famousness	109.6	6	16.9758	< 2.2e-16***
condition:language	22.4	3	6.9306	< 0.0001201***
famousness:language	4.3	2	1.9794	0.1383308
condition:naming form	1057.1	6	163.7945	< 2.2e-16***
famousness:naming form	7.1	6	1.0997	0.3599119
language:naming form	979.3	3	303.4816	< 2.2e-16***
condition:famousness:langauge	12.8	6	1.9789	0.0652170.
condition:famousness:naming form	6.3	12	0.4872	0.9234610
condition:language:naming form	205.8	6	31.8850	< 2.2e-16***
famousness:language:naming form	11.3	6	1.7470	0.1061772
condition:famousness:language:naming form	7.5	12	0.5841	0.8567756
Residuals	3272.2	3042		

As seen in Table 5.16, concerning each single-factor repeated-measure ANOVA for Section C, there are significant main effects for rating (i.e., degree of acceptability). Condition (the first-, the second-, the third mention) as the independent variable (IV) show a significant main effect on rating as the dependent variable (DV), F(3, 3042) = 137.62, p <.001***. Famousness (high, mid-high, mid-low) as the IV show a significant

[46] Significance codes: 0;'***' 0.001; '**' 0.01; '*'0.05; '.'0.1'; ' ' 1

main effect on rating as the DV, $F(2, 3042) = 21.76$, $p<.001$***. Language (English, Korean) as the IV show a significant main effect on rating as the DV, $F(1, 3042) = 226.05$, $p<.001$***. Naming form (full name, surname, pronoun, given name) as the IV show a significant main effect on rating as the DV, $F(3, 3042) = 536.71$, $p<.001$***.

The results of two-factor repeated measure ANOVA in Section C are as follows. Condition and famousness as IVs show a significant main effect on rating as the DV, and significant interaction between two IVs for rating, $F(6, 3042) = 16.98$, $p<.001$***. Condition and language as IVs show a significant main effect on rating as the DV, and significant interaction between two IVs for rating, $F(3, 3042) = 6.93$, $p<.001$***. Condition and naming form as IVs show a significant main effect on rating as the DV, and significant interaction between two IVs for rating, $F(6, 3042) = 163.79$, $p<.001$***. Famousness and naming form as IVs show no significant main effect on rating as the DV, and no significant interaction between two IVs for rating in Section C, unlike those in Section B. Language and naming form as IVs show a significant main effect on rating as the DV, and significant interaction between two IVs for rating, $F(3, 3042) = 303.48$, $p<.001$***.

The results of more than two-factor repeated measure ANOVA in Section C are as follows. Condition, famousness, and language as IVs show no significant main effect on rating as the DV, and no significant interaction between three IVs for rating in Section C, unlike those in Section B. Condition, language, and naming form as IVs show a significant main effect on rating as the DV, and significant interaction between three IVs for rating, $F(6, 3042) = 31.89$, $p<.001$***.

In addition, as seen in more than two-factor repeated measure ANOVA, condition other than factors might predict the effects on rating. Thus, this ANOVA needs to be broken down into subanalysis ANOVA by condition, i.e., the first-, the second-, and the third paragraph, as shown in Table 5.17.

Table 5.17 The results of subanalysis ANOVA test by condition for Section C

Response: rating for 1st P ~	Sum Sq	Df	F value	Pr(>F)
famousness	1.94	2	1.5952	0.2034
language	45.07	1	74.1278	<2e-16***
naming form	2004.11	3	1098.817	<2e-16***
famousness:language	1.60	2	1.3176	0.2683
famousness:naming form	1.96	6	0.5374	0.7800
language:naming form	49.21	3	26.9801	<2e-16***
famousness:language:naming form	3.05	6	0.8356	0.5424
Residuals	569.05	936		
Response: rating for 2nd P ~	Sum Sq	Df	F value	Pr(>F)
famousness	5.45	2	2.1559	0.1164
language	118.30	1	93.5593	<2e-16***
naming form	411.23	3	108.4077	<2e-16***
famousness:language	0.26	2	0.1046	0.9007
famousness:naming form	2.31	6	0.3040	0.9350
language:naming form	582.19	3	153.4756	<2e-16***
famousness:language:naming form	10.31	6	1.3590	0.2283
Residuals	1183.53	936		
Response: rating for 3rd P ~	Sum Sq	Df	F value	Pr(>F)
famousness	6.38	2	2.3687	0.09417 .
language	156.82	1	116.4185	<2e-16***
naming form	373.74	3	92.4858	<2e-16***
famousness:language	0.63	2	0.2328	0.79238
famousness:naming form	9.12	6	1.1283	0.34365
language:naming form	553.73	3	137.0259	<2e-16***
famousness:language:naming form	5.46	6	0.6751	0.66984
Residuals	1260.80	936		

As seen in Table 5.17, the results of subanalysis ANOVA by condition for Section C are as follows. For a factor, language and naming form as IVs show a statistically significant main effect on rating for each order of mention as the DV. For more than two factors, it is worth noting that there is a statistically significant interaction between language and naming form as IVs on rating for each order of mention as DVs, i.e., for the first paragraph, $F(3, 936) = 26.98$, $p<.001$***; for the second paragraph, $F(3, 936) = 153,48$, $p<.001$***; for the third paragraph, $F(3, 936) = 137.03$, $p<.001$***.

5.2.4 Section Summary

All mean values of degrees of participants' acceptability for all types of discourse anaphors in Section C (i.e., by paragraph) by native speakers of English and Korean are summarized in Table 5.18 and Figure 5.14.

Table 5.18 Summary of Section C for English and Korean

Factors / Mean by discourse anaphoric form			Mean Form			
Condition	Language	Famousness (Section)	a_FN	b_SN	c_Pro	d_GN
Q1_1st P	English	high (C1 & C4)	**5.00**[47]	2.875	1.5	1.95
		mid-high (C2 & C5)	**4.95**	2.45	1.525	1.9
		mid-low (C3 & C6)	**5.00**	2.35	1.425	1.725
	Korean	high (C1 & C4)	**4.975**	1.4	1.475	1.35
		mid-high (C2 & C5)	**4.975**	1.375	1.35	1.375
		mid-low (C3 & C6)	**5**	1.425	1.3	1.45
Q2_2nd P	English	high (C1 & C4)	3.225	**4.75**	**3.625**	3.425
		mid-high (C2 & C5)	3.225	**4.35**	**3.9**	3.05
		mid-low (C3 & C6)	3.175	**4.45**	**3.7**	2.875
	Korean	high (C1 & C4)	**4.575**	1.9	**4.25**	1.325
		mid-high (C2 & C5)	**4.55**	1.825	**3.9**	1.6
		mid-low (C3 & C6)	**4.35**	1.85	**3.875**	1.325
Q3_3rd P	English	high (C1 & C4)	3.2	**4.6**	**4.1**	3.4
		mid-high (C2 & C5)	3.5	**4.3**	**3.95**	3.425
		mid-low (C3 & C6)	3.125	**4.6**	**3.6**	3.025
	Korean	high (C1 & C4)	**4.5**	1.85	**4.175**	1.45
		mid-high (C2 & C5)	**4.5**	1.875	**3.725**	1.65
		mid-low (C3 & C6)	**4.45**	1.775	**3.85**	1.325

[47] The number in red color represents the highest mean value, while the number in blue color represents the next one in the dissertation.

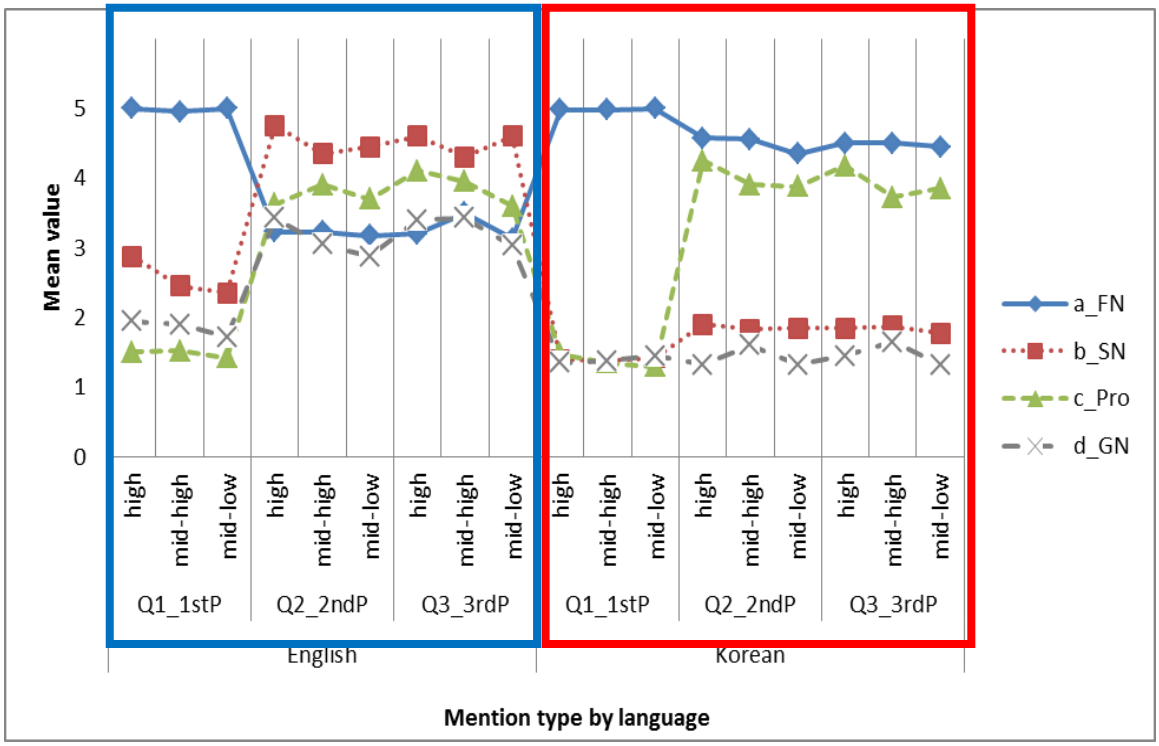


Figure 5.14 Summary of Section C for English and Korean

As seen in Table 5.18 and Figure 5.14, in Section C (discourse anaphoric patterns by paragraph), for the first paragraph, both English and Korean survey show that the mean values of the full names are the highest other than naming forms in terms of degree of acceptability. For the second paragraph, the English survey shows that surnames are the most preferred form and surnames are also acceptable, while the Korean survey shows that full names are the most preferred form and pronouns are also acceptable. For the third paragraph, the English survey shows that surnames are the most preferred form and surnames are also acceptable, while the Korean survey shows that full names are the most preferred form and pronouns are also acceptable, like the second paragraph.

5.3 Section D: Sentential Anaphors

5.3.1 New vs. Old Information

Next, the findings on the mean values of participant acceptability for all sentential anaphors in Section D are shown in terms of the givenness of information (i.e., new vs. old information) in English and Korean.

5.3.1.1 In English

First, the man values of participant acceptability for English sentential anaphor *that* and *it* are shown in Table 5.19 and Figure 5.15 in terms of the givenness of information; i.e., new vs. old information.

Table 5.19 The mean values of participant acceptability for sentential anaphors (givenness) in English

Mean by sentential anaphoric form / Factors			Mean	
			Form	
Condition		Language (Section)	a_*that*	b_*it*
Givenness	Q1_NI	English (D1)	**4.85**	2.45
	Q2_OI	(D2)	2.75	**4.95**

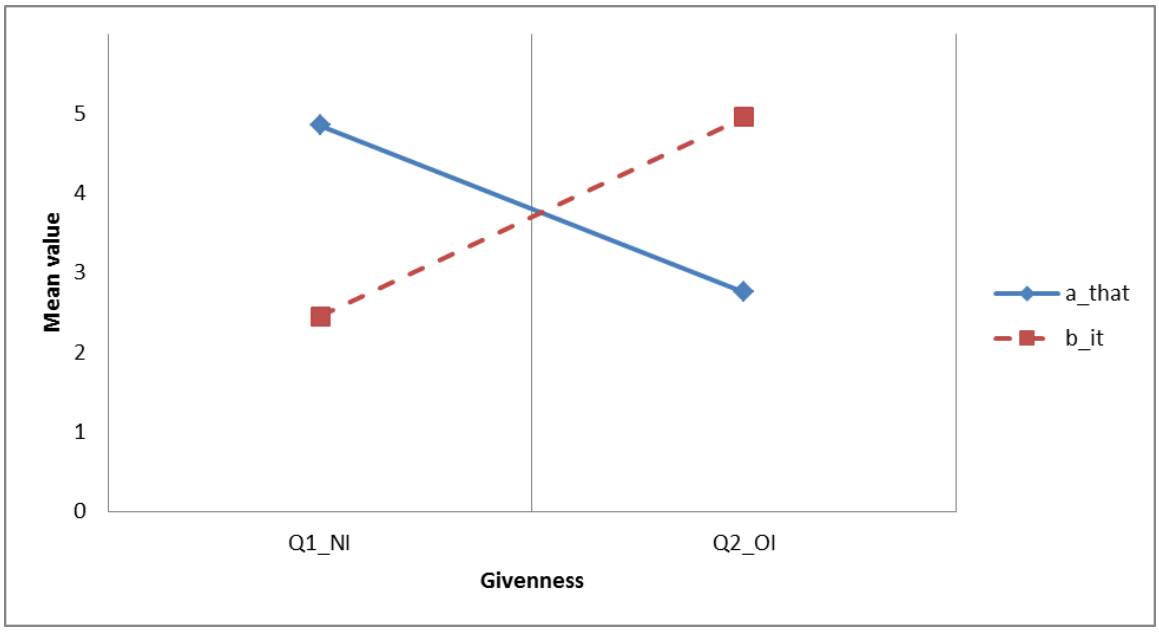


Figure 5.15 The mean values of participant acceptability for sentential anaphors (givenness) in English

As seen in Table 5.19 and Figure 5.15, for the new information, the mean values of the acceptability for the sentential anaphor *that* (M=4.85) are predominantly higher than those for the sentential anaphor *it* (M=2.45); for the old information, the mean values of the acceptability for the sentential anaphor *it* (M=4.95) are predominantly higher than those for the sentential anaphor *that* (M=2.45).

5.3.1.2 In Korean

Second, the man values of participant acceptability for Korean sentential anaphor *ku-ken* 'that' and *ku-len* 'it' are shown in Table 5.20 and Figure 5.16 in terms of the givenness of information; i.e., new vs. old information.

Table 5.20 The mean values of participant acceptability for sentential anaphors (givenness) in Korean

Factors / Mean by sentential anaphoric form			Mean	
			Form	
Condition		Language (Section)	a_*ku-ken* 'that'	b_*ku-len* 'it'
Givenness	Q1_NI	Korean (D1)	**5**	2.25
	Q2_OI	(D2)	2.65	**4.95**

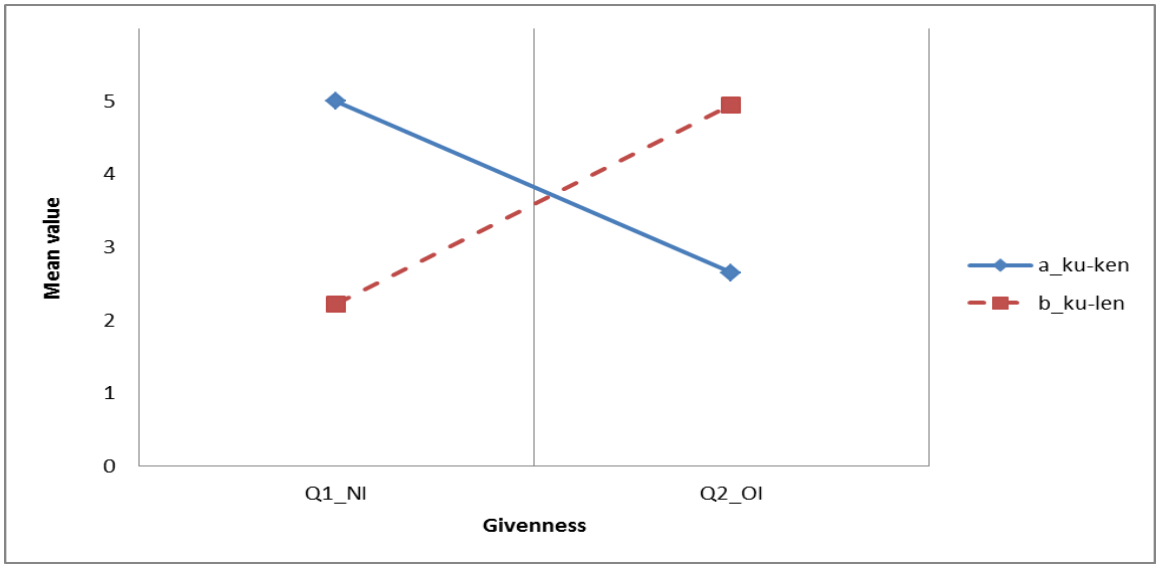


Figure 5.16 The mean values of participant acceptability for sentential anaphors (givenness) in Korean

As seen in Table 5.20 and Figure 5.16, like in English, for the new information, the mean values of the acceptability for the sentential anaphor *ku-ken* 'that' (M=5) are predominantly higher than those for the sentential anaphor *ku-len* 'it' (M=2.25); for the old information, the mean values of the acceptability for the sentential anaphor *ku-len* 'it' (M=4.95) are predominantly higher than those for the sentential anaphor *ku-ken* 'that' (M=2.65).

5.3.1.3 Section Summary

The mean values of participant acceptability for all sentential anaphors in Section D for English and Korean are summarized in terms of the givenness of information (new vs. old information) in Table 5.21.

Table 5.21 Summary of the mean values in Section D (givenness) for English and Korean

Mean by sentential anaphoric form / Factors			Mean	
			Form	
Condition		Language (Section)	a_*that*	b_*it*
			a_*ku-ken* 'that'	b_*ku-len* 'it'
Givenness	Q1_NI	English (D1)	**4.85**	2.45
	Q2_OI	(D2)	2.75	**4.95**
	Q1_NI	Korean (D1)	**5**	2.25
	Q2_OI	(D2)	2.65	**4.95**

As seen in Table 5.21, for Corpus III (the givenness of information), *that* in English and *ku-ken* 'that' in Korean are preferred for new information, whereas *it* in English and *ku-len* 'it' in Korean are preferred for old information.

5.3.2 Narrow vs. Wide Reference

The findings on the mean values of participant acceptability for all sentential anaphors in Section D are shown in terms of the width of reference (i.e., narrow vs. wide reference) in English and Korean.

5.3.2.1 In English

First, the man values of participant acceptability for English sentential anaphor *that* and *it* are shown in Table 5.22 and Figure 5.17 in terms of the width of reference; i.e., narrow vs. wide reference.

Table 5.22 The mean values of participant acceptability for sentential anaphors (width) in English

Factors / Mean by sentential anaphoric form			Mean	
			Form	
Condition		Language (Section)	a_*that*	b_*it*
Width	Q3_NR	English (D3)	**4.85**	2.05
	Q4_WR	(D4)	3.2	**4.35**

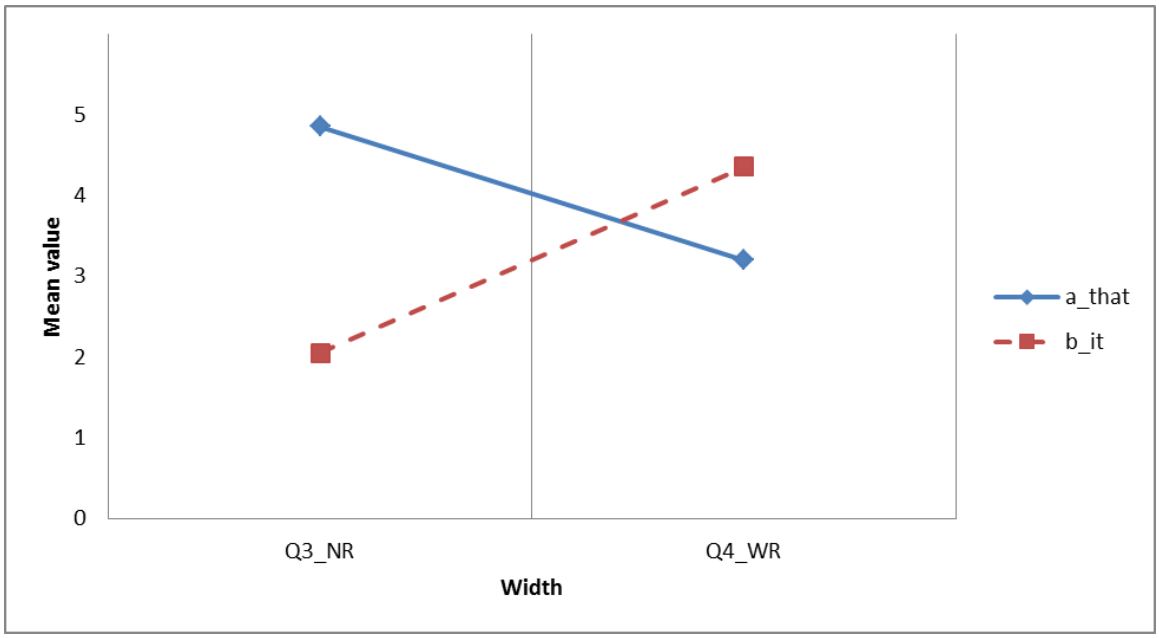


Figure 5.17 The mean values of participant acceptability for sentential anaphors (width) in English

As seen in Table 5.22 and Figure 5.17, for the narrow reference, the mean values of the acceptability for the sentential anaphor *that* (M=4.85) are predominantly higher than those for the sentential anaphor *it* (M=2.05); for the wide reference, the mean values of the acceptability for the sentential anaphor *it* (M=4.35) are relatively higher than those for the sentential anaphor *that* (M=3.2).

5.3.2.2 In Korean

Second, the man values of participant acceptability for Korean sentential anaphor *ku-ken* 'that' and *ku-len* 'it' are shown in Table 5.23 and Figure 5.18 in terms of the width of reference; i.e., narrow vs. wide reference.

Table 5.23 The mean values of participant acceptability for sentential anaphors (width) in Korean

Factors / Mean by sentential anaphoric form			Mean	
			Form	
Condition		Language (Section)	a_*ku-ken* 'that'	b_*ku-len* 'it'
Width	Q3_NR	Korean (D3)	**4.9**	2.25
	Q4_WR	(D4)	1.5	**4.9**

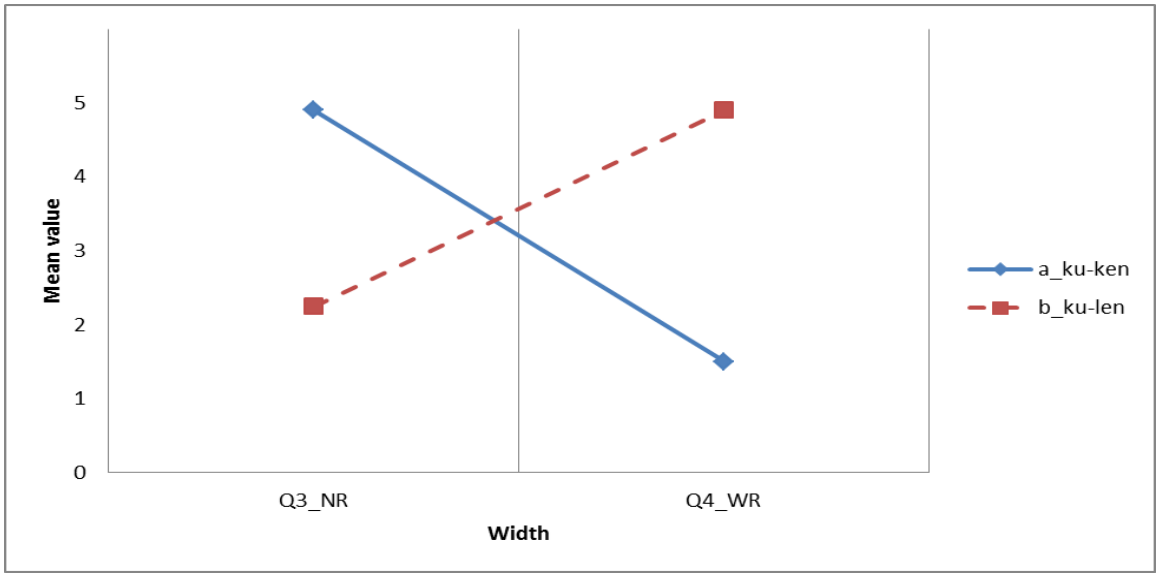


Figure 5.18 The mean values of participant acceptability for sentential anaphors (width) in Korean

As seen in Table 5.23 and Figure 5.18, like in English, for the narrow reference, the mean values of the acceptability for the sentential anaphor *ku-ken* 'that' (M=4.9) are predominantly higher than those for the sentential anaphor *ku-len* 'it' (M=2.25); for the wide reference, the mean values of the acceptability for the sentential anaphor *ku-len* 'it' (M=4.9) are predominantly higher than those for the sentential anaphor *ku-ken* 'that'

(M=1.5), in which the gap between two sentential anaphors is relatively wider than that in English in the previous section in terms of wide reference.

5.3.2.3 Section Summary

The mean values of participant acceptability for all sentential anaphors in Section D for English and Korean are summarized in terms of the width of reference (narrow vs. wide reference) in Table 5.24.

Table 5.24 Summary of the mean values in Section D (width) for English and Korean

Factors \ Mean by sentential anaphoric form			Mean	
			Form	
Condition		Language (Section)	a_*that* a_*ku-ken* 'that'	b_*it* b_*ku-len* 'it'
Width	Q3_NR	English (D3)	**4.85**	2.05
	Q4_WR	(D4)	3.2	**4.35**
	Q3_NR	Korean (D3)	**4.9**	2.25
	Q4_WR	(D4)	1.5	**4.9**

As seen in Table 5.24, for Corpus III (the width of reference), *that* in English and *ku-ken* 'that' in Korean are preferred for narrow reference, whereas *it* in English and *ku-len* 'it' in Korean are preferred for wide reference.

5.3.3 Interactions between Variables

5.3.3.1 The Givenness of Information

Now let us consider the interactions between the independent variables (IVs) and the dependent variable (DV) in the first part of Section D. The result of ANOVA test is shown in Table 5.25.

Table 5.25 The result of ANOVA test for Section D1

Response: rating ~	Sum Sq	Df	F value	Pr(>F)
condition Q1_New Information (NI) Q2_Old Information (OI)	1.406	1	2.8330	0.0944 .[48]
language English Korean	0.056	1	0.1133	0.7369
pronoun form a_*that* (*ku-ken*) b_*it* (*ku-len*)	1.056	1	2.1279	0.1467
condition:language	0.006	1	0.0126	0.9108
condition:pronoun form	232.806	1	469.0066	<.2e-16***
language:pronoun form	0.156	1	0.3148	0.5756
condition:language:pronoun form	0.506	1	1.0199	0.3142
Residuals	75.450	152		

As seen in Table 5.25, concerning each single-factor repeated-measure ANOVA in Section D1, no significant main effects for rating (i.e., degree of acceptability) was found. On the other hand, concerning more than two-factor repeated measure ANOVA in

[48] Significance codes: 0;'***' 0.001; '**' 0.01; '*'0.05; '.'0.1'; ' ' 1

Section D1, condition and pronoun form as IVs show a statistically significant main effect on rating as the DV, and significant interaction between two IVs for rating, F(1, 152) = 469.01, p<.001***, which indicates there is a marked difference between English and Korean in terms of the givenness of information, i.e., new information and old information.

5.3.3.2 The Width of Reference

Now let us consider the interactions between the independent variables (IVs) and the dependent variable (DV) in the second part of Section D. The result of ANOVA test is shown in Table 5.26.

Table 5.26 The result of ANOVA test for Section D2

Response: rating ~	Sum Sq	Df	F value	Pr(>F)
condition Q3_Narrow Reference (NR) Q4_Wide Reference (WR)	0.025	1	0.0276	0.8682350[49]
language English Korean	2.025	1	2.2369	0.1368226
pronoun form a_*that* (*ku-ken*) b_*it* (*ku-len*)	2.025	1	2.2369	0.1368226
condition:language	4.900	1	5.4128	0.0213104*
condition:pronoun form	250.000	1	276.1628	<2.2e-16***
language:pronoun form	14.400	1	15.9070	0.0001031***
condition:language:pronoun form	11.025	1	12.1788	0.0006327***
Residuals	75.450	152		

As seen in Table 5.26, concerning each single-factor repeated-measure ANOVA in Section D2, no significant main effects for rating (i.e., degree of acceptability) was also

[49] Significance codes: 0;'***' 0.001; '**' 0.01; '*'0.05; '.'0.1'; ' ' 1

found. On the other hand, concerning more than two-factor repeated measure ANOVA in Section D2, condition and pronoun form as IVs show a statistically significant main effect on rating as the DV, and significant interaction between two IVs for rating, $F(1, 152) = 276.16$, $p<.001$***, which indicates there is a marked difference between English and Korean in terms of the width of reference, i.e., narrow reference and wide reference. In addition, there are also statistically significant interactions of condition and language: $F(1, 152) = 5.41$, $p<.05$*; language and pronoun form: $F(1, 152) = 15.91$, $p<.001$***; condition, language, and pronoun form: $F(1, 152) = 12.18$, $p<.001$***.

5.4 Chapter Summary

5.4.1 Section B and C: Discourse Anaphoric Patterns in English and Korean

The results of Section B (by order of mention) and Section D (by paragraph) for discourse anaphoric patterns in English and Korean analyzed in Section 5.1 and Section 5.2 are summarized by mention types, languages, anaphoric forms, the mean values in Table 5.27 and Table 5.28.

Table 5.27 Summary of the mean values in Section B for English and Korean

Factors \ Mean by discourse anaphoric form			Mean			
			Form			
Condition	Language	Famousness (Section)	a_FN	b_SN	c_Pro	d_GN
Q1_1st M	English	high (B1 & B4)	**5.00**	2.9	1.675	2.125
		mid-high (B2 & B5)	**4.95**	2.725	1.675	2.05
		mid-low (B3 & B6)	**5.00**	2.475	1.45	1.85
	Korean	high (B1 & B4)	**4.95**	1.55	1.8	1.475
		mid-high (B2 & B5)	**4.95**	1.4	1.65	1.575
		mid-low (B3 & B6)	**4.95**	1.45	1.45	1.525
Q2_2nd M	English	high (B1 & B4)	2.775	**4.6**	**4.475**	3.7
		mid-high (B2 & B5)	2.925	**4.275**	**4.35**	3.725
		mid-low (B3 & B6)	2.9	**4.525**	**4.225**	2.925
	Korean	high (B1 & B4)	**4.325**	2.025	**4.425**	1.525
		mid-high (B2 & B5)	**4.575**	1.875	**3.875**	1.775
		mid-low (B3 & B6)	**4.3**	1.8	**3.775**	1.45
Q3_3rd M	English	high (B1 & B4)	2.85	**4.35**	**4.25**	3.75
		mid-high (B2 & B5)	3.075	**4.475**	**4.35**	3.75
		mid-low (B3 & B6)	2.65	**4.3**	**4.3**	3.35
	Korean	high (B1 & B4)	**4.375**	1.975	**4.325**	1.575
		mid-high (B2 & B5)	**4.475**	1.650	**3.95**	1.75
		mid-low (B3 & B6)	**4.4**	1.675	**3.825**	1.4

Table 5.28 Summary of the mean values in Section C for English and Korean

Mean by discourse anaphoric form / Factors			Mean			
			Form			
Condition	Language	Famousness (Section)	a_FN	b_SN	c_Pro	d_GN
Q1_1st P	English	high (C1 & C4)	**5.00**	2.875	1.5	1.95
		mid-high (C2 & C5)	**4.95**	2.45	1.525	1.9
		mid-low (C3 & C6)	**5.00**	2.35	1.425	1.725
	Korean	high (C1 & C4)	**4.975**	1.4	1.475	1.35
		mid-high (C2 & C5)	**4.975**	1.375	1.35	1.375
		mid-low (C3 & C6)	**5**	1.425	1.3	1.45
Q2_2nd P	English	high (C1 & C4)	3.225	**4.75**	**3.625**	3.425
		mid-high (C2 & C5)	3.225	**4.35**	**3.9**	3.05
		mid-low (C3 & C6)	3.175	**4.45**	**3.7**	2.875
	Korean	high (C1 & C4)	**4.575**	1.9	**4.25**	1.325
		mid-high (C2 & C5)	**4.55**	1.825	**3.9**	1.6
		mid-low (C3 & C6)	**4.35**	1.85	**3.875**	1.325
Q3_3rd P	English	high (C1 & C4)	3.2	**4.6**	**4.1**	3.4
		mid-high (C2 & C5)	3.5	**4.3**	**3.95**	3.425
		mid-low (C3 & C6)	3.125	**4.6**	**3.6**	3.025
	Korean	high (C1 & C4)	**4.5**	1.85	**4.175**	1.45
		mid-high (C2 & C5)	**4.5**	1.875	**3.725**	1.65
		mid-low (C3 & C6)	**4.45**	1.775	**3.85**	1.325

As seen in Table 5.27 for Section B (discourse anaphoric patterns by order of mention), for the first mention, both English and Korean survey show that the mean values of the full names are the highest other than naming forms in terms of degree of acceptability. For the second and the third mention, the English survey shows that surnames are the most preferred form and surnames are also acceptable, while the Korean survey shows that full names are the most preferred form and pronouns are also acceptable.

In Table 5.28 for Section C (discourse anaphoric patterns by paragraph), for the first paragraph, both English and Korean survey show that the mean values of the full names are the highest other than naming forms in terms of degree of acceptability. For the second and the third paragraph, the English survey shows that surnames are the most preferred form and surnames are also acceptable, while the Korean survey shows that full names are the most preferred form and pronouns are also acceptable.

To sum up, in Section B and Section C, the English survey shows the same selection in terms of the degree of acceptability for the preferred anaphoric forms: full names for the first mention/paragraph; surnames and pronouns for the second mention/paragraph; surnames and pronouns for the third mention/paragraph. The Korean survey also shows the same selection in terms of the degree of the acceptability for the preferred anaphoric forms: full names for all mentions/paragraphs; pronouns for the second mention/paragraph and for the third mention/paragraph.

5.4.2 Section D: Sentential Anaphors in English and Korean

The results of Section D for sentential anaphors in English and Korean analyzed in Section 5.3 are summarized by referential conditions, languages, anaphoric forms, the mean values in Table 5.29.

Table 5.29 Summary of the mean values in Section D for English and Korean

Mean by sentential anaphoric form / Factors			Mean	
			Form	
Condition		Language (Section)	a_*that* a_ku-ken 'that'	b_*it* b_ku-len 'it'
Givenness	Q1_NI	English (D1)	**4.85**	2.45
	Q2_OI	(D2)	2.75	**4.95**
	Q1_NI	Korean (D1)	**5**	2.25
	Q2_OI	(D2)	2.65	**4.95**
Width	Q3_NR	English (D3)	**4.85**	2.05
	Q4_WR	(D4)	3.2	**4.35**
	Q3_NR	Korean (D3)	**4.9**	2.25
	Q4_WR	(D4)	1.5	**4.9**

As seen in Table 5.29, for Corpus III, *that* in English and *ku-ken* 'that' in Korean are preferred for new information/narrow reference, whereas *it* in English and *ku-len* 'it' in Korean are preferred for old information/wide reference.

In the following chapter, I give a fuller discussion of the findings of the corpus patterns and the survey results, concerning the distribution and the selection of two discourse anaphoric features: discourse anaphoric patterns and sentential anaphors.

Chapter 6

Discussion and Conclusions

Based on the results and analyses in Chapter 4 (corpus) and Chapter 5 (survey), this chapter discusses the theoretical implication of discourse anaphoric patterns and referential properties of sentential anaphors in English and Korean in terms of the distribution and the selection. Section 6.1 addresses two discourse anaphoric patterns in terms of the summary of mention types (6.1.1), the distance between discourse anaphors (6.1.2), and relevant theoretical arguments such as socio-cultural, linguistic, pragmatic, and cognitive points (6.1.3). Section 6.2 discusses the sentential anaphors in terms of referential properties (6.2.1) and different properties in both languages (6.2.2). Section 6.3 generalizes those linguistic features within the purview of pragmatic heuristics, put forward by the neo-Gricean theory (6.3.1 for discourse anaphoric patterns; 6.3.2 for sentential anaphors). Section 6.4 addresses some implicational benefits of this research, based on the findings in the corpus and the survey results. Section 6.5 addresses future avenues of research that need to be explored in order to overcome some limitations.

6.1 Two Discourse Anaphoric Patterns: GAP and SAP

Discourse anaphors used in the corpus data and answers in the survey questionnaire have been examined. Let us now compare and contrast the English and Korean data in the corpus and the survey in terms of the distribution and the selection by the order of mention and those by the paragraph.

6.1.1 Mention Type and Distance between Discourse Anaphors

6.1.1.1 By Order of Mention

For the first mentions in the corpus, English corpus employs five different types of structural constructions, while Korean corpus employs eight types. In particular, English corpus contains four types of full names and pronouns, while Korean corpus all involves full names. Pronouns (only 2 occurrences) are used for the first mentions only in English corpus. Nevertheless, for the first mentions, the vast majority of English corpus deploys full name constructions with (pseudo-) titles, appositions, and nickname, while Korean corpus all deploys full name constructions with plus (pseudo-) titles, affiliations, nicknames as well. In fact, it was found that full names only in English and (pseudo-) titles plus full name constructions in Korean are preferred for the first mentions in each language. In addition, no full name plus affiliation construction was found in English, while no full name plus apposition construction was found in Korean for the first mentions. The majority of full names in Korean were also referred to with the nominative case markers.

For the first mentions in the survey, English survey shows almost the same selection. The mean values of full names were the highest among other naming forms in terms of the degree of acceptability in the first mentions. Likewise, Korean survey shows the same selection, in which full names were also the most preferred naming forms in terms of the participant acceptability.

Concerning the distribution and the selection, the predominant uses of full name constructions for the first mentions are attributed to the fact that in news articles, famous

public figures as topical referents should be identified by their full names at the beginning of a given discourse, in which they are not less accessible and less attentive at there in terms of memory and attention (Givón 1983; Ariel 1988, 1991, 1994, 2008, 2010; Gundel et al. 1988, 1989, 1990, 1993, 2001; Mulkern 1996).

For the second mentions in the corpus, English corpus employs three different types of structural constructions such as pronouns, surnames (plus titles), and lexical noun phrases, while Korean corpus employs five different types such as full names, surnames plus titles, pronouns, zero anaphors, and lexical noun phrases. No full name and zero anaphor were found in English, while no surname only use was found in Korean. The use of surname plus title in Korean is attributive to a social convention that news actors, particularly famous public figures, considered publicly honorific reference in terms of politeness, for which Lee (2010:3517) names "the semi-compulsory use of title for public figures." The majority of full names in Korean were retrieved with topical markers.

Above all, it is noteworthy that although there were uses of surname constructions, the predominant form in English is pronouns, while the predominant form in Korean is the full name constructions. The majority uses of pronouns in English and a few uses of zero anaphors in Korean can be accounted for in terms of the high accessibility of the topical referents since they are mostly referred to in the same sentences and in the next sentences by pronouns (Section 4.3.1), which conforms to Givón, Fox, Ariel, and Gundel et al.'s viewpoints. On the contrary, although there were some uses of other anaphoric forms, the vast majority uses in the Korean corpus deployed

full name constructions even with the high accessibility of the referents in the same sentences and in the previous sentences for the second mentions (Section 4.3.2), which is against the topic continuity theory, the hierarchy theory, the accessibility hierarchy, and the givenness hierarchy, put forward by Givón, Fox, Ariel, and Gundel, respectively.

For the second mentions in the survey, the English survey results showed almost the same selection, with a relatively high degree of acceptability of pronouns. However, the highest mean values indicate that the degree of acceptability for surnames was highly acceptable as well for the second mentions. On the other hand, uses in the Korean survey, like the second mentions in the corpus, show that full names were the most preferred forms. In addition, it is noteworthy that the mean values of pronouns were also higher than other forms such as surnames and given names in terms of the participant selection. In English, the most preferred anaphoric forms were surnames and the next ones were pronouns, while in Korean the most preferred forms were full names and the next ones were pronouns.

For the third mentions in the corpus, English corpus employs four different types of structural constructions such as pronouns, surnames (plus titles), and lexical noun phrases, and full names (only 2 occurrence after previous pronouns), while Korean corpus employs five different types such as full names, surnames plus titles, pronouns, zero anaphors, and lexical noun phrases. No zero anaphors were found in English, while no surname only uses were found in Korean. The majority of full names in the Korean corpus were retrieved with the topical case markers.

However, it is also noteworthy that the predominant form in English is pronouns, while the predominant form in Korean is the full name constructions. Again, the majority uses of pronouns in English and a few uses of zero anaphors in Korean can be accounted for in terms of the high accessibility of the topical referents since they are mostly referred to in the same sentences and in the next sentences by pronouns (Section 4.3.1), which conforms to Givón, Fox, Ariel, and Gundel et al.'s viewpoints. On the contrary, although there were some uses of other forms, the vast majority of uses in the Korean corpus deployed the full name constructions even with the high accessibility of the referents in the same sentences and in the next sentences for the second mentions (Section 4.3.2), which is counter to the relevant theories such as the topic continuity theory, the hierarchy theory, the accessibility hierarchy, and the givenness hierarchy.

For the third mentions in the survey, both English survey results and Korean survey results represent the same selection of discourse anaphors like the second mentions in each language. That is, in English, the degree of acceptability for surnames and pronouns was higher in order, while in Korean the acceptance rate for full names and pronouns was higher in order.

6.1.1.2 By Paragraph

The first paragraphs in the corpus basically show the same distribution since Corpus I and II were analyzed from the same corpus.

For the first paragraphs in the survey, both English and Korean survey show that full names were the most preferred forms in terms of degree of acceptability, like the first mentions/paragraphs in the corpus.

For the second paragraphs in the corpus, English corpus employs three different types of structural constructions such as surnames (plus titles), pronouns, and lexical noun phrases, while Korean corpus employs four different types such as full names, surnames plus titles, pronouns, and lexical noun phrases. No full name was found in English, while no surname only use was found in Korean. The majority of full names in Korean were retrieved with the topical case markers. It is noteworthy that the predominant form in English is the surname construction, while the predominant form in Korean is the full name constructions. The majority uses of surnames in English and a few uses of surnames in Korean can be accounted for in terms of the accessibility of the referents since they are mostly referred to in more than one sentence away by surnames (Section 4.3.1), which conforms to Givón, Fox, Ariel, and Gundel et al's standpoints. On the contrary, although there were some uses of other anaphoric forms, the vast majority of Korean corpus deploys the full name constructions even with the high accessibility of the referents in the same sentences and in the next sentences for the second paragraphs (Section 4.3.2), which is against the topic continuity model, the hierarchy model, the accessibility model, and the givenness model, put forward by Givón, Fox, Ariel, and Gundel, respectively.

For the second paragraphs in the survey, English survey shows the same selection of discourse anaphors, like the second/third mentions in the survey, in which the most preferred forms were the surnames and the next was pronouns. On the other hand, Korean survey also shows the same selection of discourse anaphors, like the second/third

mentions in the survey, in which the most preferred forms were full names and the next was pronouns.

For the third paragraphs in the corpus, English corpus employs four different types of structural constructions such as surnames (plus titles), pronouns, full names, and lexical noun phrases, while Korean corpus also employs four different types such as full names, surnames plus titles, pronouns, and lexical noun phrases. The predominant form in English is the surname constructions, while the predominant form in Korean is also the full name constructions. The majority of full names in Korean were retrieved with the topical markers. The majority uses of surnames in English and a few uses of surnames in Korean can be also accounted for in terms of the accessibility of the referents since they are mostly referred to in more than one sentence away by surnames (Section 4.3.1), which conforms to Givón, Fox, Ariel, and Gundel et al's standpoints. On the contrary, although there were some uses of other anaphoric forms, the vast majority of Korean corpus deploys the full name constructions even with the high accessibility of the referents in the same sentences and in the next sentences for the third paragraphs (Section 4.3.2), which is against the relevant four theories mentioned above.

For the third paragraphs in the survey, English survey shows the same selection of discourse anaphors, like the second/third mentions and the second paragraphs in the survey, in which the most preferred forms were the surnames and the next was pronouns. On the other hand, Korean survey also shows the same selection of discourse anaphors, like the second/third mentions and the second paragraphs in the survey, in which the most preferred forms were full names and the next was pronouns.

Thus far, we have discussed the distribution and the selection of discourse anaphors in English and Korean by the order of mention and by the paragraph both in the corpus and in the survey. The most predominant/preferred discourse anaphoric form (in red) and the next frequent/preferred form (in blue) in the corpus and in the survey are summarized in Table 6.1.

Table 6.1 The most preferred and the next form in Corpus and Survey in both languages

Lang. & Method / Mention type		English		Korean	
		Corpus (I & II)	Survey (B & C)	Corpus (I & II)	Survey (B & C)
Order of Mention	1st M ↓	**FN** >>>	**FN** >>>	**FN** >>>	**FN** >>>
	2nd M ↓	**Pro** >> **SN**	**SN** > **Pro**	**FN** >>> **SN**	**FN** > **Pro**
	3rd M ↓	**Pro** >> **SN**	**SN** > **Pro**	**FN** >>> **SN**	**FN** > **Pro**
Paragraph	1st P ↓	**FN** >>>	**FN** >>>	**FN** >>>	**FN** >>>
	2nd P ↓	**SN** >>> **Pro**	**SN** > **Pro**	**FN** >>> **SN**	**FN** > **Pro**
	3rd P ↓	**SN** >>> **Pro**	**SN** > **Pro**	**FN** >>> **SN**	**FN** > **Pro**

As seen in Table 6.1, in Corpus (I & II) and Survey (B & C), the English corpus/survey shows almost the same distribution and the selection in terms of the degree of acceptability for the preferred anaphoric forms: full names for the first

mention/paragraph; surnames and pronouns for the second mention/paragraph; surnames and pronouns for the third mention/paragraph. On the other hand, the Korean corpus/survey also shows almost the same distribution and the selection in terms of the degree of the acceptability for the preferred anaphoric forms: full names for all mentions/paragraphs; surnames and pronouns for the second mention/paragraph and for the third mention/paragraph.

6.1.1.3 Distance between Discourse Anaphors

Concerning the distance between discourse anaphors, the English corpus shows that surnames were deployed with the widest distribution and as the most independent construction in terms of frequency, implying that they are retrieved in order to refer to the topical referents in separate sentences. Next, pronouns are the most accessible anaphoric forms, mostly retrieved in the same sentences.

On the other hand, the Korean corpus shows that full names showed the widest distribution and were the most independent construction in terms of frequency, implying that they are repeatedly retrieved throughout a discourse to refer to the topical referents in the separate sentences, in lieu of being retrieved by other discourse anaphoric forms such as surnames and pronouns. Next, surnames also showed the widest distribution; however, the retrievals were not more frequent than full names, and they are mostly retrieved in the next sentences; pronouns were retrieved in the next sentences. In this respect, surnames and pronouns show similar distribution. In addition, zero anaphors were frequently retrieved only in the next sentences.

6.1.2 Two Discourse Anaphoric Patterns: GAP and SAP

6.1.2.1 Socio-cultural Point

As seen in Section 5.1.3 and Section 5.2.3 about the interactions between variables, we need to reconsider (the degree of) famousness and the famous public figures mentioned in Korean corpus (Section 3.2.1) and Korean survey (Section 3.3.2). For example, according to Wikipedia[50], Chan Ho Park was the first Korean-born baseball player in Major League Baseball (MLB) from 1994 to 2010 in the United States. He recoded the winningest Asian-born pitcher with his 124th career victory in MLB history. He played in Nippon Professional Baseball in 2011 and in Korean Professional Baseball in 2012 before he retired in December 2012. He is considered a leading exponent in Korean baseball history and a famous public figure in Korean society, so he is very well known and easily recognized and his names is already identifiable and familiar with Koreans before being a topical referent in news articles; i.e., discourse.

Likewise, Jae-Seok Yoo is publicly considered the most successful entertainer both as a comedian and as a television comedy show host in Korea. According to Wikipedia[51], in fact, he has been awarded the grand prize every year in (three major) Korean broadcasting stations since 2005. He has hosted several television variety shows in three major broadcasting stations since 1991. He has been well known with a variety of nicknames such as "Nation's Master of Ceremony (the MC of the Nation)",

[50] For Chan Ho Park, refer to "http://en.wikipedia.org/wiki/Chan_Ho_Park" in English; "http://ko.wikipedia.org/wiki/%EB%B0%95%EC%B0%AC%ED%98%B8" in Korean.
[51] For Jae-Seok Yoo, refer to "http://en.wikipedia.org/wiki/Yoo_Jae-Seok" in English; "http://ko.wikipedia.org/wiki/%EC%9C%A0%EC%9E%AC%EC%84%9D" in Korean.

"grasshopper", "class-president Yoo", "Yooruce Willis (spoof for Bruce Willis)", "Yoomes Bond (spoof for James Bond)", "Yoonunim (spoof for God in Korean), and so on because of his quick wit and modest attitude.

In other words, Chan Ho Park and Jae-Seok Yoo are very famous or socially well-known public figures in Korea because they are considered one of the best successful persons in his own area. Chan Ho Park is considered a representative in the baseball field and Jae-Seok Yoo is considered the No. 1 show host in show business. There would be a tendency that for Koreans to call them by their full names *Chan Ho Park* and *Jae-Seok Yoo* rather than surnames *Park* and *Yoo* or given names *Chan Ho* and *Jae-Seok* both in spoken discourse and in written discourse. The use of full names referring to these public figures is more preferred and tolerated than that of single names (i.e., surnames and given names) in real society.

Moreover, I argue that the use of full names for those public figures is actually reference to *title* as *a single unit*, not a reference to just a combination of "surname + given name" in written/spoken discourse as well as in everyday life. Since their achievements are already publicly acknowledged in Korean society and they are considered prototypes of success, their names are no longer personal names; instead, their names have become synonymous with their success. In this respect, I would consider this kind of use as a single unit title in Korean when we sequentially use full names throughout a discourse for referring to a particular public figure in terms of socio-cultural constraint, in lieu of being replaced by discourse anaphoric forms, such as surnames plus titles and pronouns.

Likewise, this use of full names as a single unit title can be applied to other representatively famous public figures such as Chel-Soo Ahn, a successful entrepreneur, professor, programmer, physician, programmer, and politician, considered a Korean Steve Jobs; Ji-Sung Park, the first Korean football player performing in the English Premier League in the United Kingdom; Seung-Yeop Lee, the Asian record-holder of the largest number of homeruns performing in Nippon Professional Baseball from 2004-2012, and in Korean Professional Baseball; Tae-Whan Park, the first Korean gold medalist in the Olympic swimming; Yu-Na Kim, the 2009 World champion of Women's World Figure Skating.

It is very preferred and tolerated that referring to those public figures be carried out by using the sequential pattern of full names in Korean discourse. At this point, it is worth remembering that this kind of use is a contrary to Ariel's Accessibility Hierarchy (Section 2.4.1) in that full names are not lower than any other types of proper names in terms of accessibility as least in Korean discourse because the full names for referring to those famous public figures are already familiar and identifiable between interlocutors before being introduced at beginning of a discourse.

Consequently, it is likely that the use of full names for those public figures is incorporated with a socio-cultural constraint in terms of the linguistic selection. This selection of full names from a socio-cultural constraint can affect the distribution of full names in spoken and written discourse.

6.1.2.2 Ariel's AH and Lee's Weakness

Ariel (1990:11-30) proposes that the primary function of the various referring expressions enables the addressee to mark different degrees of accessibility in memory and to find out the intended referent. In other words, Ariel claimed that the addressee is to find out among her/his mental representations an entity whose accessibility to her/him is indicated by the specific expression. Interlocutors choose their referring expressions by taking into consideration the degree of accessibility of the mental entity for the addressee.

Thus, Ariel (1990, 1994) represents the accessibility marking scale (Section 2.4.1), in which referring expressions are arranged on a scale of accessibility marking, so that each expression marks a relatively lower degree of accessibility than the member to its left, predicted by the interactions among three coding principles: informativity, rigidity and attenuation.

Concerning proper names, Ariel (1990:40-46) argues that full names are highly informative and rigid; however, they are the lowest on the accessibility hierarchy (AH). Surnames and given names follow in order, signaling relatively higher accessibility than full names. In addition, Ariel (1990, 1994) argues that in English, given names are more restricted than surnames in terms of memory and attention state they signal since given names tend to be shorter and less unique than surnames.

However, Mulkern (1996:247) argues that the opposite explanation applies to Korean (and Chinese) naming forms, in which surnames are more restricted and less unique than given names in terms of memory and attention state they signal; surnames are occasionally the same length or shorter than given names. In this respect, I argue that

Ariel's proposal should take into account cross-linguistic viewpoints including a variety of contexts in discourse and the naming convention in a particular society with a pragmatic account.

In other words, Ariel's argues that given names are more accessible than surnames in English, while Mulkern argues that surnames are more accessible than given names in Korean in terms of Ariel's AH. Consequently, this naming convention in Korean requires some change between given names and surnames in the reverse direction in Ariel's AH. In fact, the corpus results (Chapter 4) and the survey results (Chapter 5) represent that at least in English, given names, (i.e., the higher accessible marker between two) are not referred to or retrieved more frequently than surnames (i.e., the lower accessible marker). Instead, surnames were mostly retrieved in terms of frequency in the corpus and much more preferred in terms of degree of acceptability in the survey.

In addition, Lee (2010) suggests that in Korean political news, full names are all used for the first mentions; the surname and a(n) (abbreviated) title constructions are used more than 90% for the second mentions; pronouns are predominantly used for the third mentions (56%), compared with the second mentions (4%); the surname and a title construction shows the widest distribution in terms of the distance between discourse anaphors. Moreover, Lee (2010) argues that these findings are on a par with Ariel's AH in that the full name construction (i.e., the lowest accessible marker among three referring expressions) would be appropriate for the first mentions; the surname and a title construction (i.e., a relatively higher accessible marker than a full name) would be

appropriate for the second mentions; pronouns (i.e., the highest accessible marker among three) would be appropriate for the third mentions.

However, Lee's (2010) findings on English/Korean political news articles and generalization based on Ariel's AH also show a weakness in that they do not explain the distribution and the selection of discourse anaphoric patterns in Korean corpus and survey, discussed in Chapter 4, Chapter 5, and Section 6.1.1, due to the fact that famous public figures are predominantly and extensively referred to by their full names throughout a discourse, regardless of the order of mentions (i.e., the first-, the second-, and the third mentions) and the paragraph (i.e., the first-, the second-, and the third paragraphs), and the higher or lower degree of acceptability of referring expressions, encoded in that discourse anaphoric terms. Also, full names show the widest distribution in terms of the distance between discourse anaphors, which is counter to Ariel and Lee's claims.

Moreover, it was surprisingly observed that a full name is referred to and retrieved by the same full name form even within one sentence in the Korean corpus, as shown in (46).

(46) yengkwuk enlontul-i Park Ji-Sung-uy iceklwume-lul potohamyense
England media-Acc Park J-S-Gen transfer.rumor-Acc report.doing
huntul-ess-ul-ttay Park Ji-Sung-un kyengkilyek-ulo mattayunghay-ss-ta.
shake-Past-Acc-when Park J-S-Top performance-by respond-Past-Decl
'When British media reported a rumor of Ji-Sung Park's transfer with surprise,
Ji-Sung Park responded by his performance.'

(*Sports Chosun*, March 29, 2011)

As seen in (46), a referent is referred by a full name, *Ji-Sung Park*, in the subordinate clause, and this referent is also retrieved by the same full name, *Ji-Sung Park*, in the matrix clause. That is, the referent is referred to and retrieved by the full name, i.e., one of the lowest accessibility markers in Ariel's AH, in which it should be referred to or retrieved by its pronoun, i.e., one of the highest accessibility markers in AH or by the surname at least in the subordinate clause or in the matrix clause. Thus, in Korean, a topical referent is repeatedly referred to or retrieved by the same full name form even within one sentence, where a higher accessibility marker should be retrieved, which is counter Givón, Ariel, and Lee (2010)'s argument.

Concerning the semi-compulsory use of titles for public figures in Korean media, Lee (2010) argues that more than 90% of the second mentions and 40% of the third mentions are referred to by the surname plus title construction, which is due to a socio-cultural constraint in Korean political news as well as in Korean society. Concerning this issue, however, I point out that the use of the surname and a(n) (abbreviated) title construction is not described and ranked in Ariel's AH. Thus, it should be added into the accessibility scale. As seen in Chapter 4 and Chapter 5, and discussed in Section 6.1.1, where we saw that the surname and a tile construction such as *Ahn wencang* 'Dean (Dr.) Ahn' (안 원장) is mostly referred to or retrieved instead of surname only, *Ahn* (안) in Korean discourse. In a strict sense, they are different types of referring expressions from each other. In fact, it is not common for Koreans to use full names when addressing

elders such as father, professor, and doctor in particular social settings. Thus, this aspect should also be reconsidered in Ariel's AH.

Finally, it is worth recalling that Ariel's AH shows the individual expressions without implicational relations. That is, each expression type in the hierarchy is separately and individually marked for the level of accessibility the referent codes (Cornish 1999:8); therefore, each referring expression type in that scale is mutually exclusive, not inferable one from another. This fact guides us to look for an appropriate explanation to explain their distinctive properties and to indicate inferable characteristics among them. Accordingly, Ariel's AH has some limitation in that it could not explain the implicational relationship between full names and surnames within a dynamically developing discourse in terms of the distribution and the selection by interlocutors. In particular, Ariel's scale explains a full name *Chan Ho Park*, the surname *Park*, and its pronoun *he* in a distinctive way; however, it could not implicate the implicational and topical connections from one another in a given discourse. Therefore, we need to have some more relevant and plausible explanation with a comprehensive perspective.

6.1.2.3 GAP and SAP

Full names, one type of proper name, have a reference. They are not used to describe or specify characteristics of objects; however, they are logically connected to characteristics of objects to which they refer (Searle 1958:173). Full names also function as one type of referring expressions to refer to an actual entity and play a key role for linking a topical referent in a given discourse.

As discussed in Section 6.1.1, I argue that, particularly in Korean news articles, full names function as one type of discourse anaphors, being predominantly used throughout a discourse for referring to topical or attentive public figures. In addition, as discussed in Section 6.1.2.1, full names referring to particularly famous public figures in Korean news articles are not a combination of a surname and a given name, but a single unit functioning as a title.

Therefore, I proposed that according to how we refer to a particularly (famous) public figure in a given discourse, there are two main anaphoric patterns in Korean discourse: General Anaphoric Pattern (GAP) of <full name, surname/pronoun, surname/pronoun> and Sequential Anaphoric Pattern (SAP) of <full name, full name, full name> in terms of a cohesive chain or a coreferential chain (Kim 2010a, 2012). Thus, I argue that the first type of full name pattern, GAP, can explain all types of discourse anaphors in the English corpus and surveys, Ariel's AH, and Lee's (2010) findings on Korean news articles about (general) news actors, while the second type of full name pattern, SAP, can explain those in Korean news about famous public figures.

That is, in the GAP, like Lee's (2010) analysis of referring expressions in Korean news articles about news actors, after being introduced at the beginning in a given discourse for evoking an actual topical referent, a full name is followed by the surname and a title construction or pronoun, which is also followed by the surname or pronoun in the following sentences. In the SAP, like the referring expressions in Korean news articles about famous public figures, after being introduced at the beginning of a given discourse, a full name is predominantly referred to or repeatedly retrieved in the

following sentences or throughout a discourse in lieu of being replaced by surname (plus title) constructions and pronouns (or zero anaphors), functioning as a discourse anaphor.

Let us now consider the following example of the GAP in English and that of the SAP in Korean in order.

(47) a. Bill Gates reiterated Wednesday that Microsoft Corp. made "a very fair offer" to purchase Yahoo and gave no indication that his software company would increase its bid.

b. During an interview after Gates spoke to students at the University of Chicago on Wednesday evening, he called the offer of $31 a share for Yahoo "generous."

c. "They've had [the offer] for a few weeks now and I'm sure they'll continue to think about it over the next few weeks," said Gates, co-founder and chairman of Microsoft. "If they embrace it in a positive way, we think that will be a good thing for both of us."

(*Chicago Tribune*, February 21, 2008)

In (47), a full name, *Bill Gates*, is referred to as a topical referent at the beginning of this given discourse. This full name is retrieved by its pronoun *his* (i.e., a highly accessible marker) in the same sentence. This attentive referent is retrieved by its surname Gates in the next sentence/paragraph. Concerning cohesion in this discourse, the referent forms a cohesive coreferential chain such as <Bill Gates, he(/Gates), Gates(/he)>, which is considered the GAP in English.

Next, let us consider the following example of the SAP in Korean.

(48) a. 2012 MBC pangsong yenyey taysang-ul swusang-han
broadcasting entertainment the grand prize-Acc awarded-Rel
kaykumayn Yoo Jae-Seok-i uymisimcanghan swusangsokam-ul
comedian Yoo J-S-Nom very meaningful acceptance speech-Acc
palhi-ess-ta.
say-Past-Decl.
'Jae-Seok Yoo, who is awarded the grand prize at MBC TV Entertainment Awards in 2012, said in his meaningful acceptance speech.'

b. Yoo Jae-Seok-un 29il ohwu kyengki-do koyang-si ilsan
Yoo J-S-Top 29th afternoon Kyungki-province Koyang-city Ilsan
MBC Dream Center-eyse yellin sisangsik-eyse yengyey-uy
MBC Dream Center-at held award ceremony-at honor-Gen
taysang cwuinkong-i tway-ss-ta.
the grand prize winner-Nom become-Past-Decl
'Jae-Seok Yoo won the honorable grand prize at the award ceremony that was held at MBC Dream Center in Ilsan, Koyang, Kyungki.'

c. Yoo Jae-Seok-un swusangca-ro homyeng twayca kippum-ul
Yoo J-S-Top winner-as call to.be joy-Acc
kamchwuci moshamyenseto "ku enu ttaypota coysong hatanun
not.contain not.do. "than ever before sorry do
sayngkak-i tunta"-ko malmwun-ul yel-ess-ta.
thought-Nom seem"Rel saying-Acc open-Past-Decl
'Jae-Seok Yoo said without containing his joy, "I am very sorry for winning the grand prize again than ever before," when he was called his name as a winner of the grand prize.'

(*The Kyunghyang Shinmun*, December 30, 2010)

In (48), a full name, *Jae-Seok Yoo*, is referred to as a topical referent at the beginning of this given discourse. This full name is retrieved by the same full name in the second mention/paragraph, in lieu of being retrieved by the surname and its pronoun. This attentive referent is also retrieved by the same full name in the third mention/paragraph, in lieu of being retrieved by the surname and its pronoun. Concerning cohesion in this discourse, the referent forms a cohesive coreferential chain such as <Jae-Seok Yoo, Jae-Seok Yoo, Jae-Seok Yoo>, which is considered the SAP in Korean.

6.1.2.4 Cognitive Status: GHZ's GH

As represented in Section 2.4.2 and Section 2.6, with respect to cognitive status, Mulkern (1996) applies the use of naming forms to Gundel, Hedberg, Zacharski (GHZ)'s (1993) Givenness Hierarchy (GH), which represents six different cognitive statuses of referring expressions in terms of different memory and attention state that are coded in the lexical forms.

Mulkern (1996:238-240) argues that when a full name refers to an entity, it is not expected by the addressee that s/he has a representation at the beginning of discourse. Thus, it seems sufficient that the status related with a proper name be no higher than Uniquely Identifiable, similar to definite descriptions. However, it also seems necessary that the status related with a proper name be no lower than that. Single names are more ambiguous expressions than full names because they have a broader range of possible referents, so they lead the addressee to have a more restrictive cognitive status: Familiar. That is, a full name only indicates that the referent is Uniquely Identifiable, whereas a single name indicates that the writer/speaker expects the addressee to have a

representation of an individual who is already familiar to both speaker and addressee in his/her memory; therefore, a single name requires that the referent is at least Familiar, similar to demonstrative determiners *that* in discourse.

In addition, concerning the starting cognitive status and the possible landing status, Mulkern (1996:245) argues that there is an interaction between GHZ's GH with Grice's (1975) conversational maxims of quantity, from which there are implicational characteristics from the restrictive status In Focus to the least one Type Identifiable. This interaction indicates that the most appropriate expressions are the ones that exactly mark the same status a referent has (or the next restrictive status). Then Mulkern concluded that full names are generally used for referents which are Uniquely Identifiable and Familiar, whereas single names are mostly used for referents which are Familiar and Activated at an appropriate context. At this point, it is noteworthy that Mulkern (1996) argues that there are no usages of full names as In Focus, as I show Figure 6.1 (Figure 2.1).

In focus	>	Activated	>	Familiar	>	Uniquely Identifiable	> …
X		possible	←			Full Names	
X		←		Single Names			

Figure 6.1 The cognitive statuses of full name and single name in English

Based on this analysis, therefore, a full name refers to an entity that should be no higher than Uniquely Identifiable and can be extended to Familiar, like the definite description *the N* in GHZ's GH, while a single name refers to an entity that should

require at least Familiar and can be extended to Activated, like the demonstrative determiner *that N* in GHZ's GH, as presented in Section 2.6. However, that is not what was found at least in my corpus analysis and survey results of Korean, as discussed in Section 6.1.1.1 (by order of mention) and Section 6.1.1.2 (by paragraph).

Based on the results of the corpus and the survey, I proposed that there are two types of full names: i.e., the GAP (in English and in Korean) and the SAP (in Korean), functioning as discourse anaphors.

Concerning cognitive status in Korean discourse, I also propose an extension of previous theories: full names in each pattern are different from each other in terms of the starting state and the possible landing state. In particular, a full name in the GAP starts from Uniquely Identifiable and lands at Familiar, following Mulkern (1996), as shown in Figure 6.2. On the other hand, a full name in the SAP starts from Familiar and lands at In Focus, being used in lieu of the surname and its pronouns, as illustrated in Figure 6.3.

First, the cognitive status of the GAP is shown in Figure 6.2.

In focus >	Activated >	Familiar >	Uniquely Identifiable >	Referential >	Type Identifiable
X	possible	←	**Full Name**		
Ø or Pronoun	←	**Surname (+ Title)**			

Figure 6.2 The cognitive status of the GAP

Figure 6.2 shows that in the GAP, a full name is referred to for the first mentions, being in Uniquely Identifiable; the surname + a title construction is referred to for the second

retrievals, being in Familiar; a pronoun is referred to for the third retrievals, being in In Focus.

Second, the cognitive status of the SAP is shown in Figure 6.3.

In focus	>	Activated	>	Familiar	>	Uniquely Identifiable	>	Referential	>	Type Identifiable
←		←		**Full Name**						

Figure 6.3 The cognitive status of the SAP

Figure 6.3 shows that in the SAP, a full name is referred to for the first mentions, being in Familiar; the same full name is referred to for the second retrievals, being in Activated and being used in lieu of the surname + a title; the same full name is referred to for the third retrievals, being in In Focus and being used in lieu of the pronouns.

In fact, in the results of the Korean survey, we have seen that the mean values of the participants' acceptability for full names (the most preferred) and pronouns (the next) in the second mentions/paragraphs and in the third mentions/paragraphs were all higher than other discourse anaphoric forms (Section 5.1.2 and Section 5.2.2).

Therefore, the cognitive states of full names in the GAP and the SAP are summarized in terms of the starting state and the (possible) landing state, as shown in Table 6.2.

Table 6.2 The starting state vs. possible landing state of full names

	GAP	**SAP**
1. Starting state	Uniquely Identifiable	Familiar
2. (Possible) Landing state	Familiar (Activated)	In Focus

6.2 Sentential Anaphors

6.2.1 Referential Properties

I turn now to conclusions about sentential anaphors. According to Kamio and Thomas (1999), sentential anaphors are differentiated by two referential properties: the givenness of information (i.e., new vs. old information) and the width of reference (i.e., narrow vs. wide reference). For instance, English sentential anaphors *that* and *it* are differentiated by how interlocutors understand the context surrounding utterance in a given discourse in terms of their selection and distribution. Kamio and Thomas (1999) notes that there are distinctive referential properties between the sentential anaphor *that* and *it* in two respects: 1) whether a speaker knows information that a sentential anaphor refers to (i.e., new vs. old information); 2) whether a sentential anaphor points to the utterance itself or refers to the relevant context including the utterance (i.e., narrow vs. wide reference).

6.2.1.1 New vs. Old Information

First, let us consider the examples in English with respect to new vs. old information in order.

(49) New Information *that*

Pam: I was in the meeting with Jan, and,

she did say that *it could be this branch that gets the axe*.

Worker: Are you sure about *that*?

(*The Office*, Season 1, Episode #1)

In (49), the sentential anaphor *that* in the worker's utterance refers to Pam's (italicized) utterance, indicating that Pam's utterance is new information to the worker her/himself. Also, *that* points narrowly to Pam's utterance itself.

In addition, let us consider the example of old information, as shown in (50).

(50) Old Information *it*

Jim: Solitaire?

Pam: Yeah, FreeCell.

Jim: (points at the monitor) *Six on the seven*.

Pam: I know, I saw that.

Jim: So then, why didn't you do *it*?

Pam: I'm saving that, 'cause I like it when the cards go T-ts-ts-tch-tch-tch.

Jim: Who doesn't love that?

(*The Office*, Season 1, Episode #2)

In (50), when Pam is playing FreeCell, Jim walks up to her to take a jellybean. The sentential anaphor *it* in Jim's third utterance represents his utterance *(Play or Put) Six on the seven* in his previous utterance, indicating his and Pam's already-known information.

Second, let us consider the examples in Korean with respect to new vs. old information in order.

(51) New Information *ku(-ke(n))* 'that'

Miran: *Tayphwung-i-lang* *Swucin-i-lang* *sakwin-tan-ta*
Tayphwung.nom.with Swucin.nom.with date.pres.evi.decl
'*Tayphwung and Swucin are dating together*.'

Poksil: *ku-ken* ceto molla-sse-yo
that.thing me.too not.know.past.decl
'I didn't even know *that*.'

(*My Too Perfect Sons* 43th, September 5, 2009)

In (51), the sentential anaphor *ku-ken* 'that' points to Miran's (italicized) utterance, which is Poksil's new information about the utterance in that speech event. It also points narrowly to Miran's utterance itself at the same time.

In addition, let us consider the example of old information, as shown in (52).

(52) Old Information *ku-len* 'it'

Okhuy: *ne* *mayil* *chwulkunkiley* *oppaney* *kase*
you everyday going to work brother's house go.and
cokhatul *chayngkiko* *kanunke* *anunte* way malul
nephews take care of go.pres know.and why words.acc

mos-ha-ni?
not.do.Q
'I knew that *you go to your brother's home and take care of your nephews everyday*. Why don't you tell me?'

Swucin: coysonghapnita, emenim
sorry.pres.decl mother.in.law
'I am sorry, mother.'

Okhuy: nayka *ku-len-ke* moluko kyelhon helakhay-ss-ni?
I.nom *it*.thing not.know.and marriage allow.past.Q
'Did I allow your marriage without knowing *it*?'

(*My Too Perfect Sons* 53th, October 10, 2009)

In (52), the sentential anaphor *ku-len(-ke)* 'it' in Okhuy's (mother-in-law) second utterance refers to her advanced recognition and old information about that Swucin, a daughter-in-law, has gone to her brother's home to take care of nephews because her brother's wife had already passed away. It also implicates Swucin's difficult situation in that context. The use of *ku-len(-ke)* 'it' as an old information goes with the semantic meaning of the verb *alta* 'know' (*moluta* 'not.know') in Okhuy's second utterance, at which it is noteworthy that if we use *ku-ken* 'that' instead of *ku-len(-ke)* 'it', *ku-ken* points narrowly to the fact itself without evoking any relevant context and implicates Okhuy's new information about Swucin's situation.

6.2.1.2 Narrow vs. Wide Reference

First, let us consider the examples in English with respect to narrow vs. wide reference in order.

(53) Narrow Reference *that*

Oscar: And, uh, *they moved to the United States a year before I was born*.

Michael: Yeah...

Oscar: So *I grew up in the United States*.

Michael: Wow.

Oscar: And, *my parents were Mexican*.

Michael: Wow. *That* is...*that* is a great story.

That's the American dream right there, right?

(*The Office*, Season 1, Episode #1)

In (53), the sentential anaphor *that* in Michael's last utterance points narrowly to Oscar's (italicized) utterances themselves without his having any prior knowledge about Oscar. Also, it indicates that this sentential anaphor *that* represents Michael's new information about Oscar's personal information in that speech event.

In addition, let us consider the example of wide reference, as shown in (54).

(54) Wide Reference *it*

Michael: *People I respect, heroes would be, Bob Hope, um, Abraham Lincoln, definitely. Bono, and probably God, would be the fourth one. And I just think those people really, uh, helped, the world, in so many ways.*

That *it*'s, um, *it*'s really beyond words. *It*'s really incalculcable.

(*The Office*, Season 1, Episode #1)

In (54), Michael uses the sentential anaphor *it* two times. The sentential anaphor *it* in Michael's monologue refers widely to four admired people's ways of lives in addition to

their admirable accomplishments and behaviors. We can infer this fact from the phrase *so many ways* in Michael's previous utterance, for which he did not enumerate particular details, but instead he evoked some relevant contexts by using the sentential anaphor *it*.

Second, let us consider the examples in Korean with respect to narrow vs. wide reference in order.

(55) Narrow Reference *ku(-ken)* 'that'

Poksil: *ceka* *insayngese* *himtulessul-ttay* *sensayngnim-ttaymwuney*
I.hon.nom life.loc in need.past.acc.time doctor.because of
kyentilswu *issesse.yo*
endure.be.past.decl
'*When I was in trouble in my life, I can put up with it because of you.*'

Tayphwung: *ku-ken* cepeney-to malhay-ss-cana
that.thing last time.too say.past.decl
'You said *that* (to me) last time.'

(*My Too Perfect Sons* 52th, October 4, 2009)

In (55), the sentential anaphor *ku-ken* 'that' in Tayphwung's utterance points back to Poksil's (italicized) utterance itself at that time, although this information about the utterance is Tayphwung's old information. In fact, Tayphwung and Poksil are in love with each other, so Tayphwung has heard Poksil's utterance several times before this speech event.

In addition, let us consider the example of wide reference, as shown in (56).

(56) Wide Reference *ku-len* ‘it’

Senphwung: *ceto* *kulim* *coahap-ni-ta*
I.hon.too painting like.pres.decl
‘I like paining, too.’

Mwunswuk: eccem, wuli *sawinun* *mwunhak*, *si*, *kulim*
Int our son-in-law.top literature poetry painting
molununke *ep-sse*. *ku* *salamun* *cengmal*
not.know.thing not.be.decl the man really
ilpakke *molunun* *ku-len* salam-i-ya.
work.nothing but not.know *it* man.be.decl
‘Wow, *our son-in-law knows everything, such as literature, poetry, and painting*. He (my husband) is the (*it)* man who cares about nothing but his work.’

(*My Too Perfect Sons* 50th, September 27, 2009)

In (56), the sentential anaphor *ku-len* ‘it’ in Mwunswuk’s (mother-in-law) utterance refers widely to Senphwung’s (son-in-law) (italicized) utterance as well as her own (italicized) previous mention in that speech event. It evokes relevant contexts that Senphwung knows other than *kulim* ‘painting’; that is, he knows other artistic tastes such as *mwunhak* ‘literature’, *si* ‘peotry’, and so on, but her husband cares about nothing but his work. Thus, *ku-len* ‘it’ for a wide reference is useful to compare son-in-law’s artistic taste with husband’s one. Also, the sentential anaphor *ku-len* ‘it’ in Mwunswuk’s utterance indicates old-to-her information rather than new-to-her information in that

speech event. It is noteworthy that the use of *ku-ken* 'that' points to the Senphwung's utterance itself, just indicating that "Senphwung likes painting, too."

6.2.2 Different Properties of Sentential Anaphors

We have seen that referential properties of sentential anaphors are pragmatically the same in both languages. However, there appear to be innately different properties of sentential anaphors in two languages with respect to degree of deictic and anaphora in the linguistic forms.

In fact, we need to take notice that a deictic *that* in English can interchangeably be used as a sentential/nominal anaphor as well as a deictic expression, whereas not all deictic expressions function as sentential anaphors in Korean. In other words, in English a relatively more deictic item *that* is interchangeably used both deictically and anaphorically (for a nominal and a sentential referent), while a relatively more anaphoric item *it* is preferably used anaphorically (for a nominal and a sentential referent). On the other hand, in Korean a relatively more deictic *ce*, correspondent with *that* in English, is used both deictically and anaphorically (only for a nominal referent), while a relatively more anaphoric *ku*, correspondent with *it* in English, is preferably used only anaphorically (for a nominal and a sentential referent). In other words, sentential anaphors in Korean basically have the same line of reasoning as those in English in terms of pragmatic aspects and cognitive status, but only *ku* 'it', a relatively more anaphoric (i.e., less deictic) item, has a dual function as a sentential anaphor with two variants: *ku(-ke(s/n))* 'that' and *ku-len(-ke(s/n))* 'it'. Thus, in Korean, only *ku* 'it', a relatively more

anaphoric (i.e., a relatively less deictic) item, functions as a sentential anaphor with two variants: *ku(-ke(s/n))* 'that' and *ku-len(-ke(s/n))* 'it, rather than *ce* 'that' and *ku* 'it'.

In this respect, Lee (2005:65-79) notes that a sentential anaphor *ku-ke(s/n)* in Korean corresponds to *that* in English, while a sentential anaphor *ku-len(s/n)* in Korean corresponds to *it* in English in terms of referential properties. Let us consider the following example, as shown in (57).

(57) Father: ce saram com bwala. namca-ka kwikeli-lul hay-ss-ta.
that man please see.imp. man.nom earrings.acc wear.past.dec
'Look at that man! A man wears earrings!'
Daughter$_1$: appa-nun…*ku-ken* yocum huhan il-iyey-yo
father.top *that*-thing these days common thing.decl.hon
'Father, *that*'s fairly common (thing) these days'
Daughter$_2$: appa-nun…*ku-len-ken* yocum huhan il-iyey-yo
father.top *it*-thing these days common thing.decl.hon
'Father, *it*'s fairly common (thing) these days'
(Lee 2005:78-79; adapted)

In (57), *ku-ken* in Daughter$_1$'s utterance points narrowly to the utterance itself "A man wears earrings!", while *ku-len-ken* in Daughter$_2$'s utterance refers widely to the utterance, evoking the relevant contexts, such as "men wear some fashionable accessories, make up one's faces, put on their clothes like women, and so on, which are traditionally

considered women's ways of wearing earrings, making up faces, wearing colorful clothes, and wearing jewelry."

Moreover, Lee (2005:69) notes that all deictic items such as *ce(-ke)* 'that' and *i(-ke)* 'this' are not used as sentential anaphors in Korean. Let us now consider the relevant example in (58).

(58) A: hongkhongsalamtulun kankwukuy hanlywu suthatulul
Hong Kong.people.pl.top Korea.gen Hallyu star.pl.acc
nemwuna coahantay.
so much like.pres.decl
'Hong Kong people like Korean Hallyu stars so much.'
B: *ce-ke(s)*/*i-ke(s)* ta kecismaliya.
that.thing/this.thing all lie.pres.decl
'*That*'s/*This*'s all a lie.'

(Lee 2005:69: adapted)

In (58), it is not likely that *ce-ke(s)* and *i-ke(s)* refer anaphorically to the previous sentence in speaker A's utterance. Rather they just point deictically to the physical or substantial material in which the sentence itself is written; that is, "Hong Kong people like Korean Hallyu stars so much."

In short, regarding anaphoric usage in Korean, a relatively more deictic item *ce* 'that' is only used as a nominal anaphor, whereas a relatively less deictic (i.e., a relatively

more anaphoric) item *ku* ‘it’ with two variants (i.e., *ku(-ke(s/n))* and *lu-len(-ke(s/n))*) is used as a sentential anaphor as well as a nominal anaphor.

So far, we have seen the referential properties of sentential anaphors in pragmatic and cognitive terms, in particular *that* and *it* in English, in which a sentential anaphor *that* does not need to represent speaker’s prior knowledge (i.e., new information) and points narrowly to the utterance itself, whereas a sentential anaphor *it* implies speaker’s prior knowledge (i.e., old information) and refers widely to the utterance with its set of relevant contexts.

In addition, concerning cognitive status, in English, a sentential anaphor *that* points to information out of focus with Gundel’s Activated state, while a sentential anaphor *it* refers to information in focus with Gundel’s In Focus state in a given discourse, as shown Figure 6.4. On the other hand, in Korean, a sentential anaphor *ku-ken* ‘that’ points to information out of focus with Gundel’s Activated state, while a sentential anaphor *ku-len* ‘it’ refers to information in focus with Gundel’s In Focus state in a given discourse, as shown Figure 6.5.

In focus	>	Activated	>	Familiar	>	Uniquely Identifiable	>	Referential	>	Type Identifiable
{***it***}		{***that***}								

Figure 6.4 The cognitive statuses of English sentential anaphor *that* and *it*

In focus	>	Activated	>	Familiar	>	Uniquely Identifiable	>	Referential	>	Type Identifiable
{***ku-len* 'it'**}		{***ku-ken* 'that'**}								

Figure 6.5 The cognitive statuses of Korean sentential anaphor *ku-ken* 'that' and *ku-len* 'it'

Those referential properties were verified with the English corpus in the same vein. On the other hand, in Korean corpus, I found out that only *ku* 'it' functions as a sentential anaphor with two variants: *ku(-ke(s/n))*, functioning as *that*, and *ku-len(-ke(s/n))*, functioning as *it*, in terms of pragmatic and cognitive properties. In fact, I was not able to find any corresponding example of a sentential anaphor *ce* 'that', a presumably correspondent with *that* in English.

The comparative study on nominal vs. sentential anaphors in English and Korean corpus can be summarized, as shown in Figure 6.6.

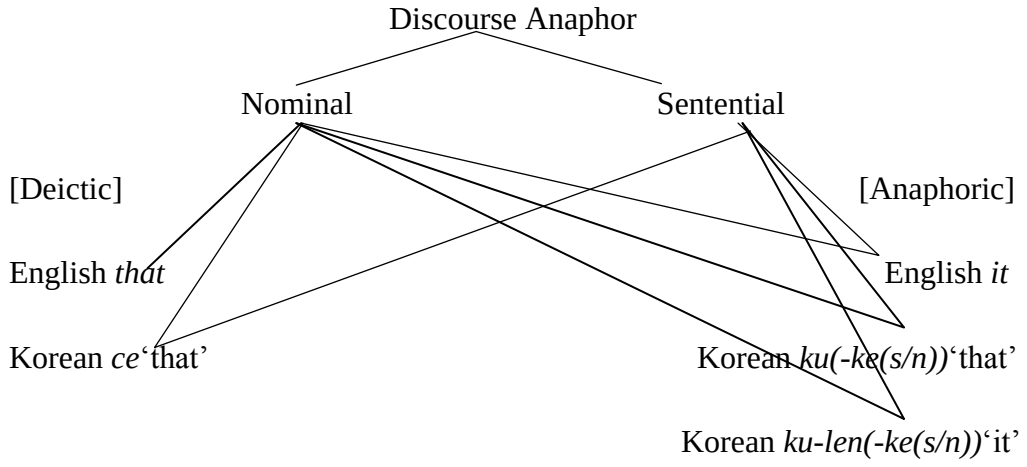


Figure 6.6 Nominal vs. sentential anaphors in English and Korean

As seen in Figure 6.6, *ce* 'that' does not function as a sentential anaphor in Korean; however, only *ku* 'it' takes its place with two variants like *ku(-ke(s/n))* 'that' and *ku-len(-ke(s/n))* 'it'. In English, there is a tendency for English *that* to be interchangeably used as a deictic, a nominal, and a sentential anaphor, while English *it* is preferred as a nominal and a sentential anaphor. However, in Korean, there is a tendency for a relatively more deictic expression *ce* 'that' to be used as a deictic and a nominal anaphor, whereas a relatively more anaphoric (i.e., a relatively less deictic) expression *ku* 'it' is used as a nominal and a sentential anaphor. That is, not every deictic expression can be used as a sentential anaphor in Korean discourse (Lee 2005:69). Then an anaphoric item *ku* 'it' can be used as a sentential anaphor with two variants: *ku(-ke(s/n))* 'that' and *ku-len(-ke(s/n))* 'it', which are considered English *that* and *it*, respectively, with the same line of pragmatic reasoning.

6.3 Pragmatic Heuristics: The Neo-Gricean Theoretical Analysis

6.3.1 GAP and SAP

Thus far, the explanation of two discourse anaphoric patterns, i.e., the GAP in English and Korean; the SAP in Korean, seems problematic in the relevant linguistic and cognitive models such as Givón's topic continuity model, Ariel's accessibility hierarchy model, and Gundel's givenness hierarchy since those models need to design another hierarchy. Accordingly, in order to explain those linguistic phenomena much more comprehensively, I applied this concept of two discourse anaphoric patterns into the neo-Gricean's pragmatic heuristics of conversational implicature.

Concerning discourse anaphora with conversational implicature, Huang (2000b: 318) notes that "the central idea of this pragmatic model is that anaphoric distribution in a discourse can largely be predicted in terms of the systematic interaction of some general pragmatic strategies such as Q-, I-, and M-principles" (e.g., Grice 1975; Horn 1984; Levinson 1995, 2000), as represented in Section 2.5.3 and summarized in Table 6.3.

In particular, Huang (2000a, 2000b, 2007) introduces the basic distributional pattern of anaphora, saying that "(i) establishment and (ii) shift of reference tends to be achieved through the use of an elaborated form, notably, lexical NP; (iii) maintenance of reference through the use of an attenuated form, notably, a pronoun or a zero anaphor".

Table 6.3 Levinson's tripartite Q-, I-, & M-principle (heuristics)

The Q-principle **(≈ Horn's Q)**	Speaker's maxim	Do not provide a statement that is informationally weaker than your knowledge of the world allows.
	Recipient's corollary	Take it that the speaker made the strongest statement consistent with what he knows.
The I-principle **(≈ Horn's R)**	Speaker's maxim	The maxim of Minimization: Say as little as necessary, produce the minimal linguistic information.
	Recipient's corollary	The Enrichment Rule: Amplify the informational content of the speaker's utterance.
The M-principle	Speaker's maxim	Indicate an abnormal, non-stereotypical situation by using marked expressions.
	Recipient's corollary	What is said in a marked message indicates a marked situation.

In Table 6.3, given the hearer-oriented Q-principle, a speaker should choose an informationally richer and morphologically more elaborated referring expression such as a proper name (with a description), while given the speaker-oriented I-principle, a speaker should choose an informationally poorer and morphologically minimal referring expression such as a pronoun or a zero anaphor. Note that the Q-principle would go against minimization in terms of meaning and expression, while the I-principle would go against recognition and identification. Thus there should be a compromise between these potentially conflicting principles, concurrently satisfying both principles in that a referring expression satisfies both recognition/identification and minimization (Huang 2000b:318-320).

In this respect, I argue that the GAP can be explained by the I-based preference for minimization, whereas the SAP in Korean can be explained by a compromise between the Q-based preference for recognition/identification and the I-based preference for minimization.

That is, on the one hand, given the speaker-oriented I-principle for the GAP in political news, a full name for the first mention should be minimized with the surname (plus a title) for the second retrieval and then this surname construction should be minimized with a pronoun for the third retrieval.

On the other hand, given the compromise between the Q- and I-principle for the SAP, a full name for all retrievals should be referred to by concurrently satisfying both

recognition/identification and minimization since a full name is being used throughout a discourse in lieu of surnames (plus titles), pronouns, and zero anaphors.

Therefore, the pragmatic heuristics explaining the GAP and the SAP are summarized in terms of the minimization and the recognition, as shown in Table 6.4.

Table 6.4 The Pragmatic heuristics for the GAP and the SAP

1. The GAP	The I-Principle (for minimal referential form, Levinson's general anaphora pattern)
2. The SAP	The compromise between the Q- and I-principle (for recognitional/identifiable referential form)

6.3.2 Referential Properties of Sentential Anaphors

As seen Section 4.4 and Section 5.3, concerning sentential anaphors in English and Korean, I analyzed the corpus data and the survey data in terms of two referential properties, i.e., the givenness of information (new vs. old information) and the width of reference (narrow vs. wide reference), put forward by Kamio and Thomas (1999).

However, it seems that in Kamio and Thomas (1999), those two different properties of (nominal and) sentential anaphors are firmly encoded in the linguistic items such as *that* and *it* in English and *ku-kun* 'that' and *ku-len* 'it' in Korean. In other words, it looks likely that there are distinctive relationships between sentential anaphoric forms in terms of two referential properties, without any implicational relationship between them. In particular, if a speaker represents some new information or points narrowly to a

referent, s/he should use the sentential anaphor *that* in English and *ku-ken* 'that' in Korean, while if s/he represents some old information or refers widely to a referent, s/he should use the sentential anaphor *it* in English and *ku-len* 'it' in Korean.

In order to solve this problem, I applied those properties into GHZ's GH, in which particular linguistic items with corresponding cognitive statuses have the implicational relationship between them based on (neo-)Gricean conversational maxims of quantity, as illustrated in Section 6.2.2. Furthermore, I applied those notions into neo-Gricean's heuristics of conversational implicature, particularly (Horn's and Levinson's) Q-scalar implicature, in order to specifically combine those linguistic items in one scale according to the referential property, as schematized in (59).

(59) Q-scalar implicature

a. The givenness of information

<old, new> (i.e., <*it*, *that*>; <*ku-len* 'it', *ku-ken* 'that'>)

(i) S(old) ├ S(new)

(ii) S(new) +> $Q_{\text{-scalar}}$ ~S(old)

b. The width of reference

<wide, narrow> (i.e., <*it*, *that*>; <*ku-len* 'it', *ku-ken* 'that'>)

(i) S(wide) ├ S(narrow)

(ii) S(narrow) +> $Q_{\text{-scalar}}$ ~S(wide)

As seen in (59), in the scale of the givenness of information, *it* and *ku-len* for old information is strong expressions, while *that* and *ku-ken* for new information is weak expressions. Likewise, in the scale of the width of reference, *it* and *ku-len* for wide reference is strong expressions, while *that* and *ku-ken* for narrow reference is weak expressions. Concerning the Q-implicature, the utterance with the sentential anaphor for old information entails the one for new information, while the utterance with the sentential anaphor for wide reference entails the one for narrow reference. On the other hand, the utterance with the sentential anaphor for new information $Q\text{-}_{scalar}$ implicates the negation of the one for old information, while the utterance with the sentential anaphor for narrow reference $Q\text{-}_{scalar}$ implicates the negation of the one for wide reference. Thus, we can infer utterances with particular sentential anaphors in terms of the implicational relationship within conversational implicature.

6.4 Benefits of Research

Keeping these conclusions in mind, I now turn to the potential benefits of the findings. First of all, this combined theoretical and empirical study of discourse anaphora is expected to contribute to the comprehensive research on referring expressions in pragmatics from a cross-linguistic perspective. In fact, referring expressions have been a theoretically arguable issue in pragmatics regarding how a discourse anaphor is referentially linked to the antecedent in a given discourse such as constructed texts and naturally occurring discourses within grammatical, linguistic, cognitive, and pragmatic framework. However, as Lee (2010) notes, little attention has been paid to exploring how discourse anaphors are deployed in written discourse such as news articles and TV drama

scripts from a contrastive and cross-linguistic perspective. In particular, discourse anaphoric patterns and sentential anaphors in written discourse have been empirically unexplored fields of studies in pragmatics. In fact, I have noticed that there have been remarkably frequent uses of full names as discourse anaphors instead of using other discourse anaphoric forms in newspaper articles, particularly in Korean. On the other hand, I have noticed that little attention has been given to comparative and contrastive research on sentential anaphors, while particular anaphoric forms as nominal anaphors have been more frequently studied at the intra-sentential level and the inter-sentential level in terms of grammatical, linguistic, cognitive, and pragmatics points. In this respect, the research on these two types of discourse anaphora studied in this dissertation, viz., discourse anaphoric patterns and sentential anaphors, fills a needed research gap.

In addition, as a quantitative approach, the empirical research conducted in this dissertation on discourse anaphora plays a significant role in synchronically and cross-linguistically understanding different types of discourse anaphoric patterns and different properties of sentential anaphors in terms of the anaphoric distribution in authentic discourse and the selection by actual language users. The combination of the data collection and the survey complementarily serves as a verifiable factor to validly explain the distribution and the selection of discourse anaphora in English and Korean. By this combination of research methods, the shared features and the different properties of discourse anaphora in two languages are empirically demonstrated and verified in terms of bringing pragmatic factors into the theoretical study of discourse anaphora.

Finally, as a qualitative approach, the relevant theoretical models discussed so far in pragmatics play a key role in explaining and understanding discourse anaphora by means of grammatical, linguistic, cognitive and pragmatic perspective. Meanwhile, those theoretical models have prescriptively focused on economy or accessibility of linguistic form in the process of unfolding those discourses anaphoric forms in a given discourse. However, if we face prescriptively inappropriate uses of discourse anaphors such as the SAP in Korean, we find that we need to reverse the scale suggested in each theoretical model or to make another scale to explain that linguistic phenomenon. If so, we can increase a level of descriptive adequacy to explain different types of discourse anaphora from a cross-linguistic perspective; but we can decrease a degree of explanatory adequacy in terms of Occam's razor. In this respect, neo-Gricean pragmatic heuristics of conversational implicature serves as a complementary window to comprehensively explaining comparative and contrastive research on discourse anaphora in terms of the distribution and the selection without modifying or changing that framework.

6.5 Future Avenues of Research

This dissertation has focused on a synchronic comparative and contrastive study of discourse anaphoric patterns and different properties of sentential anaphors in authentic discourse in English and Korean. It also has indicated that theoretical frameworks should be verified by empirical research such as the data collection form naturally occurring texts and the use of a survey, and vice versa; that is, newly observed linguistic phenomena used by language users should be explained within the relevant theoretical models.

Nevertheless, there might be critical suggestions for the future avenues of this research in some respects as follows:

1. For discourse anaphoric patterns;

a. How will those discourse anaphoric patterns be distributed and selected in a variety of genres of discourse in English and Korean?

b. How will those discourse anaphoric patterns differently be distributed and selected in terms of diachronic research, i.e., historic change, in addition to this synchronic research?

c. How will actual cognitive status on different anaphoric forms differ by the mention type in a given discourse? In particular, how will the degree of acceptability of full names and pronouns in the SA in Korean differ in more experimental and psychological research? For example, an experimental analysis on the eye movement for those anaphoric forms in certain mention types by using the eye tracker or on the response time for them by using DMDX.

2. For sentential anaphors;

a. What will be comparative referential properties of other sentential anaphors and the categorical relation between those expressions such as *this* in English and other variants of *ku-ken* ‘that’ and *ku-len* ‘it’ in Korean, in addition to those discussed in this dissertation?

b. How will actual cognitive status of different sentential anaphors differ by the referential properties in a given discourse in English and Korean by means of more psychological and experimental research?

Without any doubt, further research on these discourse anaphoric features suggested above needs to be conducted in terms of pragmatics in cooperation with other subfields of linguistics. Hopefully, this limited, but focused research conducted in this dissertation could be another stepping stone for those future avenues of research in comparative study of pragmatics.

Appendix A

The Survey Questionnaire in English

APPROVED

MAR 2 6 2012

Institutional Review Board

INFORMED CONSENT

PRINCIPAL INVESTIGATOR NAME:
Kim, Sok-Hun

TITLE OF PROJECT:
A Corpus-based Analysis of Discourse Anaphora in English and Korean: A Neo-Gricean Pragmatic Approach (A Survey of Word Choice: Naming Forms and Pronouns)

INTRODUCTION:
You are being asked to participate in a research survey about word choice; i.e., naming forms and pronouns. Your participation is anonymous and voluntary. Refusal to participate or withdrawal from the survey at any time will involve no penalty or loss of benefits to which you are otherwise entitled. Please ask questions if there is anything you do not understand.

PURPOSE:
The purpose of this study is to investigate possible associations between one's degree of preference for naming forms and degree of famousness of particular public figures in texts, and those between one's degree of preference for two pronouns in conversations and different contexts.

DURATION:
The survey session will take approximately 30 minutes to complete.

NUMBER OF PARTICIPANTS:
The survey will be conducted by a maximum of 25 native speakers of English and a maximum of 25 native speakers of Korean, depending on availability and willingness of potential participants.

PROCEDURES:
1. Recruiting and Two screening questions
You will be recruited directly by me by canvassing public areas of the UT Arlington campus to participate in this research survey. After being notified that the survey is anonymous and voluntary, and being asked about your willingness to participate, you will be asked to answer two screening questions regarding native language (i.e., English & Korean) and age (i.e., over 18). If you are willing to participate in this research survey and meeting these screening criteria, you will be presented with the survey's questions of your native language.
2. Survey questionnaire for each native speaker of each language

You will complete a word choice questionnaire. The form includes questions related to your natural uses of naming forms and pronouns in texts. During the survey, you will see four sections of questions. In section A, you will be asked to complete questions about yourself (native language, sex, education, and age). In section B, you will be asked to complete questions about your preference (from 5-1) for naming forms (such as Steve Jobs, Jobs, He, Steve, and other possible forms) in texts by the first three mentions. In section B, you will be asked to complete the same type of questions about your preference (from 5-1) for naming forms in texts, but by the first subject position in each paragraph. In section D, you will be asked to complete questions about your preference (from 5-1) between two pronouns (*that* and *it* in English; *kuken* 'that' and *kulen* 'it' in Korean) in conversations and different contexts.

POSSIBLE BENEFITS:
There are no direct benefits to you for participating in this survey other than the possibility for considering an association between naming forms and famousness of a public figure, an association between pronouns and different contexts in your native language.

POSSIBLE RISKS/DISCOMFORTS:
There are no foreseeable risks or discomforts involved in your participation in this survey since this survey is anonymous and voluntary. In addition, as no unique personal identifiers (i.e. SSN, picture ID, and voice) will be solicited, and as the data to be analysis will be aggregated, there are no anticipated risks to your legal, physical, emotional, or social well-being.

COMPENSATION:
You will receive $ 5 cash if you participate in this research survey and complete this survey questionnaire. The cash will be given at the end of this survey.

VOLUNTARY PARTICIPATION:
Participation in this research survey is anonymous and voluntary. You have the right do decline participation in any or all survey procedure or quit at any time at no consequences. Refusal to participate or withdrawl from the research survey will involve no penalty or loss of benefits to which you are otherwise entitled.

CONTACT FOR QUESTIONS:
Questions about this survey or your rights as a research subject may be directed to Sok-Hun Kim at (682)-234-2750 or at e-mail: sokhun.kim@mavs.uta.edu. You may contact the chairperson of the UT Arlington Institutional Review Board at (817)-272-3723 in the event of a research-related injury to the subject.

Word Choice Questionnaire

SECTION A: General Questionnaire

Date (mm/dd/yr)

		/			/	2	0	1	2

Native Language

English ☐	Korean ☐
Other:	

Sex

Male ☐	Female ☐

Education

High School ☐	Masters (MA) ☐	Other:
Bachelor (BA) ☐	PhD/MD/JD ☐	

Age

~19	☐
20~29	☐
30~39	☐
40~49	☐
50~59	☐
60~	☐

SECTION B: Naming Form Type I

DIRECTION

The following text is taken from a newspaper article.

Please read the paragraph, (i) rate your level of acceptability for **all naming forms (a, b, c, d) in the first three mentions** on a scale of five to one (5-1), (ii) write **other possible naming form(s) (e)** with the level of acceptability, and (iii) answer the question.

SECTION B (1)

Q1>_____ is an American baseball shortstop who has played 17 years in Major League Baseball (MLB) for the New York Yankees. A twelve-time All-Star and five-time World Series champion, **Q2>**_____ has been a central figure of the Yankees during their success of the 1990s and 2000s due to his clubhouse presence, on-field leadership, hitting ability, and baserunning. **Q3>**_____ is the Yankees' all-time career leader in hits (3,088), games played (2,426), stolen bases (329), and at bats (9,868). His accolades include four Silver Slugger Awards and five Gold Glove awards. (*Wikipedia*)

Q1.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Derek Jeter	5	4	3	2	1
b. Jeter	5	4	3	2	1
c. He	5	4	3	2	1
d. Derek	5	4	3	2	1
e. Other:	5	4	3	2	1

Q2.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Derek Jeter	5	4	3	2	1
b. Jeter	5	4	3	2	1
c. He	5	4	3	2	1
d. Derek	5	4	3	2	1
e. Other:	5	4	3	2	1

Q3.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Derek Jeter	5	4	3	2	1
b. Jeter	5	4	3	2	1
c. He	5	4	3	2	1
d. Derek	5	4	3	2	1
e. Other:	5	4	3	2	1

Q4. Please indicate the level of famousness for this person in the United States.

very well-known				very unknown
5	4	3	2	1

SECTION B (2)

Q1>_____ is an American singer-songwriter and occasional actress. In 2002, **Q2**>_____ launched her solo music career with the release of the commercially successful and critically acclaimed album *Come Away With Me*, which was certified a diamond album in 2002, selling over 20 million copies. **Q3**>_____ earned five Grammy Awards, including the Album of the Year, Record of the Year, and Best New Artist. Her subsequent studio albums, *Feels Like Home*, released in 2004, *Not Too Late*, released in 2007, the same year she made her film debut in *My Blueberry Nights*, and her 2009 release *The Fall* all gained Platinum status after selling over a million copies and were generally well received by critics. (*Wikipedia*)

Q1.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Norah Jones	5	4	3	2	1
b. Jones	5	4	3	2	1
c. She	5	4	3	2	1
d. Norah	5	4	3	2	1
e. Other:	5	4	3	2	1

Q2.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Norah Jones	5	4	3	2	1
b. Jones	5	4	3	2	1
c. She	5	4	3	2	1
d. Norah	5	4	3	2	1
e. Other:	5	4	3	2	1

Q3.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Norah Jones	5	4	3	2	1
b. Jones	5	4	3	2	1
c. She	5	4	3	2	1
d. Norah	5	4	3	2	1
e. Other:	5	4	3	2	1

Q4. Please indicate the level of famousness for this person in the United States.

very well-known				very unknown
5	4	3	2	1

SECTION B (3)

Q1>_____, born 1949 in Seattle, Washington, is the President and Publisher of *Farrar, Straus and Giroux*, one of the eight major publishers in New York.

Q2>_____ began his publishing career at Houghton Mifflin in Boston, moved to Random House in New York, and finally, to Farrar, Straus & Giroux. **Q3**>_____ joined FSG as executive editor in 1985, after being fired from Random House. Two years later, he was named editor-in-chief, and is now President and Publisher. (Wikipedia)

Q1.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Jonathan Galassi	5	4	3	2	1
b. Galassi	5	4	3	2	1
c. He	5	4	3	2	1
d. Jonathan	5	4	3	2	1
e. Other:	5	4	3	2	1

Q2.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Jonathan Galassi	5	4	3	2	1
b. Galassi	5	4	3	2	1
c. He	5	4	3	2	1
d. Jonathan	5	4	3	2	1
e. Other:	5	4	3	2	1

Q3.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Jonathan Galassi	5	4	3	2	1
b. Galassi	5	4	3	2	1
c. He	5	4	3	2	1
d. Jonathan	5	4	3	2	1
e. Other:	5	4	3	2	1

Q4. Please indicate the level of famousness for this person in the United States.

very well-known				very unknown
5	4	3	2	1

SECTION B (4)

Q1>_____ was an American businessman and inventor widely recognized as a charismatic pioneer of the personal computer revolution. **Q2**>_____ was co-founder, chairman, and chief executive officer of Apple Inc. **Q3**>_____ was co-founder and previously served as chief executive of Pixar Animation Studios; he became a member of the board of directors of The Walt Disney Company in 2006, following the acquisition of Pixar by Disney. (Wikipedia)

Q1.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Steve Jobs	5	4	3	2	1
b. Jobs	5	4	3	2	1
c. He	5	4	3	2	1
d. Steve	5	4	3	2	1
e. Other:	5	4	3	2	1

Q2.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Steve Jobs	5	4	3	2	1
b. Jobs	5	4	3	2	1
c. He	5	4	3	2	1
d. Steve	5	4	3	2	1
e. Other:	5	4	3	2	1

Q3.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Steve Jobs	5	4	3	2	1
b. Jobs	5	4	3	2	1
c. He	5	4	3	2	1
d. Steve	5	4	3	2	1
e. Other:	5	4	3	2	1

Q4. Please indicate the level of famousness for this person in the United States.

very well-known				very unknown
5	4	3	2	1

SECTION B (5)

Q1>_____ is an American late-night talk show host and stand-up comedian.

Q2>_____ hosted *The Tonight Show with Jay Leno* from 1992-2009, then hosted his own show, *The Jay Leno Show* for one year due to a host controversy, before returning to *The Tonight Show* in 2010. **Q3**>_____ is known for his clean, America-friendly act and his characteristically protruding chin. He has won an Emmy Award.

(Biography)

Q1.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Jay Leno	5	4	3	2	1
b. Leno	5	4	3	2	1
c. He	5	4	3	2	1
d. Jay	5	4	3	2	1
e. Other:	5	4	3	2	1

Q2.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Jay Leno	5	4	3	2	1
b. Leno	5	4	3	2	1
c. He	5	4	3	2	1
d. Jay	5	4	3	2	1
e. Other:	5	4	3	2	1

Q3.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Jay Leno	5	4	3	2	1
b. Leno	5	4	3	2	1
c. He	5	4	3	2	1
d. Jay	5	4	3	2	1
e. Other:	5	4	3	2	1

Q4. Please indicate the level of famousness for this person in the United States.

very well-known				very unknown
5	4	3	2	1

SECTION B (6)

Q1>_____ (born September 8, 1947) is an American short story writer and novelist. **Q2**>_____ has received an award for excellence from the American Academy and Institute of Arts and Letters and a PEN/Bernard Malamud Award for excellence in the short story form. The work has been compared to that of Alice Adams, J.D. Salinger, John Cheever, and John Updike. **Q3**>_____ holds an undergraduate degree from American University and a masters degree from the University of Connecticut.

(Wikipedia)

Q1.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Ann Beattie	5	4	3	2	1
b. Beattie	5	4	3	2	1
c. She	5	4	3	2	1
d. Ann	5	4	3	2	1
e. Other:	5	4	3	2	1

Q2.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Ann Beattie	5	4	3	2	1
b. Beattie	5	4	3	2	1
c. She	5	4	3	2	1
d. Ann	5	4	3	2	1
e. Other:					

Q3.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Ann Beattie	5	4	3	2	1
b. Beattie	5	4	3	2	1
c. She	5	4	3	2	1
d. Ann	5	4	3	2	1
e. Other:					

Q4. Please indicate the level of famousness for this person in the United States.

very well-known				very unknown
5	4	3	2	1

SECTION C: Naming Form Type II

DIRECTION

The following text is taken from a newspaper article. The three paragraphs are from the same text and are arranged in order. That is, paragraph A comes at the beginning of the text. Paragraph B follows it. And paragraph C follows that.

Please read the paragraphs, (i) rate your level of acceptability for **all naming forms (a, b, c, d) in each paragraph** on a scale of five to one (5-1), (ii) write **other possible naming form(s) (e)** with the level of acceptability, and (iii) answer the question.

SECTION C (1)

Paragraph A

Q1>_____ is the shortstop and captain of the New York Yankees. He has been an all-star, a most valuable player and one whose name is spoken in the adoring tones reserved mostly for players with names like Ruth, Gehrig or Mantle. In the long trip from phenom to aging veteran, his skills may have begun to wane, but his popularity has not.

Paragraph B

Q2>_____ has been a team leader who never gets in trouble in an era rife with athlete misbehavior and discontent. He has also been the most marketable athlete in New York — as safe for an advertiser to use as a deceased icon.

Paragraph C

Q3>_____ has played for no other major league team, which puts him in a revered class with seven Yankees Hall of Famers, as well as Thurman Munson and Don Mattingly, his predecessors as team captain. For the current generation of Yankees fans, he is the equivalent of other Yankees-for-life.

(*The New York Times*, *Times Topics*)

Q1.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Derek Jeter	5	4	3	2	1
b. Jeter	5	4	3	2	1
c. He	5	4	3	2	1
d. Derek	5	4	3	2	1
e. Other:	5	4	3	2	1

Q2.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Derek Jeter	5	4	3	2	1
b. Jeter	5	4	3	2	1
c. He	5	4	3	2	1
d. Derek	5	4	3	2	1
e. Other:	5	4	3	2	1

Q3.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Derek Jeter	5	4	3	2	1
b. Jeter	5	4	3	2	1
c. He	5	4	3	2	1
d. Derek	5	4	3	2	1
e. Other:	5	4	3	2	1

Q4. Please indicate the level of famousness for this person in the United States.

very well-known				very unknown
5	4	3	2	1

SECTION C (2)

Paragraph A

Q1>_____ is an American singer-songwriter and occasional actress.

Paragraph B

Q2>_____ launched her solo music career in 2002 with the release of the commercially successful and critically acclaimed album *Come Away With Me*, which was certified a diamond album in 2002, selling over 20 million copies. The record earned Jones five Grammy Awards, including the Album of the Year, Record of the Year, and Best New Artist. Her subsequent studio albums, *Feels Like Home*, released in 2004, *Not Too Late*, released in 2007, the same year she made her film debut in *My Blueberry Nights*, and her 2009 release *The Fall* all gained Platinum status after selling over a million copies and were generally well received by critics.

Paragraph C

Q3>_____ has won nine Grammy Awards and was *Billboard* magazine's 60th-best-selling music artist of the 2000–2009 decade. Throughout her career, Jones has won numerous awards and has sold over 40 million albums worldwide. Billboard magazine named her the top Jazz artist of the 2000–2009 decade, establishing herself as one of the best-selling artists of her time.

(Wikipedia)

Q1.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Norah Jones	5	4	3	2	1
b. Jones	5	4	3	2	1
c. She	5	4	3	2	1
d. Norah	5	4	3	2	1
e. Other:	5	4	3	2	1

Q2.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Norah Jones	5	4	3	2	1
b. Jones	5	4	3	2	1
c. She	5	4	3	2	1
d. Norah	5	4	3	2	1
e. Other:	5	4	3	2	1

Q3.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Norah Jones	5	4	3	2	1
b. Jones	5	4	3	2	1
c. She	5	4	3	2	1
d. Norah	5	4	3	2	1
e. Other:	5	4	3	2	1

Q4. Please indicate the level of famousness for this person in the United States.

very well-known				very unknown
5	4	3	2	1

SECTION C (3)

Paragraph A

Q1>_____ is the President and Publisher of *Farrar, Straus and Giroux*, one of the eight major publishers in New York. He began his publishing career at Houghton Mifflin in Boston, moved to Random House in New York, and finally, to Farrar, Straus & Giroux. He joined FSG as executive editor in 1985, after being fired from Random House. Two years later, he was named editor-in-chief, and is now President and Publisher.

Paragraph B

Q2>_____ is also a translator of poetry and a poet himself. He has translated and published the poetic works of the Italian poets Giacomo Leopardi and Eugenio Montale. His honors as a poet include a 1989 Guggenheim Fellowship, and his activities include having been poetry editor for *The Paris Review* for ten years, and being an honorary chairman of the Academy of American Poets. He has published poems in literary journals and magazines including *Threepenny Review, The New Yorker, The Nation* and the Poetry Foundation website.

Paragraph C

Q3>_____ graduated from Phillips Exeter Academy where he became interested in poetry, writing and literature, and from Harvard College in 1971. He was a Marshall Scholar at Christ's College, Cambridge. He realized while attending Christ's College that he wanted a career in book publishing. Galassi was born in Seattle (his father worked as an attorney for the Justice Department), but he grew up in Plympton, Massachusetts. He lives in Brooklyn. He was married to Susan Grace, with whom he had two daughters. The couple divorced in late 2011. (Wikipedia)

Q1.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Jonathan Galassi	5	4	3	2	1
b. Galassi	5	4	3	2	1
c. He	5	4	3	2	1
d. Jonathan	5	4	3	2	1
e. Other:	5	4	3	2	1

Q2.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Jonathan Galassi	5	4	3	2	1
b. Galassi	5	4	3	2	1
c. He	5	4	3	2	1
d. Jonathan	5	4	3	2	1
e. Other:	5	4	3	2	1

Q3.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Jonathan Galassi	5	4	3	2	1
b. Galassi	5	4	3	2	1
c. He	5	4	3	2	1
d. Jonathan	5	4	3	2	1
e. Other:	5	4	3	2	1

Q4. Please indicate the level of famousness for this person in the United States.

very well-known				very unknown
5	4	3	2	1

SECTION C (4)

Paragraph A

Q1>_____ was an American businessman and inventor widely recognized as a charismatic pioneer of the personal computer revolution. He was co-founder, chairman, and chief executive officer of Apple Inc. Jobs was co-founder and previously served as chief executive of Pixar Animation Studios; he became a member of the board of directors of The Walt Disney Company in 2006, following the acquisition of Pixar by Disney.

Paragraph B

Q2>_____ directed the Apple II sits aesthetic design of the Apple II series in the late 1970s, marketing along with A.C. "Mike" Markkula, Jr. and others. In the early 1980s, Jobs was among the first to see the commercial potential of Xerox PARC's mouse-driven graphical user interface, which led to the creation of the Apple Lisa (engineered by Ken Rothmuller and John Couch) and, one year later, creation of Apple employee Jef Raskin's Macintosh.

Paragraph C

Q3>_____ left Apple after losing a power struggle with the board of directors in 1985 and founded NeXT, a computer platform development company specializing in the higher-education and business markets. In 1986, he acquired the computer graphics division of Lucasfilm Ltd, which was spun off as Pixar Animation Studios. He was credited in *Toy Story* (1995) as an executive producer. (Wikipedia)

Q1.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Steve Jobs	5	4	3	2	1
b. Jobs	5	4	3	2	1
c. He	5	4	3	2	1
d. Steve	5	4	3	2	1
e. Other:	5	4	3	2	1

Q2.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Steve Jobs	5	4	3	2	1
b. Jobs	5	4	3	2	1
c. He	5	4	3	2	1
d. Steve	5	4	3	2	1
e. Other:	5	4	3	2	1

Q3.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Steve Jobs	5	4	3	2	1
b. Jobs	5	4	3	2	1
c. He	5	4	3	2	1
d. Steve	5	4	3	2	1
e. Other:	5	4	3	2	1

Q4. Please indicate the level of famousness for this person in the United States.

very well-known				very unknown
5	4	3	2	1

SECTION C (5)

Paragraph A

Q1>_____ was born James Douglas Muir Leno, on April 28, 1950, in New Rochelle, New York (raised in Andover, MA), to Angelo Leno, and insurance salesman, and Cathryn Leno, a homemaker.

Paragraph B

Q2>_____ displayed his comic tendencies with pranks and practical jokes, as a student in grade school. His fifth-grade teacher's report card comments - "if Jay spent as much time studying as he does trying to be a comedian, he'd be a big star" - turned out to be prophetic.

Paragraph C

Q3>_____ is best known for being one of the busiest performers in comedy; for years he booked over 300 appearances annually. He's also famous for his clean, observational brand of humor and a cartoon-like facial features, which he made reference to in the title of his 1996 autobiography, *Leading With My Chin*.

(Biography)

Q1.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Jay Leno	5	4	3	2	1
b. Leno	5	4	3	2	1
c. He	5	4	3	2	1
d. Jay	5	4	3	2	1
e. Other:	5	4	3	2	1

Q2.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Jay Leno	5	4	3	2	1
b. Leno	5	4	3	2	1
c. He	5	4	3	2	1
d. Jay	5	4	3	2	1
e. Other:	5	4	3	2	1

Q3.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Jay Leno	5	4	3	2	1
b. Leno	5	4	3	2	1
c. He	5	4	3	2	1
d. Jay	5	4	3	2	1
e. Other:	5	4	3	2	1

Q4. Please indicate the level of famousness for this person in the United States.

very well-known				very unknown
5	4	3	2	1

SECTION C (6)

Paragraph A

Q1>_____ (born September 8, 1947) is an American short story writer and novelist. She has received an award for excellence from the American Academy and Institute of Arts and Letters. Her work has been compared to that of Alice Adams, J.D. Salinger, John Cheever, and John Updike. She holds an undergraduate degree from American University and a masters degree from the University of Connecticut.

Paragraph B

Q2>_____ grew up in Chevy Chase, Washington, D.C., born in Washington, D.C. She gained attention in the early 1970s with short stories published in *The Western Humanities Review*, *Ninth Letter*, the *Atlantic Monthly*, and *The New Yorker*. Critics have praised her writing for its keen observations and dry, matter-of-fact irony which chronicle disillusionments of the upper-middle-class generation that grew up in the 1960s.

Paragraph C

Q3>_____ has taught at Harvard College and the University of Connecticut and presently teaches at the University of Virginia, where she is the Edgar Allan Poe Chair of the Department of English and Creative Writing. In 2005 she was selected as winner of the Rea Award for the Short Story, in recognition of her outstanding achievement in that genre. (Wikipedia)

Q1.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Ann Beattie	5	4	3	2	1
b. Beattie	5	4	3	2	1
c. She	5	4	3	2	1
d. Ann	5	4	3	2	1
e. Other:	5	4	3	2	1

Q2.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Ann Beattie	5	4	3	2	1
b. Beattie	5	4	3	2	1
c. She	5	4	3	2	1
d. Ann	5	4	3	2	1
e. Other:					

Q3.

acceptability / naming form	perfectly acceptable				absolutely unacceptable
a. Ann Beattie	5	4	3	2	1
b. Beattie	5	4	3	2	1
c. She	5	4	3	2	1
d. Ann	5	4	3	2	1
e. Other:					

Q4. Please indicate the level of famousness for this person in the United States.

very well-known				very unknown
5	4	3	2	1

SECTION D: Pronoun

DIRECTION

The following conversations are taken from a mockumentary comedy. Please read each conversation, (i) rate your level of acceptability for **all pronouns (a, b) in each conversation** on a scale of five to one (5-1) and (ii) write **other possible pronoun(s) (c)** with the level of acceptability.

Conversation A

Pam: I was in the meeting with Jan, and, she did say that *it could be this branch that gets the axe*.

Worker: Are you sure about **Q1**>_____? (*The Office*, Season 1, Episode #1)

Conversation B

Jim: Solitaire?

Pam: Yeah, Freecell.

Jim: (points at the monitor) *Six on the seven*.

Pam: I know, I saw that.

Jim: So then, why didn't you do **Q2**>_____?

Pam: I'm saving that, 'cause I like it when the cards go T-ts-ts-tch-tch-tch.

Jim: Who doesn't love that? (*The Office*, Season 1, Episode #2)

Q1.

acceptability / pro-form	perfectly acceptable				absolutely unacceptable
a. that	5	4	3	2	1
b. it	5	4	3	2	1
c. Other:	5	4	3	2	1

Q2.

acceptability / pro-form	perfectly acceptable				absolutely unacceptable
a. that	5	4	3	2	1
b. it	5	4	3	2	1
c. Other:	5	4	3	2	1

Conversation C

Oscar: And, uh, *they moved to the United States a year before I was born.*

Michael: Yeah...

Oscar: So *I grew up in the United States.*

Michael: Wow.

Oscar: And, *my parents were Mexican.*

Michael: Wow. **Q3**>_____'s a great story. That's the American dream right there, right?

(*The Office*, Season 1, Episode #1)

Conversation D

Michael: *People I respect, heroes would be, Bob Hope, um, Abraham Lincoln, definitely. Bono, and probably God, would be the fourth one. And I just think those people really, uh, helped, the world, in so many ways.* That it's, um, *it*'s really beyond words. **Q4**>_____'s really incalculcable.

(*The Office*, Season 1, Episode #1)

Q3.

acceptability / pro-form	perfectly acceptable				absolutely unacceptable
a. that	5	4	3	2	1
b. it	5	4	3	2	1
c. Other:	5	4	3	2	1

Q4.

acceptability / pro-form	perfectly acceptable				absolutely unacceptable
a. that	5	4	3	2	1
b. it	5	4	3	2	1
c. Other:	5	4	3	2	1

COMMENTS ON SECTION B (Naming Form Type I)

● If you have any ideas or comments on these questions and this survey, please share them.

COMMENTS ON SECTION C (Naming Form Type II)

● If you have any ideas or comments on these questions and this survey, please share them.

COMMENTS ON SECTION D (Pronoun)

● If you have any ideas or comments on these questions and this survey, please share them.

Thanks for your time and participation!

Appendix B

The Survey Questionnaire in Korean

APPROVED

MAR 2 6 2012

Institutional Review Board

INFORMED CONSENT

PRINCIPAL INVESTIGATOR NAME:
Kim, Sok-Hun

TITLE OF PROJECT:
A Corpus-based Analysis of Discourse Anaphora in English and Korean: A Neo-Gricean Pragmatic Approach (A Survey of Word Choice: Naming Forms and Pronouns)

INTRODUCTION:
You are being asked to participate in a research survey about word choice; i.e., naming forms and pronouns. Your participation is anonymous and voluntary. Refusal to participate or withdrawal from the survey at any time will involve no penalty or loss of benefits to which you are otherwise entitled. Please ask questions if there is anything you do not understand.

PURPOSE:
The purpose of this study is to investigate possible associations between one's degree of preference for naming forms and degree of famousness of particular public figures in texts, and those between one's degree of preference for two pronouns in conversations and different contexts.

DURATION:
The survey session will take approximately 30 minutes to complete.

NUMBER OF PARTICIPANTS:
The survey will be conducted by a maximum of 25 native speakers of English and a maximum of 25 native speakers of Korean, depending on availability and willingness of potential participants.

PROCEDURES:
1. Recruiting and Two screening questions
You will be recruited directly by me by canvassing public areas of the UT Arlington campus to participate in this research survey. After being notified that the survey is anonymous and voluntary, and being asked about your willingness to participate, you will be asked to answer two screening questions regarding native language (i.e., English & Korean) and age (i.e., over 18). If you are willing to participate in this research survey and meeting these screening criteria, you will be presented with the survey's questions of your native language.
2. Survey questionnaire for each native speaker of each language

You will complete a word choice questionnaire. The form includes questions related to your natural uses of naming forms and pronouns in texts. During the survey, you will see four sections of questions. In section A, you will be asked to complete questions about yourself (native language, sex, education, and age). In section B, you will be asked to complete questions about your preference (from 5-1) for naming forms (such as Steve Jobs, Jobs, He, Steve, and other possible forms) in texts by the first three mentions. In section B, you will be asked to complete the same type of questions about your preference (from 5-1) for naming forms in texts, but by the first subject position in each paragraph. In section D, you will be asked to complete questions about your preference (from 5-1) between two pronouns (*that* and *it* in English; *kuken* 'that' and *kulen* 'it' in Korean) in conversations and different contexts.

POSSIBLE BENEFITS:
There are no direct benefits to you for participating in this survey other than the possibility for considering an association between naming forms and famousness of a public figure, an association between pronouns and different contexts in your native language.

POSSIBLE RISKS/DISCOMFORTS:
There are no foreseeable risks or discomforts involved in your participation in this survey since this survey is anonymous and voluntary. In addition, as no unique personal identifiers (i.e. SSN, picture ID, and voice) will be solicited, and as the data to be analysis will be aggregated, there are no anticipated risks to your legal, physical, emotional, or social well-being.

COMPENSATION:
You will receive $ 5 cash if you participate in this research survey and complete this survey questionnaire. The cash will be given at the end of this survey.

VOLUNTARY PARTICIPATION:
Participation in this research survey is anonymous and voluntary. You have the right do decline participation in any or all survey procedure or quit at any time at no consequences. Refusal to participate or withdrawl from the research survey will involve no penalty or loss of benefits to which you are otherwise entitled.

CONTACT FOR QUESTIONS:
Questions about this survey or your rights as a research subject may be directed to Sok-Hun Kim at (682)-234-2750 or at e-mail: sokhun.kim@mavs.uta.edu. You may contact the chairperson of the UT Arlington Institutional Review Board at (817)-272-3723 in the event of a research-related injury to the subject.

어휘 선택 질문지 (Word Choice Questionnaire)

SECTION A: 일반사항

날짜 (월/일/연도) (Date (mm/dd/yr))

		/			/	2	0	1	2

모국어 (Native Language)

영어 (English) ☐	한국어 (Korean) ☐
기타 (Other):	

Sex (성별)

남성(Male) ☐	여성(Female) ☐

교육 (Education)

고등학교 (High School) ☐	석사 (Masters (MA)) ☐	기타 (Other):
학사 (Bachelor (BA)) ☐	박사 (PhD/MD/JD) ☐	

연령 (Age)

~19	☐
20~29	☐
30~39	☐
40~49	☐
50~59	☐
60~	☐

SECTION B: 이름형태 유형 I

지시사항

다음의 글은 신문 기사에서 발췌된 내용입니다.

빈칸에 들어갈 **a, b, c, d 모두의 이름 형태에 대하여 각각의 알맞은 정도**

(수용가능성, acceptability)를 **5** (매우 적절함)에서 **1** (매우 부적절함)중에 하나로 표시해 주시면 됩니다.

사용가능한 다른 이름의 형태와 그 알맞은 정도도 **기타(other) (e)**란에 작성해 주시고 질문에 답하여 주시면 됩니다.

SECTION B (1)

Q1>_____ (37.피츠버그 파이리츠)은/는 미국프로야구 메이저리그에서 124 승째를 수확하며 아시아 출신 선수 통산 최다승 신기록을 세운기쁨을 감추지 않았다. **Q2>**_____은/는 2 일 플로리다 말린스와 경기에서 3 이닝 무실점 호투로 통산 124 승째를 거둔 뒤 메이저리그 공식 홈페이지 'MLB.com'에 올라온 인터뷰에서 "메이저리그에서 124 승이 그리 위대한 것은 아니지만 (나에게는) 매우 특별하다"고 운을 뗐다. **Q3>**_____은/는 이어 "처음 미국에 왔던 17 년 전이 떠오른다"면서 "나를 여기에 데려오고 도와줬던 사람들이 생각난다"며 감사 인사를 전했다. (*Yonhap News*, October 2, 2010)

Q1.

수용가능성 / 이름 형태	매우 적절함				매우 부적절함
a. 박찬호	5	4	3	2	1
b. 박	5	4	3	2	1
c. 그	5	4	3	2	1
d. 찬호	5	4	3	2	1
e. 기타:	5	4	3	2	1

Q2.

수용가능성 / 이름 형태	매우 적절함				매우 부적절함
a. 박찬호	5	4	3	2	1
b. 박	5	4	3	2	1
c. 그	5	4	3	2	1
d. 찬호	5	4	3	2	1
e. 기타:	5	4	3	2	1

Q3.

수용가능성 / 이름 형태	매우 적절함				매우 부적절함
a. 박찬호	5	4	3	2	1
b. 박	5	4	3	2	1
c. 그	5	4	3	2	1
d. 찬호	5	4	3	2	1
e. 기타:	5	4	3	2	1

Q4. 위의 글에 언급된 이름의 사람이 한국에서 어느정도 유명한지를 표시해 주세요.

매우 유명함				전혀 유명하지 않음
5	4	3	2	1

SECTION B (2)

싱어송라이터 **Q1>**_____ 이/가 '세이 예스-필링/힐링(SAY YES-FEELING/HEALING)'이란 타이틀로 올해 첫 단독 공연을 개최한다. **Q2>**_____은/는 11 월 26 일, 27 일 양일간에 걸쳐 KT&G 상상마당 라이브홀에서 두 가지 버전의 색다른 무대를 꾸밀 예정이다. **Q3>**_____ 은/는 '사랑해 사랑해', '언젠가는', '공무도하가', '새' '비밀의 화원' 등 대중에게 사랑받은 대표곡을 두가지 버전으로 들려줄 계획이다.

(*Herald Biz*, November 19, 2011)

Q1.

수용가능성 / 이름 형태	매우 적절함				매우 부적절함
a. 이상은	5	4	3	2	1
b. 이	5	4	3	2	1
c. 그	5	4	3	2	1
d. 상은	5	4	3	2	1
e. 기타:	5	4	3	2	1

Q2.

수용가능성 / 이름 형태	매우 적절함				매우 부적절함
a. 이상은	5	4	3	2	1
b. 이	5	4	3	2	1
c. 그	5	4	3	2	1
d. 상은	5	4	3	2	1
e. 기타:	5	4	3	2	1

Q3.

수용가능성 / 이름 형태	매우 적절함				매우 부적절함
a. 이상은	5	4	3	2	1
b. 이	5	4	3	2	1
c. 그	5	4	3	2	1
d. 상은	5	4	3	2	1
e. 기타:	5	4	3	2	1

Q4. 위의 글에 언급된 이름의 사람이 한국에서 어느정도 유명한지를 표시해 주세요.

매우 유명함				전혀 유명하지 않음
5	4	3	2	1

SECTION B (3)

소설가 **Q1>**_____(67) 이/가 암투병을 통해 자신을 성찰한 글을 천주교 서울대교구 '서울주보'에 연재해 잔잔한 반향을 일으키고 있다. **Q2>**_____ 은/는 1 월 1 일자 주보 '말씀의 이삭' 코너에 실린 '지금 이 세상 어디선가 누군가 울고 있다'를 시작으로 매주 글을 싣고 있으며 다음 달까지 모두 9 편의 글을 연재할 예정이다. **Q3>**_____ 은/는 2008 년 침샘 부근 암 수술을 한 뒤 모진 항암 치료를 겪으며 소설 '낯익은 타인들의 도시'를 집필해 화제를 모으기도 했다.

(*Yonhap News*, January 10, 2012)

Q1.

수용가능성 / 이름 형태	매우 적절함				매우 부적절함
a. 최인호	5	4	3	2	1
b. 최	5	4	3	2	1
c. 그	5	4	3	2	1
d. 인호	5	4	3	2	1
e. 기타:	5	4	3	2	1

Q2.

수용가능성 / 이름 형태	매우 적절함				매우 부적절함
a. 최인호	5	4	3	2	1
b. 최	5	4	3	2	1
c. 그	5	4	3	2	1
d. 인호	5	4	3	2	1
e. 기타:	5	4	3	2	1

Q3.

수용가능성 / 이름 형태	매우 적절함				매우 부적절함
a. 최인호	5	4	3	2	1
b. 최	5	4	3	2	1
c. 그	5	4	3	2	1
d. 인호	5	4	3	2	1
e. 기타:	5	4	3	2	1

Q4. 위의 글에 언급된 이름의 사람이 한국에서 어느정도 유명한지를 표시해 주세요.

매우 유명함				전혀 유명하지 않음
5	4	3	2	1

SECTION B (4)

Q1>_____ 서울대 융합과학기술대학원장이 6 일 '안철수재단'(가칭) 설립 계획을 발표했다. **Q2>**_____은/는 이날 프레스센터에서 열린 재단 설립 발표 기자회견 인사말을 통해 "재단은 사회적으로 편중되어 있던 기회의 격차를 해소하는데 주력하겠다"며 "현재 가장 시급한 것이 일자리 문제, 소외 계층 교육, 세대간 소통이라고 판단해 우선 중점 사업으로 정했다"고 밝혔다. **Q3>**_____은/는 "재단의 작은 시작이 더불어 희망을 품고 살아가는 사회를 만드는데 의미있는 일이 되기를 바란다"며 "재단이 앞으로 우리 사회의 중요한 문제들을 좀 더 창조적이고 혁신적인 방법으로 해결해나가는데 기여할 수 있기를 바란다"고 말했다.

(*The Kyunghyang Shinmun*, February 6, 2012)

Q1.

수용가능성 / 이름 형태	매우 적절함				매우 부적절함
a. 안철수	5	4	3	2	1
b. 안	5	4	3	2	1
c. 그	5	4	3	2	1
d. 철수	5	4	3	2	1
e. 기타:	5	4	3	2	1

Q2.

수용가능성 / 이름 형태	매우 적절함				매우 부적절함
a. 안철수	5	4	3	2	1
b. 안	5	4	3	2	1
c. 그	5	4	3	2	1
d. 철수	5	4	3	2	1
e. 기타:	5	4	3	2	1

Q3.

수용가능성 / 이름 형태	매우 적절함				매우 부적절함
a. 안철수	5	4	3	2	1
b. 안	5	4	3	2	1
c. 그	5	4	3	2	1
d. 철수	5	4	3	2	1
e. 기타:	5	4	3	2	1

Q4. 위의 글에 언급된 이름의 사람이 한국에서 어느정도 유명한지를 표시해 주세요.

매우 유명함				전혀 유명하지 않음
5	4	3	2	1

SECTION B (5)

대상의 영광은 '국민 MC' **Q1>**_____에게 돌아갔다. **Q2>**_____은/는 30 일 오후 9 시 서울 등촌동 SBS 공개홀에서 열린 2011 SBS 연예대상에서 대상 수상의 영예를 안았다. **Q3>**_____은/는 '일요일이 좋다-런닝맨'에 출연해 게스트와 멤버간의 조화를 이끌어내고 캐릭터와 상황을 설정하는 등 탁월한 능력을 선보이고 있다. '런닝맨'은 초반에 포맷이 안정화되지 않아 부진했으나, 최근 긴박감 넘치는 에이스로 20%에 육박하는 시청률을 기록하며 큰 사랑을 받고 있다.

(*Money Today*, December 31, 2011)

Q1.

수용가능성 / 이름 형태	매우 적절함				매우 부적절함
a. 유재석	5	4	3	2	1
b. 유	5	4	3	2	1
c. 그	5	4	3	2	1
d. 재석	5	4	3	2	1
e. 기타:	5	4	3	2	1

Q2.

수용가능성 / 이름 형태	매우 적절함				매우 부적절함
a. 유재석	5	4	3	2	1
b. 유	5	4	3	2	1
c. 그	5	4	3	2	1
d. 재석	5	4	3	2	1
e. 기타:	5	4	3	2	1

Q3.

수용가능성 / 이름 형태	매우 적절함				매우 부적절함
a. 유재석	5	4	3	2	1
b. 유	5	4	3	2	1
c. 그	5	4	3	2	1
d. 재석	5	4	3	2	1
e. 기타:	5	4	3	2	1

Q4. 위의 글에 언급된 이름의 사람이 한국에서 어느정도 유명한지를 표시해 주세요.

매우 유명함				전혀 유명하지 않음
5	4	3	2	1

SECTION B (6)

작가 **Q1>**_____(48)이/가 돌아왔다. 올 4 월 장편소설 <엄마를 부탁해> 영문판이 일으킨 선풍 이후 북미와 유럽, 중동까지 지구 반바퀴를 돌아온 참이다. 책이 나온 10 개국의 여러 도시마다 며칠씩 머물다 떠나기를 4 개월여 반복했고, 아직 그 여정이 끝나지 않았다. 1 년간의 미국 컬럼비아대 방문학자 생활을 마치고 귀국했으나 곧 다시 호주와 일본으로 출국해야 한다. **Q2>**_____은/는 아직 시차에 적응하지 못해 29 일 기자간담회에서 몹시 피곤한 상태였다. **Q3>**_____은/는 "번역은 여행인 것 같다. 평소 여행을 즐기는 편은 아닌데 책과 함께 온 세계를 돌아다녔다. 새로운 언어로 책이 나올 때마다 그곳에서 나는 신인작가가 됐다. 3 박 4 일 동안 13 개 인터뷰를 한 적도 있다. 즐겁고 영광스러웠지만, 동시에 힘든 시간이었다"고 말했다.

(*The Kyunghyang Shinmun*, August 29, 2011)

Q1.

수용가능성 / 이름 형태	매우 적절함				매우 부적절함
a. 신경숙	5	4	3	2	1
b. 신	5	4	3	2	1
c. 그	5	4	3	2	1
d. 경숙	5	4	3	2	1
e. 기타:	5	4	3	2	1

Q2.

수용가능성 / 이름 형태	매우 적절함				매우 부적절함
a. 신경숙	5	4	3	2	1
b. 신	5	4	3	2	1
c. 그	5	4	3	2	1
d. 경숙	5	4	3	2	1
e. 기타:					

Q3.

수용가능성 / 이름 형태	매우 적절함				매우 부적절함
a. 신경숙	5	4	3	2	1
b. 신	5	4	3	2	1
c. 그	5	4	3	2	1
d. 경숙	5	4	3	2	1
e. 기타:					

Q4. 위의 글에 언급된 이름의 사람이 한국에서 어느정도 유명한지를 표시해 주세요.

매우 유명함				전혀 유명하지 않음
5	4	3	2	1

SECTION C: 이름형태 유형 II

지시사항

다음의 글은 신문 기사에서 발췌된 내용입니다. 본문은 단락 A, 단락 B, 단락 C 의 순서로 이루어진 하나의 글입니다.

각 단락의 빈칸에 들어갈 **a, b, c, d 모두의 이름 형태에 대하여 각각의 알맞은 정도** (수용가능성, acceptability)를 **5** (매우 적절함)에서 **1** (매우 부적절함)중에 하나로 표시해 주시면 됩니다.

사용가능한 다른 이름의 형태와 그 알맞은 정도도 **기타(other) (e)**란에 작성해 주시고 질문에 답하여 주시면 됩니다.

SECTION C (1)

단락 A

Q1>_____ (37.피츠버그 파이리츠)은/는 미국프로야구 메이저리그에서 124 승째를 수확하며 아시아 출신 선수 통산 최다승 신기록을 세운 기쁨을 감추지 않았다.

단락 B

Q2>_____은/는 2 일 플로리다 말린스와 경기에서 3 이닝 무실점 호투로 통산 124 승째를 거둔 뒤 메이저리그 공식 홈페이지 'MLB.com'에 올라온 인터뷰에서 "메이저리그에서 124 승이 그리 위대한 것은 아니지만 (나에게는) 매우 특별하다"고 운을 뗐다.

단락 C

Q3>_____은/는 이어 "처음 미국에 왔던 17 년 전이 떠오른다"면서 "나를 여기에 데려오고 도와줬던 사람들이 생각난다"며 감사 인사를 전했다.

(*Yonhap News*, October 2, 2010)

Q1.

수용가능성 / 이름 형태	매우 적절함				매우 부적절함
a. 박찬호	5	4	3	2	1
b. 박	5	4	3	2	1
c. 그	5	4	3	2	1
d. 찬호	5	4	3	2	1
e. 기타:	5	4	3	2	1

Q2.

수용가능성 / 이름 형태	매우 적절함				매우 부적절함
a. 박찬호	5	4	3	2	1
b. 박	5	4	3	2	1
c. 그	5	4	3	2	1
d. 찬호	5	4	3	2	1
e. 기타:	5	4	3	2	1

Q3.

수용가능성 / 이름 형태	매우 적절함				매우 부적절함
a. 박찬호	5	4	3	2	1
b. 박	5	4	3	2	1
c. 그	5	4	3	2	1
d. 찬호	5	4	3	2	1
e. 기타:	5	4	3	2	1

Q4. 위의 글에 언급된 이름의 사람이 한국에서 어느정도 유명한지를 표시해 주세요.

매우 유명함				전혀 유명하지 않음
5	4	3	2	1

SECTION C (2)

단락 A

싱어송라이터 **Q1>**_____ 이/가 '세이 예스-필링/힐링(SAY YES-FEELING/HEALING)'이란 타이틀로 올해 첫 단독 공연을 개최한다.

단락 B

Q2>_____은/는 11 월 26 일, 27 일 양일간에 걸쳐 KT&G 상상마당 라이브홀에서 두 가지 버전의 색다른 무대를 꾸밀 예정이다.

단락 C

Q3>_____ 은/는 '사랑해 사랑해', '언젠가는', '공무도하가', '새' '비밀의 화원' 등 대중에게 사랑받은 대표곡을 두가지 버전으로 들려줄 계획이다.

(*Herald Biz*, November 19, 2011)

Q1.

수용가능성 / 이름 형태	매우 적절함				매우 부적절함
a. 이상은	5	4	3	2	1
b. 이	5	4	3	2	1
c. 그	5	4	3	2	1
d. 상은	5	4	3	2	1
e. 기타:	5	4	3	2	1

Q2.

수용가능성 / 이름 형태	매우 적절함				매우 부적절함
a. 이상은	5	4	3	2	1
b. 이	5	4	3	2	1
c. 그	5	4	3	2	1
d. 상은	5	4	3	2	1
e. 기타:	5	4	3	2	1

Q3.

수용가능성 / 이름 형태	매우 적절함				매우 부적절함
a. 이상은	5	4	3	2	1
b. 이	5	4	3	2	1
c. 그	5	4	3	2	1
d. 상은	5	4	3	2	1
e. 기타:	5	4	3	2	1

Q4. 위의 글에 언급된 이름의 사람이 한국에서 어느정도 유명한지를 표시해 주세요.

매우 유명함				전혀 유명하지 않음
5	4	3	2	1

SECTION C (3)

단락 A

소설가 **Q1>**_____이/가 암투병을 통해 자신을 성찰한 글을 천주교 서울대교구 '서울주보'에 연재해 잔잔한 반향을 일으키고 있다.

단락 B

Q2>_____ 은/는 1 월 1 일자 주보 '말씀의 이삭' 코너에 실린 '지금 이 세상 어디선가 누군가 울고 있다'를 시작으로 매주 글을 싣고 있으며 다음 달까지 모두 9 편의 글을 연재할 예정이다.

단락 C

Q3>_____ 은/는 2008 년 침샘 부근 암 수술을 한 뒤 모진 항암 치료를 겪으며 소설 '낯익은 타인들의 도시'를 집필해 화제를 모으기도 했다.

(*Yonhap News*, January 10, 2012)

Q1.

수용가능성 / 이름 형태	매우 적절함				매우 부적절함
a. 최인호	5	4	3	2	1
b. 최	5	4	3	2	1
c. 그	5	4	3	2	1
d. 인호	5	4	3	2	1
e. 기타:	5	4	3	2	1

Q2.

수용가능성 / 이름 형태	매우 적절함				매우 부적절함
a. 최인호	5	4	3	2	1
b. 최	5	4	3	2	1
c. 그	5	4	3	2	1
d. 인호	5	4	3	2	1
e. 기타:					

Q3.

수용가능성 / 이름 형태	매우 적절함				매우 부적절함
a. 최인호	5	4	3	2	1
b. 최	5	4	3	2	1
c. 그	5	4	3	2	1
d. 인호	5	4	3	2	1
e. 기타:					

Q4. 위의 글에 언급된 이름의 사람이 한국에서 어느정도 유명한지를 표시해 주세요.

매우 유명함				전혀 유명하지 않음
5	4	3	2	1

SECTION C (4)

단락 A

Q1>_____ 서울대 융합과학기술대학원장이 이달 중순 마이크로소프트(MS) 공동 창업주로 기부 전도사인 빌 게이츠를 만난다.

단락 B

Q2>_____은/는 오는 8 일 융합과학기술대학원의 신임 교수 채용을 목적으로 동료 교수들과 미국으로 출국할 예정이다.

단락 C

Q3>_____ 은/는 자신이 보유한 안철수연구소 지분의 절반을 사회에 환원키로 하였으며 이번 출장에서 현지시각 11 일 빌 & 멜린다 게이츠 재단을 설립한 게이츠를 만나기로 해 눈길을 끌고 있다.

(*Yonhap News*, January 5, 2012)

Q1.

수용가능성 / 이름 형태	매우 적절함				매우 부적절함
a. 안철수	5	4	3	2	1
b. 안	5	4	3	2	1
c. 그	5	4	3	2	1
d. 철수	5	4	3	2	1
e. 기타:	5	4	3	2	1

Q2.

수용가능성 / 이름 형태	매우 적절함				매우 부적절함
a. 안철수	5	4	3	2	1
b. 안	5	4	3	2	1
c. 그	5	4	3	2	1
d. 철수	5	4	3	2	1
e. 기타:	5	4	3	2	1

Q3.

수용가능성 / 이름 형태	매우 적절함				매우 부적절함
a. 안철수	5	4	3	2	1
b. 안	5	4	3	2	1
c. 그	5	4	3	2	1
d. 철수	5	4	3	2	1
e. 기타:	5	4	3	2	1

Q4. 위의 글에 언급된 이름의 사람이 한국에서 어느정도 유명한지를 표시해 주세요.

매우 유명함				전혀 유명하지 않음
5	4	3	2	1

SECTION C (5)

단락 A

'국민 MC' **Q1>**_____이/가 'SBS 연예대상'에서 대상을 수상했다.

단락 B

Q2>_____은/는 '국민 MC'라는 타이틀에 걸맞게 프로그램을 위해 성실함과 리더십을 보이며 시청자들에게 큰 재미와 웃음을 선사해 왔다.

단락 C

Q3>_____은/는 시상식에서 "굉장히 기쁘고 뭐라 말로 표현할 수 없는 여러가지 감정이 든다"며 "지금까지 함께 와 준 스태프와 멤버들에게 이 모든 영광을 돌린다"고 말했다.

(*CCToday*, December 31, 2011)

Q1.

수용가능성 / 이름 형태	매우 적절함				매우 부적절함
a. 유재석	5	4	3	2	1
b. 유	5	4	3	2	1
c. 그	5	4	3	2	1
d. 재석	5	4	3	2	1
e. 기타:	5	4	3	2	1

Q2.

수용가능성 / 이름 형태	매우 적절함				매우 부적절함
a. 유재석	5	4	3	2	1
b. 유	5	4	3	2	1
c. 그	5	4	3	2	1
d. 재석	5	4	3	2	1
e. 기타:	5	4	3	2	1

Q3.

수용가능성 / 이름 형태	매우 적절함				매우 부적절함
a. 유재석	5	4	3	2	1
b. 유	5	4	3	2	1
c. 그	5	4	3	2	1
d. 재석	5	4	3	2	1
e. 기타:	5	4	3	2	1

Q4. 위의 글에 언급된 이름의 사람이 한국에서 어느정도 유명한지를 표시해 주세요.

매우 유명함				전혀 유명하지 않음
5	4	3	2	1

SECTION C (6)

단락 A

작가 **Q1>**_____(48)이/가 돌아왔다. 올 4 월 장편소설 <엄마를 부탁해> 영문판이 일으킨 선풍 이후 북미와 유럽, 중동까지 지구 반바퀴를 돌아온 참이다. 책이 나온 10 개국의 여러 도시마다 며칠씩 머물다 떠나기를 4 개월여 반복했고, 아직 그 여정이 끝나지 않았다. 1 년간의 미국 컬럼비아대 방문학자 생활을 마치고 귀국했으나 곧 다시 호주와 일본으로 출국해야 한다.

단락 B

Q2>_____은/는 아직 시차에 적응하지 못해 29 일 기자간담회에서 몹시 피곤한 상태였다.

단락 C

Q3>_____은/는 "번역은 여행인 것 같다. 평소 여행을 즐기는 편은 아닌데 책과 함께 온 세계를 돌아다녔다. 새로운 언어로 책이 나올 때마다 그곳에서 나는 신인작가가 됐다. 3 박 4 일 동안 13 개 인터뷰를 한 적도 있다. 즐겁고 영광스러웠지만, 동시에 힘든 시간이었다"고 말했다.

(*The Kyunghyang Shinmun*, August 29, 2011)

Q1.

수용가능성 / 이름 형태	매우 적절함				매우 부적절함
a. 신경숙	5	4	3	2	1
b. 신	5	4	3	2	1
c. 그	5	4	3	2	1
d. 경숙	5	4	3	2	1
e. 기타:	5	4	3	2	1

Q2.

수용가능성 / 이름 형태	매우 적절함				매우 부적절함
a. 신경숙	5	4	3	2	1
b. 신	5	4	3	2	1
c. 그	5	4	3	2	1
d. 경숙	5	4	3	2	1
e. 기타:					

Q3.

수용가능성 / 이름 형태	매우 적절함				매우 부적절함
a. 신경숙	5	4	3	2	1
b. 신	5	4	3	2	1
c. 그	5	4	3	2	1
d. 경숙	5	4	3	2	1
e. 기타:					

Q4. 위의 글에 언급된 이름의 사람이 한국에서 어느정도 유명한지를 표시해 주세요.

매우 유명함				전혀 유명하지 않음
5	4	3	2	1

SECTION D: 대명사

지시사항

다음 대화들은 드라마에서 발췌된 내용입니다.
각 대화(A-D)의 빈칸에 들어갈 **a, b 모두의 대명사(대용표현)에 대하여 각각의 알맞은 알맞은 정도** (수용가능성, acceptability)를 **5** (매우 적절함)에서 **1** (매우 부적절함)중에 하나로 표시해 주시면 됩니다.
사용가능한 다른 대명사 형태와 그 알맞은 정도도 **기타(other) (c)**란에 작성해 주시면 됩니다.

대화 A

미란: 대풍이랑 수진이랑 사귄단다.

복실: **Q1>**_____ 저도 몰랐어요.

([솔약국집 아들들], 제 43 회, 09/05/2009)

대화 B

옥희: 근데 왜 그래? (왜 자꾸 그러니 응? 니가 이러면 이럴수록 너나 나나 불편하고 또 남이 보면 내가 무슨 아주 나쁜 시어머닌 줄 알아요. 아니?) 너 맨날 출근길에 오빠네 들리는거 아는데.. 그냥 어머니 저 오늘 먼저 나가서 조카들 좀 챙기고 출근 할께요. 하면 될 것을 왜 자꾸 나보고 시계보고, 나보고 시계를 봐야!

수진: 죄송해요.

옥희: (뭐가 죄송해.) 언제는 **Q2>**_____ 모르고 결혼 허락 했니?

([솔약국집 아들들], 제 53 회, 10/10/2009)

Q1.

수용가능성 / 대명사	매우 적절함				매우 부적절함
a. 그건	5	4	3	2	1
b. 그런건	5	4	3	2	1
c. 기타:	5	4	3	2	1

Q2.

수용가능성 / 대명사	매우 적절함				매우 부적절함
a. 그거	5	4	3	2	1
b. 그런거	5	4	3	2	1
c. 기타:	5	4	3	2	1

대화 C

복실: 제가 제 인생에 있어 가장 힘들었을 때 엄마 보내고 가장 처참하고 외로웠을 때 선생님(하고 선생님 가족분들 때문에) 견딜수 있었어요.

대풍: 아, 알어. **Q3>**_____ 저번에도 말했잖어.

([솔약국집 아들들], 제 52 회, 10/04/2009)

대화 D

선풍: 저도 그림 좋아합니다, 장모님.

문숙: 어쩜 우리 송서방은 문학이면 문학 시면 시! 그림이면 그림! 모르는게 없어. 우리 그이를 봐아! 내가 언제 그림 전시회 한번 가자고 해도 가는가? 그 사람은 정말 일밖에 모르는 **Q4>**_____ 재미없는 사람이야!

([솔약국집 아들들], 제 50 회, 09/27/2009)

Q3.

수용가능성 / 대명사	매우 적절함				매우 부적절함
a. 그건	5	4	3	2	1
b. 그런건	5	4	3	2	1
c. 기타:	5	4	3	2	1

Q4.

수용가능성 / 대명사	매우 적절함				매우 부적절함
a. 그거	5	4	3	2	1
b. 그런	5	4	3	2	1
c. 기타:	5	4	3	2	1

의견사항 (SECTION B: 이름형태 유형 I)

● 위의 질문이나 설문(지)에 대해 귀하의 생각이나 의견을 말씀해 주시면 감사하겠습니다.

의견사항 (SECTION C: 이름형태 유형 II)

● 위의 질문이나 설문(지)에 대해 귀하의 생각이나 의견을 말씀해 주시면 감사하겠습니다.

의견사항 (SECTION D: 대명사)

● 위의 질문이나 설문(지)에 대해 귀하의 생각이나 의견을 말씀해 주시면 감사하겠습니다.

귀중한 시간 내 주셔서 감사합니다. 좋은 하루 되세요!

Appendix C

UT Arlington IRB Documentation

THE UNIVERSITY
OF TEXAS
AT ARLINGTON

Office of Research Administration
Box 19188
202 E. Border St., Suite 214
Arlington, Texas
76019-0188

T 817.272.3723
F 817.272.1111

http://www.uta.edu/research
Expertise at UT Arlington
http://www.uta.edu/expertise

December 19, 2011

Sok-Hun Kim
Dr. Laurel Stvan
Linguistics
Box 19259

Protocol Title: *A Corpus-based Analysis of Discourse Anaphora in English and Korean: A Neo-Gricean Pragmatic Approach (A Survey for Word Choice)*

RE: Exempt Approval Letter

IRB No.: **2012-0250e**

The UT Arlington Institutional Review Board (UTA IRB) Chair (or designee) has reviewed the above referenced study and found that it qualified as exempt from coverage under the federal guidelines for the protection of human subjects as referenced at Title 45 Part 46.101(b)(2). You are therefore authorized to begin the research as of November 23, 2011.

Please be advised that as the principal investigator, you are required to report local adverse (unanticipated) events to this office within 24 hours. In addition, pursuant to Title 45 CFR 46.103(b)(4)(iii), investigators are required to, "promptly report to the IRB ***any*** proposed changes in the research activity, and to ensure that such changes in approved research, during the period for which IRB approval has already been given, are **not initiated without IRB review and approval** except when necessary to eliminate apparent immediate hazards to the subject."

All investigators and key personnel identified in the protocol must have documented Human Subject Protection (HSP) Training or *CITI Training* on file with this office. The UT Arlington Office of Research Administration Regulatory Services appreciates your continuing commitment to the protection of human research subjects. Should you have questions or require further assistance, please contact Robin Dickey by calling (817) 272-9329.

Sincerely,

Patricia Turpin

Digitally signed by Patricia Turpin
DN: postalCode=76019, o=The University of Texas at Arlington, street=701 South Nedderman Drive, st=TX, l=Arlington, c=US, cn=Patricia Turpin, email=pturpin@uta.edu
Date: 2011.12.21 13:47:43 -06'00'

Patricia G. Turpin, PhD, RN, NEA-BC
Clinical Associate Professor
UT Arlington IRB Chair

THE UNIVERSITY
OF TEXAS
AT ARLINGTON

Office of Research
Administration
Box 19188
202 E. Border St., Suite 214
Arlington, Texas
76019-0188

T 817.272.3723
F 817.272.1111

http://www.uta.edu/research
Expertise at UT Arlington
http://www.uta.edu/expertise

BeA Maverick

March 30, 2012

Sok-Hun Kim
Dr. Laurel Stvan
Linguistics
The University of Texas at Arlington
Box 19559

IRB No.: **2012-0250**

RE: Minor Modification Approval Letter

Title: *A Corpus-based Analysis of Discourse Anaphora in English and Korean: A Neo-Gricean Pragmatic Approach (A Survey for Word Choice)*

The UT Arlington Institutional Review Board (UTA IRB) Chair (or designee) reviewed and approved the modification(s) to this protocol on **March 26, 2012** in accordance with Title 45 CFR 46. 110(b)(2). Therefore, you are authorized to conduct your research. The modification(s), indicated below, was/were deemed minor and appropriate for expedited review.

- Change the length of the survey
- Include participant compensation
- Add a consent document to reflect these changes

Pursuant to Title 45 CFR 46.103(b)(4)(iii), investigators are required to, "promptly report to the IRB ***any*** proposed changes in the research activity, and ensure that such changes in approved research, during the period for which IRB approval has already been given, **are not initiated without IRB review and approval** except when necessary to eliminate apparent immediate hazards to the subject."

The modification approval will additionally be presented to the convened board on April 10, 2012 for full IRB acknowledgment [45 CFR 46.110(c)]. All investigators and key personnel identified in the protocol must have documented Human Subjects Protection (HSP) training, *CITI* Training, or other approved training on file with the UT Arlington Office of Research Administration; Regulatory Services.

The UT Arlington Office of Research Administration appreciates your continuing commitment to the protection of human research subjects. Should you have questions or require further assistance, please contact Robin Dickey at robind@uta.edu or you may contact the Office of Regulatory Services at 817-272-3723.

Sincerely,

Patricia Turpin

Digitally signed by Patricia Turpin
DN: postalCode=76019, o=The University of Texas at Arlington, street=701 South Nedderman Drive, st=TX, l=Arlington, c=US, cn=Patricia Turpin, email=pturpin@uta.edu
Date: 2012.04.03 19:15:32 -05'00'

Patricia Turpin, Ph.D., RN, NEA, BC
Clinical Associate Professor
UT Arlington IRB Chair

APPROVED

MAR 2 6 2012

Institutional Review Board

INFORMED CONSENT

PRINCIPAL INVESTIGATOR NAME:
Kim, Sok-Hun

TITLE OF PROJECT:
A Corpus-based Analysis of Discourse Anaphora in English and Korean: A Neo-Gricean Pragmatic Approach (A Survey of Word Choice: Naming Forms and Pronouns)

INTRODUCTION:
You are being asked to participate in a research survey about word choice; i.e., naming forms and pronouns. Your participation is anonymous and voluntary. Refusal to participate or withdrawal from the survey at any time will involve no penalty or loss of benefits to which you are otherwise entitled. Please ask questions if there is anything you do not understand.

PURPOSE:
The purpose of this study is to investigate possible associations between one's degree of preference for naming forms and degree of famousness of particular public figures in texts, and those between one's degree of preference for two pronouns in conversations and different contexts.

DURATION:
The survey session will take approximately 30 minutes to complete.

NUMBER OF PARTICIPANTS:
The survey will be conducted by a maximum of 25 native speakers of English and a maximum of 25 native speakers of Korean, depending on availability and willingness of potential participants.

PROCEDURES:
1. Recruiting and Two screening questions
You will be recruited directly by me by canvassing public areas of the UT Arlington campus to participate in this research survey. After being notified that the survey is anonymous and voluntary, and being asked about your willingness to participate, you will be asked to answer two screening questions regarding native language (i.e., English & Korean) and age (i.e., over 18). If you are willing to participate in this research survey and meeting these screening criteria, you will be presented with the survey's questions of your native language.
2. Survey questionnaire for each native speaker of each language

You will complete a word choice questionnaire. The form includes questions related to your natural uses of naming forms and pronouns in texts. During the survey, you will see four sections of questions. In section A, you will be asked to complete questions about yourself (native language, sex, education, and age). In section B, you will be asked to complete questions about your preference (from 5-1) for naming forms (such as Steve Jobs, Jobs, He, Steve, and other possible forms) in texts by the first three mentions. In section B, you will be asked to complete the same type of questions about your preference (from 5-1) for naming forms in texts, but by the first subject position in each paragraph. In section D, you will be asked to complete questions about your preference (from 5-1) between two pronouns (*that* and *it* in English; *kuken* 'that' and *kulen* 'it' in Korean) in conversations and different contexts.

POSSIBLE BENEFITS:
There are no direct benefits to you for participating in this survey other than the possibility for considering an association between naming forms and famousness of a public figure, an association between pronouns and different contexts in your native language.

POSSIBLE RISKS/DISCOMFORTS:
There are no foreseeable risks or discomforts involved in your participation in this survey since this survey is anonymous and voluntary. In addition, as no unique personal identifiers (i.e. SSN, picture ID, and voice) will be solicited, and as the data to be analysis will be aggregated, there are no anticipated risks to your legal, physical, emotional, or social well-being.

COMPENSATION:
You will receive $ 5 cash if you participate in this research survey and complete this survey questionnaire. The cash will be given at the end of this survey.

VOLUNTARY PARTICIPATION:
Participation in this research survey is anonymous and voluntary. You have the right do decline participation in any or all survey procedure or quit at any time at no consequences. Refusal to participate or withdrawl from the research survey will involve no penalty or loss of benefits to which you are otherwise entitled.

CONTACT FOR QUESTIONS:
Questions about this survey or your rights as a research subject may be directed to Sok-Hun Kim at (682)-234-2750 or at e-mail: sokhun.kim@mavs.uta.edu. You may contact the chairperson of the UT Arlington Institutional Review Board at (817)-272-3723 in the event of a research-related injury to the subject.

References

Ariel, Mira. 1988. Referring and accessibility. Journal of Linguistics 24. 67-87.

Ariel, Mira. 1990. Accessing non-phrase antecedents. London and New York: Routledge.

Ariel, Mira. 1994. Interpreting anaphoric expressions: a cognitive versus a pragmatic approach. Journal of Linguistics 30(1). 3-42.

Ariel, Mira. 1996. Referring expressions and the +/- coreference distinction. Reference and referent accessibility, ed. by T. Fretheim and J. K. Gundel, 12-35. Amsterdam: John Benjamins.

Ariel, Mira. 2008. Pragmatics and grammar. Cambridge: Cambridge University Press.

Ariel, Mira. 2010. Defining pragmatics. Cambridge: Cambridge University Press.

Bowker, Lynne, and Pearson, Jennifer. 2002. Working with specialized language: a practical guide to using corpora. London and New York: Routledge.

Brown, Gillian and Yule, George. 1983. Discourse analysis. Cambridge: Cambridge University Press.

Bühler, Karl. 1982. The deictic field of language and deictic words. Extracts in English from Bühler (1934) Sprachtheorie. Speech, place and action: studies in deixis and related topics, eds. by Jarvella, Robert. J. and Klein, Wolfgang, 9-30. Chichester: John Wiley.

Cheshire, Jenny. 1996. That jacksprat: an interactional perspective on English that. Journal of Pragmatics 25(3). 369-393.

Chomsky, Noam. 1981. Lectures on government and binding. Dordrecht: Foris.

Chomsky, Noam. 1995. The minimalist program. Cambridge, MA: The MIT Press.

Cornish, Francis. 1999. Anaphora, discourse, and understanding: evidence from English and French. Oxford: Clarendon Press.

Cornish, Francis. 2008. How indexicals function in texts: discourse, text, and one neo-Gricean account of indexical reference. Journal of Pragmatics 40. 997-1018.

Clark, Herbert H., and Marshall, Catherine C. 1981. Definite reference and mutual knowledge. Elements of discourse understanding, eds. by Joshi, Aravind K., Webber, Bonnie Lynn, and Sag, Ivan, 10-63. Cambridge: Cambridge University Press.

Creswell, John W. 2009. Research design: qualitative, quantitative, and mixed methods approaches (3rd edition). Los Angeles, CA: Sage.

Ehlich, Konrad. 1982. Anaphora and deixis: same, similar, or different? Speech, place and action: studies in deixis and related topics, ed. by Jarvella, Robert. J. and Klein, Wolfgang, 313-388. Chichester: John Wiley.

Fox, Barbara A. 1987. Discourse structure and anaphora: written and conversational English. Cambridge: Cambridge University Press.

Givón, Talmy. 1983. Topic continuity in discourse: an introduction. Topic continuity in discourse: a quantitative cross-language study, ed. by Tamly Givón, 1-42. Amsterdam: John Benjamins.

Green, Georgia M. 1996. Pragmatics and natural language understanding. Mahwah, NJ: Erlbaum.

Grice, H. P. 1975. Logic and conversation. Syntax and semantics 3: speech acts, ed. by P. Cole and J. Morgan, 41-58. New York: Academic Press.

Grice, H. P. 1978. Further notes on logic and conversation. Syntax and semantics 9: pragmatics, ed. by P. Cole, 113-128. New York: Academic Press.

Grimes, Joseph. 1975. The thread of discourse. The Hague: Mouton.

Grosz, Barbara, Pollack, Martha, and Sinder, Candace. 1989. Discourse. Foundations of Cognitive Science, ed. by Posner, Michael, 437-468. Cambridge, MA: MIT Press.

Gundel, Jeanette K. 1985. 'Shared knowledge' and topicality. Journal of Pragmatics 9.83-107.

Gundel, Jeanette K. 1988. Universals of topic-comment structure. Studies in syntactic typology, eds. by M. Hammond, E. Moravcsik, and J. Wirth, 209-239. Amsterdam: John Benjamins.

Gundel, Jeneatte K. 1996. Relevance theory meets givenness hierarchy: an account of inferrables. Reference and referent accessibility, ed. by T. Fretheim and Jeaneatte K Gundel, 141-153. Amsterdam: John Benjamins.

Gundel, Jeneatte K. 2003. Information structure and referential givenness/newness: how much belongs in the grammar? Journal of Cognitive Science 4: 177-199.

Gundel, Jeneatte K., Hedberg, Nancy and Zacharski, Ron. 1988. On the generation and interpretation of demonstrative expressions. Proceedings of the International Conference on Computational Linguistics 12. 216-221.

Gundel, Jeneatte K., Hedberg, Nancy, and Zacharski, Ron. 1989. Givenness, implicature and demonstrative expressions in English discourse. Papers from the 25th Regional Meeting. Chicago Linguistic Society (CLS), 89-103.

Gundel, Jeneatte K., Hedberg, Nancy, and Zacharski, Ron. 1990. Givenness, implicature, and the form of referring expressions in discourse. Papers from the 16th Regional Meeting. Berkeley Linguistic Soceity (BLS), 442-253.

Gundel, Jeneatte K., Hedberg, Nancy, and Zacharski, Ron. 1993. Cognitive status and the form of referring expressions in discourse. Language 69 (2). 274-307.

Gundel, Jeneatte K., Hedberg, Nancy, and Zacharski, Ron. 2001. Definite descriptions and cognitive status in English: why accommodation is unnecessary. Journal of English Language and Linguistics 5 (2): 273-295.

Halliday, M. A. K., and Hasan, Ruqayia. 1976. Cohesion in English. London: Longman.

Hinds, John. 1978. Anaphora in discourse. Edmonton: Linguistic Research Inc.

Hinds, John. 1979. Organizational patterns in discourse. Syntax and semantics 12: discourse and syntax, ed. by Tamly Givón, 135-157. London: Academic Press.

Horn, Laurence R. 1984. Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. Meaning, form, and use in context: linguistic applications, ed. by Schiffrin, Deborah, 11-42. Washington, DC: Georgetown University Press.

Horn, Laurence R. 1989. A natural history of negation. Chicago: University of Chicago Press.

Horn, Laurence R. 2004. Implicature. Handbook of pragmatics, ed. by L. Horn and G. Ward, 3-28. Oxford: Blackwell.

Huang, Yan. 1991. A neo-Gricean pragmatic theory of anaphora. Journal of Linguistics 27. 301-335.

Huang, Yan. 1994. The syntax and pragmatics of anaphor. Cambridge: Cambridge University Press.

Huang, Yan. 2000a. Discourse anaphora: four theoretical models. Journal of Pragmatics 32. 151-176.

Huang, Yan. 2000b. Anaphora: a cross-linguistic study. Oxford: Oxford University Press.

Huang, Yan. 2007. Pragmatics. Oxford: Oxford University Press.

Kamio, Akio and Margaret, Thomas. 1999. Some referential properties of English it and that. Function and structure: in honor of Susumu Kuno, ed. by A. Kamio and K. Takami, 289-315. Amsterdam: John Benjamins.

Kim, Sok-Hun. 2010a. Pragmatic aspects of full name pattern in Korean and English. Harvard Studies in Korean Linguistics XIII, ed. by Kuno, Susumu et al., 349-363. Harvard University, Cambridge, MA, USA: Harvard-Yenching Institute.

Kim, Sok-Hun. 2010b. A corpus-based pragmatic study on sentential anaphors: English *that*/*it* and Korean *ku* 'it'. Proceedings of SICOL-2010: The 2010 Seoul International Conference on Linguistics. The Linguistic Society of Korea, Seoul, Korea.

Kim, Sok-Hun. 2012. Referring expressions in Korean sports News: a reply to Lee (2010). Harvard Studies in Korean Linguistics XIV, ed. by Kuno, Susumu et al., 139-152. Harvard University, Cambridge, MA, USA: Harvard-Yenching Institute.

Kittay, Jeffrey. 1988. On anaphora and antecedence. Semiotica 72(3-4). 205-234.

Lambrecht, Knud. 1994. Information structure and sentence form: topic, focus, and the mental representations of discourse referents. Cambridge: Cambridge University Press.

Lee, Hye-Kyung. 2010. Referring expressions in English and Korean political news. Journal of Pragmatics 42. 2506-2518.

Lee, Sungbom. 2005. Understanding English referring expressions: a cognitive and pragmatic approach (in Korean). Seoul: Boobooks.

Levinson, Stephen C. 1993. Pragmatics. Cambridge: Cambridge University Press.

Levinson, Stephen C. 1987a. Minimization and conversational inference. The pragmatic perspective, ed. by Verschueren, J and Bertuccelli-Papi, M, 61-129. Amsterdam: John Benjamins.

Levinson, Stephen C. 1987b. Pragmatics and the grammar of anaphora: a partial pragmatic reduction of binding and control phenomena. Journal of Linguistics 23.379-434.

Levinson, Stephen C. 1991. Pragmatic reduction of the binding conditions revisited. Journal of Linguistics 27. 107-161.

Levinson, Stephen C. 1995. Three levels of meaning. Grammar and meaning: essays in honor of Sir John Lyons, ed. by F. Palmer, 90-115. Cambridge: Cambridge University Press.

Levinson, Stephen C. 2000. Presumptive meanings: the theory of generalized conversational implicature. Cambridge, MA: MIT Press.

Linde, Charlotte. 1979. Focus of attention and the choice of pronouns in discourse. Syntax and Semantics 12: Discourse and Syntax, ed. by Givón, 337-353. New York: Academic Press.

Lyons, John. 1975. Deixis as the source of reference. Formal semantics of natural languages, ed. by Keenan, Edward L., 61-83. Cambridge: Cambridge University Press.

Lyons, John. 1977. Semantics. Vol, 2. Cambridge: Cambridge University Press.

Lyons, John. 1979. Deixis and anaphora. The development conversation and discourse, ed. by T. Myers, 88-103. Edinburgh: Edinburgh University Press.

Longacre, Robert E. 1979. The paragraph as a grammatical unit. Syntax and semantics 12: discourse and syntax, ed. by Tamly Givón, 115-134. London: Academic Press.

Mann, William C. and Sandra A. Thompson. 1987. Rhetorical structure theory: a theory of text organization. Reprinted from The structure of discourse, ed. by L. Polanyi.1-81. ISI: Information Sciences Institute, Los Angeles, CA, ISI/RS-87-190. (This book did not actually appear, and the article later appeared in Text 8 (3)).

Mann, William C. and Sandra A. Thompson. 1988. Rhetorical structure theory: toward a functional theory of text organization. Text 8.243-281.

Mulkern, Ann E. 1996. The game of the name. Reference and referent accessibility, ed. by T. Fretheim and J. K. Gundel, 235-250. Amsterdam: John Benjamins Publishing.

Prince, Ellen F. 1981. Toward a taxonomy of given-new information. Radical pragmatics, ed. by Cole Peter, 223-255. New York: Academic Press.

Reichman, Rachel. 1981. Plain speaking: a theory and grammar of spontaneous discourse. Cambridge, MA: Bolt, Beranek and Newman.

Sanford, A. J., K. Moar, and S. C. Garrod. 1988. Proper names as controllers of discourse focus. Language and Speech 31. 43-56.

Schiffrin, Deborah. 1994. Approaches to discourse. Oxford: Blackwell.

Searle, John. 1958. Proper names. Mind, New Series 67. No. 266. 166-173.

Tai, James H. Y. 1978. Anaphoric constraints in Mandarin Chinese narrative discourse. Anaphora in discourse, ed. by John Hinds, 279-338. Edmonton: Linguistic Research Inc.

Verschueren, Jef. 1999. Understanding pragmatics. London: Arnold Publisher.

Walker, Marilyn A. 1998. Centering, anaphora resolution and discourse structure. Centering theory in discourse, ed. by Walker, M. A, Josh, A., and Prince, E, 401-435. Oxford: Calrendon Press.

Zipf, George K. 1949. Human behavior and the principle of least effort: an introduction human ecology. New York: Addison-Wesley.

Biographical Information

Sok-Hun Kim graduated Summa Cum Laude with a BA in English Language and Literature in 2002 (1994-2002) from Hanyang University and an MA in English Linguistics in 2004 (2002-2004) from Sogang University, Korea. After studying English Linguistics in the PhD program at Sogang University for two years (2005-2007), he started his graduate studies in the Department of Linguistics and TESOL at The University of Texas at Arlington in Fall 2007. He earned his PhD in Linguistics from UT Arlington in May 2013. His research interests include semantics/pragmatics, discourse analysis, and corpus linguistics, focusing mainly on neo-Gricean theories, reference, anaphora, implicature, pragmatic inference, language and thought, and experimental pragmatics.