

Binding Conditions of English Reflexives and Pronouns in the ICE-USA

Ji-Hye Kim^{1†}, Soojin An² & Ahreum Jung¹

¹Korea National University of Education, ²Myeongji Middle School (teacher)

ABSTRACT

This study investigates how the syntactic and pragmatic conditions on the distributions of English reflexives and pronouns are reflected in the International Corpus of English (ICE)-USA corpus. The study attempted to compare different theoretical approaches to binding phenomena – Standard Binding Theory (SBT), HPSG-based Binding Theory and Logophoricity theory – in their predictions. 1,500 English sentences were extracted from ICE-USA corpus to be encoded with 8 syntactic-pragmatic factors related to the binding conditions of reflexives/pronouns. The results demonstrated the following: i) English reflexives and pronouns were mostly in complementary distribution; ii) English reflexives showed apparent cases of LD-exempt/discourse binding; iii) Structural constraints such as c-command and binding domain were rarely violated with reflexives having sentential antecedents; iv) Outside the syntactic domain, proper logophoric conditions for licensing the LD-reflexives were observed. The results call for incorporation of discourse-pragmatic analysis with syntactic approaches to explain various binding phenomena.

Keywords: binding theory, reflexive, anaphor, pronoun, pronominal, ICE-USA corpus, binding in American English

1. Introduction: Binding Theory and Conditions on Binding

Since Chomsky (1980, 1981, 1986) proposed Standard Binding Theory (SBT, henceforth), the distributions of different nominals – anaphors (i.e., reflexives and reciprocals) and pronouns (i.e., pronominals) – are known to be in complementary distributions. The conditions on binding have also been discussed in theoretical approaches other than Generative Grammar – such as Head-driven Phrase Structure Grammar (HPSG, Pollard and Sag, 1992) and Argument Structure-based Binding Theory (Jackendoff, 1972) and discourse-pragmatic approaches to binding (Sells,

* Acknowledgment: We appreciate Dr. Yong-hun Lee for his professional help and advice for statistical analysis and interpretations and also wish to express our deepest gratitude to the three anonymous reviewers who have kindly provided a great deal of advice with lots of valuable comments.

† Corresponding author: psychlg@gmail.com, jkim@knue.ac.kr

1987; Kuno, 1987), etc. Büring (2005:11) proposed the binding conditions that can be applied to the cross-theoretical definitions, as shown in (1).

(1) Binding Conditions (Büring, 2005, 11p)¹⁾

An NP of class must (not) be coindexed with an x-commanding NP within its domain.

In SBT, the structural condition of c-command between the antecedent and the anaphor/pronoun is crucial for defining binding: An anaphor must be bound by its c-commanding antecedent, while a pronoun must not, as in (2).

- (2) a. Mary's brother_i tortured himself_i / *him_i.
 a. Mary_i's brother tortured *herself_i / her_i.

As for the domain for binding, Chomsky (1981) defined the local domain as 'Governing Category (GC)', which is defined as the conjunction of the two Opacity Conditions (Chomsky, 1980) – Specified Subject Conditions (SSC, i.e., the binding across a specified subject is banned) and Tensed S Conditions (TSC, i.e., the binding out of a finite clause is not allowed). While violation of SSC robustly results in ungrammaticality for anaphors, as in (3a-c), whether TSC serves as a robust binding domain is controversial, as shown in (3d-e).²⁾

- (3) a. John_i loves himself_i / *him_i.
 b. John_i says [that Bill hates *himself_i/him_i].
 c. John_i loves [Mary_j's article about *himself_i/him_i].
 d. John_i thinks that [*himself_i is to blame].
 e. John_i said that [pictures of himself_i/him_i are on display].

With respect to binding domain (GC), Büring (2005:11) explained the binding conditions for anaphors and pronouns as in (4).

-
- 1) The underlined parts can be regarded as variables: for 'class', three types of nominals (i.e., anaphors, pronouns, R-expressions) are applied; for 'x-commanding', c-command in SBT, o-command in HPSG, and Θ -command in Argument Structure-based Binding Theory can be applied. The domain is considered usually within a clause (or a sentence) cross-theoretically, and has been the most extensively discussed in detail within the framework of SBT.
- 2) TSC is not counted as binding domain in HPSG, since Pollard & Sag (1992) claims that the sentences violating TSC is ungrammatical with a reason other than binding – that is, English does not have nominative anaphors, which makes the sentences with TSC-violation bad.

- (4) a. Condition A: {class=anaphor; must be coindexed; c-commanding NP;
domain = Governing Category}

An anaphor must be bound within its Governing Category.

- b. Condition B: {class=pronoun; must not be coindexed; c-commanding NP;
domain = Governing Category}

A pronoun must be free within its Governing Category.

Though the SBT (Chomsky, 1980, 1981, 1986) provided an explanatorily adequate framework for the distributions of anaphors and pronouns, the theory has been weakened by the evidence of cross-linguistic differences in binding domain (Yang, 1983; Manzini & Wexler, 1987; Huang & Liu, 2001; Cole, Hermon & Huang, 2001; Pollard & Xue, 2001) and form-function correlations in local vs. long-distance (LD) binding found in languages with multiple anaphors (Cole, Hermon & Sung, 1990; Reinhart & Reuland, 1993; Reuland, 2011, 2017). Most crucially, an apparent existence of English anaphors that are bound LD outside the GC was found (Zribi-Hertz, 1989; Pollard & Sag, 1992; Reinhart & Reuland, 1993; Runner & Kaiser, 2005; Pollard, 2005), which provided counter-examples to the original assumption of SBT – the complementary distributions between anaphors and pronouns.

Other scholars (Sells, 1987; Kuno, 1987; Zribi-Hertz, 1989; Pollard & Sag, 1992; Huang & Liu 2001, etc.) have reached an alternative approach for treating such exceptional cases of binding, by focusing on pragmatic properties of LD-binding (Sells, 1987; Kuno, 1987; Zribi-Hertz, 1989) or by proposing that there are different types of anaphors – core/grammatical anaphors (occurring within GC under syntactic constraints) and exempt anaphors/logophors (occurring outside the GC) that are licensed by discourse-pragmatic conditions (Pollard & Sag, 1992; Huang & Liu, 2001, etc.). Especially, Pollard & Sag (1992) came up with syntactic theory of binding within HPSG framework to explain binding phenomena mentioned in SBT, and further extended the theory to the distributions of exceptional types of anaphors. According to Pollard & Sag (1992), the anaphors with a superior co-argument antecedent must be syntactically bound within binding domain, whereas there are some anaphors that are exempt from syntactic binding constraints – ‘exempt anaphors’. As noted by Pollard & Sag (1992), exempt anaphors can be bound outside the local GC (i.e., violating SSC and TSC) like pronouns, as shown in (5a-b); they

can be unbound or discourse-bound (cf. 5c) and do not need c-commanding antecedent within its GC (i.e., the picture NP) (cf. 5d).

- (5) a. Bill_i remembered [that the Times had printed [a picture of himself_i/him_i] in its Sunday edition].
- b. John_i thinks [that [an article written by himself_i/him_i] caused the uproar].
- c. Physicist like yourself_i/you_i are a godsend.
- d. [Incriminating pictures of himself_i] worry Bill_i.

Pollard & Sag (1992) claimed that in English, anaphors must be bound syntactically when a syntactically accessible (co-argument) antecedent is present. However, when there is no potential antecedent within the binding domain, the anaphor becomes exempt and can be bound outside the local domain.³⁾ Instead of being free from syntactic constraints, the exempt anaphors are licensed by discourse-pragmatic conditions (Pollard & Sag, 1992; Huang & Liu, 2001).

As for the discourse-pragmatic conditions of these discourse-bound/LD-anaphors, many scholars have proposed different theories (e.g., Kuno (1987) for empathy, Sells (1987) for logophoricity, Huang & Liu (2001) for *de se* restriction, etc.) to explain pragmatic properties of LD-anaphors. Especially, Sells (1987) attempted to cover all kinds of discourse-related bindings and proposed a set of logophoric components – instantiated by three roles (logophoric centers) – that give the antecedents prominence in the discourse, as shown in (6)⁴⁾.

- (6) a. SOURCE: the agent communicating the propositional content
- b. SELF: one whose mental state or attitude the content of the proposition describes
- c. PIVOT: one with respect to whose (space-time) location the content of the proposition is evaluated

Sells (1987) further claimed that there is a canonical hierarchy for the above three

3) This assumption by Pollard & Sag (1992) can be falsified, since English sometimes allow LD-binding even in the presence of locally accessible syntactic antecedent (i.e., SSC-violation, see Results section for some examples.)

4) As an anonymous reviewer mentioned, though the long-distance binding was mainly discussed in Sells (1987), as long as an argument can be identified as one of the logophoric centers, there is no way to block for it to be co-valued with a logophor in a local context. For example, PIVOT in Sells (1987) overlaps the point of views/empathy in Kuno (1987), which can be applied to the local binding contexts as well. (We were aware of this point though we coded the logophoric roles with the case of LD-binding limitedly.) It should be also noted, on the other hand, Huang & Liu (2001) focused mainly on SELF (i.e., *de se* interpretation), which is applied to LD-binding only.

roles: SOURCE > SELF > PIVOT, where SOURCE is usually the informant that can be center of the discourse (i.e., PIVOT) and SELF represents the view of the experiencer, which can also be PIVOT as well. Such hierarchy with different logophoric roles can be seen in (7), as mentioned in Kim & Yoon (2009). In (7), the structural distance and relation between the antecedent and the anaphor are identical (i.e., no c-command/backward binding), but a clear degree of contrast is shown in terms of acceptability of the sentences.

- (7) a. [Incriminating pictures of himself_i published in the Times] have been mentioned by John_i in his talk. (SOURCE)
- b. [Incriminating pictures of himself_i published in the Times] have been worrying John_i for some time. (SELF)
- c.? [Incriminating pictures of himself_i published in the Times] have all but eliminated John_i's chances of being promoted. (PIVOT)
- d.*? [Incriminating pictures of himself_i published in the Times] accidentally fell on John_i's head. (Kim & Yoon, 2009)

The judgments in (7) reflect that 'John' can be identified as a logophoric center – by being a SOURCE or SELF (or PIVOT) – in (7a-c), whereas in (7d), it can nearly represent any of them (if it could, by any chance, a very weak version of PIVOT).

The studies with exempt anaphors (or LD-bound anaphors) have been done through some experimental studies (Runner et al., 2003, 2006; Runner & Kaiser, 2005, etc.) as well as some studies using corpus data (Zribi-Hertz, 1989; Lange, 2006; Song, 2017; Kim et al., 2018; Kim, et al., 2020). Though corpus data can serve the best to represent real use of English anaphors, most of previous studies based on corpus data merely provided some sentences extracted from corpus to explain certain aspects of binding phenomena (Zribi-Hertz, 1989; Lange, 2006), which calls for more systematic corpus-based analysis focusing on binding phenomena – and possibly within some theoretical frameworks of binding.

So far, only a handful of studies have taken corpus-based approach based on specific theoretical framework (Kim et al., 2018; Kim et al., 2020). Kim et al. (2018) attempted to investigate the distribution of English pronouns and reflexives based on Standard Binding Theory (SBT, Chomsky, 1980, 1981, 1986), by focusing on 4 syntactic conditions – c-command, 2 conditions for binding domain (TSC and SSC violations) and the distance between the antecedents and the anaphors/pronouns. They extracted 1,000 sentences including reflexives and pronouns in International

Corpus of English – Great Britain (ICE-GB) (i.e., corpus for British English) and manually coded the extracted sentences in terms of the 4 binding related factors, so as to conduct a frequency-based statistical analysis of the distributions by comparing reflexives and pronouns. The results demonstrated that the reflexives and the pronouns in their study showed complementary distributions overall; nonetheless, they also found apparent cases of exempt anaphors bound outside the binding domain. Kim et al. (2020) attempted a multi-factorial analysis including more linguistic factors (26 factors – including binding-related factors and other linguistic factors that were not necessarily related to binding) using the same corpus (i.e., ICE-GB for British English).

Though the studies (Kim et al., 2018; Kim et al., 2020) are meaningful, in that they started corpus-based approach to test assumptions and claims of specific theoretical frameworks in terms of binding phenomena, their study had some limitations: First of all, when Kim et al. (2018) extracted pronouns, they did not discriminate solid pronouns (i.e., pronominals in Chomsky's term in SBT) and possessive pronouns. Since the syntactic-semantic behaviors of possessive pronouns are different from those of pronominals (i.e., they can be bound-variable pronouns or co-referential pronouns), they should have excluded possessive pronouns to examine the distributions of usual pronouns that are under the control of Binding Condition B. Including the possessive pronouns together with usual pronouns in their analysis must have affected overall analyses of the result patterns of pronouns.⁵⁾

Furthermore, the analyses of the previous studies (Kim et al., 2018; Kim et al., 2020) were limited in examining only syntactic factors though the study reported the apparent cases of exempt binding that are explained pragmatically. Given that the corpus-study of exempt anaphors has not been done systematically yet, it would be proper to conduct a study that also deals with discourse-pragmatic factors comprising exempt binding of anaphors.

Finally, Kim et al. (2018) and Kim et al. (2020) were based on British English using ICE-GB. However, Zribi-Hertz (1989)'s examples of corpus data from British literature showed binding domain violations that are not allowed in SBT, which implies possible difference(s) between British and American varieties of English – in terms of the frequency of long-distance binding. In other words, it is possible

5) In the later study (Kim et al., 2020), possessive pronouns seemed to be excluded, but the study concerned mainly on which factors were significant in determining the well-formedness of binding.

that the former accepts long-distance (exempt) binding or discourse binding more frequently than the latter. Therefore, it is necessary to conduct the similar line of corpus study that was previously done with British English, this time with American English to examine the possible differences and further comparisons.

In these reasons, the current study was based on the similar paradigm of Kim et al. (2018) and Kim et al. (2020), by implementing manual coding of syntactic binding factors in the sentences extracted from corpus data and conducting frequency-based statistical analysis to compare the distributions of reflexives and pronouns – this time with American English. Furthermore, the current study investigated pragmatic as well as syntactic conditions of the binding by adding discourse-pragmatic factors (i.e., logophoricity) that have been mentioned in previous theoretical studies (Sells, 1987, etc.).

In the present study, ICE-USA, the American component of International Corpus of English (ICE), was used. ICE was designed for cross-sectional analyses of world Englishes: There are 20 sub-corpora of ICE in total (e.g., Australia, Cameroon, Canada, Caribbean (Jamaica), East Africa (Kenya, Malawi, Tanzania), Fiji, Ghana, Great Britain, Hong Kong, India, Ireland, New Zealand, Nigeria, Philippines, Sierra Leone, Singapore, South Africa, and United States of America), which makes possible of the comparisons of regional variations of English across different English-speaking countries (as mother tongue or as official language). Each 20 sub-corpora of ICE were composed of 1,000,000 words and the ICE-USA have 200 written texts⁶⁾ composed of approximately 2,000 words. We used this corpus to make further comparisons with ICE-GB possible in the future study.

The organization of the paper is as follows: Next section will introduce research questions and describe the method of the study. The following section will present the results of the distributions of the reflexives/pronouns through statistical analyses and discuss them in terms of Binding Theory. The final section will provide a conclusion and further discussions.

6) The 200 written texts represent various sources of writings (academic writing (40), non-academic writing (40), reportage (20), instructional writing (20), persuasive writing (10), creative writing (20)) that also include non-professional writing (20) and correspondence (30).

2. Method

2.1. Research Questions & Procedures

The goal of this study is to investigate the distributions of English reflexives and pronouns using ICE-USA and examine whether and how the binding conditions proposed in earlier theories, such as SBT (Chomsky, 1980, 1981, 1986), HPSG (Pollard & Sag, 1992) and logophoricity (Sells, 1987), are reflected well in the corpus data. The main research questions are the following.

- 1) Do English reflexives and pronouns represented in ICE-USA show complementary distributions as proposed in Standard Binding Theory (Chomsky, 1980, 1981, 1986)?
- 2) Are syntactic conditions (i.e., c-commanding and conditions for GC) in SBT (Chomsky, 1980, 1981, 1986) well-represented in the distributions of reflexives and pronouns in ICE-USA, when they are bound within the GC?
- 3) When the reflexives are LD-bound/discourse-bound outside the syntactic GC, do they represent exempt conditions and discourse-pragmatic conditions, as proposed in Pollard & Sag (1992) and Sells (1987)?

2.2. Procedures

The procedures of extracting sentences and manual coding of the factors were as follows. First, a total of 1,500 English sentences (600 containing reflexives and 900 pronouns) were extracted from ICE-USA. Then, the sentences were manually coded in terms of 8 different factors related to binding – 5 factors representing syntactic binding and 3 factors representing pragmatic (logophoric) binding. The selection of the syntactic factors was based on Kim et al. (2018), which are, i) the distance (i.e., number of embedding(s)) between the antecedent and the NP (reflexive or pronoun)⁷⁾, ii) TSC (i.e., if the NP is bound outside the finite clause), iii) SSC (i.e., if the NP is bound across an intervening Subject), iv) if the antecedent c-commands the NP. In addition, one more condition was added as syntactic factor

7) As one of the reviewers kindly pointed out, the factor ‘Number of embedding’ is closely related to the factor TSC, and indeed, the presence/absence of clausal boundary between the antecedent and the NP is more important than how far those two NPs are located from each other. Therefore, the result will be focused on TSC rather than the number of embedding to eliminate unnecessary redundancy.

– the Grammatical Relation (GR) of the antecedent – which was designed to investigate if any of the GRs of the antecedent is predominant in the binding of English.⁸⁾

On the other hand, the selection of the discourse-pragmatic factors was mainly based on Pollard & Sag (1992) and Sells (1987), which are, i) whether the NP is exempt anaphor (i.e., if the position for syntactically accessible antecedent is empty), ii) if the antecedent is logophoric, iii) in case the antecedent is logophoric, what is the logophoric role represented in the given sentence. The encoded factors are represented in the Table 1 below.

The 8 factors were submitted to manual coding procedures. Since the target sentences for analysis should be construed as representing binding relations, several

Table 1. The encoded factors

Type	ID Tag	Comment	ID-tag level ⁹⁾
Syntactic	Embedding	Number of clauses between the antecedent and the NP	0, 1, 2, 3, n/a
	Command	Structural relation between the antecedent and the NP	c-command, sub-command, backward binding, No-command, n/a
	TSC	TSC-violation (GC)	Yes, no
	SSC	SSC-violation (GC)	Yes, no
	Antecedent orientation	The GR of the antecedent	Subject, Object, DO, IO, Possessor of subject, Possessor of object, n/a
Discourse-pragmatic	Exempt	The absence of potential syntactic antecedent	Yes, no, n/a
	Logophoric	If the antecedent meets logophoricity	Yes, no
	Logophoric roles	Logophoric roles of the antecedent	SOURCE, SELF, PIVOT, n/a

8) Though the antecedent orientation is important in some other theoretical frameworks, it is not the focus of the current study; therefore, we will not include the result of this factor in our discussion.
9) As for the ID-tag levels, some sentences were difficult to be coded in one specific level due to complications (e.g., binding phenomena mixed with control theory, the antecedents and the target NPs presented across an adjunct clause, discourse binding, etc.) or debatable across scholars – which were left not coded and marked as ‘n/a’. Most of the ‘n/a’s in within-sentence factors were from the cases of discourse binding where the calculation of the antecedent could not be made within the given sentences. However, since discourse binding and sentential binding were generally analyzed separately, n/a’s did not make significant influence on the analysis of the overall results.

sentence types were excluded during the coding procedures. For instance, the cases of emphatic reflexives and generic reflexives (i.e., oneself) were excluded because they do not fit to represent Binding Condition A. Likewise, the sentences with possessive pronouns were also excluded because possessive pronouns are unlike usual pronouns, thus do not fit for the Binding Condition B. As a result of screening, only 277 sentences with reflexives and 378 sentences with pronouns remained for analysis. Furthermore, since the cases of syntactic binding and pragmatic binding should be mutually exclusive, à la Pollard & Sag (1992), the discourse-pragmatic factors were encoded mainly in the cases where the sentences did not represent syntactic binding.

After the coding was completed, the sentences were analyzed for each case of the encoded factors – through Pearson's Chi-square (χ^2) tests, in which the frequency of the occurrences between the reflexives and the pronouns in the given conditions were compared according to different syntactic-pragmatic factors. If the result of the χ^2 -test represents statistical significance, then it would mean that the distributional difference between the reflexives and the pronouns is meaningful.

3. Results and Discussion

3.1. Sentential Antecedent vs. Discourse Antecedent

Before encoding the syntactic vs. pragmatic binding factors, the sentences were first coded under the division as to whether the NPs (anaphors/pronouns) have sentential antecedent or discourse antecedent. The cases in which both the antecedents and the reflexives/pronouns were present in the same sentence were encoded as sentential antecedent, while those where the reflexives/pronouns did not find the antecedents within the same sentence were encoded as discourse antecedent. While pronouns are natural when they are found with discourse antecedents, as for the reflexives, the cases are interesting, though rare. Some examples of discourse-bound reflexives found from ICE-USA are shown in (8) below.

- (8) a. Part of that stems from the unbelievable lack of communication between myself_i and the hundreds of wonderful Obies. <ICE-USA: W1B-004#97:1>
 b. Like myself_i, these women are white and all are in their mid-thirties to early forties. <ICE-USA: W2A-020#32:1>

The results showed different rates of sentential vs. discourse antecedent with respect to the NP types, as shown in Figure 1. While the pronouns found discourse antecedents in more cases (115 out of 378: 30.4%), the reflexives showed only 20 cases for discourse antecedents out of 277 sentences (7.2%).

The Pearson’s Chi-square test with Yates’ continuity correction revealed that the distributions of the pronouns and the reflexives used in this study were significantly different from each other ($\chi^2 = 75.154$, $df = 1$, $p < .001$), with respect to finding sentential vs. discourse antecedent. This result indicates apparent cases of discourse-bound anaphors, which implies that binding of reflexives is not restricted to within-sentence domain but is extended to discourse. This supports the exempt anaphor approach (Pollard & Sag, 1992; Reinhart & Reuland, 1993; Huang & Liu, 2001, etc.).

3.2. Syntactic Binding Conditions

The syntactic binding conditions such as c-command (and embedding) were mostly limited to the case of sentential binding shown earlier in Figure 1. On the other hand, the two conditions for defining GC (i.e., TSC and SSC) were calculated by including discourse binding, since discourse binding in itself can be considered TSC and/or SSC violations.

As for the structural relation of c-command between the antecedent and the reflexives/pronouns, the two NPs again showed opposite pattern of distributions from each other. While the reflexives were found with their c-commanding antecedents in most cases (88.2%), the pronouns showed only 36.9% of the cases

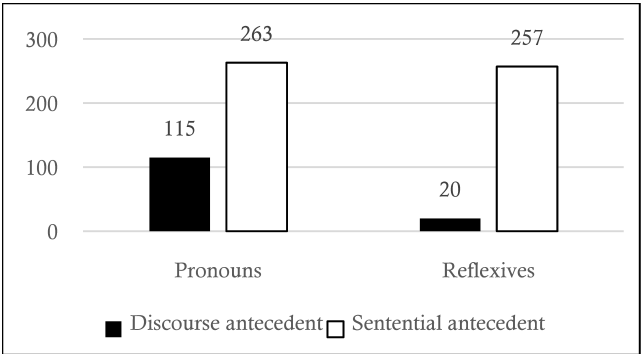


Figure 1. The cases of discourse-antecedent and sentential antecedent

with the c-commanding antecedents (and even in those cases of c-command, it was outside the binding domain which they found their c-commanding antecedents from). Other than that, the pronouns found their antecedents in various structural conditions: sub-commanding antecedents (i.e., where the possessor of the c-commanding NP binds the pronouns) up to 5.1% and non-c-commanding antecedents up to 20.4%.¹⁰⁾ Such patterns of distributions are shown in Table 2 below.

The Pearson's Chi-square test revealed that the distribution of the reflexives and the pronouns was statistically different ($\chi^2 = 125.58$, $df = 4$, $p < .001$) in terms of C-command. This pattern of results supports SBT (Chomsky, 1980, 1981, 1986), given that the reflexives were generally c-commanded by their antecedents. What is also interesting about the distribution of the reflexives in this study is that they seemed to show some cases of exempt binding – with backward binding and discourse antecedents (8.4%), which goes well with the exempt anaphor approach (Pollard & Sag, 1992, etc.).

In addition, the sentences representing two Opacity Conditions (Chomsky, 1980) for GC also showed the similar pattern of complementary distributions between the reflexives and the pronouns. Table 3 below shows that the reflexives were inclined not to violate TSC (TSC-violation of the reflexives: 12.6%), while the pronouns violated TSC in most cases (TSC-violation of the pronouns: 95.5%). The χ^2 -test showed that the difference between the reflexives and the pronouns was significant ($\chi^2 = 508.36$, $df = 1$, $p < .001$), which supports the original assumption of the complementary distributions, proposed by SBT.

Though TSC and SSC violations overlap in many cases, we could find some examples of reflexives violating only TSC from ICE-USA, as shown in (9).

Table 2. C-command between the antecedent and the NP

	C-command	Sub-command	Backward binding	No-command	n/a	Sum	χ^2
Pronouns	114	16	1	63	115	309	125.58***
Reflexives	240	0	3	0	29	272	
Sum	354	16	4	63	144	581	

10) We agree with the anonymous reviewers that the pronouns can co-refer with the individual via coreference and the co-indexed patterns with pronouns in this study may not all represent the case of binding. Nevertheless, we provided the results of the reflexives in comparison with the pronouns, so as to represent the distributional differences.

Table 3. TSC-violation

	No	Yes	Sum	χ^2
Pronouns	17	361	378	508.36***
Reflexives	242	35	277	
Sum	259	396	655	

- (9) a. It can perhaps best be explained as an example of Aristotle_i,
accommodating himself_i to the religion of his day. <ICE-USA: W2A-008#34:1>
- b. It is only the grander passions of poetry_i, allying themselves_i with forms
more abstract and permanent," that, De Quincey said, could last. <ICE-USA: W2B-003#15:1>

On the other hand, some examples of SSC-violating reflexives extracted from ICE-USA are shown in (10).

- (10) a. ...but they do not make it just as they please; they_i do not make it under
circumstances chosen by themselves_i, but under circumstances directly
encountered, given and transmitted from the past. <ICE-USA: W1A-007#4:3>
- b. After all, there is a risk of failure in committing myself_i to this situation,
but not in planning for the future. <ICE-USA: W1B-012#127:1>

As for SSC-violation, not to mention the case of reflexives, even the case of pronouns reported lower rate of violating SSC compared to the cases of TSC-violation (cf. Table 3 from above). However, though the pronouns tended to violate SSC less than TSC, the rates of SSC-violation between the pronouns and the reflexives were still different (SSC-violation: Pronouns = 39%, Reflexives = 9.0%). The results with SSC-violation is shown in Table 4.

The χ^2 -test revealed that there was a significant difference between the pronouns and the reflexives in terms of the violation of SSC ($\chi^2 = 108.53$, $df = 1$, $p < .001$). The results with SSC-violation is interesting, in that the apparent existence of the sentences with SSC-violating anaphors was not predicted either in SBT or HPSG. This indicates some gaps between theory and reality.¹¹⁾

In summary, overall patterns of the results with syntactic binding factors described

Table 4. SSC-violation

	No	Yes	Sum	χ^2
Pronouns	231	148	379	108.53***
Reflexives	252	25	277	
Sum	483	173	656	

so far imply that reflexives and pronouns are mostly in complementary distribution with each other, which goes well with the main assumption of SBT (Chomsky, 1980, 1981, 1986) as well as syntactic analysis of binding in HPSG (Pollard & Sag, 1992). At the same time, some exceptional cases of the reflexive distribution (including LD-binding and discourse binding) seem to strengthen the exempt anaphor approach in HPSG (Pollard & Sag, 1992, etc.)¹². As for SSC-violating anaphors, neither of the original theories could predict them; they would rather be explained in purely pragmatic approaches to binding (Zribi-Hertz, 1989) or some extended versions of theories on Long-Distance Anaphors (LDAs) (Sells, 1987; Kuno, 1987, etc.).

3.3. Discourse-Pragmatic Conditions for Binding

The first discourse-pragmatic factor (i.e., Exempt condition) was coded with the presence/absence of potential syntactic antecedent of the target NP. If the NP had a potential antecedent available in the syntactically accessible position (i.e., c-commanding the reflexive and not violating TSC and SSC), the sentence was coded ‘no’ for Exempt condition. In contrast, when the NP had no syntactically accessible antecedent, it was coded ‘yes’ to be considered a case of exempt binding. Some examples of exempt anaphors extracted from ICE-USA are shown in (11) below.

- (11) a. Like myself_i, these women are white and all are in their mid-thirties to early forties. <ICE-USA: W2A-020#32:1>
b. Long-term earthquake prediction is one of the major goals of seismologists, like myself_i. <ICE-USA: W2B-025#59:1>

11) Later, Pollard (2005) introduced these cases of SSC-violating anaphors by mentioning Runner & Kaiser (2005)’s examples. Cole et al. (2001) also mentioned such case of SSC-violating English anaphors.
12) This does not mean that SBT completely excluded the possibility of exceptional binding of the exempt anaphors; Chomsky (1986) mentioned some cases of exempt anaphors (such as picture nouns, etc.) with accessibility conditions, but did not deal with exempt anaphors thoroughly.

c. Everyone, including myself, considered her a kind of goddess. <ICE-USA: W2F-005#103:1>

The result with the relevant sentences showed that only 10.1% of the reflexives were found to be in the exempt position, whereas the pronouns showed the distribution of 75.6% without accessible syntactic antecedents, as shown in Table 5 below.

The difference between the reflexives and the pronouns with respect to Exempt condition (i.e., without syntactically accessible antecedent) was significant, as was revealed by Pearson's χ^2 -test ($\chi^2 = 321.78$, $df = 2$, $p < .001$). This pattern of results is in line with Pollard & Sag (1992) that there are anaphors that can be exempt from syntactic constraints with no syntactically accessible antecedent.

On the other hand, the coding of the Logophoric condition and logophoricity-related factors was done with only reflexives, since English pronouns are different from typical logophoric pronouns that are observed in specific languages other than English as reported in Sells (1987) and Maling (1984). Logophoric condition was encoded as to whether the antecedent of the reflexive requires a salient discourse-pragmatic role or not.¹³⁾ Table 6 below shows the results with Logophoric condition of the reflexives that the reflexives did not require salient discourse antecedents in most cases (Logophoric condition of the reflexives: 8.3%).

Table 5. Exempt condition (i.e., without syntactically accessible antecedent)

	Yes	No	n/a	Sum	χ^2
Pronouns	286	4	88	378	321.78***
Reflexives	28	21	228	277	
Sum	314	25	316	655	

13) Being an exempt anaphor may require the long-distance/discourse antecedent to be pragmatically-salient, but not vice versa. For example, the anaphors in adjunct PPs are considered exempt and they require logophoric antecedents. However, not all logophoric antecedents are considered exempt – for example, SSC-violating anaphors are rare, but they have the syntactically-accessible subject position filled with another intervening subject and are bound across the intervening subject, as follows.

a. But Rupert_i was unduly worried about Peter's opinion of himself_i. (Cole, Hermon & Huang, 2001)

This can serve as a counter-example to exempt anaphor approach (Pollard & Sag, 1992), not to mention SBT. In this reason, the present study distinguished logophoric condition from exempt condition.

Table 6. Logophoric condition

	Yes	No	Sum
Reflexives	35 (8.3%)	254 (91.7%)	277 (100%)

With Logophoric condition of ‘yes’, the logophoric roles of the LD-antecedents of the reflexives (i.e., SOURCE, SELF, PIVOT) were encoded. In the coding procedures, SOURCE was coded with the sentences where the LD-antecedent of the anaphor is the agent of ‘verbs of saying’, while SELF was coded with the cases where the LD-antecedent represents the experiencer of the psychological event. As for PIVOT, the cases were encoded where LD-antecedent represents discourse-salient entity that were SELF, nor SOURCE. The examples of the antecedents with different logophoric roles from ICE-USA are the following, shown in (12).

- (12) a. SOURCE: After all, there is a risk of failure in committing myself_i to this situation, but not in planning for the future. <ICE-USA: W1B-012#127:1>
b. SELF: It was really bizarre to see myself_i in a situation that I normally see as happening to people when they're older. <ICE-USA:W1B-008#120:1>
c. PIVOT: Hurt herself_i? <ICE-USA: W2F-020#47:1>

As shown in Table 7 below, PIVOT got the highest rates (7.2%) for the LD-antecedents of the reflexives among other roles. On the contrary, as for the rates of SOURCE and SELF antecedents, the reflexives reported higher rate for SOURCE compared with SELF (SOURCE: 4.3%, SELF: 1.1%).

To summarize, overall results with discourse-pragmatic factors of the pronouns and the reflexives are in line with exempt anaphor approach (Pollard & Sag, 1992, etc.) and the theory of logophoricity (Sells, 1987), in that the LD-antecedents of

Table 7. Logophoric role of the antecedent¹⁴⁾

	SOURCE	SELF	PIVOT	n/a	Sum
Reflexives	12 (4.3%)	3 (1.1%)	20 (7.2%)	242 (87.4%)	277 (100%)

14) The rate of ‘n/a’ was reported with the cases where no specific logophoric role was required with the antecedents of target NPs, most of which were the cases where the antecedents were found very locally.

the exempt anaphors represented canonical logophoric roles. On the other hand, the pattern of the results with logophoric roles did not represent the canonical hierarchy of logophoric centers proposed by Sells (1987): The most frequent logophoric roles found for the LD-antecedents of exempt anaphors were PIVOT, which takes the lower position than SELF or SOURCE in Sells (1987)' hierarchy.¹⁵⁾ This shows that the corpus data with logophoric roles can merely show which type of predicates is more frequently used; therefore, it is difficult to see the hierarchy represented in any way through corpus data. Given that SOURCE and SELF are limited to certain types of predicates, whereas PIVOT is not limited to the embedded attitude complement, it would be natural that the distribution of PIVOT is much freer than the other two roles and is thus found to make the highest frequency in the limited number of data. Moreover, since three logophoric roles do not block one another (i.e., the presence of the SOURCE does not necessarily block the binding with a PIVOT), it may be natural to have difficulty seeing Sells (1987)' hierarchy in our data.¹⁶⁾

4. Conclusion

This study took a corpus-based approach to examine the distribution of English reflexives and pronouns by using ICE-USA. The study specially focused on comparing different theoretical frameworks on binding phenomena – as to whether and how the predictions of the different previous theoretical claims (Chomsky, 1980, 1981, 1986; Pollard & Sag, 1992; Sells, 1987, etc.) are borne out through the actual use of English. The study was based on the method of extracting English sentences representing binding from ICE-USA – for manual encoding of the factors (i.e., theoretical constructs related to binding phenomena in different approaches), à la Kim et al. (2018) and Kim et al. (2020), but made up for the limitations of those

15) However, this does not undermine the theory of logophoricity. As one of the anonymous reviewers kindly pointed out, the logophoric hierarchy is conceptual: For instance, the 'verbs of saying' induces an indirect speech context, which makes the speaker identified as SOURCE at the same time; from the modal perspective, it should also be SELF, considering the *de se* interpretation in Huang & Liu (2001)'s term. On the other hand, in the case of psychological predicate like 'believe', the author of the attitude context just identifies as a SELF. Since SOURCE and SELF both have to do with specific types of attitude predicates, it is hard to say that which one occupies higher position in the hierarchy.

16) We deeply appreciate the anonymous reviewer who kindly provided us with the suggestion of such interpretation with the relevant patterns of results.

studies. Different from previous corpus studies on binding, which focused mostly on the syntactic properties of binding, the present study incorporated exempt binding and discourse-pragmatic factors for binding, in order to discuss cross-theoretical implications regarding binding phenomena.

The summary of the results are as follows. First of all, English reflexives and pronouns represented in ICE-USA still showed overall complementary distributions with sentential binding, as proposed in Standard Binding Theory (Chomsky, 1980, 1981, 1986) and HPSG (Pollard & Sag, 1992). Besides, there were some cases representing LD-exempt anaphors and discourse-bound anaphors that are in line with exempt anaphor approach of HPSG (Pollard & Sag, 1992, etc.). Furthermore, when the reflexives were bound within their GC (i.e., binding domain), syntactic constraints (i.e., c-commanding and conditions for GC) were not likely to be violated often, as predicted in both SBT (Chomsky, 1980, 1981, 1986) and HPSG (Pollard & Sag (1992). On the other hand, when the antecedents of the reflexives were bound outside the syntactic GC, they tended to satisfy discourse-pragmatic conditions (Pollard & Sag, 1992) – by being a discourse-salient logophoric centers, as Sells (1987) proposed.

On the contrary to the original assumption of Pollard & Sag (1992) that exempt binding would become available only when there is no syntactically available antecedent, the corpus data in the current study presented apparent examples of SSC-violating anaphors (in which the syntactically accessible antecedent position is not empty), as also shown in Runner & Kaiser (2005) and Pollard (2005). This implies that syntactic binding conditions and the conditions on pragmatic/logophoric binding may not be mutually exclusive – like what Pollard & Xue (2001) once proposed with the case of Chinese reflexives.

These SSC-violating LD anaphors could only be explained by pragmatic binding theory (Sells, 1987; Kuno, 1986, etc.). Therefore, the results of this corpus-based study indicate that binding phenomena cannot be discussed only with syntactic conditions, but should be extended to the thorough discussion of discourse-pragmatic conditions as well. That is, the approach that takes only syntactic analysis is not sufficient to explain various types of binding, which implies the needs for incorporating discourse-pragmatic properties to discuss binding phenomena. More future studies are called for to explain such cases (i.e., LD-binding with SSC violation, etc.), which can provide more refined theoretical explanations.

The current corpus study attempted to include pragmatic factors of LD/discourse reflexives as well as well-known syntactic factors that are crucial for binding of the

reflexives. However, as LD-exempt anaphors – especially, SSC-violating anaphors – are usually rare in English, the cases of LD/discourse-bound reflexives were found in low frequency in this corpus-based study. Nonetheless, future corpus-based studies with larger sample size and more comprehensive as well as exhaustive coding of syntactic and pragmatic factors may enable us to see more cases for LD-exempt binding to discuss further issues such as the theory of logophoricity (Sells, 1987). Also, if there is/are any frequency-based difference(s) in binding between American and British English – as Zribi-Hertz (1989) and Pollard (2005) mentioned – future studies should also focus on figuring out the differences by comparing the two corpora.

References

- Büring, D. (2005). *Binding theory*. Cambridge: Cambridge University Press.
- Chomsky, N. (1980). On binding. *Linguistic Inquiry*, 11, 1-46.
- Chomsky, N. (1981). *Lectures in government and binding*. Dordrecht: Foris.
- Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin and Use*. New York: Praeger.
- Cole, P., Hermon, G. & Huang, C.-T. J. (Eds.). (2001). Introduction: Long-distance reflexives: The state of the art. *Syntax and Semantics*, 33, xiii-xlvii.
- Cole, P., Hermon, G. & Sung, L.-M. (1990). Principles and parameters of long-distance reflexives. *Linguistic Inquiry*, 21(1), 1-22.
- Huang, C. T. J., & Liu, C. S. L. (2001). Logophoricity, attitude, and ziji at the interface. *Syntax and Semantics*, 33, 141-195.
- Kim, J.-H. & Yoon, J. H.-S. (2009). Long-distance bound local anaphors in Korean-An empirical study of the Korean anaphor caki-casin. *Lingua*, 119(5), 733-755.
- Kim, J.-H., Lee, Y. & Kim, D. E. (2018). A study on binding conditions of reflexives and pronouns in the ICE-GB corpus. *Modern Studies of Grammar*, 99, 143-160.
- Kim, J.-H., Lee, Y., Kim, J., & Kim, Y. (2020). A multifactorial analysis of binding conditions in British English using ICE-GB corpus. *Studies in Linguistics*, 55, 83-107.
- Kuno, S. (1987). *Functional Syntax: Anaphora, discourse and empathy*. Chicago: Chicago University Press.
- Lange, C. (2006). Reflexivity and intensification in Irish English and other new Englishes. *Celtic Englishes*, IV, 259-282.
- Maling, J. (1984). Non-clause bounded reflexives in Modern Icelandic. *Linguistics and Philosophy*, 7, 211-41.
- Manzini, R., & Wexler, K. (1987). Parameters, binding theory and learnability. *Linguistic Inquiry*, 18, 413-444.

- Pollard, C. J., & Sag, I. A. (1992). Anaphors in English and the scope of binding theory. *Linguistic Inquiry*, 23, 261-303.
- Pollard, C. J., & Xue, P. (2001). Syntactic and non-syntactic constraints on long-distance reflexives. In P. Cole, H. Gabriella, & J. Huang (Eds.), *Long distance reflexives. Syntax and Semantics Series*. New York: Academic Press, 317-342.
- Pollard, C. (2005). Remarks on binding theory. In *Proc. 12th Int'l. Conf. on HPSG, Dept. of Informatics, Univ. of Lisbon*, ed. by S. Müller, 561-577. Stanford, CA: CSLI Publications.
- Reinhart, T., & Reuland, E. (1993). Reflexivity. *Linguistic Inquiry*, 24, 657-720.
- Reuland, E. (2011). *Anaphora and Language Design*. Cambridge, MA: MIT Press.
- Reuland, E. (2017). Why is reflexivity so special? Understanding the world of reflexives, *Studia Linguistica*, 71(1-2). 12-59.
- Runner, J. T., Sussman, R. S. & Tanenhaus, M. K. (2003). Assignment of reference to reflexives and pronouns in picture noun phrases: evidence from eye-movements. *Cognition*, 89, B1-B13.
- Runner, J. T. & Kaiser, E. (2005). Binding in picture noun phrases: implications for binding theory. In A. Branco et al., eds., 53-58.
- Runner, J. T., Sussman, R. S. & Tanenhaus, M. K. (2006). Processing reflexives and pronouns in picture noun phrases. *Cognitive Science*, 30, 193-241.
- Sells, P. (1987). Aspects of logophoricity. *Linguistic Inquiry*, 18(3), 445-479.
- Song, S. (2017). A corpus study of unbound reflexive pronouns in English. *Korean Journal of English Language and Linguistics*, 17(2), 275-305.
- Yang, D.-W. (1983). The extended binding theory of anaphors. *Language Research*, 19, 169-192.
- Zribi-Hertz, A. (1989). Anaphor binding and narrative point of view: English reflexive pronouns in sentence and discourse. *Language*, 65, 695-727.

Ji-Hye Kim

Professor

Department of English Education

Korea National University of Education

250 Taeseongtapyeon-ro, Heungduk-gu, Cheongju 28173, Korea

jkim@knue.ac.kr

Soojin An

Teacher (Former Graduate Student in KNUE)

Myeongji Middle School

54, Myeongjigukje 6-ro 318beon-gil, Gangseo-gu, Busan 46726, Korea

english0130@myeongji.ms.kr

Ahreum Jung
Graduate Student (Ph D Track)
Department of English Education,
Korea National University of Education
250 Taeseongtapyeon-ro, Heungduk-gu, Cheongju 28173, Korea
ggamjje@naver.com

Received: July 04, 2020

Revised version received: August 07, 2020

Accepted: October 29, 2020