Korean Null Pronouns: Classification and Annotation

Na-Rae Han
University of Pennsylvania
nrh@ling.upenn.edu

Abstract

This paper discusses an annotation scheme for Korean null pronouns, which were used in annotating three kinds of Korean text corpora including Penn Korean Treebank. In annotating the corpora, null pronouns and their antecedents were marked up for their type and reference, with coreference relation tracked by numeric identifiers. Based on the annotation scheme, an outline of a potential pronoun resolution strategy is also proposed. The resulting dataset of annotated text is rather small at 11,834 words; we hope the null pronoun classification and annotation scheme proposed in this study will serve as a basis in developing a large-scale annotated corpus in the future.

1 Introduction

Korean is one of the class of languages which permits extensive use of null arguments. These zero arguments are believed to be a substitute for pronouns[1] and have received terms that reflect such a view throughout the literature, including *zero anaphor*, *zero pronominal*, and more theory–specific *pro*. I adopt the term *null pronoun* in my paper.

Use of null pronouns in a discourse has been believed to be largely governed by discourse principles. In such uses of null pronouns, *saliency* of their referents in the discourse context is said to play a major role. However, null pronouns are not always anaphoric in terms of the types of reference. In a categorization proposed by Kameyama (1985) for Japanese zero anaphora, three subcategories belong to the type of "text independent use", which cover non-anaphoric usage of zero pronouns. This classification is given in Table 1.

The first category, "discourse anaphoric zero anaphor", is where the null pronoun derives its referents from what has been said earlier in

---

[1] Kameyama (1985) chapter 3.3

text dependent	discourse anaphoric zero anaphor
text independent	zero deictic and zero indexical
	indefinite personal zero anaphor
	general situational zero anaphor

Table 1: Classification of Japanese zero anaphor by Kameyama (1985)

the text and is motivated by discourse effects. The second main category, which represents the text independent use of null pronouns, is non-anaphoric in nature: the entities referred to by null pronouns in the category are not verbally expressed in the discourse.

Korean null pronouns show the same kind of usage: it is not difficult to construct examples for these categories in Korean, as shown in examples 1-3. Zero deictic and zero indexical pronouns (in example 1) are null pronouns that directly refer to an entity in the given spatiotemporal context. Indefinite personal zero anaphors (in 2) are usually a subject and refer to the people in general, as in some uses of English *they*, *one* and *you*. General situational zero anaphors (in 3) are zero subjects that refer to time, weather, and a general situation like English dummy subject *it*.

(1) a. (looking at an elephant)
    ϕ 엄청 크네.
    ϕ emcheng ku-ney.
    SBJ very big-EXCLAM
    '(It)'s very big!'

   b. ϕ 집에 가니?
    ϕ cip-ey ka-ni?
    SBJ home-to go-Q?
    'Are (you) going home?'

(2) ϕ 산에 가야 ϕ 호랑이를
   ϕ san-ey ka-ya ϕ holangi-lul
   SBJ mountain-to go-only if SBJ tiger-ACC

잡지.
cap-ci.
catch-PRES_DEC
'Only if (one) goes to the mountains (one) catches a tiger.'

(3) ϕ 벌써   열시다.
    ϕ pelsse   yel-si-ngi-ta.
    SBJ already ten-o'clock-COP-DEC
    '(It) is 10 o'clock already.'

This paper presents a result of preliminary studies on Korean null pronouns, based on findings from real corpora. The main purpose is to establish a suitable classification scheme for Korean null pronouns and propose an annotation scheme which can serve as a guideline for producing an annotated corpus which can eventually be used in developing a pronoun resolution system.

2   The Corpora

For this study, I used three distinct sources of corpora: 7 pieces of news and magazine articles, the novel *Little Prince* chapters 1 through 8, and part of the Penn Korean Treebank corpus[2]. The entire collection of the corpora consisted of 11,834 words[3], with a total of 1,299 sentences. The average length of sentences is 9.11 words; the sentences in the article corpus proved to be the longest, and those in the novel *Little Prince* were the shortest. There were 1,028 tokens of null pronouns identified in total. The article corpus shows the highest number of null pronouns per sentence, which is largely due to its longer sentence length overall. When the frequency of the null pronoun occurrence is measured with the word as the unit, the article corpus in fact scores the lowest among the three corpus. The statistic facts about the three corpora are summarized in Table 2 and Table 3.

Taking the number of null pronoun per word as the indicator, the frequencies of null pronouns represented in the three corpus are in the following order: articles > *Little Prince* > Penn Korean Treebank. The Penn Korean Treebank corpus entirely consists of dialogue sequences, and the novel *Little Prince* is intended for a younger audience with many dialogues scattered throughout the text. From this, one

can stipulate that the more informal and colloquial the text is the higher the frequency of null pronouns is.

3   Korean Null Pronouns: A Modified Classification

From examining the null pronoun cases in the corpora, I found the classification system in Table 1 proposed for Japanese to be largely suitable for Korean. However, I would like to introduce a further subcategorization for the class 'Discourse-anaphoric' and propose a revision on the class 'Anonymous', which is shown in Table 4.

3.1   Propositional vs. NP anaphoric

First, I suggest that the 'Discourse-anaphoric' class be subdivided into 'Propositional anaphoric' and 'Noun Phrase anaphoric". The distinction is a semantic one: Propositional null pronouns are those that refer to a propositional content or an event represented by a sentence, while NP anaphoric null pronouns are those that are coreferential with other noun phrase arguments mentioned in the context. The following are examples of each type:

(4)   **propositional anaphoric**

a. "정치적인   허위 소문은   사흘
    cengchicekin hewi somwun-un sahul
    political   false rumor-Top 3days
    동안만   신뢰를   받으면
    dongan-man sinloy-lul   pat-umyen
    while-Only   credit-Acc get-If
    커다란   도움이   된다."
    khetalan towun-i   toy-nta.
    big   help-Nom become-PresDec
    'False political rumors are very useful if it is credited for as short as 3 days.'

b. ϕ 쿠데타의   명인이던
    ϕ khwuteytha-uy   myengin-i-ten
    SBJ coup-d'e'tat-Gen master-Co-PastRel
    디 메디치의   말이다.
    ti meytichi-uy mal-i-ta.
    Di Medici-Gen words-Co-PastDec
    '(That) is Di Medici's words, a master of cout d'e'tat.'

(5)   **NP-anaphoric**

a. 보아뱀은   먹이를
    poapaym-un   meki-lul
    boa-snake-Top prey-Acc
    씹지않고   통째로
    ccip-ci-anh-ko   thongccaylo
    chew-Aux-Neg-And as-whole

---

[2]Each file in the Penn Korean Treebank has a dialogue sequence followed by a random selection of sentences. Only the dialogue sections from each file were examined for this study.

[3]Punctuations and symbols were not segmented for this count.

corpus	characteristics	words	sentences	null pronouns
Articles	news and magazine articles	4098	340	330
Little Prince	novel; Korean translation	3513	462	311
Penn Korean Treebank	military domain; dialogues	4223	497	387
all		11834	1299	1028

Table 2: Summary of the corpora

corpus	words/sentence	null pronoun/sentence	null pronoun/word
Articles	12.05	0.97	0.080
Little Prince	7.60	0.67	0.089
Penn Korean Treebank	8.50	0.78	0.092
all	9.11	0.79	0.087

Table 3: Statistics of the three corpora

text dependent use	discourse anaphoric	propositional
		NP anaphoric
text independent use	zero deictic and zero indexical	
	indefinite null pronoun	generic (non-specific)
		specific
	general situational null pronoun	

Table 4: Classification of Korean null pronouns

삼킨다.
samkhi-nta.
swallow-Pres-Dec
'Boa constrictors swallow their prey whole, without chewing.'

b. φ 그런 다음 먹이가
   φ Kule-n taum meki-ka
   SBJ do-so-Rel next prey-Nom
소화될 때까지 잠을
sohwatoy-l ttay-kkaci cam-ul
digest-Rel time-until sleep-Acc
잔다.
ca-n-ta.
sleep-Pres-Dec
'After that (they) sleep until the prey is digested'

The rationale behind this distinction comes not only from their apparent semantic differences: they show different patterns in terms of their anaphoric tendencies. From the example above, the sentence (4a) serves as the antecedent of the subject null pronoun in (4b), whereas in the case of NP-anaphoric pronouns, the subject null pronoun is coreferential with the subject NP of the preceding sentence. Under an approach that attempts to provide an explanation on their behavioral patterns such as the Centering Theory, these two types are subject to different sets of rules. We discuss this further in a later section.

3.2 Generic vs. Specific Indefinite Null Pronouns

Not all indefinite null pronouns are generic, e.g. refer to the people in general. Sometimes, it is understood from the context that there has to be a specific entity that a null pronoun refers to, although it is not mentioned anywhere in the text or not locally accessible. Take the following sentence, for example:

(6)   **indefinite specific null pronouns**

φ '큐브'는
φ kyupu-nun
'The Cube'-Top
공상과학영화의 틀을
kongsangkwahakyenghwa-uy thul-ul
SFmovie-Gen frame-Acc
빌려 미로의 탈출이라는
pilli-e milo-uy thalchwul-ilanun
adopt-Aux labyrinth-Gen escape-Quote
신화를 영화로 만든
sinhwa-lul yenghwa-lo mantu-n
myth-Acc movie made-Rel
작품이다.
cakphwum-i-ta.
work-Cop-Dec.

'Cube is a work that (someone) made the myth of an escape from a labyrinth into a movie adopting a SFmovie framework.'

It is implied that there exists someone who is the maker of the movie(the director), which is

the referent of the subject null pronoun. However, a mention of the director occurs only at a later part of the text, after several intervening paragraphs. Contrast this case with the following example, where the null pronoun is referring to an indefinite entity:

(7) **indefinite generic (non-specific) null pronouns**

φ 밤의　　　어둠　　　속에서　길을
φ pam-uy　　etwum　　sok-eyse kil-ul
φ night-Gen darkness midst-at way-Acc
잃을　　　때　지리는　　　　매우
ilh-l　　ttay cili-nun　　　　maywu
lose-Rel when geography-Top very
편리하다.
phenliha-ta.
convenient-Dec

'When (one) gets lost in the darkness of the night, geography comes in very handy.'

Notably, the indefiniteness of such null pronouns is closely tied to the semantic interpretation of the sentence in which they occur. The sentence often involves modal environment such as conditionals and irrealis tense. From a theoretical linguistic point of view, this type of use of null pronouns can be treated as involving quantification on the pronoun by an operator residing outside of the pronoun itself.

Distinguishing such different types of indefinite null pronouns lets us extend the well-studied specific/non-specific dichotomy of the English indefinite article *a* to the case of Korean null pronouns.

4 Annotating the Data

The three corpora were annotated using the proposed classification system for Korean null pronouns; null pronouns were identified and categorized in the text, with their potential antecedents identified where applicable. Some issues arose in the course of studying the corpus, which were reflected in designing the annotation scheme used for this study.

4.1 Where to Find the Invisibles

Null pronouns, being an invisible linguistic construct, present a fundamental problem to be addressed prior to their treatment, namely their identification. In fact, the problem of identification has direct consequences on the theory of null pronouns and can be also approached in a way that facilitates treatment of a particular problem of null pronouns, for example, null pronoun resolution.

In order for an unexpressed element to be identified, it has to be predictable from the structure of the sentence in which it is contained. Therefore, identification of null pronouns crucially relies on the specific representation of the clause and sentence structures. The most significant indicator is, arguably so in any theory, the predicate. Predicates have their own subcategorization structure, and when an argument is missing from this scheme, one can reasonably induce the presence of a null pronoun.

For the purpose of annotation, the syntactic annotation guidelines built for the Penn Korean Treebank were assumed. The decision in part was influenced by the fact that the Treebank corpus used in this study is already annotated according to the standards.

4.1.1 Situational Null Pronouns

In determining syntactic structures of a sentence, there are cases which require judgment that draws on a particular linguistic theory one employs. One such case is the situational null pronouns. Given they are both phonetically and semantically null, is it necessary to posit a null syntactic element for them in the first place? One can subscribe to a syntactic theory for Japanese or Korean which, unlike languages such as English, simply do not require verbs to have a syntactic subject. I will not pursue this matter further in this paper, but instead offer an argument in support of situational null pronouns from a more practical perspective. For an application dealing with null pronouns, it is necessary at one stage or another to examine all possible argument slots of a predicate and determine whether it is filled or not, before deciding whether the null pronoun is semantically present. This is true because the subject slot of those predicates with situational null subjects is under no syntactic obligation that they should be empty, as seen in example below:

(8) 지금　시간이　　　벌써
cikum sikan-i　　　pelsse
now　time-NOM already
열시다.
yel-si-ngi-ta.
ten-o'clock-COP-DEC
'The hour is 10 o'clock already.'

Therefore, from a practical point of view, it is more suitable for an application to start with every argument position that is unfilled and determine whether there is a null element that has a semantic content. This, in my opinion, can

be achieved without committing to any specific theory of syntax.

4.1.2 Null Pronouns with a Co-sentential Antecedent

Having the null pronoun as a valid construct in a syntactic theory offers conflicting takes on two or more predicates that seemingly share an argument. A sentence with two predicates which share one subject can receive two distinct structural analyses: one with VP coordination with a shared subject or the other with two coordinated clauses with a null pronoun posited as the subject in one of them, probably coindexed with the subject of the other clause.

(9) **Syntactic Structures for 'We had dinner and came back.'**

a. **VP coordination: shared NP subject**

```
(S (NP-SBJ 우리/NPN+는/PAU)
   (VP (VP (NP-OBJ 저녁/NNC+을/PCA)
           먹/VV+고/ECS)
       (VP 돌아오/VV+았/EPF+다/EFN))
   ./SFN)
```

b. **S coordination: null pronoun represented**

```
(S (S (NP-SBJ 우리/NPN+는/PAU)
      (VP (NP-OBJ 저녁/NNC+을/PCA)
          먹/VV+고/ECS))
   (S (NP-SBJ *pro*)
      (VP 돌아오/VV+았/EPF+다/EFN))
   ./SFN)
```

The Penn Korean Treebank offers strictly defined guidelines for annotating such constructions. It allows for coordination at all levels: VV, VP and S, whilst recommending coordination at the lowest level possible for a given sentence. [4]

Still a structural ambiguity remains even after two full clausal units are assigned to such sentences: it is yet to be decided which predicate has an overt subject and which one a null subject. The decision will lead to two distinct syntactic trees: one with a coordinated S structure and the other with a subordinate S modifying the main clause VP, as illustrated in the following example (this time with a similar connective -고서 -*kose*: 'and then').

In addition to the coordinated S structure where the overt subject is associated with the first predicate of the sequence, associating the overt subject with the final VP will place the null subject within an intervening subordinate clause which modifies the main clause VP. [5]

(10) **Syntactic Structures for 'We had dinner and came back.'**

a. **null pronoun in main clause**

```
(S (S (NP-SBJ 우리/NPN+는/PAU)
      (VP (NP-OBJ 저녁/NNC+을/PCA)
          먹/VV+고서/ECS))
   (S (NP-SBJ *pro*)
      (VP 돌아오/VV+았/EPF+다/EFN))
   ./SFN)
```

b. **null pronoun in subordinate clause**

```
(S (NP-SBJ 우리/NPN+는/PAU)
   (VP (S (NP-SBJ *pro*)
          (VP (NP-OBJ 저녁/NNC+을/PCA)
              먹/VV+고서/ECS))
       (VP 돌아오/VV+았/EPF+다/EFN))
   ./SFN)
```

One important implication of these two distinct structures is the relative placement of the overt NP and the null pronoun in terms of clausal depth. Given that the final verb in Korean projects the matrix clause, the structure in (10a) places the overt NP subject in the subordinate clause and the null pronoun in the main clause, while the relation is reversed in (10b). If one is to view the overt subject NP as the antecedent of the null pronoun, the analysis in (10b) will appeal to many since the antecedent is occupying a structurally higher position than the bound null pronoun. Adding more support to this position, clausal depth of an argument plays a significant role in any theory which attempts to represent level of saliency or discourse accessibility of discourse entities. From this theoretical stance, it seems more reasonable to place the full NP subject at the main clause level for an easy back reference in subsequent discourse.

---

[4]It is noteworthy that Penn Korean Treebank recognizes only two morphemes out of a large number of Korean connective verbal endings as capable of heading a VP coordination: '-고 (-*ko*)' and '-거나 (-*kena*)', standard conjunctive and disjunctive connectives in Korean. Clauses headed by all other connectives, if the option of VV coordination is unavailable, will have to be analyzed as involving an S coordination structure and, therefore, a null pronoun in the subject position in one of the clauses. This annotation decision partly contributed to a large number of subject null pronouns in a coordinated S structure with intra-sentential coreference.

[5]There is an obvious third possibility, where the null pronoun linearly precedes the overt NP subject. As logical it may be, this option seems highly unintuitive and therefore is not considered here.

The Guidelines for Penn Korean Treebank recommends an "incremental parsing" approach (Chapter 22.3), i.e. associating the overt NP subject with the lefthand side (subordinate clause) VP and assigning a null pronoun to the latter (main-clause) VP. However, it also notes that this is just a guideline to be followed when the annotator is in doubt. In a nutshell, I assumed the structure in (10b) wherever possible, i.e. in the other two corpora that have not been pre-annotated with syntactic structures as the Penn Korean Treebank is.

4.2 Determining the Categories

Once a null pronoun is identified, the next step in corpus annotation is to determine its type, according to the classification scheme previously presented in Table 4. In this section, I will discuss some issues that arise in categorization, which eventually lead to some design decisions made for the annotation scheme, to be given in next section.

4.2.1 Deictic or Anaphoric

Deictic pronouns are those that directly refer to an entity that is present in the scene of the discourse, and anaphoric pronouns are those whose reference can be determined via an antecedent expression that is found within the text. Upon a close look, however, these definitions are not mutually exclusive: a null pronoun can refer to an entity present in the discourse scene and have a coreferential antecedent at the same time. This is especially true for first and second person pronouns, whose referents are always a part of the discourse scene. Distinguishing between the two distinct usages of such null pronouns is not always straightforward. Consider the subject null pronoun in the second sentence. Should it be anaphoric or deictic?

(11) 나는$_5$ ϕ_{i5} 좀 똑똑해 보이는 사람을 만날 때
마다, ϕ_{i5} 항상 품고 다니던 내$_5$ 그림 제 1호
를 꺼내 그를 시험해 보곤 했다. ϕ_{i5} 그가 정
말 이해력 있는 사람인가 알고 싶었던 것이
다.

Whenever (I$_{i5}$) met a smart-looking person, I$_5$ tested him whipping out my$_5$ drawing No.1 which (I$_{i5}$) always carried around. (I$_{i5}$) wanted to know whether he really possessed understanding.

Rather than trying to look into the underlying principle that motivated the use of each deictic/anaphoric null pronoun, my choice was simply to allow a null pronoun to be anaphoric and deictic at the same time. Whenever a null pronoun refers to an entity that is understood to be present in the discourse scene, mostly in first person singular/plural and a second person, it will bear a tag linking it to a discourse entity, in addition to a number which coindexes it with an antecedent NP in the text, if present.

4.2.2 Coreference in Generic Null Pronouns

Consider the following examples:

(12) 바오밥나무는 ϕ 자칫 늦게
paopapnamwu-nun ϕ cachic nuckey
paopap-tree-Top SBJ per-chance late
손을 쓰면 그땐 정말 ϕ
son-ul ssu-myen kuttay-n cengmal ϕ
hand-Acc use-If then-Top really SBJ
처치할 수 없게
chechiha-l swu eps-key
remove-Rel possibility not-exist-AuxEnd
된다.
twoy-nta.
become-Dec.
'Baobab tree becomes impossible (for one) to remove if (one/he) treats it too late.'

(13) ϕ ϕ 할 일을 뒤로
ϕ ϕ ha-l il-ul twi-lo
SBJ SBJ do-Rel task-Acc later-To
미루는 것
milu-nun kes
postpone-Rel Comp
'(For one) to postpone a task (for one) to do.'

In these examples, all the null pronouns refer to generic "people" or anonymous "someone". Therefore they can be categorized as an indefinite generic use of a null pronoun. However, note that the two pronouns within each sentence are understood to be coreferential: in example (12), the persons who wishes to catch a tiger and having to go to the mountains are the same one, whoever it might be. This fact is expressed in the English translation for each sentence, where the second occurrence of *one* is substitutable with *he*. Semantically, one of the two null pronouns is a bound variable, whose referent is determined through the other one.[6] This shows that two or more indefinite generic null pronouns can share a referent, and therefore we need a means to express this fact. It can be achieved by assigning each of such null

---

[6] Alternatively, both are variables bound by an operator which resides higher up in the tree, giving them generic interpretation.

pronoun tokens a shared index, in the same fashion as antecedent–anaphor resolution is expressed with coindexation. Therefore, both of such generic null pronouns will bear two indexes, one for their generic category and the other expressing coreference relation, as shown below.

(14) ϕ_{g7} 호랑이를   잡으려면 ϕ_{g7}
     ϕ_{g7} holangi-lul cap-ulyemyen ϕ_{g7}
     SBJ tiger-Acc   catch-IntendIf SBJ
     산에   가야한다.
     san-ey   ka-ya-ha-nta.
     mountain-To go-must-PresDec.
     'If (one) wishes to catch a tiger, (one/he) must go to the mountains.'

4.3   The Annotation Scheme

Annotation of the three corpora on null pronoun reference is done in the following manner. For each null pronoun identified within a text, all noun phrases with a same referent, in either a full NP form, a pronominal or even the null form with a same referent, are identified across the text and given a unique numerical index. All non-anaphoric types and propositional type null pronouns are marked with their category. For reasons explained in the previous sections, deictic and generic null pronouns can also bear numerical indices, further indicating a coreference relation between them and other NPs. For anaphoric propositional null pronouns, their potential antecedents are not an NP but rather a clausal unit, which were identified and marked with an index.

The following summarizes the coding scheme used in annotation:

(15)   **A coding system for null pronoun reference**

  (i) NP-anaphoric relations are indicated by a numerical index

  (ii) All other categories are marked with an alphabetic index:
    - deictic speaker: **i**
    - deictic hearer: **y**
    - deictic speaker and hearer: **w**
    - indefinite generic: **g**
    - indefinite specific: **s**
    - situational: **x**
    - anaphoric propositional: **p**

Example (16) shows a sample piece of coreference annotation using the scheme. The file is pre-annotated with syntactic information in Ko-

rean Treebank style. For each null pronoun, the coreference indices are given on its phrasal projection node. Noun phrases which are in a coreference relation with any other null pronouns are also given a numerical index.

(16) ;;A;06:1: 자, 김 특무장, 앉으라.
              All right, 1SG KIM, sit down.
     (S (INTJ 자/IJ)
        ,/SCM
        (NP-VOC#2 김/NPR 특무장/NNC)
        ,/SCM
        (S (NP-SBJ#y2 *pro*)
           (VP 앉/VV+으라/EFN))
        ./SFN)

   ;;A;06:2: 뭐 좀 먹었나?
              Did you eat something?
     (S (NP-SBJ#y2 *pro*)
        (VP (NP-OBJ#3 뭐/NPN)
           (VP (ADVP 좀/ADV)
              (VP 먹/VV+었/EPF+나/EFN)))
        ?/SFN)

   ;;B;06:3: 예, 먹었습니다.
              Yes, I ate.
     (S (INTJ 예/IJ)
        ,/SCM
        (S (NP-SBJ#i2 *pro*)
           (VP (NP-OBJ#3 *pro*)
              먹/VV+었/EPF+습니다/EFN))
        ./SFN)

5   A Reference Resolution System: a Preview

In most approaches to the resolution of English pronouns, which usually concern the third person pronouns only, a successful resolution means locating an antecedent NP in the surrounding text. However, in the case of Korean null pronouns, the diverse nature of of the null pronouns makes it necessary for a reference resolution system to deal with two disparate tasks: classification of the null pronoun and identification of the referent. Moreover, each type of null pronoun has a different definition of successful resolution.

For some types, the resolution stops at classification: once a null pronoun is classified as an **x**-type, i.e. an expletive null pronoun, there is no further step to be taken. Other types involve identification of the referent in some way. Deictic null pronouns need to be resolved to an entity present in the discourse scene, in most cases a discourse participant. Anaphoric pronouns can be treated in the same manner as English pronouns — by resolving to a coreferential noun phrase in the text, if it is NP-anaphoric;

or by finding a clausal unit which it refers to if it is propositional.

The cases of indefinite null pronouns are less clear. From its definition, it is unlikely to be able to recover the referent of specific indefinites from the environment. One can also argue that the referent of generic indefinites need not be further resolved, as they can be left as vague as required. However, the referent of a generic indefinite is often restricted by the domain of discourse, which is often signified in a nearby context. For now, we will adopt a simple approach that stipulates that no further resolution is needed for these types, leaving further discussion to future studies. The resolution strategies for each type are summarized in (17):

(17)   **Resolution strategy per null pronoun type**

   i. NP anaphoric: find an NP antecedent in surrounding text

   ii. propositional anaphoric: find a clausal antecedent in surrounding text

   iii. deictic: find the referent among discourse entities, usually discourse participants

   iv. expletive: no further resolution

   v. specific indefinite: no further resolution

   vi. generic indefinite: no further resolution

We have seen earlier in sections 4.2.1 and 4.2.2 that deictic and generic null pronouns can be non-anaphoric and anaphoric at the same time. For such cases, a successful resolution can occur if such a pronoun is successfully resolved in either one of the two categories. Take, for example, the null pronoun subject of the third sentence in example (16). We can have a resolution strategy which views its resolution successful (a) if the pronoun is resolved to i (the speaker), *or* (b) if the pronoun is resolved to the subject null pronoun of preceding sentence 2, which is coindexed with the pronoun.

6   Conclusion

In this paper, I reviewed the cases of dropped arguments in Korean found in naturally occurring data, and proposed a classification system and a corresponding annotation scheme. They are aimed to serve as a basis in creating an annotated corpus, which can be used in developing a reference resolution system for Korean null

pronouns. In particular, a large-scale reference-annotated corpus built to conform to the syntactic annotation standards of the Korean Treebank will be a valuable resource in developing a resolution system which employs various statistical techniques.

References

Marilyn W. Friedman Brennan, Susan E and Carl J. Pollard. 1987. A centering approach to pronouns. In *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*, pages 155–162. ACL.

Chung-hye Han and Na-Rae Han. 2001. Part of speech tagging guidelines for penn korean treebank. Technical report, IRCS, University of Pennsylvania.

Na-Rae Han Han, Chung-hye and Eon-Suk Ko. 2001. Bracketing guidelines for penn korean treebank. Technical report, IRCS, University of Pennsylvania.

Peter A. Heeman and Graeme Hirst. 1995. Collaborating on referring expressions. *Computational Linguistics*, 21(3).

Megumi Kameyama. 1985. *Zero Anaphora: The Case of Japanese*. Ph.D. thesis, Stanford University. Available from University Microfilms at the University of Michigan.

Megumi Kameyama. 1997a. Intrasentential centering: A case study. In A. Joshi Walker, M. and E. Prince, editors, *Centering Theory in Discourse*. Oxford University Press.

Megumi Kameyama. 1997b. Recognizing referential links: An information extraction perspective. Technical report, AI Center, SRI International.

Shalom Lappin and Herbert J. Leass. 1994. An algorithm for pronominal anaphora. *Computational Linguistics*, 20(4).

Dania Egedi Park, Hyun-Seok and Martha Palmer. 1994. Recovering empty arguments in korean. In *Proceedings of the Joint Conference of the 8th Asian Conference on Language, Information, and Computation and the 2nd Pacific Asia Conference on Formal and Computational Linguistics*.

Michael Sturbe and Udo Hahn. 1999. Functional centering– grounding referential coherence in information structure. *Computational Linguistics*, 25(309–344).

Masayo Iida Walker, Marilyn and Sharon Cote. 1994. Japanese discourse and the process of centering. *Computational Linguistics*, 20(2):193–232.