Zero Object Resolution in Korean

Arum Park
Dept. of German Linguistics
& Literature, Sungkyunkwan
University /
25-2, Sungkyunkwan-Ro,
Jongno-Gu,
Seoul, Korea
remin2@skku.edu

Seunghee Lim
Dept. of German Linguistics
& Literature, Sungkyunkwan
University /
25-2, Sungkyunkwan-Ro,
Jongno-Gu,
Seoul, Korea
rusilen21@skku.edu

Munpyo Hong*
Dept. of German Linguistics
& Literature, Sungkyunkwan
University /
25-2, Sungkyunkwan-Ro,
Jongno-Gu,
Seoul, Korea
skkhmp@skku.edu

Abstract

Korean is one of the well-known 'pro-drop' languages. When translating Korean zero object into languages in which objects have to be overtly expressed, the resolution of zero object is crucial. This paper proposes a machine learning method to resolve Korean zero object. We proposed 8 linguistically motivated features for ML (Machine Learning). Our approach has been implemented with WEKA 3.6.10 and evaluated by using 10-fold cross validation method. The accuracy of the proposed method reached 73.37%.

1   Introduction

Korean is one of the so-called pro-drop languages. Certain pronouns may be omitted and the omitted pronouns are often called zero pronouns. This kind of pronoun also occurs in other languages, such as Japanese or Spanish. The omitted pronouns in Korean can appear in subject and object position, whereas in Spanish or Italian, they can appear only in subject position. A zero subject is the most frequent type of anaphoric expressions in Korean. Hong (2000) reported that about 57% of the pronouns are a zero subject pronoun in pronoun occurrences in Korean spoken text.

Zero object is the second most frequent zero pronoun type in Korean spoken text. Despite of the frequent use of zero objects, most of the previous works do not deal with the zero objects in Korean. In this work, we focus on Korean spoken texts, since zero pronouns occur more frequently than in a written text. Ryu (2001) showed that a zero pronoun rarely appears in written texts when it is compared with spoken texts in Korean. For this reason, we conduct a study for Korean zero object in Korean spoken text and try to find the linguistic clues for the zero object resolution.

In the context of machine translation, the resolution of Korean zero objects could be one of the most important issues in order to translate them into the target language like English and German. One of the reasons that zero objects in Korean is a problem in MT is that the omitted objects in Korean have to be translated into overt objects in target languages. Unfortunately, the majority of MT systems do not deal with this problem, because most of the current commercial MT systems do not treat the linguistic phenomena that go beyond a sentence level. To illustrate this issue, let's take a look at the following example (1).

(1) MT results from Korean(a) to German(b)

(a) Korean

A: 여권ᵢ을   분실했습니다.

---

* Corresponding author

a passport-OBJ   lost
(yekwenul   punsilhaysssupnita.)

"I lost a passport."

B: ø_i 다시   발급받으셔야 합니다.
   again   have to issue
   (tasi   palkuppatusyeya hapnita.)

"You have to issue a passport again."

(b1) German - Systran translator

A: Verlor den Pass.
B: Fragen wiederholt.

(b2) German - Google translator

A: Ich habe meinen Pass verloren.
B: Wir müssen neu aufgelegt zu werden.

The omitted object is represented by the symbol ø. In this example, the Korean object is not overtly expressed in the sentence B and it refers to '여권'(yekwen, "Passport") in sentence A. To translate the omitted object into German correctly, the gender and number of the antecedent 'yekwen' has to be considered. Since the morphological information of 'yekwen' is 'masculine' and 'singular' in German, the omitted object has to be translated as 'ihn' considering its case. However, the object of the sentence B is not translated in German in either MT system. Then, the results would be ungrammatical in German. Therefore, the resolution of Korean zero objects is crucial in MT systems with Korean as a source language, when translating them into languages in which objects have to be overtly expressed.

In section 2 we present the related works about anaphora resolution and their limitation. Section 3 explains zero objects phenomenon in Korean. We suggest the machine learning (ML) method for Korean zero object resolution and propose 8 features for ML method in section 4. In addition, the effect of using ML is evaluated. Finally, the conclusion is presented in section 5.

2 Related Works

Zero pronouns have already been studied in other languages, such as Japanese (e.g. Nakaiwa and Shirai, 1996; Okumura and Tamura, 1996) and Spanish (Park and Hong, 2014; Palomar et al., 2001; Ferrández and Peral, 2000). These studies are based on the researches about anaphora resolution. It has been a wide-open research field since 1970 focusing on English. Regardless of languages, similar strategies for anaphora resolution have been applied. Using linguistic information is the most representative technique; constraints and preferences methods are distinguished in the related works (Baldwin, 1997; Lappin and Leass, 1994; Carbonell and Brown, 1988).

Constraints discard possible antecedents and are considered as absolute criteria. Preferences being proposed as heuristic rules tend to be relative. After applying constraints, if there are still unresolved candidate antecedents, preferences set priorities among candidate antecedents. Nakaiwa and Shirai (1996) focus on semantic and pragmatic constraints such as cases, modal expressions, verbal semantic attributes and conjunctions in order to determine the reference of Japanese zero pronouns. However, they proposed constraints focusing on zero subjects mainly. Therefore, it is hard to apply their approach on zero object resolution.

Centering theory (Grosz et al., 1995) is one of the approaches using heuristic rules. It is claimed that certain entities mentioned in an utterance are more central than others, and this property has been applied to determine the antecedent of the anaphor. Walker et al. (1994) applied the centering model on zero pronoun resolution in Japanese. Roh and Lee (2003) proposed a generation algorithm of zero pronouns using a Cost-based Centering Model which considers the inference cost. It is known that the most salient element of the given discourse is likely to be realized as a zero pronoun. We take this into account in selecting the features for ML.

Current anaphora resolution methods rely mainly on constraint and preference heuristics, which employ morpho-syntactic information or shallow semantic analysis. These methods are a deterministic algorithm which always produces the same output in a given particular condition. However, even if the condition is applied, the

output can be wrong. ML methods which are a non-deterministic algorithm have been studied on anaphora resolution (Connolly et al., 1994; Paul et al., 1999). Since ML learns from data and makes predictions of the most likely candidate on the data, it can overcome the limitation of the deterministic method.

Park and Hong (2014) proposed a hybrid approach to resolve Spanish zero subjects that integrates heuristic rules and ML in the context of Spanish to Korean MT. Since Spanish zero subjects can be restored from the verb ending, they use morphological flections for verbs. After that, ML is utilized for some ambiguous cases. Unlike this work, our work deals with Korean zero object. Morphological information cannot be utilized for Korean because of the difference of the two languages. For this reason, we use ML method alone to determine the antecedent of the zero objects in spoken Korean.

3   Zero object phenomenon in Korean

A prominent phenomenon in Korean is the prevalence of zero pronouns. Unlike English, zero pronouns occur very frequently in Korean. In Korean, a zero subject is the most frequent type of anaphoric expression. The second most frequent type is zero objects, especially when the direct object is omitted. According to Hong (2000), when the direct object does not occur in spoken Korean, the rate becomes 19.1%.

To resolve zero object, centering theory can be utilized. The centering theory endeavors to identify the antecedent of a (zero) pronoun using the idea of the most central entity that a sentence concerns which tends to be expressed by a (zero) pronoun. There are some studies attempting to apply the centering theory to anaphora resolution (Choi and Lee, 1999; Hong, 2000; Hong, 2011). The forward looking center rankings for Korean are defined differently in the studies. Following Hong (2011)'s discussion, we accept the forward looking center ranking for Korean as follows:

· Forward looking center ranking for Korean (Hong, 2011)

TOPIC > SUBJECT > OBJECT > ADVERB > OTHERS

Given the hierarchy of the forward looking center ranking, a zero object can be interpreted as the topic which is the most salient discourse entity. The topic of the sentence contributes to discourse salience and maintains discourse coherence by preferring the CONTINUE transition state. The topic of the sentence can be detected easily in Korean using the topic markers '은'(eun), '는'(nun) and delimiters such as '도'(to), '만'(man). Therefore, it is likely a candidate antecedent is the antecedent of the zero object if the candidate has one of the topic markers or delimiters. We can see some examples in the following table.

Speaker	Korean dialogue
A	음식 주문$_1$을 ordering a food-OBJ (umsik cumunul 어떻게 하는 거죠? how can ettehkey hanun kecyo?) "How can I order a food?"
B	저 기계$_2$에서 메뉴$_3$를 that machine-ADV menu-OBJ (ce kikyeyeyse menyulul 선택한 후 식권$_4$을 after selecting a meal ticket-OBJ senthaykhan hu sikkwenul 뽑으세요. buy ppopuseyyo.) "You can buy a meal ticket after selecting menu from that vending machine."
A	아침 식사$_5$는 11 시까지만 breakfast-TOP until 11 o'clock (achim siksanun 11sikkaciman 되는 건가요? is possible to toynun kenkayo?) "Is it possible to have breakfast until 11 o'clock?"
B	네, 지금 Yes, right now (ney, cikum 정확히 11 시니까

because it's 11 o'clock
cenghwakhi 11sinikka

ø 원하신다면 ø 해 드릴게요.
  if you want   can serve
wenhasintamyen hay tulilkeyyo.)

"Yes, if you want, I can serve you a breakfast because it's 11 o'clock right now."

Speaker	
A	감사합니다. thank you (kamsahapnita.) "Thank you."

Table 1 dialogue example including topic markers

In table 1, the dialogue's omitted object is represented by the symbol ø. There are 5 candidate antecedents: 1. '음식 주문' (umsik cumun, "order"), 2. '기계' (kikyey, "machine"), 3. '메뉴' (menyu, "menu"), 4. '식권' (sikkwen, "meal ticket"), 5. '아침 식사' (achim siksa, "breakfast"). The first, third and fourth candidates occur in the object position, the second candidate is in the adverb position, and the last candidate has a topic marker 'nun'. Since the topic is the highest position of the forward looking center ranking, the last candidate is likely to be the antecedent of the zero object.

Speaker	Korean dialogue
A	무슨 일 있으신가요? what happened (musun il issusinkayo?) "What happened?"
B	화장실 $_1$ 에 휴지 $_2$ 가 in toilet-ADV toilet tissue-TOP (hwacangsiley hucika 없어서요. is not eppseseyo.) "There is no toilet paper in the restroom."
A	잠시만요. wait a minute (camsimanyo.)

"Wait a minute."

Speaker	
A	제가 ø 꺼내 드릴게요. I-SUBJ will give (ceyka kkenay tulilkeyo.) "I'll give it to you."
B	알겠습니다. all right (alkeysssupnita.) "All right."

Table 2 dialogue example of syntactic function

In this case, there are 2 candidate antecedents for the zero object which are '화장실' (hwacangsil, "restroom") and '휴지' (huci, "toilet tissue"). Since the second candidate has the higher raking in the forward looking center ranking than the first one, it can be the antecedent of the zero object, and this is actually the case. As the syntactic function of the candidate antecedents is important to resolve Korean zero object, we utilize this information.

Property-sharing constraint can also be the clue to resolve Korean zero object. Kameyama (1986) suggested property-sharing constraint of zero pronouns in Japanese. She claimed that if a zero pronoun is the subject of a verb, the antecedent is perhaps a subject in the antecedent's sentence. In addition, if a zero pronoun is an object, the antecedent is highly likely an object. Since Japanese and Korean share many of their linguistic properties, we can apply this constraint to resolve Korean zero object. The following table shows an example of property-sharing constraint.

Speaker	Korean dialogue
A	제 애완동물 $_1$ 을 my pet-OBJ (cey aywantongmulul 잃어버렸습니다. have lost ilhepelyesssupnita.) "I have lost my pet."

B	ø$_1$ 어디서 잃어버리셨나요?
	  where have lost
	  (etise ilhepelisyessnayo?)

	"Where have you lost her?"
A	객실 $_2$ 에서 나오면서 room-ADV when come out of (kayksileyse naomyense 사라졌습니다. disappeared salacyesssupnita.) "She disappeared when I came out of the room."
B	모두 찾아보셨나요? everywhere have been looking for (motu chacaposyessnayo?) "Have you been looking for her everywhere?"
A	관리실 $_3$ 빼고는 except management office-ADV (kwanlisil ppaykonun 다 찾아봤습니다. all have been looking for ta chacapwasssupnita.) "I have been looking for her everywhere except for the management office."
B	그럼 저희 직원들 $_4$ 이 then our staff-SUBJ (kulem cehuy cikwentuli ø$_2$ 찾아보겠습니다. will look for chacapokeyssupnita.) "Then our staff will look for her there."

Table 3 dialogue example
of property-sharing constraint

In the above examples, there are 4 candidate antecedents for the second zero object. From the candidate antecedents, the first candidate '애완동물' (aywantongmul, "pet") is the antecedent of the zero object. Even if there is an entity which has ranked higher in the forward looking center ranking, the farthest candidate which is in the object position as the zero object is the antecedent of the second zero object. This is one of examples showing the property-sharing constraint. Therefore, the parallelism of syntactic function between a zero object and a candidate antecedent can be utilized.

The semantic relation between the predicate of a zero object and a candidate antecedent is another property of Korean zero object. When the semantic of the predicates correlates between a zero object and a candidate antecedent, the candidate preferred to be the antecedent of the zero object.

Speaker	Korean dialogue
A	어디 가시나요? where are you going (eti kasinayo?) "Where are you going?"
B	콘서트 $_1$ 를 관람하러 갑니다. concert-OBJ go to watch (khonsethulul kwanlamhale kapnita.) "I go to (watch) the concert."
A	이미 콘서트 광장 $_2$ 은 already the concert hall-TOP (imi khonsethu kwangcangun 사람 $_3$ 이 많아서 people-SUBJ many salami manhase 들어가실 수 없습니다. can't enter to tulekasil su epssupnita.) "You can't enter the concert hall because there are already too many people."
B	저도 좀 ø$_1$ 관람하고 싶습니다. I-SUBJ want to watch (ceto com kwanlamhako sipsupnita.) "I want to watch the concert, too."
A	좀 일찍 오셨더라면 ø$_2$ earliy if you have come (com ilccik osyesstelamyen 볼 수 있었을 겁니다. could see pol su issessul kepnita.) "If you had come earlier, then you could have seen the concert."

Table 4 dialogue example (1)
including semantic relation of predicates

Table 4 shows the importance of utilizing semantic between the predicate of the candidate

antecedents and the zero objects. In this case, there are three candidate antecedents: 1. '콘서트' (khonsethu, "concert"), 2. '콘서트 광장' (khonsethu kwangcang, "concert hall"), 3. '사람' (salam, "people"). Even though the last two candidates have the higher syntactic function than the first one, the first candidate 'khonsethu' is the antecedent of the zero objects, because the first candidate and the first zero object have the same predicate '관람하다' (kwanlamhata, "watch").

The antecedent of the second zero object is the first candidate antecedent. The meaning of predicates 'kwanlamhata' and '보다' (pota, "see") is similar. Therefore, we consider the semantic of predicates between a candidate antecedent and a zero object as one of the important indicators to resolve Korean zero object. The opposite meaning of predicates can also be the clue in the following table example.

Speaker	Korean dialogue
A	죄송합니다, 기내 $_1$ 에서는 sorry, in flight-ADV (coysonghapnita, kinayeysenun 휴대전화 $_2$ 를 the phone-OBJ hyutaycen-hwalul 꺼 주셔야 합니다. have to turn off kke cusyeya hapnita.) "Sorry, you have to turn off the cell-phone during the flight."
B	그런가요, 알겠습니다. all right (kulenkayo, alkeyssssupnita.) "All right."
A	비행기 $_3$ 가 완전히 flight-SUBJ completely (pihayngkika wancenhi 이륙한 후에는 ø takes off after ilyukhan hueynun 키셔도 됩니다. can turn on kisyeto toypnita.)

"After the machine completely takes off, you can turn on the cell-phone."

Table 5 dialogue example (2) including semantic relation of predicates

The above table dialogue has 3 candidate antecedents. From these candidates, the second candidate is the antecedent of the zero object. The antecedent and the zero object have predicates '끄다' (kkuta, "turn off") and '켜다' (khyeta, "turn on"), respectively. The predicates are in an antonym relation which is much more important than the syntactic function. This is one of the reasons why we consider the semantic of predicates between the candidate antecedents and the zero object as a clue for Korean zero object resolution.

Like WordNet in English, Korean dictionary can give information whether predicates are identical or different or opposite in meaning. Sejong electronic dictionary[1] and KorLex[2] are one of the Korean dictionaries which are available to extract information. Sejong electronic dictionary includes information about word meaning relation such as synonyms and antonyms. KorLex is another dictionary based on WordNet. This dictionary is constructed by translating WordNet and then modifying for Korean. Using these dictionaries, meaning relation of predicates between the candidate antecedent and the zero object can be automatically compared.

4   Experiments

4.1 Feature sets

In this paper, we employ a machine learning method to deal with the zero objects phenomenon. In order to apply a machine learning method, 8 features are proposed as presented in table 6. The following table explains the functions of each feature with their value.

	Feature	Value
f1	Syntactic function of candidate antecedent	*top*, *sub*, *obj*, *adv*, *comp*,

[1] https://ithub.korean.go.kr/
[2] http://klpl.re.pusan.ac.kr/

		poss
f2	Parallelism of syntactic function	*para*, *diff*
f3	Semantic relation between predicates	*sim*, *same*, *oppo*, *diff*, *loc*[3]
f4	Sentence distance	*loc*, $0, \ldots n$
f5	Sentence distance based on Speaker of zero object	$-n \ldots 0 \ldots n$
f6	Headedness of candidate antecedent	*head*, *not*
f7	The most salient candidate antecedent	*1*, *0*
f8	Gold referential relation	*yes*, *no*

Table 6 Features for ML

In this paragraph we explain feature 1 and 2 in detail. Among feature 1 values, if the value *top* is assigned to an entity, it has given preferential treatment to make them antecedent. In Korean, the markers 'eun', 'nun', 'to', 'man' show which entity is a topic or delimiter.

Feature 2 encodes whether the syntactic function of the candidate antecedent and the zero object are equal. When the syntactic functions are different, the value is *diff*. When a candidate antecedent has the same syntactic function as the zero object, it is more likely to be an antecedent. This is one of the reasons why we introduce feature 2.

Feature 3 represents the semantic relation of predicates between the candidate antecedent and the zero object. In Korean, it tends to be correlated for meaning between the predicate of the candidate antecedent and the zero object. The values of feature 3 encode this tendency.

Feature 4 is about the sentence distance between the zero object and the candidate antecedent. The value *loc* indicates that the pronoun and the potential antecedent are in the same local clause. When the pronoun and the potential antecedent occur in the same sentence but not in the same clause, the value becomes *0*. Higher values indicate larger distances. Candidates, which appear on the first sentence from the complex sentence or the sentence before the current sentence, are more preferred to be the antecedent than the other candidates.

---

[3] If the candidate antecedent and the zero object occur in the same clause, the value *loc* is assigned.

Since we deal with spoken text form, there is a chance to have some difficulties in applying the methods in the previous studies focusing on written sentences. Because of this reason, we introduce feature 5 which reflect the properties of spoken texts. Feature 5 encodes the sentence distance between the zero object and the candidate antecedent based on the speaker of the zero object. We assumed that considering the sentence distance based on the speaker of the zero object can reflect the original aim to introduce sentence distance for one of the features for ML. Unlike feature 4, the value of feature 5 can be negative according to the consistency of the speaker between the zero object and the candidate antecedent. If the speaker of them is not the same, then the value of this feature will be negative.

Feature 6 represents the headedness of the candidate antecedent. When a candidate antecedent NP occurs in the head of the NP, then it can be considered as the likeliest antecedent than the candidates which are not the head of the NP.

Feature 7 is based on the framework of centering theory. In the previous literature, it is argued that a salient entity recoverable by inference from the context is frequently omitted (Walker et al., 1994; Iida, 1998; Hong, 2000). Therefore, we utilize the forward looking center ranking for Korean, assuming the most salient candidate antecedent which is marked as a value *1* is likely to be the antecedent of the zero object.

Feature 8 encodes the gold referential relation between the candidate antecedent and the zero object. It takes the value *yes* if the noun phrase is in fact an antecedent of the zero object, and *no* if it is not.

4.2 Experiment

To evaluate the effect of machine learning method, we use 'WEKA' system (3.6.10 version). Since SVM (Support Vector Machine) algorithm has shown good performance in various tasks in NLP (Kudo and Matsumoto, 2001; Isozaki and Kazawa, 2002), SVM algorithm is selected for evaluation. We collect spoken texts about tourism containing Korean zero objects. 1123 coreferential pairs are extracted from the corpus; 308 pairs are positive, and 824 pairs are negative.

The experiment result was obtained by splitting the data set in ten parts equally for 10-fold cross validation. Each training set contains 90% of the total number of pairs, and the remaining 10% are assigned to the test sets. Using 8 features, we have found 73.37% of accuracy. It may not be quite fair comparison if we compare our result with the results of other studies on Korean written texts. Therefore, we set a baseline by choosing the most salient candidate in the discourse according to the forward looking center ranking in Hong (2011) for comparison. As shown in Table 7, the proposed method can improve the accuracy up to 62% which is above the baseline.

	Baseline	Experiment	remark
Accuracy	11.66%	73.37%	61.71% improved

Table 7 The result of experiment

Ranking	Feature	
1	f4	Sentence distance
2	f3	Semantic relation between predicates
3	f7	The most salient candidate antecedent
4	f5	Sentence distance based on Speaker of zero object
5	f1	Syntactic function of candidate antecedent
6	f2	Parallelism of syntactic function
7	f6	Headedness of candidate antecedent

Table 8 The ranking of features

Table 8 shows the ranking of the features selected by 'InfoGainAttribute Evaluator'. As table 8 shows, feature 5 has ranked top. Sentence distance is commonly utilized in other works on anaphora resolution, because candidate antecedents from the previous clause or sentence are preferred. McEnery et al. (1997) examined the distance of pronouns and their antecedent, and concluded that the antecedents of pronouns do exhibit clear patterns of distribution. The result of the feature ranking reflects the importance of the role of sentence distance.

As the table 8 shows, feature 3 ranked second. In the previous works, subcategorization information is utilized for semantic constraints. For example, if a zero is the subject of 'eat,' the antecedent is probably a person or an animal, and so on. However, feature 3, which is different from the semantic constraints from the other studies, is first introduced in this work for zero object resolution. From the result, we can assume that this feature plays very important role to zero object resolution in Korean.

We can also verify that the centering theory is crucial to resolve Korean zero object. According to the theory, there is a tendency that the most salient candidate antecedent is realized as a zero pronoun. Since feature 7 reflects this property and this feature ranked third among the features we proposed, the tendency is proven to be significant for zero object resolution.

5 Conclusion

In this paper, we proposed a ML method to resolve Korean zero object in spoken texts. Determining an antecedent of a zero object is crucial in developing MT systems with Korean as a source language. In case of translating Korean into target languages like English and German, the omitted object has to be resolved in order to generate overt objects in target languages. In order to utilize ML, 8 features were suggested for Korean zero object resolution. An experiment was conducted to test the feasibility for our method. The accuracy was 73.37% which was higher than the baseline, when 8 features were used for the ML. Currently, we are increasing the size of the training corpus, and are planning to validate our model in depth with the new training corpus.

Acknowledgments

References

Baldwin, B. 1997. CogNAC: high precision coreference with limited knowledge and linguistic resources. In Proceedings of a Workshop on Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts. Association for Computational Linguistics, pp. 38-45.

Carbonell, J. G., & Brown, R. D. 1988. Anaphora resolution: a multi-strategy approach. In Proceedings of the 12th conference on Computational linguistics, Volume 1. Association for Computational Linguistics, pp. 96-101.

Choi, J. & Lee, M. 1999. Focus – Formal Semantics and description of Korean. Seoul: Hanshin publishing company. (in Korean)

Connolly, D., Burger, J. D., & Day, D. S. 1997. A machine learning approach to anaphoric reference. In New Methods in Language Processing. pp. 133-144.

Ferrández, A., & Peral, J. 2000. A computational approach to zero-pronouns in Spanish. In Proceedings of the 38th Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics. pp. 166-172.

Grosz, B. J., Weinstein, S., & Joshi, A. K. 1995. Centering: A framework for modeling the local coherence of discourse. Computational linguistics, 21(2), pp. 203-225.

Hong, M. 2000. Centering Theory and Argument Deletion in Spoken Korean. The Korean Journal Cognitive Science. Vol. 11(1). pp. 9-24. (in Korean)

Hong, M. 2002. A review on zero anaphora resolution theories in Korean. Studies in Modern Grammar, 29, pp. 167-186.

Hong, M. 2011. Zero subject Resolution in Korean for machine translation into German. German linguistics, 24, pp. 417-439. (in Korean)

Iida, M. 1998. Discourse coherence and shifting centers in Japanese texts. Centering theory in discourse, pp. 161-180.

Isozaki, H., & Kazawa, H. 2002. Efficient support vector classifiers for named entity recognition. In Proceedings of the 19th international conference on Computational linguistics, Volume 1. Association for Computational Linguistics. pp. 168-184.

Kameyama, M. 1986. A property-sharing constraint in centering. In Proceedings of the 24th annual meeting on Association for Computational Linguistics.

Association for Computational Linguistics, pp. 200-206.

Kim, L. K. 2010. Korean Honorific Agreement too Guides Null Argument Resolution: Evidence from an Offline Study. University of Pennsylvania Working Papers in Linguistics, 16(1), 12. pp. 101-108.

Kudo, T., & Matsumoto, Y. 2001. Chunking with support vector machines. In Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies. Association for Computational Linguistics, pp. 1-8.

Lappin, S., & Leass, H. J. 1994. An algorithm for pronominal anaphora resolution. Computational linguistics, 20(4), pp. 535-561.

Lee, D. 2002. Discourse Representation Methods and Korean Dialogue. In Proceedings of the 2002 Winter Linguistic Society of Korea Conference. Seoul National University, pp. 88-104.

Okumura, M., & Tamura, K. 1996. Zero pronoun resolution in Japanese discourse based on centering theory. In Proceedings of the 16th conference on Computational linguistics, Volume 2. Association for Computational Linguistics. pp. 871-876.

McEnery et al. 1997. Corpus annotation and reference resolution. In Proceedings of the ACL Workshop on Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts, pp. 67-74.

Nakaiwa, H., & Shirai, S. 1996. Anaphora resolution of Japanese zero pronouns with deictic reference. In Proceedings of the 16th conference on Computational linguistics, Volume 2. Association for Computational Linguistics. pp. 812-817.

Palomar, M., Ferrández, A., Moreno, L., Martínez-Barco, P., Peral, J., Saiz-Noeda, M., & Muñoz, R. 2001. An algorithm for anaphora resolution in Spanish texts. Computational Linguistics, 27(4), pp. 545-567.

Park, A., & Hong, M. 2014. Hybrid Approach to Zero Subject Resolution for multilingual MT-Spanish-to-Korean Cases. In Proceedings of the 28th Pacific Asia Conference On Language Information and Computing. pp.254-261.

Paul, M., Yamamoto, K., & Sumita, E. 1999. Corpus-based anaphora resolution towards antecedent preference. In Proceedings of the Workshop on Coreference and its Applications. Association for Computational Linguistics. pp. 47-52.

Roh, J. E., & Lee, J. H. 2003. An empirical study for generating zero pronoun in Korean based on Cost-

based Centering Model. In Proceedings of Australasian Language Technology Association, pp. 90-97.

Ryu, B. R. 2001. Centering and zero anaphora in the Korean discourse. Seoul National University, MS Thesis.

Walker, M., Cote, S., & Iida, M. 1994. Japanese discourse and the process of centering. In Proceedings of Computational linguistics, 20(2), pp. 193-232.