Feasibility Study for Ellipsis Resolution in Dialogues by Machine-Learning Technique

YAMAMOTO Kazuhide and SUMITA Eiichiro
ATR Interpreting Telecommunications Research Laboratories
E-mail: yamamoto@itl.atr.co.jp

Abstract

A method for resolving the ellipses that appear in Japanese dialogues is proposed. This method resolves not only the subject ellipsis, but also those in object and other grammatical cases. In this approach, a machine-learning algorithm is used to select the attributes necessary for a resolution. A decision tree is built, and used as the actual ellipsis resolver. The results of blind tests have shown that the proposed method was able to provide a resolution accuracy of 91.7% for indirect objects, and 78.7% for subjects with a verb predicate. By investigating the decision tree we found that topic-dependent attributes are necessary to obtain high performance resolution, and that indispensable attributes vary according to the grammatical case. The problem of data size relative to decision-tree training is also discussed.

1 Introduction

In machine translation systems, it is necessary to resolve ellipses when the source language doesn't express the subject or other grammatical cases and the target must express it. The problem of ellipsis resolution is also troublesome in information extraction and other natural language processing fields.

Several approaches have been proposed to resolve ellipses, which consist of endophoric (intrasentential or anaphoric) ellipses and exophoric (or extrasentential) ellipses. One of the major approaches for endophoric ellipsis in theoretical basis utilizes the centering theory. However, its application to complex sentences has not been established because most studies have only investigated its effectiveness with successive simple sentences.

Several studies of this problem have been made using the empirical approach. Among them, Murata and Nagao (1997) proposed a scoring approach where each constraint is manually scored with an estimation of possibility, and the resolution is conducted by totaling the points each candidate receives. On the other hand, Nakaiwa and Shirai (1996) proposed a resolving algorithm for Japanese exophoric ellipses of written texts, utilizing semantic and pragmatic constraints. They claimed that 100% of the ellipses with exophoric referents could be resolved, but the experiment was a closed test with only a few samples. These approaches always require some effort to decide the scoring or the preference of provided constraints.

Aone and Bennett (1995) applied a machine-learning technique to anaphora resolution in written texts. They attempted endophoric ellipsis resolution as a part of anaphora resolution, with approximately 40% recall and 74% precision at best from 200 test samples. However, they were not concerned with exophoric ellipsis. In contrast, we applied a machine-learning approach to ellipsis resolution (Yamamoto et al., 1997). In this previous work we resolved the agent case ellipses in dialogue, with a limited topic, and performed with approximately 90% accuracy. This does not sufficiently determine the effectiveness of the decision tree, and the feasibility of this technique in resolving ellipses by each surface case is also unclear.

We propose a method to resolve the ellipses that appear in Japanese dialogues. This method resolves not only the subject ellipsis, but also the object and other grammatical cases. In this approach, a machine-learning algorithm is used to build a decision tree by selecting the necessary attributes, and the decision tree is used as the actual ellipsis resolver.

Another purpose of this paper is to discuss how effective the machine-learning approach is

in the problem of ellipsis resolution. In the following sections, we discuss topic-dependency in decision trees and compare the resolution effectiveness of each grammatical case. The problem of data size relative to the decision-tree training is also discussed.

In this paper, we assume that the detection of ellipses is performed by another module, such as a parser. We only considered ellipses that are commonly and clearly identified.

2 When to Resolve Ellipsis in MT ?

As described above, our major application for ellipsis resolution is in machine translation. In an MT process, there can be several approaches about the timing of ellipsis resolution: when analyzing the source language, when generating the target language, or at the same time as translating process. Among these candidates, most of the previous works with Japanese chose the source-language approach. For instance, Nakaiwa and Shirai (1996) attempted to resolve Japanese ellipsis in the source language analysis of J-to-E MT, despite utilizing target-dependent resolution candidates.

We originally thought that ellipsis resolution in the MT was a generation problem, namely a target-driven problem which utilizes some help, if necessary, of source-language information. This is because the problem is output-dependent and it relies on demands from a target language. In the J-to-Korean or J-to-Chinese MT, all or most of the ellipses that must be resolved in J-to-E are not necessary to resolve.

However, we adopted source-language policy in this paper, with the necessity that we consider a multi-lingual MT system TDMT (Furuse et al., 1995), that deals with both J-to-E and J-to-German MT. English and German grammar are not generally believed to be similar.

3 Ellipsis Resolution by Machine Learning

Since a huge text corpus has become widely available, the machine-learning approach has been utilized for some problems in natural language processing. The most popular touchstone in this field is the verbal case frame or the translation rules (Tanaka, 1994). Machine-learning algorithm has also been attempted to solve some

Table 1: Tagged Ellipsis Types

Tag	Meaning
$\langle 1sg \rangle$	first person, singular
$\langle 1pl \rangle$	first person, plural
$\langle 2sg \rangle$	second person, singular
$\langle 2pl \rangle$	second person, plural
$\langle g \rangle$	person(s) in general
$\langle a \rangle$	anaphoric

discourse processing problems, for example, in discourse segment boundaries or discourse cue words (Walker and Moore, 1997). This section describes a method to apply a decision-tree learning approach, which is one of the machine-learning approaches, to ellipsis resolution.

3.1 Ellipsis Tagging

In order to train and evaluate our ellipsis resolver, we tagged some ellipsis types to a dialogue corpus. The ellipsis types used to tag the corpus are shown in Table 1. Each ellipsis marker is tagged at the predicate. We made a distinction between first or second person and person(s) in general. Note that 'person(s) in general' refers to either an unidentified or an unspecified person or persons. In Far-Eastern languages such as Japanese, Korean, and Chinese, there is no grammatically obligatory case such as the subject in English. It is thus necessary to distinguish such ellipses.

We also made a tag '$\langle a \rangle$' which means the mentioned ellipsis is anaphoric; in case we need to refer back to the antecedent in the dialogue. In this paper we are not concerned with resolving the antecedent that such ellipses refer to, because it is necessary to have another module to deal with the context for resolving such endophoric ellipses, and the main target of this paper is the exophoric ellipses.

3.2 Learning Method

We used the $C4.5$ algorithm by Quinlan (1993), which is a well-known automatic classifier that produces a binary decision tree. Although it may be necessary to prune decision trees, no pruning is performed throughout this experiment, since we want to concentrate the discussion on the feasibility of machine learning. As shown in the experiment by Aone and Ben-

Table 2: Number of training attributes

Attributes	Num.
Content words (predicate)	100
Content words (case frame)	100
Func. words (case particle)	9
Func. words (conj. particle)	21
Func. words (auxiliary verb)	132
Func. words (other)	4
Exophoric information	1
Total	367

nett (1995), which attempted to discuss pruning effects on the decision tree, no more conclusions are expected other than a trade-off between recall and precision. We leave the details of decision-tree learning research to itself.

3.3 Training Attributes

The training attributes that we prepared for Japanese ellipsis resolution are listed in Table 2. The training attributes in the table are classified into the following three groups:

- Exophoric information:
  Speaker's social role.

- Topic-dependent information:
  Predicates and their semantic categories.

- Topic-independent information:
  Functional words which express tense, modality, etc.

There is one approach that only uses topic-independent information to resolve ellipses that appear in dialogues. However, we took the position that both topic-dependent and -independent information should have different knowledge. Thus, approaches utilizing only topic-independent knowledge must have a performance limit for developing an ellipsis resolution system. It is practical to seek an automatically trainable system that utilizes both types of knowledge.

The effective use of exophoric information, i.e., from the actual world, may perform well for resolving an ellipsis. Exophoric information consists of a lot of elements, such as the time, the place, the speaker, and the listener of the utterance. However, it is difficult to become aware of some of them, and some are rather difficult to prescribe. Thus we utilize one element, the speaker's social role, i.e., whether the speaker is the customer or the clerk. The reason for this is that it must be an influential attribute, and it is easy to detect in the actual world. Many of us would accept a real system such as a spoken-language translation system that detects speech with independent microphones.

It is generally agreed that attributes to resolve ellipses should be different in each case. Thus although we have to prepare them on a case by case basis, we trained a resolver with the same attributes.

Because we must deal with the noisy input that appears in real applications, the training attributes, other than the speaker's social role, are questioned on a morphological basis. We give each attribute its positional information, i.e., search space of morphemes from the target predicate. Positional information can be one of five kinds: before, at the latest, here, next, and afterward. For example, a case particle is given the position of 'before', the search position of a prefix 'o-' or 'go-' is the 'latest', and an auxiliary verb is 'after' the predicate. The attributes of predicates, and their semantic categories are placed in 'here'.

For predicate semantics, we utilized the top two layers of *Kadokawa Ruigo Shin-Jiten*, a three-layered hierarchical Japanese thesaurus.

4 Discussion

In this section we discuss the feasibility of the ellipsis resolver via a decision tree in detail from three points of view: the amount of training data, the topic dependency, and the case difference. The first two are discussed against 'ga(v.)' case (see subsection 4.3).

We used F-measures metrics to evaluate the performance of ellipsis resolution. The F-measure is calculated by using recall and precision:

$$F = \frac{2 \times P \times R}{P + R} \qquad (1)$$

where P is precision and R is recall. In this paper, F-measure is described with a percentage (%).

1430

Table 3: Training size and performance

Dial.	Samp.	$\langle 1sg \rangle$	$\langle 2sg \rangle$	$\langle a \rangle$	Total
25	463	71.0	55.6	66.2	59.0
50	863	76.4	69.7	71.5	67.2
100	1710	82.1	76.4	77.0	73.2
200	3448	85.1	79.8	79.7	76.7
400	6906	84.7	81.1	82.0	78.7

4.1 Amount of Training Data

We trained decision trees with a varied number of training dialogues, namely 25, 50, 100, 200 and 400 dialogues, each of which included a smaller set of training dialogues. The experiment was done with 100 test dialogues (1685 subject ellipses), and none were included in the training dialogues.

Table 3 indicates the training size and performance calculated by F-measure. This illustrates that the performance improves as the training size increases in all types of ellipses. Although it is not shown in the table, we note that the results in both recall and precision improve continuously as well as those in F-measure.

The performance difference of all ellipsis types by training size is also plotted in Figure 1 on a semi-logarithmic scale. It is interesting to see from the figures that the rate of improvement gradually decelerates and that some of the ellipsis types seem to have practically stopped improving at around 400 training dialogues (6806 samples). Aone and Bennett (1995) claimed that the overall anaphora resolution performance seems to have reached a plateau at around 250 training examples. This result, however, indicates that $10^4 \sim 10^5$ training samples would be enough to train the trees in this task.

The chart gives us more information that performance limitation with our approach would be 80% $\sim$ 85% because each ellipsis type seems to approach the similar value, in particular for those in large training samples $\langle 1sg \rangle$ and $\langle 2sg \rangle$. Greater performance improvement is expected by conducting more training in $\langle 2pl \rangle$ and $\langle g \rangle$.

4.2 Topic Dependencies

It is completely satisfactory to build resolution knowledge only with topic-independent information. However, is it practical? We will discuss this question by conducting a few experi-
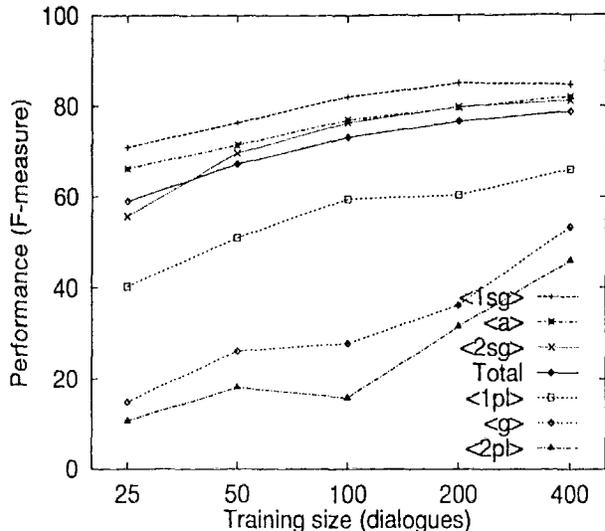

Figure 1: Training size and performance

ments.

We utilized the ATR travel arrangement corpus (Furuse et al., 1994). The corpus contains dialogues exchanged between two people. Various topics of travel arrangements such as immigration, sightseeing, shopping, and ticket ordering are included in the corpus. A dialogue consists of 10 to 30 exchanges. We classified dialogues of the corpus into four topic categories:

H_1 Hotel room reservation, modification and cancellation

H_2 Hotel service inquiry and troubleshooting

H_R Other hotel arrangements, such as hotel selection and an explanation of hotel facilities

R Other travel arrangements

Fifty dialogues were chosen randomly from the corpus in the topic category H_1, H_2, R, and the overall topic $T(= H_1 + H_2 + H_R + R)$ as training dialogues. We used 100 unseen dialogues as test samples again, which were the same as the samples used in the training-size experiment.

Table 4 shows the topic-dependency of each topic category that we provide with the F-measure. For instance, the first figure in the '$T/$' row (73.4) denotes that the accuracy with the F-measure is 73.4% against topic H_1 test samples when training is conducted on T, i.e., all topics. Note that the second row of the table indicates the ingredient of each topic in the test samples (and thus, the corpus).

1431

Table 4: Topic dependency

Train/Test (%)	/H_1	/H_2	/H_R	/R	Total
	20.1	27.7	11.2	40.9	100.0
H_1/	78.1	55.9	65.3	61.6	63.7
H_2/	71.3	67.0	62.6	62.6	65.6
R/	75.1	61.7	61.1	75.4	69.9
T/	73.4	62.5	62.6	66.2	66.2
$T - H_R$/	73.7	61.9	59.5	63.9	64.8

Table 5: Performance of major types in case

Case	$\langle$1sg$\rangle$	$\langle$2sg$\rangle$	$\langle$a$\rangle$	Total
ga(adj.)	58.3	68.1	85.9	79.7
wo	66.7	—	97.7	95.6
ni	95.2	95.7	81.9	91.7
ga(v.)	84.7	81.1	82.0	78.7

The results illustrate that very high accuracy is obtained when a training topic and a test topic coincide. This implies the importance not to train dialogues of unnecessary topics if the resolution topic is imaginable or restricted, in order to obtain higher performance. Among four topic subcategories, topic R shows the highest accuracy (69.9%) in total performance. The reason is not that topic R has something important to train, but that topic R contains the most test dialogues chosen at random.

The table also illustrates that a resolver trained in various kinds of topics ('T/') demonstrates higher resolving accuracy against the testing data set. It performs with better than average accuracy in every topic compared to one which is trained in a biased topic. By looking at some examples it may be possible to build an all-around ellipsis resolver, but topic-dependent features are necessary for better performance. The '$T - H_R$/' resolver shows the lowest performance (59.5%) against '/H_R' test set. This result is more evidence supporting the importance of topic-dependent features.

4.3 Difference in Surface Case

We applied a machine-learned resolver to agent case ellipses (Yamamoto et al., 1997). In this paper, we discuss whether this technique is applicable to surface cases.

We examined the feasibility of a machine-learned ellipsis resolver for three principal surface cases in Japanese, 'ga', 'wo', and 'ni'[1]. Roughly speaking, they express the subject, the direct object, and the indirect object of a sentence respectively. We classified the 'ga' case into two samples: a predicate of a sentence with a 'ga' case ellipsis that is a verb or an adjective.

In other words, this distinction corresponds to whether a sentence in English is a *be*-verb or a general-verb sentence. Henceforth, we call them 'ga(v.)' and 'ga(adj.)' respectively.

The training attributes provided are the same in all surface cases. They are listed in Table 2. In the experiment, 300 training dialogues and 100 unseen test dialogues were used. The following results are shown in Table 5[2]. The table illustrates that the ga(adj.) resolver has a similar performance to the ga(v.) resolver, whereas the former has a distinctive tendency toward the latter in each ellipsis type. The ga(adj.) case resolver produces unsatisfactory results in $\langle$1sg$\rangle$ and $\langle$2sg$\rangle$ ellipses, since insufficient samples appeared in the training set.

In the 'wo' case, more than 90% of the samples are tagged with $\langle$a$\rangle$, thus they are easily recognized as anaphoric. Although it may be difficult to decide the antecedents in the anaphoric ellipses by using information in Table 2, the results show that it is possible to simply recognize them. After recognizing that the ellipsis is anaphoric, it is possible to resolve them in other contextual processing modules, such as centering.

It is important to note that a satisfactory performance is presented for the 'ni' case (mostly indirect object). One reason for this could be that many indirect objects refer to exophoric persons, and thus an approach utilizing a decision tree that makes a selection from fixed decision candidates is suitable for 'ni' resolution.

5 Inside a Decision Tree

A decision tree is a convenient resolver for some kinds of problems, but we should not regard it as a black-box tool. It tells us what attributes are important, whether or not the attributes are

---

[1] We cannot investigate other optional cases due to a lack of samples.

[2] The result of the ga(v.) case is the same as '400' in Table 3.
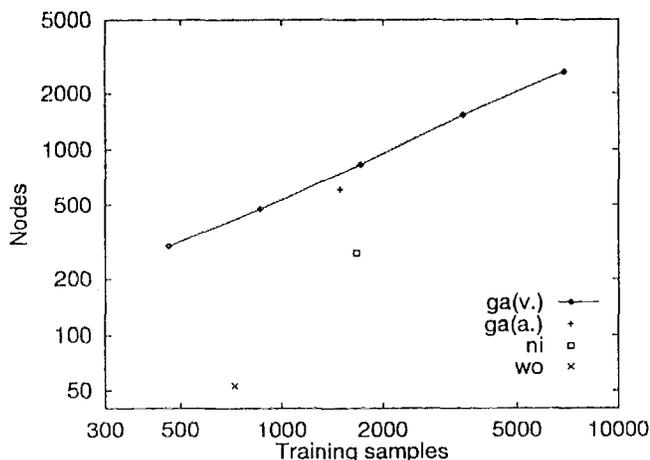
Figure 2: Training samples vs. nodes

Table 6: Depth and maximum width of decision tree

	$ga/25$	$/100$	$/400$	ga(adj.)	wo	ni
Depth	27	34	49	28	10	18
Width	26	58	146	52	10	28

sufficient, and sometimes more. In this section, we investigate decision trees and discuss them in detail.

5.1 Tree Shape

The relation between the number of training samples and the number of nodes in a decision tree is shown logarithmically in Figure 2. It is clear from the chart that the two factors of 'ga(v.)' case are logarithmically linear. This is because no pruning is conducted in building a decision tree. We also see that a more compact tree is built in the order of 'wo', 'ni', 'ga(adj.)' and 'ga(v.)'. This implies that the 'wo' case is the easiest of the four cases for characterizing the individuality among the ellipsis types.

Table 6 shows node depth and the maximum width in the decision trees we have built. By studying Table 5 and Table 6, we can see that the shallower the decision tree is, the better the resolver performs. One explanation for this may be that a deeper (and maybe bigger) decision tree fails to characterize each ellipsis type well, and thus it performs worse.

5.2 Attribute Coverage

We define a factor '$coverage$' for each attribute. Attribute coverage is the rate of the samples used to reach a decision about the samples used to build a decision tree. If an attribute is used at the top node of a decision tree, the attribute coverage is 100% in the definition, because all samples use it (first) to reach their decision. From this, we can learn the participation of each attribute, i.e., each attribute's importance.

Some typical attribute-coverages are expressed in Table 7. Note that '$ga(25)$' denotes the results of 'ga(v.)' with 25-dialogue training. A glance at the table will reveal that the coverage is not constant with an increasing number of training dialogues. Here we build a hypothesis from the table that more general attributes are preferred with a increase in training size.

The table illustrates that the topic-independent attributes increase with a rise in training size, such as '$-tekudasaru$' or '$-teitadaku$' (both auxiliary verbs which express the hearer's action toward the speaker with the speaker's respect). The table shows in contrast that the topic-dependent attributes decrease, such as ':before 72' (a category in which words concerned with $intention$ are included before the predicate mentioned) or ':before 94'. There are also some topic-independent words such as '$-ka$' (a particle that expresses that the sentence is interrogative) or ':before $41/43$'[3] which are still important regardless of the training size. This indicates the advantages of a machine-learning approach, because difficulties always arise in differentiating these words in manual approaches.

Table 8 also contrasts typical coverage in surface cases. It illustrates that there is a distinct difference between 'ga(v.)' and 'ga(adj.)'. The resolver of the 'ga(adj.)' case is interested in another cases, such as '$-de$' or contents of another case ':before $16/34$', whereas 'ga(v.)' case resolver checks some predicates and influential functional words. Coverage of each attribute in the 'ni' case has similar tendencies to those in the 'ga(v.)' case, except for a few attributes.

6 Conclusion and Future Work

This paper proposed a method for resolving the ellipsis that appear in Japanese dialogues. A machine-learning algorithm is used as the ac-

---

[3] We practically regard them as topic-independent words, because expressing the speaker's intention/thought is topic-independent.

Table 7: Training Size vs. Coverage

Attribute	$ga/25$	$ga/100$	$ga/400$
:here 43(intention)	100.0	100.0	100.0
:here 41(thought)	72.8	84.8	86.5
'-*ka*'(question)	53.1	83.2	66.3
'-*tekudasaru*'(polite)	9.1	49.1	49.8
honorific verbs	—	39.9	36.8
'-*teitadaku*'(polite)	—	33.2	33.9
'-*suru*' (to do)	4.1	22.0	26.1
:before *72*(facilities)	55.1	0.5	3.8
:before *94*(building)	28.5	9.8	7.7
:before *83*(language)	25.1	1.1	1.3
Speaker's role	11.7	9.1	20.5

Table 8: Case vs. Coverage

Attribute	$ga/400$	ga(adj.)	ni
'-*gozaimasu*'(polite)	—	100.0	—
:before 16(situation)	5.1	68.5	0.5
:before 34(statement)	5.3	59.0	11.2
'-*de*'(case particle)	5.2	23.9	1.9
'-*o*/-*go*'	46.4	7.0	100.0
:here 43(intention)	100.0	—	49.8
:here 41(thought)	86.5	—	43.5
Speaker's role	20.5	33.1	28.0

tual ellipsis resolver with this approach. The results of blind tests have proven that the proposed method is able to provide a satisfactory resolution accuracy of 91.7% in indirect objects, and 78.7% in subjects with verb predicates.

We also discussed training size, topic dependency and difference in grammatical case in a decision tree. By investigating decision trees, we conclude that topic-dependent attributes are also necessary for obtaining higher performance, and that indispensable attributes depend on the grammatical case to resolve.

Although this paper limits its scope, the proposed approach may also be applicable to other problems, such as referential property and the number of nouns, and in other languages such as Korean. In addition, we will explore contextual ellipses in the future, since it was found that most of the ellipses that appeared in spoken dialogues are found to be anaphoric in the '*wo*' case.

References

C. Aone and S. W. Bennett. 1995. Evaluating Automated and Manual Acquisition of Anaphora Resolution Strategies. In *Proc. of 33rd Annual Meeting of the ACL*, pages 122–129.

O. Furuse, Y. Sobashima, T. Takezawa, and N. Uratani. 1994. Bilingual Corpus for Speech Translation. In *Proc. of AAAI'94 Workshop on the Integration of Natural Language and Speech Processing*, pages 84–91.

O. Furuse, J. Kawai, H. Iida, S. Akamine, and D.-B. Kim. 1995. Multi-lingual Spoken-Language Translation Utilizing Translation Examples. In *Proc. of Natural Language Processing Pacific-Rim Symposium (NLPRS'95)*, pages 544–549.

M. Murata and M. Nagao. 1997. An Estimate of Referents of Pronouns in Japanese Sentences using Examples and Surface Expressions. *Journal of Natural Language Processing*, 4(1):87–110. written in Japanese.

H. Nakaiwa and S. Shirai. 1996. Anaphora Resolution of Japanese Zero Pronouns with Deictic Reference. In *Proc. of COLING-96*, pages 812–817.

J. R. Quinlan. 1993. *C4.5: Programs for Machine Learning*. Morgan Kaufmann.

H. Tanaka. 1994. Verbal Case Frame Acquisition from a Biliungual Corpus: Gradual Knowledge Acquisition. In *Proc. of COLING-94*, pages 727–731.

M. Walker and J. D. Moore. 1997. Empirical Studies in Discourse. *Computational Linguistics*, 23(1):1–12, March.

K. Yamamoto, E. Sumita, O. Furuse, and H. Iida. 1997. Ellipsis Resolution in Dialogues via Decision-Tree Learning. In *Proc. of Natural Language Processing Pacific-Rim Symposium (NLPRS'97)*, pages 423–428.

決定木による日本語格要素省略補完：概要

山本 和英　　　　　隅田 英一郎

ATR 音声翻訳通信研究所

E-mail: yamamoto@itl.atr.co.jp

特 徴

日本語対話文における格要素の省略補完について述べる。主語や目的語などの表示が義務的でない日本語の言語処理においては、これら省略される (非明示の)[1] 格要素を補う処理が重要である。格要素の省略は日本語に特有の現象ではなく、例えば韓国語、中国語などにも認められる。これら省略のある言語から英語やドイツ語など必須格を持つ言語への翻訳処理を行なう際には、補完処理 (省略内容の推定処理) は重要な処理となる。

本論文では、省略補完知識の決定木 (decision tree) による表現、及び省略情報の正解付きコーパスから言語現象と補完すべき省略の関係を帰納的に機械学習し、これによって日本語対話文の格省略を補完する手法を提案する。本論文で扱う問題は外界省略 (exophoric ellipsis) の補完と文脈省略 (endophoric ellipsis) の認知という問題である。対話文においては省略全体に占める外界省略の割合が高いため、これらの人称、数を決定することを問題とし、文脈省略に対する補完は本論文の対象外とする。

近年多くのテキストやシソーラスなどが機械可読化されてきており、多くの場合これらの言語資源は入手が可能となっている。本研究では、頑健性及び他話題への適用性を考慮して、形態素分割されて品詞と省略情報が付与されたコーパス、及びシソーラスのみを用いた汎用性の高い手法を提案する。提案手法は、特定のコーパス、品詞体系、シソーラスをいずれも仮定しないため、大量の知識を作成、保守する必要性がなく、手作業による補完規則やパラメータの調整を行なう必要もない。また本手法では、構文解析も仮定しないため、構文解析の手法や精度とは独立である。

結 論

日本語対話文の格要素省略に対して、決定木による補完処理の表現および機械学習によって補完知識を獲得する手法を提案した。補完に必要な知識として、内容語の意味属性、機能語の存在、話者知識の三種類を使用した。本手法により獲得した決定木で未学習文に対してテストを行なった結果、「が」格と「に」格に対しては十分な精度で省略された人称を補完することを確認した。「を」格に関しては、その補完内容が照応的であるという認知を行なうのに有効であることを確認し、本手法の有効性を確認することができた。

また提案手法に関して、処理の有効性を学習量、話題依存性、使用属性との関係の3点から議論した。本研究で得られた主な知見を以下にまとめる。

- 「が」格 (動) や「に」格は、尊敬を示す機能語などを重要視する。「が」格 (形) は他の格要素の情報によって補完を試みる傾向がある。

- 当該問題に対する学習量は全体として $10^4 \sim 10^5$ サンプルで十分である。この時の補完精度の上限は 80% $\sim$ 85% と予想される。

- 対話の話題が既知もしくは予測可能な時は、その話題のみによる学習が最高の性能を示す。話題が未知の場合は、可能な限り広範な話題に対して学習するのが最も効果的である。

- 学習量増加に伴い、決定木は機能語などのより一般的な属性を採用する。

本論文では日本語対話文における格要素の補完処理に限定して述べてきたが、提案手法の有効性はこれだけにとどまらない。例えば韓国語は日本語などと同様に格要素の省略が観察される。韓国語などにおける省略補完処理も本手法の応用によって可能になると考えられる。

---

[1] そもそも省略ではなく非存在とする解釈もあるが、ここでは格要素が明示されていないものすべてを「省略」と呼び、本論文の研究対象とする。