Centering as an Anaphora Generation Algorithm: A Language Learning Aid Perspective

Mitsuko Yamura-Takei†, Miho Fujiwara‡ and Teruaki Aizawa§

†Graduate School of Information Sciences
§Faculty of Information Sciences
Hiroshima City University
3-4-1 Ozuka-higashi, Asaminami-ku,
Hiroshima 731-3194, JAPAN
yamuram@nlp.its.hiroshima.cu.ac.jp
aizawa@its.hiroshima-cu.ac.jp

‡Department of Japanese and Chinese
Willamette University
900 State Street, Salem, OR
97301, USA
mfujiwar@willamette.edu

Abstract

This paper describes our attempt to develop an NLP-enhanced CALL program for Japanese learners. This CALL program focuses on zero anaphora acquisition. Based upon our previous work (Yamura-Takei et al., 2001), this paper will analyze the revision-based zero anaphora generation task. We first discuss the problems students face in treating zero anaphora by providing some empirical data from a language classroom. We then propose an anaphora generation algorithm within the Centering framework and evaluate its feasibility by comparing the algorithm output with human judgments. The evaluation result provides insights for the empirical validity of the algorithm and its future extensions.

1 Introduction

Natural Language Processing (NLP), including its subfield Natural Language Generation (NLG), is an emerging technology with a variety of real-world applications. Computer-Assisted Language Learning (CALL) is one such area that NLP techniques can be applied to.

Among a range of techniques that NLP can offer, anaphora resolution/generation is of particular importance to many NLP applications and has attracted extensive research efforts, mainly in machine translation (MT). It is rather surprising that CALL has not yet seen much interest in anaphora. Zero anaphora is a hard nut to crack for learners but it has not drawn much attention in Japanese as a Second Language (JSL) instruction, as discussed in Section 2.

Although there are many programs available that focus on lexico-grammatical aspects, discourse-level learning aids are not as prevalent. It is probably true that vocabulary plays a central role in both reading and writing. But in order to improve reading fluency, it is also essential to explore discourse structure. Furthermore, when writing, more linguistic sophistication can be achieved by constructing discourse in an appropriate fashion.

Centering Theory, originating as a discourse structure model, has served as one of the major anaphora resolution algorithms. The Centering model is immensely language-independent: an individual language is modified by setting its language-specific parameters within the model. The parameters for Japanese were proposed by Kameyama (1985) and Walker et al. (1990, 1994). Since then, this Japanese model has been experimented with by many researchers, and has proven to be reasonably feasible.

Based on previous research, we have designed Zero Checker, an NLP-enhanced CALL program. Zero Checker, created within the Centering framework, supports JSL learners in understanding zero anaphora. In our previous study (Yamura-Takei et al., 2001), we proved the feasibility of Centering as an adequate zero anaphora understanding aid model. In this paper, we will provide more empirical evidence to evaluate our zero anaphora understanding aid model and further investigate its validity.

In Section 2, we discuss the challenge that zero anaphora poses for JSL learners by presenting some language classroom data. Section 3 provides an overview the Centering formalism and its modification for this study. Also in Section 3 we propose the anaphora generation algorithm. In Section 4, we present the evaluation results of the algorithm on our JSL data and discuss its feasibility as a language learning aid model and potential extensions that our findings suggest. Section 5 concludes with a discussion of possible future work.
2 Zero Anaphora for JSL Learners

In English, every verb’s argument must be overtly expressed in a sentence. The sentence becomes ungrammatical otherwise. In contrast, Japanese allows such arguments to be freely omitted when they are recoverable from a given context. These unexpressed elements are called zero anaphora or zero pronouns (henceforth zeros, following Iida, 1998).

This striking contrast posits a major challenge not only for Japanese-English MT developers but also for JSL learners who have English or another explicit language as their first language. Very few JSL textbooks, however, have a section addressing formal instruction or include intensive exercises on this ellipsis mechanism. As a result, many JSL teachers rely heavily on their intuition about naturalness, rather than depending upon systematic knowledge, when they explain zeros. Intuition is a conventional tool in teaching one’s native language, but from a students’ perspective, a well-developed systematic theory can be more convincing. This pedagogical discrepancy is the motive for building an autonomous learning aid that is specially designed for understanding passages with zeros and producing contextually appropriate zeros.

The following section presents empirical data from a JSL classroom to verify that zero anaphora is one of the critical issues students face when writing Japanese.

2.1 Underusing Zeros

For those whose first language does not permit sentence parts to be omitted, it would be a perplexing task to identify which elements can be safely omitted and what contexts allow omission. It is easily assumed that learners tend to underuse zeros by using the strategy of avoidance, which often results in unnaturalness caused by the redundant use of full Noun Phrases (NPs). In order to examine this assumption, let us present an intermediate student’s writing sample in (1).

1. nomin-ga itsumo tozoku-ni kome-ya okane-o nusumarete
   farmer-SUB often thief-OBJ rice-and money-OBJ be-robbed
   “The farmers were often robbed of rice and money by thieves, and”

2. nomin-ga tozoku-o taosu tame ni,
   farmer-SUB thief-OBJ beat in-order-that
   “in order that the farmers beat the thieves,”

3. 7-nin-no samurai-o yatoimashita.
   Ø-SUB 7 samurai-OBJ hired
   “Ø hired 7 samurais.”

Japanese language teachers would recommend deleting the subject in the second utterance (nomin “the farmers”) for more natural Japanese discourse.

As illustrated by this example (as well as other examples found in our data), such omissions are often advised in order to avoid unnaturalness caused by redundancy.

2.2 Overusing Zeros

Japanese is an elliptic language. Learners understand that Japanese quite freely permits sentence parts to be omitted, but what triggers such ellipsis is not as easily understood. This potentially creates overuse of zeros. Let us look at the next sample written by a lower-intermediate student in (2).

1. mukashi-mukashi aru mura-ni aru kodomo-ga ita.
   once-upon-a-time a village-LOC a child-SUB was
   “Once upon a time, there was a child in a village.”

2. ookami-ga kita-to sakenda
   wolf-SUB came-COMP Ø-SUB shouted
   “Ø shouted that a wolf came. ”

3. murabito-ga kita ga, ookami-wa nakatta.
   villager-SUB came but wolf-SUB was not
   “The villagers came, but there was no wolf.”

4. nandomo kurikaeshite, murabito-ga okotta.
   many times Ø-SUB Ø-OBJ repeated, villager-SUB got-angry
   “Ø repeated Ø many times, and the villagers got angry.”

5. konakunatta.
   Ø-SUB came-not
   “Ø did not come any more.”

The four zeros in the passage cause ambiguity. It is not clear who shouted, who repeated what many times and who did not come any more. In this example, teachers would advise not to omit zeros in order to avoid ambiguity.

The use of zeros is a double-edged sword precariously balanced on a thin line. Underuse of zeros causes redundancy while overuse of zeros causes ambiguity. However fine the line may be, there needs to be some theoretical guidelines about where to draw it.
3 Centering Model

In order to use a computer-based approach to solve the problems presented in the previous section, we decided to use the Centering model as our theoretical framework. We expect this model not only to produce the final outcome, but also to provide learners with its zero resolving/generating process, e.g., by highlighting antecedent candidates with their preference order in output text.

This section presents the basic definitions and assumptions in Centering discussed in the literature (e.g., Grosz et al., 1995; Walker, 1998). We discuss modifications for Japanese and propose a zero generation1 of zeros.

3.1 Overview of Centering

Centering is a theory of discourse structure that models the interrelationships between focus, the choice of referring expressions, and perceived coherence of utterances.

A discourse segment consists of a sequence of utterances2 Ut, … Ut. Each utterance Ui evokes a set of discourse entities, the FORWARD-LOOKING CENTERS, Cf(Ui). The highest ranked entity in Cf(Ui) realized in Ui is the BACK-WORD-LOOKING CENTER, Cb(Ui). The highest ranked member of the Cf set is the PREFERRED CENTER, Cp(Ui). The Cp is predicted to be Cb in the following utterance. The members of the Cf list are ranked as in (3), in case of Japanese (Walker et al., 1994):

(3) Cf ranking for Japanese

GRAMMATICAL OR ZERO) TOPIC > EMPATHY > SUBJECT > OBJECT2 > OBJECT > OTHERS

Four types of transitions, reflecting a degree of local coherence, are defined, as shown in Table 1.

### Table 1: Centering Transitions

<table>
<thead>
<tr>
<th>Cb(Ui) = Cb(Ui-1)</th>
<th>Cb(Ui) ≠ Cb(Ui-1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONTINUE</td>
<td>SMOOTH-SHIFT</td>
</tr>
</tbody>
</table>

On top of these definitions, Centering also includes two rules in (4).

(4) Centering Rules

- **Rule 1**: If some element of Cf(Ui-1) is realized as a pronoun in Ui then so is Cb(Ui).
- **Rule 2**: Transition states are ordered. CONTINUE is preferred to RETAIN, which is preferred to SMOOTH-SHIFT, which is preferred to ROUGH-SHIFT.

Pronominalization (or zero pronominalization) is predicted to comply with Rule 1 that is called Pronoun Rule, on the assumption that discourse is coherent by following the preference ordering given in Rule 2. The Pronoun Rule and the choice of referring expressions over transition types play a significant role in the development of (zero) anaphora resolution/generation algorithms.

Centering’s advantage as a language learning aid model lies mainly in its simplicity. The syntactic (rather than thematic) criterion for ranking potential antecedents makes it easier for learners to track them down.

3.2 Global Model

Centering, as described in the previous subsection, is a theory of local focusing. The antecedents of zeros are assumed to be in the immediately preceding utterance. A study of naturally occurring texts, however, reveals abundant cases that contradict this assumption in various languages. For example, the use of zeros in RETAIN and ROUGH-SHIFT transition states is not rare in Japanese, as data presented by Iida (1998) and Yamura-Takei et al. (2000, 2001) indicate.

These cases are further classified into two kinds: (i) the antecedent of zero (Ui) is an element other than Cp(Ui) in Cf(Ui-1); or (ii) no appropriate antecedent can be found in Cf(Ui). The latter case needs to further search for potential antecedents beyond the local boundary.

Iida (1998) introduced a global Cb (GCB) list. This list is comprised of former Cbs ordered by recency, as an expanded search area. Yamura-Takei et al. (2000) introduced an alternative global center structure: a Global Topic (GT) list that is a list of former topics (marked by topic marker −wa) ordered by recency.3 Between these two global models, there is no significant difference in their resolution performance (ibid). Theoretically and computationally speaking, GCB, which shares the same center with a local model, is more reasonable,

---

1. The program takes a learner’s written text as input and provides the learner with Centering-based revision advice as output.
2. Following Kameyama (1998), Walker (1998) and Iida (1998), we defined utterance as a finite clause with one predicate, serving as a center updating unit.
3. They also proposed Global Topic Hypothesis (GTH) to compensate for a global Cb list model in treating utterances with multiple zeros. In GTH, each utterance is assigned one and only one Global Topic (GT) with a proximate grammatical topic. Global Transition: GT(Ui)=Cp(Ui) is applied when one or more zeros, in an utterance with multiple pronouns, are retrieved globally.
but from a learners’ point of view, GT is more recognizable. This is the case with Japanese language because the topic is explicitly marked with a specific particle – *wa*.

These models of global focusing have primarily been developed in the context of anaphora resolution, but their implications should also contribute to the generation task.

### 3.3 Centering Algorithm

Centering has been widely researched from the point of view of interpreting anaphora, and it has also been applied to the tasks of coherence-based text planning (e.g., Kibble and Power, 1999), essay evaluation (e.g., Miltsakaki and Kukich, 2000), and referring expressions generation (e.g., Mittal *et al.*, 1998).

In this paper, we attempt to examine how the principles underlying the constraints and rules of the theory can be facilitated as a revision-based microplanning operator for generating *zeros* in an appropriate fashion. In an attempt of fitting the principles into the generation task, we first defined the hypothesis as in (5), simply assuming that *zeros* are used when their antecedents are locally recoverable, in Yamura-Takei *et al.* (2001).

(5) Zero Generation Hypothesis (local version)

<table>
<thead>
<tr>
<th>CONTINUE/SMOOTH-SHIFT</th>
<th>RETAIN/ROUGH-SHIFT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ZERO</strong></td>
<td><strong>FULL NP</strong></td>
</tr>
<tr>
<td><strong>7</strong></td>
<td><strong>2</strong></td>
</tr>
</tbody>
</table>

The six cases of full NPs in CONTINUE/SMOOTH-SHIFT transition states indicate the underuse of *zeros*, and the algorithm returns “message1” to suggest revision by zero-pronominalizing the NP. The three cases of *zeros* in RETAIN/ROUGH-SHIFT transition states, on the other hand, denote possible overuse. The algorithm returns “message2” to recommend revision by replacing *zeros* with overt anaphors unless they are either locally or globally co-referential.

### 4 Data Analysis and Evaluation

#### 4.1 Zero Anaphora in JSL Writing

In order to test the feasibility of our generation algorithm, two upper-intermediate students’ essays were analyzed, and their choice of anaphors were examined. The results are summarized in Table 2.

<table>
<thead>
<tr>
<th>CONTINUE/SMOOTH-SHIFT</th>
<th>RETAIN/ROUGH-SHIFT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ZERO</strong></td>
<td><strong>FULL NP</strong></td>
</tr>
<tr>
<td><strong>6</strong></td>
<td><strong>2</strong></td>
</tr>
</tbody>
</table>

The six cases of full NPs in CONTINUE/SMOOTH-SHIFT transition states indicate the underuse of *zeros*, and the algorithm returns “message1” to suggest revision by zero-pronominalizing the NP. The three cases of *zeros* in RETAIN/ROUGH-SHIFT transition states, on the other hand, denote possible overuse. The algorithm returns “message2” to recommend revision by replacing *zeros* with overt anaphors unless they are either locally or globally co-referential.

#### 4.2 Evaluation

##### 4.2.1 Method

CALL systems are, in a sense, substitutes for human teachers and they are expected to help generate appropriate and natural-sounding texts. We compared the algorithm output with JSL teachers’ judgments to evaluate the performance of the generation algorithm.

The human judges consist of nine JSL teachers who are native speakers of Japanese. Given the JSL
learners’ writing samples, they were asked to decide (i) which full NPs should be zero-pronominarized to avoid redundancy, and (ii) which zeros should be overtly expressed to avoid ambiguity, as they do in their classrooms. This method provides insights of both strengths and shortcomings of the algorithm, as noted in Reiter and Dale (2000). These insights will improve future development of our program.

4.2.2 Results

Before we discuss the results, we should note that there is considerable disagreement among our human judges. This is often the case, as pointed out in Dale and Mellish (1998). The choice of anaphors is not a grammatical or “right or wrong” issue, but rather an issue of discourse-level perceived naturalness about coherence. Each individual’s intuition has a great influence on his/her decision. We will discuss what this disagreement means in the zero generation task, in addition to the overall evaluation of the algorithm.

Five out of six cases of possible underuse of zeros, (presented in 4.1), matched at least some human judgments. These results are from a total of 29 NPs. All three possible overuse cases found in a total of 30 zeros were supported by human judges. This is evidence to assess the performance of the algorithm as a substitute human tutoring.

Now, let us further examine the underuse cases. The algorithm suggests revision by zero-pronominarizing NPs. Table 3 shows how many (out of nine total) human judges match this decision.

<table>
<thead>
<tr>
<th>Case</th>
<th>Anaphor</th>
<th># of human judges</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) CONTINUE</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>(2) CONTINUE</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>(3) SMOOTH-SHIFT</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>(4) SMOOTH-SHIFT</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>(5) SMOOTH-SHIFT</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>(6) SMOOTH-SHIFT</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

A closer examination of the students’ samples and comments from our team of human judges lead us to an explanation of this variety in number. Cases (1), (2) and (3) seem to be affected by their positions within a discourse. Cases (1) and (2) are in discourse-initial, while (3) in discourse-final. The tendency for explicitness in the beginning and end of a Japanese narrative might affect some human judgments. Cases (4), (5) and (6) involve definite NPs, i.e. demonstrative + noun (e.g., *sono hito* “the person”). It can be assumed that definite description, unlike repetitive NP anaphors, does not create a sense of redundancy. Also, (4) is preceded by a subordinated clause, while (5) and (6) by independent clauses (sentences). Connectives between adjacent utterance pairs might affect choice of anaphors.

Next, let us examine three possible overuse cases. The algorithm suggests these zeros be overtly expressed. Table 4 shows how many (out of nine total) humans made the same decision as the algorithm.

<table>
<thead>
<tr>
<th>Case</th>
<th>Anaphor</th>
<th># of human judges</th>
</tr>
</thead>
<tbody>
<tr>
<td>(7) RETAIN</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>(8) RETAIN</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>(9) ROUGH-SHIFT</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

Cases (8) and (9) are supported by the majority, while in (7) some world knowledge seems to cause two distinct views.

We have presented evidence for the algorithm performance and our explanations for the inter-rater disagreement. Another point of interest is that humans suggest further revisions outside of our algorithm assumption, as summarized in Table 5.

<table>
<thead>
<tr>
<th>Case</th>
<th>Anaphor</th>
<th># of humans</th>
</tr>
</thead>
<tbody>
<tr>
<td>(10) ROUGH</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>(11) CONTINUE</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>(12) CONTINUE</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>(13) CONTINUE</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>(14) CONTINUE</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>(15) CONTINUE</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>(16) CONTINUE</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>(17) SMOOTH</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>(18) SMOOTH</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

In (10), three human judges recommend the use of zeros in ROUGH-SHIFT. This is another apparent world knowledge effect on anaphor choice. In (11) through (18), the human judges suggest that zeros be lexically expressed. Seven of these cases, (12) through (18), involve multiple zeros in subject and object positions. Double zeros are likely to create more ambiguity and excessive inference load on the hearer/reader. This probably motivates human judges to suggest revision. Case (11) also involves double zeros in a complex sentence: subjects in both main and subordinate clauses are omitted, which causes ambiguity.

The result of this comparison between the algorithm output and human judgments demonstrate both a satisfactory performance of the algorithm as a baseline model and insights for its enhancement.

4.3 Discussion

This paper has highlighted some implications of Centering Theory for supporting language learning in respect to understanding zeros in Japanese. Evaluating the algorithm on our empirical data, Centering Theory has functioned satisfactorily as a language learning aid model in the context of revision-based zero generation tasks. Comparison of the algorithm
output with human performance reveals the complexity of human zero processing mechanism and some limitations of the algorithm itself.

Future issues to consider include:

(a) positions within a discourse,  
(b) connectives between utterance pairs, 
(c) multiple zeros, 
(d) non-zero anaphors, and 
(e) world knowledge.

These issues could be further investigated for additional zero checking criteria in our CALL program.

5 Future Work

A full study will require a larger empirical corpus from a variety of proficiency levels: more JSL writing samples should be analyzed to access the performance of the algorithm. In particular, we need more samples with zeros recoverable from global center lists. Pedagogically speaking, zero generation based on the local center model is helpful at the rudimentary stage of learning. Globally recoverable zeros should be introduced at higher proficiency levels.

We should further refine the evaluation methodology. Feedback from our judges indicates that choosing whether zeros should or should not be used is often a hard task. More choices (e.g., “either is fine”) reflect human judgment more accurately.

This preliminary study provided us with valuable insight. These findings should be incorporated into our CALL program, Zero Checker. This program consists of four components: assessing one’s understanding of zeros, providing aids to understand zeros, reading aid and writing revision aid. Zero Checker is an interactive tutorial system. In its writing revision aid component, for example, Centering-based revision advice will be given along with supplementary instructional comments. In this way, learners and teachers can learn ellipsis processing mechanisms.

Acknowledgments

We would like to thank the following JSL teachers for their contributions: Yasuyo Ebara, Mariko Gomez, Mariko Moroishi, Noriko Okamoto, Mari Takada, Takae Tsujioka, Tomoharu Yanagimachi, and Momoko Yokoi. We are also grateful to the three anonymous reviewers who enlightened us with their valuable comments.

References


6 Details of Zero Checker will be discussed in Fujiwara and Yamura-Takei (in preparation).