

Part II. (Zero) Pronominal Anaphora Resolution

1. Introduction

1.1. The Issue

Pronominal anaphora resolution here means to find the antecedents (the closest preceding co-referential expressions) for (zero) pronouns among the precedent morphemes. Let us look at the following example in Japanese.

Example:

「今日のお昼にうな丼を学食で食べた。」 "I ate *unadon* for lunch today."

「それ、おいしかった？」 "Was *it* good?"

Morpheme sequence:	今日	
	の	
	お昼	
	に	
	うな丼	← Antecedent
	を	
	学食	
	で	
	食べ	
	た	
	それ	← Pronoun
	...	

In this example, the anaphora resolution for the pronoun 「それ」 is to select the proper antecedent 「うな丼」 in the morpheme sequence.

Various kinds of clues for anaphora resolution or antecedent selection have been proposed.¹¹ For example, one can expect that if the other conditions are the same, the closer the morpheme is to the pronoun, the more likely it is to be the antecedent. If an antecedent candidate is in the same sentence the pronoun is in, there are syntactic constraints on the candidacy. The Centering Theory (Grosz, Joshi, and Weinstein 1995) claims that grammatical functions such as Subject and Object have to do with the selection of the antecedent. Semantic affinities between the pronoun and morphemes should play a role too. For example, the antecedent of the pronoun 「かれ」 (he) would have the masculine gender and a singular number.

In this report, the case frame is considered to be an important key for resolving anaphora. Since a pronoun and the antecedent are co-referential, if the antecedent is a nominal, the replacement of the pronoun with the antecedent in the case frame (sentence) should be semantically fine. To take the example above, since 「それ」 is the subject in the case frame 「おいしい」, the antecedent should also be able to be in the subject position for 「おいしい」; i.e., the expression 「うな丼はおいしい」 should be semantically fine.

1.2. The Current Approach

A simple filtering program which judges if pairs of pronoun case frames and antecedent candidates would

previous to the utterance where the pronoun occurs and recedes along the time line (in the experiments, only inter-sentential anaphora resolution was tested). In each utterance, kinds of intra-sentential ordering of the candidates are tested (see 3.3).

For the judgment of the appropriateness of the pronoun-antecedent candidate pairs, a statistical method was used: the probabilities of the pronoun-antecedent candidate pairs and the replacement of antecedent candidates into the pronoun case frames were calculated from the statistics in a corpus.

2. The Scope of the Experiments

In this section, the 1) anaphora data for statistics and evaluation, 2) qualifications on the experiments, and 3) auxiliary data employed will be explained.¹²

2.1. Anaphora Corpora

The main data for the experiments is a corpus of travel conversations tagged with the usage and anaphoric relations on pronouns and zero pronouns. 375 dialogues in the ATR-ITL Speech and Language DataBase (SLDB) tagged for this purpose were used.

2.2. Qualifications

Not all of the anaphoric data in the corpus was used. The following is the list of qualifications with regard to the data:

- Only anaphoric pronouns are used: those for deictic or indexical use, for example, are excluded.¹³
- Only inter-sentential anaphora are examined: intra-sentential anaphora are excluded.
- Only those with single antecedents are used. Moreover, the pronouns "どちら," "どっち," "いずれ," "どれ," and "それぞれ" are excluded, as they are assumed to have multiple antecedents.
- Only those with nominal antecedents are used: those with predicative antecedents, for example, are excluded. The nominal in this report is defined by the following parts of speech:

〈普通名詞〉 〈固有名詞〉 〈代名詞〉 〈名詞句〉 〈接尾辞〉 〈サ変名詞〉
〈数詞〉 〈日時〉 〈副助詞〉 〈人名〉 〈住所名〉 〈副詞的名詞〉

2.3. Auxiliary Data:

▪ Case Analysis Corpus

As mentioned above, case frames play important roles in the experiments. The case frames were taken from the case analysis corpus created at ATR-ITL for the SLDB dialogues.

▪ Dictionaries for Semantic Information

To make anaphora resolution robust, semantic abstraction by semantic codes (semcodes, hereafter) and semantic features (HUM, LOC, TLOC) is used.

Semcodes:

A semcode list has been prepared at Department 3 of ATR-ITL following the system of The Kadokawa Thesaurus (大野 et al. 1981). 〈日時〉 (temporal) expressions not included in the list are added manually.

3. Details

Although the basic idea for the experiments is the same for pronouns and zero pronouns, there are minor differences. Below, details unique to regular pronouns are explained first, and then those unique to zero pronouns are explained. Finally, details common to both regular pronouns and zero pronouns are explained.

3.1. Pronoun Anaphora Resolution

3.1.1. The pronoun-antecedent candidate pair

As mentioned earlier, the filter program judges if pairs of pronoun case frames and antecedent candidates are appropriate to be anaphoric. The following is an example pair and the format of pronoun-antecedent candidate pairs. Although the antecedent candidate is associated with a case frame, its predicative (verb or adjective) was not actually used.

A sample pair:

TCC22071-0150-1, 15, TCC22071-0140-1, 11,
 (((それ (を <格助詞>) (作れ <本動詞> 一段))
 ((ごはん <普通名詞>) (を <格助詞>) (炊 <本動詞> 五段力)) OBJE -), C

3.1.2.2. Screening with Case Frame Database

If the substitution of the antecedent for the pronoun in a case frame does not appear enough in the case frame database compared to the occurrence of the case frame without the pronoun part, the pair is screened out. Because sparseness is expected, semantic features and parts of speech are used instead of morphemes.

Example: If the case frame ((TLOC +)) (へ <格助詞>) (着 <本動詞> 五段力)) does not appear enough in the case frame database compared to the frame without the pronoun part ((へ <格助詞>) (着 <本動詞> 五段力)), it is screened out.

3.1.2.3. Constraints

The following non-statistic constraints were used to judge certain kinds of pronoun-antecedent candidate pairs as inappropriate to be anaphoric.

Nouns in Adjective Use Check

Nouns used as adjectives such as “特別”, “別々”, “本当” are screened out from the antecedent candidates with a hand-made dictionary.

Pronoun-Antecedent Feature-Conflict Check

Feature conflicts between pronouns and antecedent candidates are handled by hard coding (↓). (These constraints can be learned and the hand-coding will turn out to be unnecessary [see 5. Results]. The constraints are set weak so that they do not screen valid combinations out.)

```
(case (first (first pattern))
      ((あそこ ここ そこ) '((HUM +) (LOC -) (TLOC +)))
      ((あちら あっち) '((TLOC +) (HUM +)))
      ((おたく) '((TLOC +)))
      ((それ) '((HUM +)))
      ((これ どれ) '((HUM +) (TLOC +)))
      ((それぞれ) NIL)
      ((こちら どちら) NIL)
      ((そちら) '((HUM +) (TLOC +)))
      ((こっち そっち どっち) '((HUM +)))
      ((彼) '((HUM -) (LOC +) (TLOC +)))
      ))
```

LISP code describing conflicts between pronouns and features

3.1.2.4. Heuristics

The following heuristics were used to pass certain kinds of pronoun-antecedent candidate pairs as appropriate to be anaphoric. (The heuristics turned out to be ineffective in the experiments.)

Same Pronoun Check

If the antecedent candidate is a pronoun of the same form, it is judged to be the antecedent.

COND in the Previous TFQ/Confirmation Check

If 1) the pronoun has the thematic role COND (CONDition) and 2) the antecedent candidate is in the previous utterance and of the type TFQ (True-False Question) or Confirmation-Question and has the thematic role COND, then the candidate is judged to be the antecedent.

3.2. Zero Pronoun Anaphora Resolution

A sample pair:

TAG22013-0090-1, 16, TAG22013-0020-1, 5,
 (((願 <本動詞> 五段ワ) OBJE)
 ((予約 <サ変名詞>) (を <格助詞>) (願 <本動詞> 五段ワ))) +)

The format:

1. the utterance ID of the pronoun
2. the sub-utterance index of the pronoun
3. the utterance ID of the candidate
4. the sub-utterance index of the candidate
5. (((predicative thematic-role)
 (first-case-frame-component
 second-case-frame-component
 third-case-frame-component))
 polarity) ; "+" if antecedent, else "-"

3.2.2. Screening**3.2.2.1. Screening by Learning**

Pairs of zero pronouns with their case frames and their antecedent candidates can be used to learn which pairs are likely to be zero pronoun-antecedent pairs. The pairs with actual antecedents are positive examples and those without actual antecedents are negative examples.

Zero Pronoun-Antecedent Combination Check

If the combination of the thematic role of the zero pronoun and the candidate often appears in the corpus of zero pronoun-antecedent candidate pairs but mostly as a negative example, then the combination is screened out. The candidate is matched with its form, semcodes, semantic features, or parts of speech.

Example patterns:

(DEST (金曜 <日時>))
 (DEST ("151d"))
 (DEST ((TLOC -)))
 (DEST <日時>)

Case Frame Check

If the embedding of the antecedent to the zero position of the case frame often appears in the corpus of zero pronoun-antecedent candidate pairs, but mostly as a negative example, then the combination is screened out. A pattern matched with morphemes, semcodes, semantic features, or parts of speech.

Example patterns:

((金曜 <日時>) DEST (着 <本動詞> 五段力))	Morpheme Matching
("151d") DEST ("223" "302b" "314b" "386c"))	Semcode Matching
((TLOC -)) DEST (着 <本動詞> 五段力))	Sem. Feature Matching
<日時> DEST (着 <本動詞> 五段力))	Part-of-Speech Matching

3.2.2.2. Screening with Case Frame Database

A pair of a zero-pronoun case frame and an antecedent candidate is screened out if the embedding of the antecedent to the zero does not appear enough in the case frame database compared to the occurrence of the case

(〈日時〉 DEST (着 〈本動詞〉 五段力)) does not appear enough in the case frame database compared to the frame (DEST (着 〈本動詞〉 五段力)), it is screened out.

3.2.2.2. Constraints

Nouns in Adjective Use Check

As with regular pronouns, nouns that are used as adjectives such as “特別”, “別々”, “本当” are screened out from the antecedent candidates with a hand-made dictionary.

Zero Pronoun-Antecedent Feature Conflict Check

Feature conflicts between zero pronouns (case slots) and antecedent candidates were handled by hard coding. (These constraints can be learned, but full learning requires more data. The constraints are set weak so that they do not screen valid combinations out.) (Ref. ↓)

```
(case (second (first pattern))
  ((AGEN) ' ((TLOC +)))
  ((CONT) ' ((TLOC +) (HUM +) (LOC +)))
  ((DEPT DEST LOCT) ' ((HUM +) (TLOC +)))
  ((EXPR) ' ((TLOC +)))
  ((MUTL) ' ((TLOC +)))
  ((GOAL) ' ((HUM +)))
  ((RECP) ' ((TLOC +)))
  ((ROUT) ' ((HUM +) (TLOC +)))
))
```

LISP code describing conflicts between zero pronouns and features

3.3. Intra-sentential Ordering

Three kinds of intra-sentential ordering were tested. In the standard setting if one condition did not bring about ordering, then the next condition was used for ordering.

3.3.1. Ordering by Semantic Affinity

In this ordering, priority is given to a pair where the substitution of the antecedent candidate for the pronoun (for a zero pronoun, the embedding of the antecedent candidate to the zero position of the frame) occurs more in the case frame database.

3.3.2. Ordering by Grammatical Function

In this ordering, priority is given based on the Centering Theory¹⁵. Namely, 係助詞 (TOPIC) > を > に > 格助詞 (Other Case Markers) > Others. Zero pronouns (see 3.4) get a higher priority than other elements (i.e., AGEN > OBJE > ZERO > 係助詞) under the assumption that zero pronouns are used to refer to the most salient objects.

3.3.3. Ordering by the Order of Occurrence

In this ordering, priority is given to an element that occurs later (i.e., closer to the pronoun). This is the default ordering.

3.4. Data Reliability & Thresholds

of the calculated probability can be. Thus, the experiments were set so that screening would be done only when the denominators are larger than certain thresholds (which depended on the types of screening). The thresholds were empirically determined.

4. Results

4.1. Pronoun Anaphora Resolution

Below, A-E each signifies a different way for 5-way cross validation tests. The scores are success rates from anaphora resolution. The "Total" scores on the right are cross validation summations. The nine sets of scores following "The Best Score" are obtained by removing one screening process or heuristic or adding one kind of intra-sentential ordering from "The Best Score" anaphora resolution process. The last set of scores shows the baseline, namely, the score obtained when the program judges the closest precedent nominals to be the antecedents of pronouns. By default, intra-sentential ordering is done only by the occurrence order (ordering by semantic affinity & grammatical functions is not used).

- **The Best Score**

A	B	C	D	E	Total
.5821 (39/67)	.6552 (38/58)	.6491 (37/57)	.6301 (46/73)	.5733 (43/75)	.6152 (203/330)

- **Pronoun-Antecedent Combination Check (Case-Based) Off**

A	B	C	D	E	Total
.5224 (35/67)	.5345 (31/58)	.5439 (31/57)	.6164 (45/73)	0.52 (39/75)	.5485 (181/330)

- **Case-Frame Check (Case-Based) Off**

A	B	C	D	E	Total
.5373 (36/67)	.6379 (37/58)	0.614 (35/57)	.6164 (45/73)	.5867 (44/75)	.5970 (197/330)

- **Screening with Case Frame Database Off**

A	B	C	D	E	Total
.5821 (39/67)	.6724 (39/58)	.6491 (37/57)	.6301 (46/73)	0.56 (42/75)	.6121 (202/330)

- **Nouns in Adjective Use Check Off**

A	B	C	D	E	Total
.5672 (38/67)	.6552 (38/58)	.6491 (37/57)	.6301 (46/73)	.5733 (43/75)	.6121 (202/330)

- **Pronoun-Antecedent Feature Conflict Check Off**

A	B	C	D	E	Total
.5821 (39/67)	.6724 (39/58)	.5965 (34/57)	.6027 (44/73)	.5733 (43/75)	.6030 (199/330)

- **Same Pronoun Check Off**

A	B	C	D	E	Total
.5821 (39/67)	.6552 (38/58)	.6491 (37/57)	.6301 (46/73)	.5733 (43/75)	.6152 (203/330)

- **COND in the Previous TFQ/Confirmation Check Off**

A	B	C	D	E	Total
.5522 (37/67)	.6552 (38/58)	.6491 (37/57)	.6301 (46/73)	.5733 (43/75)	.6091 (201/330)

- **Intra-Sentential Ordering by Semantic Affinity On**

A	B	C	D	E	Total
.5672 (38/67)	0.5 (30/58)	.667 (38/57)	.6164 (45/73)	0.6 (45/75)	.5900 (195/330)

- **Non-Stochastic Checks** [Nouns in Adj., Feat. Conflict, Same Pron. & COND-TFQ] **Off**

A	B	C	D	E	Total
.5373 (36/67)	.6724 (39/58)	.5965 (34/57)	.6027 (44/73)	.5733 (43/75)	.5939 (196/330)

- **Stochastic Checks Off**

A	B	C	D	E	Total
.4776 (32/67)	.5517 (32/58)	.4912 (28/57)	.5890 (43/73)	.4933 (37/75)	.5212 (172/330)

- **Baseline: All Checks Off** (Linear Order Only)

A	B	C	D	E	Total
.4328 (29/67)	.5517 (32/58)	.4035 (23/57)	.5342 (39/73)	.4933 (37/75)	.4848 (160/330)

- **Closed Test**

The success rate for the closed test with the same conditions as the cross validation (open experiment) with the best score is 0.7273 (240/330).

- **Summary**

Due to redundancies in screening, removing one kind of screening does not bring about much deterioration of the result. The **Case-Frame Check** (by a Case-Frame Database) and the **Same Pronoun Check** can be safely removed. The efficacy of the **Nouns in Adjective Use Check**, the **Pronoun-Antecedent Feature Conflict Check**, and the **COND in the Previous TFQ/Confirmation Check** seems small (if any).

Intra-sentential reordering by semantic affinity or grammatical functions has negative effects in the experiments above.

- **Success Rates with regard to Pronouns** (the best cross validation score)

Pron.	Success Rate
あそこ	0.0 (0/1)
あちら	1.0 (1/1)
ここ	.857 (6/7)
こちら	0.5 (51/102)
こっち	0.0 (0/1)
これ	.652 (15/24)
そこ	0.64 (32/50)
そちら	.606 (20/33)
そっち	0.0 (0/1)
それ	.704 (76/108)
彼	0.5 (1/2)

4.2. Zero Pronoun Anaphora Resolution

A-E each signifies a different way for 5-way cross validation tests. The scores are success rates from anaphora resolution. The "Total" scores on the right are cross validation summations. The seven sets of scores following "The Best Score" are obtained by removing one screening process, heuristic, or one kind of

.5354 (257/480)	.5131 (216/421)	.4322 (236/546)	.4953 (261/527)	.4959 (241/486)	.4923 (1211/2460)
-----------------	-----------------	-----------------	-----------------	-----------------	-------------------

- **Zero Pronoun-Antecedent Combination Check (Case-Based) Off**

A	B	C	D	E	Total
.4792 (230/480)	.4988 (210/421)	.3773 (206/546)	.4611 (243/527)	.4753 (231/486)	.4553 (1120/2460)

- **Case-Frame Check (Case-Based) Off**

A	B	C	D	E	Total
.4813 (231/480)	.4608 (194/421)	.3938 (215/546)	.4497 (237/527)	.4383 (213/486)	.4431 (1090/2460)

- **Screening with Case Frame Database Off**

A	B	C	D	E	Total
.5333 (256/480)	.5036 (212/421)	.4121 (225/546)	.4820 (254/527)	.4753 (231/486)	.4789 (1178/2460)

- **Nouns in Adjective Use Check Off**

A	B	C	D	E	Total
.5333 (256/480)	.5107 (215/421)	.4322 (236/546)	.4934 (260/527)	.4959 (241/486)	.4911 (1208/2460)

- **Zero Pronoun-Antecedent Feature Conflict Check Off**

A	B	C	D	E	Total
.5333 (256/480)	.5131 (216/421)	.4322 (236/546)	.4934 (260/527)	.4959 (241/486)	.4915 (1209/2460)

- **Intra-Sentential Ordering by Semantic Affinity Off**

A	B	C	D	E	Total
.5188 (249/480)	.4964 (209/421)	.4267 (233/546)	.4877 (257/527)	.4856 (236/486)	.4813 (1184/2460)

- **Intra-Sentential Ordering by Grammatical Function Off**

A	B	C	D	E	Total
.5375 (258/480)	.5202 (219/421)	.4231 (231/546)	.4972 (262/527)	.4877 (237/486)	.4907 (1207/2460)

- **Non-Stochastic Checks [Nouns in Adj., Feat. Conflict] Off**

A	B	C	D	E	Total
.5312 (255/480)	.5107 (215/421)	.4322 (236/546)	.4915 (259/527)	.4959 (241/486)	.4902 (1206/2460)

- **Stochastic Checks (+ Intra-Sentential Ordering by Semantic Affinity) Off**

A	B	C	D	E	Total
.2708 (130/480)	.3017 (127/421)	.2161 (118/546)	.2638 (139/527)	.2778 (135/486)	.2638 (649/2460)

- **Baseline: All Checks Off (Linear Order Only)**

A	B	C	D	E	Total
.1750 (84/480)	.2399 (101/421)	.1465 (80/546)	.1841 (97/527)	.1872 (91/486)	.1841 (453/2460)

- **Closed Test**

The success rate for the closed test with the same conditions as the cross validation (open experiment) with the best score is 0.7663 (1885/2460).

- **Summary**

Due to redundancies in screening, removing one kind of screening does not bring about much deterioration of the result. Unlike in the pronoun experiments, intra-sentential-ordering by semantic affinity or gram-

quired to attain more. As for heuristics, it may be of help to detect parallelism in the preceding utterances.

References for Part II

- Abraços et al. (1994) "Extending DRT with a Focusing Mechanism for Pronominal Anaphora and Ellipsis Resolution," In Proceedings of COLING'94, pp. 1128-1132.
- 荒川直哉 (1995) "代名詞の先行詞推定に関するセンタリング理論の評価," ATR-ITL Technical Report TR-IT-0141.
- Grosz B., Joshi A. and Weinstein S. (1995) "Centering: A Framework for Modeling the Coherence of Discourse," *Computational Linguistics*, Vol. 21, #2, pp. 203-225.
- Lappin S. and Leass H. (1994) An Algorithm for Pronominal Anaphora Resolution, *Computational Linguistics*, Vol. 20, #4, pp. 535-561.
- 村田真樹、長尾真 (1998) "表層表現と用例を用いた照応省略解析手法," 信学技報 (Technical Report of IECE), NLC97-56, pp. 9-16.
- 中岩浩巳 (1998) "日英対訳コーパス中のゼロ代名詞とその指示対象の自動認定," 自然言語処理 123-5, pp. 33-40.
- Nasukawa T. (1994) "Robust Method of Pronoun Resolution Using Full-Text Information," In Proceedings of COLING'94, pp. 1157-1163.
- 大野晋、浜西正人 (1981) *角川類語新辞典*, 角川書店, Tokyo, 932pp.
- Takada S. and Doi N. (1994) "Centering in Japanese: A Step Towards Better Interpretation of Pronouns and Zero-Pronouns," In Proceedings of COLING'94, pp. 1151-1156.
- 田村浩二、奥村学 (1995) "センター理論による日本語談話の省略解析," 自然言語処理 107-12, pp. 91-96.
- Walker M., Iida M. and Cote S. (1994) "Japanese Discourse and the Process of Centering," *Computational Linguistics*, Vol. 20 #2, pp. 193-231.
- Yamamoto K. and Sumita E. (1998) "Feasibility Study for Ellipsis Resolution in Dialogues by Machine-Learning Technique," In Proceedings of COLING'98, pp. 1428-1477.