Pronominal Anaphora Resolution in KANTOO English-to-Spanish Machine Translation System[1]

Teruko Mitamura, Eric Nyberg, Enrique Torrejon, David Svoboda, Kathryn Baker

Language Technologies Institute
Carnegie Mellon University
Pittsburgh, PA 15213
teruko@cs.cmu.edu

Abstract

We describe the automatic resolution of pronominal anaphora using KANT Controlled English (KCE) and the KANTOO English-to-Spanish MT system. Our algorithm is based on a robust, syntax-based approach that applies a set of restrictions and preferences to select the correct antecedent. We report a success rate of 89.6% on a training corpus with 289 anaphors, and 87.5% on held-out data containing 145 anaphors. Resolution of anaphors is important in translation, due to gender mismatches among languages; our approach translates anaphors to Spanish with 97.2% accuracy.

Keywords

Pronominal Anaphora Resolution, English to Spanish Machine Translation

Introduction

It has been claimed that encoding all the linguistic and domain knowledge required for successful pronominal anaphora resolution is a difficult and time-consuming task. Recent research in anaphora resolution seeks a simple, rapid, yet reliable approach which does not require extensive syntactic and semantic knowledge (Mitkov, 1998; Nasukawa 1994). In this work, we explore the extension of KANT Controlled English to include pronominal anaphors, and present an algorithm that effectively resolves anaphors while preserving the accuracy of the translated text. In contrast with statistical methods, our approach does not require a large bilingual aligned corpus for training.

Our resolution algorithm is syntax-based, but we also follow the trend toward a simple, rapid approach. Our approach draws from linguistic approaches in earlier work (Carbonell and Brown 1998; Lappin and McCord 1990; Lappin and Leass 1994; Ferrandez et al. 1998). These approaches require a full parse or partial parse of the input sentence. Our approach is a sequential, rule-based, domain-independent procedure, which has been implemented and integrated into the KANTOO English-Spanish MT system. The success rate for anaphora resolution is 89.6% on our training corpus and 87.5% on held-out data. The accuracy rate for translation of pronouns to Spanish is 97.2% on the held-out data.

Encoding the world knowledge and deep inference mechanisms required for selecting the right antecedent is a bottleneck in reaching 100% coverage in unrestricted texts for both "knowledge-poor" approaches (Dagan and Itai 1990; Kennedy and Boguraev 1996; Mitkov 1998; Nasukawa 1994) and linguistic approaches. Nasukawa pointed out the difficulty of achieving a success rate of over 90% without this type of knowledge, and our results support this claim.

Our algorithm is syntax-based, utilizing the f-structure that results from a full parse using an analysis grammar. The algorithm applies a set of well-known heuristics (constraints and preferences) used in "knowledge-poor" systems. However, it differs from previous approaches in that it does not calculate weights for the heuristics in order to choose the right antecedent; rather it applies heuristics in a sequential manner until one candidate antecedent remains. Since our evaluation indicates performance comparable to that of score-based, knowledge-poor systems, it can be inferred that adding more linguistic knowledge reduces the need for scoring procedures to prune incorrect antecedents. If necessary, semantic knowledge can be used once syntactic rules have been exhausted.

In the next sections, we explain the details of our resolution algorithm, present the results of an evaluation on texts drawn from technical manuals, and discuss some implications for current and future work.

KANTOO Anaphora Resolution

KANTOO is a knowledge-based, interlingual machine translation system, the most recent implementation of the

[1] Proceedings of MT Summit VIII, Santiago de Compostela, Spain, September 18-22, 2001.

original KANT MT system (Mitamura et al. 1991). KANTOO accepts Controlled English as input (Mitamura and Nyberg, 1995); the current input specification is referred to as KANT Controlled English (KCE). KCE places some restrictions on vocabulary and grammar; although some of the sentences in this study were rewritten to conform to KCE, we did not edit pronominal anaphors or any other constituents relevant to the anaphor resolution process.

Identification of Possible Antecedents

Possible antecedents for a given pronoun are identified according to a set of pre-defined constraints:

1. The candidate antecedent must be a noun, unit, tag, or conjoined NP.

2. If the antecedent is in same sentence as the pronoun, it must precede the pronoun.

3. If the antecedent is a conjoined NP, it must conjoin NPs with *and* or *or*.

4. Prune any antecedent that is a part of a coordination.

5. The pronoun and candidate antecedent must agree in number (a conjunction is implicitly considered plural).

6. If the pronoun is a verb argument, the antecedent may not be an argument of the same verb (note: we do not consider reflexives such as 'itself'.)

7. If the pronoun is the object of a prepositional phrase or relative clause modifying a noun, then that noun may not be a valid antecedent.

Ordered Heuristics for Antecedent Selection

After identifying the set of valid candidates, we apply the following heuristics, in order to select the preferred candidate. After each heuristic is applied, if the set of valid candidates contains only a single antecedent, it is selected, otherwise the next heuristic is applied. It is important to note that not every heuristic is tried for each anaphor, and sequential ordering is used to rank the heuristics. This is in direct contrast with approaches that try all heuristics on every anaphor, and use a weighted-sum scoring technique to make the final selection.

1. Prefer an antecedent that is also an anaphor.

2. Prefer an antecedent that is not a tag.

3. If two antecedents occur in this form: <np1> of <np2>, prefer <np1>. But if <np1> is one of "type/length/size/part", prefer <np2>.

4. Collocation: Prefer antecedents that attach to the same syntactic constituent as the pronoun.

5. Syntactic Parallelism: Prefer antecedents that attach to the same part of speech as the pronoun.

6. Syntactic Parallelism: Prefer antecedents that fill the same grammatical function as the pronoun.

7. Prefer antecedents that are conjunctions.

8. Definiteness-1: Prefer nominal antecedents that have a determiner, quantifier, or possessor, or are the value of a tag.

9. Definiteness-2: Prefer nominal antecedents that have a definite determiner.

10. Closeness: Prefer the last (most recent) antecedent.

Evaluation and Results

In order to tune the algorithm, we selected a training corpus from electronic product manuals. The corpus consists of 221 sentences containing 289 third person pronouns (*it, they*, and *them*) in inter- and intra-sentential positions. Roughly 27% of the pronouns are inter-sentential and 73% intra-sentential. The average number of candidate antecedents (noun, noun phrases, and pronouns) considered for the anaphora resolution in the corpus is 3.

The algorithm resolved the anaphors in the training corpus with a success rate of 89.6%, which is comparable to Mitkov (1998) and Nasukawa (1994).

	Correct	Total	Success rate
IT	194	219	88.5%
THEY	24	24	100 %
THEM	41	46	89.1%
Total	259	289	89.6%

Table 1: Results of Training Corpus

As shown in the Table 1, the anaphor *it* was correctly resolved with a success rate of 88.5%, the anaphor *them* with a success rate of 89.1% and all instances of the anaphor *they* were correctly resolved. The success rate for intra-sentential *it, they* and *them* is 89%, and for inter-sentential anaphors in general the success rate is 91%.

Results from Held-Out Data

We tested the algorithm on held-out data in a corpus containing 134 sentences with 145 pronouns. The distribution of intersentential and intrasentential pronouns was similar (25% and 75% respectively). The sentences were rewritten in KCE where necessary, making sure that this editing did not affect any pronoun or candidate antecedent needed for the evaluation. The average number of candidate antecedents in this corpus was also 3.

The algorithm resolved the anaphors in this corpus with a success rate of 87.5%, which is similar to the result obtained for the training corpus. As shown in the Table 2,

the anaphor *it* was correctly resolved with a success rate of 84.7%, and the anaphors *them* and *they* with a success rate of 91.6%. The success rate for intra-sentential anaphors is 86.1%, and for inter-sentential anaphors, 91.8%.

	Correct	Total	Success rate
IT	72	85	84.7%
THEY	22	24	91.6%
THEM	33	36	91.6%
Total	127	145	87.5%

Table 2: Results of Test Corpus

Evaluation of Spanish Translation Results

We also carried out an evaluation of the Spanish translation accuracy for these anaphors, to measure the postediting effort that would be imposed on translators if pronouns were admitted in KCE but incorrectly translated. Mismatches in gender systems among languages are of particular concern for MT accuracy. Spanish, for example, distinguishes between third person masculine and third person feminine pronouns. The pronoun *it* in direct object position is translated as *lo* if the antecedent is masculine, and *la* if it is feminine. Also, Spanish has a tendency to drop pronouns in subject position, if the antecedent can be inferred unambiguously from the context.

For the held-out data in the test corpus, the Spanish translation accuracy rate is 97.2%. There are 14 anaphors that are correctly translated, even though an incorrect antecedent was chosen by the resolution algorithm. This occurs when the selected antecedent happens to have the same gender in Spanish as the correct one, or when the anaphor is in subject position and gender is irrelevant when the translated pronoun is dropped (although the gender information must be correct if the verb phrase contains an attribute which must agree with the subject). For example, in the sentence below

Spots and smudges appear in the background areas of transparencies when ***they*** *are projected on the screen.*

suppose the anaphor *they* is wrongly associated with *spots and smudges* as the antecedent. The correct antecedent is *transparencies*. However, both antecedents are feminine, and the anaphor is correctly translated. Although it is dropped in the Spanish translation, the participle *projected* must agree in gender and number with the subject *they*. This is the case in the translation rendered by KANTOO:

Las marcas y las manchas aparecen en las áreas de fondo de transparencias cuando son proyectadas en la pantalla.

where *proyectadas*, feminine plural, agrees with the dropped subject of the verb *son*.

Many commercial Spanish MT systems translate every third person pronoun as masculine, without trying to identify the correct antecedent. If we consider the nouns listed in the Spanish Thesaurus compiled by Julio Casares (1996), we see that almost 40% of them are feminine. The implication is that translating always to a default choice (masculine) can at best achieve a success rate of 60% in similar cases.

The 60% (masculine) vs. 40% (feminine) gender distribution of the Spanish language is almost exactly the distribution found in the test corpus. We found that 40.1% of the correct antecedents in the corpus are feminine, and 59.8% of them are masculine. Since the success rate of KANTOO's algorithm is 97.2% when used with Spanish MT, the postediting effort required to fix wrong translations of anaphors is significantly less.

Comparison with Other Approaches

Since both Mitkov (1998) and Nasukawa (1994) tested their knowledge-poor algorithms on similar technical documentation (a printer manual and a computer user's guide respectively), we feel it is appropriate to compare these approaches with ours.

Nasukawa reports a success rate of 93.8% for a corpus containing 112 third-person pronouns. First, the algorithm applies constraints such as number and gender agreement. A set of preference scores is then generated for appropriate candidates. The candidate receiving the highest score is chosen. The algorithm implements three preferences: (1) the existence of collocation patterns with the modifiee of a pronoun in the source text; (2) the frequency of repetition of the candidates in the 10 previous sentences; and (3) the closest candidate to the pronoun in the same sentence or in the previous sentence.

In Nasukawa's approach the collocation pattern plays a decisive role, since it is set to a constant value of 3. While our algorithm makes use of collocation patterns in the same sentence or in the previous sentence, it does not use values or weights to calculate the right antecedent. In our algorithm, preference is implemented via rule ordering. Because of this strict rule ordering, there are cases where the rules can select the wrong antecedent, as shown below:

Before you begin printing envelopes, you can receive faxes to memory and then print ***them*** *after you have reloaded the normal paper.*

In this example, the correct antecedent for *them* is *faxes*, not *envelopes*.

Nasukawa's closeness preference relies on syntactic position. A higher score is assigned to an antecedent closer to the anaphor. This simple preference accounts for 82.1% of the correct resolutions, without using information on frequency of repetition or collocation patterns. This is significant, since Mitkov reports a much lower success rate (65.95%). The higher utility of this preference in Nasukawa's approach contributes to his reported success rate of 93.8%.

The KANTOO algorithm has a similar preference for the closest antecedent in the current or previous, when more than one candidate remains after applying the other heuristics. In the training corpus, this closeness preference plays a role in deciding 10 sentences (3.4%). This percentage is the "frequency of use" (Mitkov, 1998)[2]. If we also consider Mitkov's "discriminative power"[3] measurement, the discriminative power of Closeness in KANTOO is 100%, since Closeness selects the correct antecedent whenever it is applied. Mitkov reports 98.9% frequency of use for his version of Closeness, with 34.4% discriminative power. In Table 3, we compare the results of the three algorithms.

	Frequency of Use	Discriminative Power
Nasukawa	100.0 %	82.1%
Mitkov	98.9%	34.4%
KANTOO	3.4%	100.0%

Table 3: Closeness/Referential Distance

Mitkov (1998) also reports a high success rate of 89.7% in a corpus containing 104 pronouns from two different manuals. This is very similar to the KANTOO success rate and comparable to results reported for another syntax-based approach (Lappin and Leass, 1994), for which the success rate is 86%. Mitkov developed a robust "knowledge-poor" anaphor resolution algorithm, that uses a part-of-speech tagger and simple noun phrase rules, and applies a set of "antecedent indicators" with various scores. The candidate with the best aggregate score is chosen. A distance of 2 sentences is considered when finding candidate antecedents.

The success of this approach, according to the author, is due to the fact that 10 antecedent indicators are taken into account and "no factor is given absolute preference". It is clear that the success of this approach hinges upon the weights assigned to these indicators. For instance, there are two indicators that have the highest discriminative power in the evaluation performed by Mitkov: "collocation pattern" and "immediate reference". The former assigns a score of 2 to candidates that have an identical collocation pattern with a pronoun, and it has a discriminative power of 90.9%. The latter also assigns a score of 2 to the noun phrase immediately after a verb in a particular structure. This indicator has a discriminative power of 100%, that is, every time it is applied the antecedent is successfully identified. He also reports a frequency of use of 31.1% for this indicator, which means that this indicator alone helps to resolve a third of the 104 anaphors in the two corpora used in the evaluation. However, we find that the closeness preference, as implemented in KANTOO, can resolve the following sentence presented by Mitkov; KANTOO picks *printer* as the antecedent:

*To print the paper, you can stand the printer up or lay **it** flat.*

KANTOO performs well without implementing a scoring procedure for the set of preferences, and without using preferences like "indicating verb", "immediate reference" or "term preference" that would be too time-consuming to implement for a large technical domain.

On the other hand, KANTOO has difficulty in some cases where syntactic parallelism would select the wrong candidate, whereas the closeness heuristics or the definiteness heuristics would have a better chance of selecting the right antecedent. Such conflicts are unavoidable when sequential heuristics are used instead of a weighted sum. For instance, in the sentence below:

*The contrast setting affects the lightness or darkness of an outgoing fax as **it** is being sent.*

KANTOO's algorithm wrongly selects *the contrast setting* as the antecedent because of the strong preference that is placed on syntactic parallelism. Notice that the definiteness heuristics would also rule out the right antecedent, *fax*. The only heuristic that could choose *fax* is the closeness heuristic, but this heuristic is applied after syntactic parallelism. Interestingly enough, if Mitkov's algorithm is applied, the same wrong antecedent may be selected. This is because *fax* is penalized due to the "non-prepositional noun phrase" preference and "definiteness" preference.

A better strategy for resolving this example might be to implement semantic preferences or domain knowledge. However, in some cases, the preference for the closest antecedent resolved cases where some semantic preference would be necessary to pick the right antecedent. For instance, in the sentence,

*If the page is longer, the scanner continues to scan, but **it** only keeps the first 14 inches of data on the page and discards the rest.*

Pronoun *it* is correctly resolved to *scanner* using this preference. There is no need to implement domain information, such the fact that only scanners, printers, or fax machines can keep certain amount of data. However, as Mitkov points out, the longer and more complex the sentence, the higher the probability that this preference will fail.

Conclusion

Roughly 10% of antecedents chosen by KANTOO remain incorrect, and these examples would require domain knowledge or world knowledge for successful resolution. It seems apparent that both score-based, knowledge-poor systems and syntax-based systems cannot raise this ceiling of 90% accuracy unless world knowledge is somehow incorporated into the algorithm. However, our current results provide us with very high accuracy in Spanish translation, and the effort in post-editing is not

[2] "Number of non-zero applications" divided by "Number of anaphors"

[3] "Number of successful antecedent identifications when this indicator was applied" divided by "Number of applications of this indicator"

significantly affected, even though KANT Controlled English has been extended to allow pronouns.

References

Carbonell, J. and R. Brown (1988). Anaphora Resolution: A Multi-Strategy Approach. In Proceedings of the 12th International Conference on Computational Linguistics (COLING'88). Budapest, Hungary.

Dagan, Ido and A. Itai (1990). Automatic Processing of Large Corpora for the Resolution of Anaphora References. In Proceedings of the 13th International Conference on Computational Linguistics (COLING'90). Vol. 3, 330-332, Helsinki, Finland.

Diccionario Ideológico de la Lengua Española de Julio Casares (1996) Editorial Gustavo Gili, Barcelona.

Ferrández, A., M. Palomar, and L. Moreno (1998). Anaphora Resolution in Unrestricted Texts with Partial Parsing. In Proceedings of the 17th International Conference on Computational Linguistics (COLING-ACL'98), Montreal, Canada.

Kennedy, C. and B. Boguraev (1996). Anaphora for Everyone : Pronominal Anaphora Resolution without a Parser. In Proceedings of the 16th International Conference on Computational Linguistics (COLING'96). 113-118. Copenhagen, Denmark.

Lappin, S. and H.J. Leass (1994). An Algorithm for Pronominal Anaphora Resolution. Computational Linguistics, 20(4), 535-561.

Lappin, S. and M. McCord (1990). Anaphora Resolution in Slot Grammar. Computational Linguistics, 16 :4, 197-212.

Nasukawa, T. (1994). Robust Method of Pronoun Resolution Using Full-Text Information. In Proceedings of the 15th International Conference on Computational Linguistics (COING-94), 1157-1163, Kyoto Japan.

Mitkov, R. (1998). Robust Pronoun Resolution with Limited Knowledge. In Proceedings of the 17th International Conference on Computational Linguistics (COLING-ACL'98), Montreal, Canada.

Mitamura, T., E. Nyberg, and J. Carbonell (1991). An Efficient Interlingua Translation System for Multi-lingual Document Production. In Proceedings of the Third Machine Translation Summit, Washington, D.C.

Mitamura, T. and E. Nyberg (1995). Controlled English for Knowledge-Based MT : Experience with the KANT system. In Proceedings of the Sixth International Conference on Theoretical and Methodological Issues in Machine Translation. Leuven, Belgium.