A dependency-based approach to anaphora annotation

Eckhard Bick

University of Southern Denmark, eckhard.bick@mail.dk

Abstract. The paper describes a novel approach to the resolution of pronominal anaphora, where a hierarchy of linguist-written Constraint Grammar rules are used to add relational tags to anaphoric tokens in running text. Pro-drop subjects are covered as well as multi-sentence anaphoric chains. The system exploits function-marked dependency trees provided by a CG parser, and performs semantic tagging of pronouns as an intermediate step, based on verbo-nominal selection restrictions harvested from a collocation corpus. News text evaluation results are provided and compared across different anaphora types.

Keywords: Anaphora, rule-based annotation, dependency grammar, NLP, Constraint Grammar, semantical tagging of pronouns

1 Introduction

Anaphora fall into the realm of structural, rather than morphosyntactic annotation. Though subject to morphological restraints such as number and gender agreement between referent and antecedent, anaphora are hard to resolve without both a structural analysis and a semantic knowledge base. In an effort to capture relational tendencies rather than rules, most automatic systems for anaphora resolution use so-called salience weights, as suggested by Lappin & Leass (1994), which allow the computation of co-reference likelihood from local feature sets. In one such system, Ferrández, Palomar & Moreno (1998) use a Spanish tagger and their own SUG parser (Slot Unification Grammar) to provide syntactic information, reporting a pronominal anaphor recall of 83%. Though we agree in principle on the choice of relevant features and functions, we believe that local and structural salience criteria, as well as semantic features, should be exploited explicitly in a rule-based fashion, optimally using syntactic dependency links[1] and to-be-assigned anaphoric relations in parallel, within one and the same formalism. The dependency-based anaphora resolution method described in this paper is an extension of the Constraint Grammar (CG) formalism (Karlsson et al 1995), implemented with a modified version of GrammarSoft's open source CG3 compiler. The anaphora module builds on automatically analyzed output from the PALAVRAS parser (Bick 2000), which

1 We prefer dependency links to the otherwise information-equivalent constituent tree structures, because the former provide a more direct linking of tokens, and are thus structurally more akin to the phenomenon of anaphora as such. Also, token-to-token links are easier to follow across sentence boundaries.

provides both morphosyntactic tags and full dependency trees.

One of the most quoted anaphora resolution systems, and one that is fully automatic, with no need of input correction, is Mitkov's MARS system (Mitkov et al. 2002), reported to recover between 55% and 87% (upper bound) of pronominal co-referents, depending on text type. Though the MARS system is described as "knowledge-poor", the newest version does employ a Constraint Grammar-related dependency parser (FDG, Tapanainen & Järvinen, 1997) to instantiate these different criteria on antecedent candidates. Thus, the difference between MARS and our own CG approach does not so much reside in the criteria used, or in the depth of input analysis, but in the way anaphoric relations are *assigned:* Our grammar does not only exploit dependency edge labels (syntactic functions), but follows dependencies with explicit rules and – most important – generalizes the dependency formalism as a special type of relation, adding anaphoric relations of different kinds with the same kind of rule apparatus we employ for dependency and ordinary syntactic Constraint Grammar. Mitkov's principles have been adapted for Brazilian Portuguese in the RAPM system (Chaves & Rino 2008), with a reported success rate of 67% for 3rd person pronouns on raw text.

2 Anaphora types

The prototypical anaphor-relation is a direct individual relation between a pronoun and an np, but recent research has also focused on indirect and associative anaphor, where Vieira et al. (2007) report 30% correct resolution for Portuguese, zero-anaphora or abstract pronominal anaphora (Navaretta & Olsen 2008).

In our present work, we have focused on pronominal anaphora rather than np-co-reference, with machine translation (MT) in mind, that depends on anaphora resolution, for instance in order to generate English surface pronouns when translating zero-subject Portuguese verbs. Peral & Ferrández (2002) addressed this issue for the Spanish-English language pair, reporting 89% detection and 81% resolution for Spanish zero—anaphora, and 80% / 82% correctly resolved 3rd person personal pronouns for the two languages, respectively.

Pronominal classes differ with regard to syntactic reach, suggesting different resolution strategies. We treat the following (no. 5 only experimentally):

1. **Relative pronouns:** relevant for MT pronoun generation and propagation of semantic constraints to the relative clause
2. **Reflexive pronouns:** almost always intra-clausal, often zero-linked
3. **Possessive pronouns:** less local (and harder) than relatives and reflexives.
4. **Personal pronouns:** often cross-sentence. 'o' may refer to actions/events
5. **Non-prenominal demonstratives**, e.g. appositions or anaphoric subjects
6. **Adverbial pronouns:** Exploit the same rules as type-1 relatives. Among the deictics, 'lá' may have a textually-manifest, local or temporal, antecedent.

In our actual annotation scheme, co-reference is marked as token based ID-reference

tags, in the same fashion dependency relations are marked in the PALAVRAS input parses we use. Thus, “ID:14 R:poss:5” means that token 14 (ID:14) has an anaphoric relation (R:) of type 'possessive' (poss) to token number 5. The four non-adverbial classes have their own type marker ('R:rel', R:refl', 'R:poss', 'R:dem' and the default 'R:ref'), while adverbial pronouns may be either 'rel' (relative) or 'ref' (default). External “antecedents”, e.g. 2nd person, are attached to the null-token 0. So far, we only tag one non-pronominal type of co-reference between surface tokens – predicatives (“R:pred”). As a pro-drop language, Portuguese allows for non-expressed subjects, which can be regarded as a kind of zero-form pronouns, incorporated into the finite verb. In these cases, we introduce a link between the verb and its extra-clausal surface subject antecedent. The link is called r:subj from subject to verb, and r:e-subj (ellipsed subject) in the direction from verb to subject. The resulting anaphora links are potentially very long, since the surface antecedent may lie several sentences back, with several sentence-root verbs sharing the same subject referent.

3 The grammar

Our anaphora grammar is implemented as a separate CG module to be run on CG-compatible, syntactically analyzed input with dependency links. The existing PALVRAS input grammar can thus be regarded as a black box, and could in principle be replaced with another live system – or a treebank. The current grammar has about 200 rules, divided into a preparatory section (1/3) and the anaphora section itself (2/3). The main function of the preparatory section is to add information (tags) central to anaphora resolution, marking for instance:

- definiteness of noun phrases (£np-def and £np-idf)
- top-subjects (£top-subj, the subject immediately governed by the root-verb)
- animacy traces, more precisely ±HUM tags (£hum/£non-hum), for pronouns

Since PALAVRAS provides semantic prototype tags for nouns, as well as sophisticated named-entity annotation, animacy information is readily available for nouns, while animacy information for personal, possessive and reflexive pronouns has to be recovered through a combination of verbal selection restrictions and dependency links. PALAVRAS provides some indirect clues, such as certain valency classes, e.g. <vq> for que-governing cognitive verbs, and a not-fully implemented <vH> marking for human-subject potentiality. In order to include object and prepositional complements, and to achieve broad lexical coverage, we used dependency collocations from the DeepDict lexicon (Bick 2009) to extract *Framenet*-like selection restriction information, generalizing the semantic prototype class of collocates into a few major classes that were then used to create verbal selection tags, for instance <fACC/food:30>, meaning that a given verb has a 30% probability of a direct object (ACC) of the food class. These tags allow us to classify verbs, and to add animacy traces to their pronoun complements. In rule (1), a +human-threshold of minimum 70% (H>70) triggers a £hum tag for the pronoun dependent (PERS @P<) of 'com' (PRP-COM) in a prepositional object (@PIV), as in “falava com ela.”:

(1) ADD (£hum) TARGET PERS + @P<
(p @PIV LINK 0 PRP-COM LINK p (<fPRP-com/H>70>)) ;

The rules of the main section are ordered so that more specific contexts for a given anaphoric relation override more general ones. For instance, semantically restricted rules generally precede definiteness restricted rules and syntactically, subject-searching rules and top-level rules have hight priority. Distance weighting is implicit by scanning antecedent candidates right to left until one fulfills all conditions in the relevant rule. Currently, the maximum scan scope includes up to 2 sentences left of the target sentence. Relations longer than this limit can usually be recovered by relation propagation, using antecedents that are themselves pronouns, or subject-incorporating verbs, as stepping stones. Rule 2a sets an anaphoric relation (ref) for a 3rd person personal pronoun (PERS3) in the nominative (NOM) if the grammar has already marked it as human (£hum) and if is a direct dependent (p) of a declarative (STAtement) top verb. The TO field searches for the antecedent left of the sentence boundary (*-1 >>> LINK **1W) and defines it to be either a definite noun phrase head, a top-level subject or a subject-less top-level verb, whatever is found closest and fulfills the conditions of gender-number-agreement ($$GN) and human animacy[2].

(2a) SETRELATION (ref) TARGET @SUBJ + $$GN
(0 PERS3 + NOM) (0 (£hum)) (p @FS-STA)
TO (*-1 >>> LINK **-1W DEF-HEAD + $$GN OR TOP-SUBJ + $$GN
OR (<nosubj> @FS-STA) LINK 0 N-HUM-PERSON OR V-HUM) ;

Rule 2b covers elliptic-subject-anaphora and adds two relation tags, one on verb (e-subj), one on the subject antecedent (subj). The rule links a top-level verb with the closest top-level subject – if it is a +HUM noun phrase (N-HUM) of the right number.

(2b) SETRELATIONS (e-subj) (subj) TARGET IND + $$NUMBER
(NONE c @SUBJ&) (0 (<fmc>)) (0 V-HUM)
TO (*-1W TOP-SUBJ + $$NUMBER LINK 0 N-HUM) ;

Rules2c-d, demonstrate propagation from a pronoun antecedent to that pronoun's own antecedent (2c), or via a verbal “stepping-stone” to an elliptic-subject (2d)[3]:

(2c) SETRELATION (ref) TARGET PRON TO (r:ref PRON LINK r:* (*)) ;
(2d) SETRELATION (pred) TARGET PRON TO (r:pred V LINK r:e-subj (*)) ;

A special challenge were external referents, such as impersonal “subject” use of “se” or addressee-referring 3rd person pronouns in interviews/dialogue (“seu”), where the antecedent position “0” was used, as well as abstract anaphora with clausal antecedents, where the antecedent marker was tagged on verbs.

---

2 The rule follows standard CG3 shorthand for positions and LINKing, with 0 meaning “same token” and '-' meaning 'left'. Ordinarily a global search (*) would stop if any condition fails, but **W will continue to search the whole window span until all conditions are found true.

3 R:* means 'any relation type', and (*) means “any token”, so the propagation rules have no added conditons on the final, real antecedent – they just follow the stepping-stone link.

4 Evaluation

Annotation systems with hand-written rules have a slow learning curve, and while performance evaluation can aid development by identifying problematic areas, it can be difficult to judge exactly when the method's potential is sufficiently realized for a final evaluation. Thus,our anaphora system does now cover a wide range of anaphora types robustly on raw input, but on the other hand we have identified many problems that we think our method can solve, but which we haven't had the time to address yet.

We used internet text, news and Wikipedia for development, and three VEJA texts (~ 3,000 tokens) for the final evaluation, with a fourth 3000+ token sample, where only personal pronouns were evaluated. Since pronouns and head verbs are robustly PoS tagged, and hence easily identifiable for the anaphora rules, false positives are rare, meaning that performance can be reliably measured by recall alone, which was roughly identical to precision and F-score, the exception being nominal predicatives.

n=3064	Text 1 Rec. %	Text 2 Rec.%	Text 3 Rec. (/ Pr.) %	Text 4 Rec.	All Rec. (/ Pr.) %
Personal pronouns (n=38)	100	(n=0)	85.7	77,7	83.7
Possessives[4] (n=34)	100	58.3	89.4	-	79.4
Relative pronouns[5] (n=35)	90	81.8	100	-	91.4
Pronoun *se* (n=22)	77.7	66.6	85.7	-	77.2
Zero-subjects (n=74)	54.5.	68.4	82.3	-	70.6
Predicatives (n=86)	77.4	70	88.5 / 81.5	-	80.2 / 77.5

Two words of caution are in order: First, a “soft” evaluation method was used (output inspection by one evaluator only). Second, the relatively low incidence of relevant anaphora types (partially compensated by using extra data for personal pronouns), and the considerable cross-text variance in performance limits the statistical representativeness of our preliminary results, which also seem to be quite text type dependent. For example, the first text was an interview, with 1st and 2nd person forms, and syntactically “false” subjects from pre-added interviewer/-ee names, interfering with subject-involving anaphora. Still, results are encouraging, and comparable to figures quoted for the English (MARS) and Portuguese (RAPM) systems cited above.

A closer look at individual pronoun types reveals that relative pronouns were easiest, and zero-subject-anaphora most difficult, probably due to their respective link lengths. It has to be born in mind that syntactic/dependency ambiguities and ensuing disambiguation errors may often be the true reason for anaphoric resolution failure, e.g. when relatives are preceded by chains of postnominal prepositional phrases.

Finally, though this is a hardware-dependent measure, it should be added that our

4 This category includes both “seu” possessives, and Brazilian postnominal “dele”, “dela” etc.

5 Adverbial relatives with a relative antecedent were included in this category

system is fairly fast, with 250 words per second on a server-grade unix machine.

5. Conclusion

We have shown that a Constraint Grammar-based method for anaphora resolution can be very effective not only in exploiting, in a malleable and precise way, various types of syntactic and semantic information from a parser, but also in adding and adapting such information to better suit the needs of an anaphora annotator. Given the fact that rules can be molded to cater for a wide variety of structures and text types, and that existing semantic information could be used much more specifically, we believe that there is substantial room for grammar improvement. Computed as an average of all types, accuracy in our pilot evaluation was an encouraging 81.3% for non-demonstrative pronominal anaphora (86.8 when counting only surface pronouns), but future work should include a thorough evaluation against a larger, revised corpus.

Acknowledgements. Implementation of our anaphora grammar called for several changes in the open source CG3 compiler, and we would like to thank Tino Didriksen (GrammarSoft), who programmed the necessary code additions and has reacted swiftly and competently to numerous suggestions and bug reports.

References

1. Bick, Eckhard: The Parsing System Palavras - Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Famework. Aarhus: Aarhus University Press (2000)
2. Bick, Eckhard (2009). DeepDict - A Graphical Corpus-based Dictionary of Word Relations". *Proceedings of NODALIDA 2009.* NEALT Proceedings Series Vol. 4. pp. 268-271. Tartu: Tartu University Library. ISSN 1736-6305
3. Chaves, A, Rino, L: The Mitkov Algorithm for Anaphora Resolution in Portuguese. Proceedings of PROPOR 2008, Aveiro. Berlin: Springer. LNCS serues, pp. 51-60 (2008)
4. Ferrández, A., Palomar, M., Moreno, L.: Anaphor Resolution in Unrestricted Texts with Partial Parsing. In: Proceedings of COLING-ACL 1998: pp. 385-391 (1998)
5. Karlsson et al.: Constraint Grammar - A Language-Independent System for Parsing Unrestricted Text. Natural Language Processing, No 4. Berlin: Mouton de Gruyter (1995)
6. Lappin, S., Leass, H.J.: An Algorithm for Pronominal Anaphora Resolution. Comutational Lingistics 20 (1994). pp. 536-561
7. Mitkov, R., Evans, R., Orasan, C.: A New, Fully Automatic Version of Mitkov's Knowledge-Poor Pronoun Resolution Method Source Lecture Notes In Computer Science; Vol. 2276. pp. 168-186 (2002)
8. Navaretta, C., Olsen, S.: Annotating abstract pronominal anaphora in the DAD project. Proceedings of LREC 2008 (7 pages).
9. Peral, J., Ferrández, A.: Pronominal anaphora generation in an English-Spanish MT approach. Lecture Notes In Computer Science; Vol. 2276. pp 187-196 (2002)
10. Tapanainen, P., Järvinen, T.: A Non-Projective Dependency Parser. In: Proceedings of the 5th Conference of Applied Natural Language Processing. Pp 64-71, ACL (1997)
11. Vieira, R.: Semantic tagging for resolution of indirect anaphora. Proceedings of 7th SIGdial Workshop on Discourse and Dialogue. Pp76-79. Sydney (2006)