Ambiguity in Anaphora Workshop Proceedings

ESSLLI 2006, Málaga, Spain, 7–11 August 2006

Edited by Ron Artstein and Massimo Poesio
(who also organized the workshop)

http://cswww.essex.ac.uk/anaphora/

Contents

Ambiguity in Anaphora: Introduction

Massimo Poesio and Ron Artstein

Our original idea for a workshop on Ambiguity in Anaphora was to create a dialogue between computational linguists, theoretical linguists and psycholinguists. The mainstream work in the computational community assumes that anaphora resolution is a problem of identifying a unique (or best) interpretation for each anaphoric expression. We feel that this view overlooks cases where reference is genuinely ambiguous, and the ARRAU project set out to investigate anaphoric ambiguity using both corpus annotation and eye-tracking experiments (Poesio et al., in press). This workshop brings together researchers who work on ambiguity from a variety of perspectives, with the hope that ideas can find their way from one field to another.

Indeed, the workshop represents a variety of theoretical and methodological approaches. Two papers are very close to our own corpus annotation research: Mareile Knees uses corpus annotation to identify multiple references of the temporal anaphor *danach* in German, and Yannick Versley shows that expressions whose referents are vague are more likely to cause disagreement among annotators. Four papers report on experimental psycholinguistic studies, using a variety of methodologies: Laura Kertz, Andrew Kehler, and Jeffrey Elman use a question-answering task to argue for a model of coherence for pronoun resolution in English; Arnout Koornneef, Frank Wijnen, and Eric Reuland use eye tracking to determine preferences for bound variable or coreferent readings of pronouns in Dutch; Gerlof Bouma and Holger Hopp use a web-based preference task to determine that grammatical function rather than linear order determines pronoun interpretation in German; and Claire Delle Luche, Roger van Gompel, Frédérique Gayraud, and Bruno Martinie use a preference task to show that markedness rather than informativeness determines the interpretation of relative pronouns in French, thus also determining relative clause attachment. One paper looks at the implications of psycholinguistic models on computational implementations: Lucas Champollion shows that modeling the different psychological behavior of pronouns in main clauses and subordinate clauses has a very small impact on the performance of anaphora-resolution systems. And finally, three papers look at the theory behind anaphoric behavior: Laia Mayol presents a game-theoretical model for the interpretation of pronouns in Catalan; Maia Duguine develops an information structure account of asymmetries between strict and sloppy readings of pronouns in English and Basque; and Alexandre Denis, Guillaume Pitel, and Matthieu Quignard give a model for the interpretation of plural references using reference domains.

We thus have an impressive array of papers for this workshop, which should result in much fruitful discussion. We wish to thank our speakers for submitting their work and coming to Màlaga to present it. We also want to use this opportunity to invite the workshop participants to submit their work to a special issue of *Research on Language and Computation* which we are editing; see announcement at the end of this volume.

Acknowledgments

We thank the following individuals and institutions who have made this workshop possible: the ESSLLI 2006 program committee and local organizers for all their assistance in hosting

this workshop; our reviewers Jennifer Arnold, Chris Barker, Kees van Deemter, Paul Elbourne, Ruth Filik, Alan Garnham, Klaus von Heusinger, Véronique Hoste, Christer Johansson, Elsi Kaiser, Ruslan Mitkov, Christoph Müller, Rebecca Passonneau, Tanya Reinhart, Tony Sanford, Patrick Sturt, and Bonnie Webber; and EPSRC grant GR/S76434/01 "ARRAU" (Anaphora Resolution and Underspecification) for financial support.

References

Massimo Poesio, Patrick Sturt, Ron Artstein, and Ruth Filik. Underspecification and anaphora: Theoretical issues and preliminary evidence. *Discourse Processes*, in press. Distributed as Technical Report CSM-438, University of Essex Dept. of Computer Science, October 2005.

Ambiguity in Anaphora Workshop Proceedings, pages 5–12
ESSLLI 2006, Málaga, Spain, 7–11 August 2006
Edited by Ron Artstein and Massimo Poesio
http://cswww.essex.ac.uk/anaphora/

Effects of Word Order and Grammatical Function on Pronoun Resolution in German

Gerlof Bouma and Holger Hopp
Center for Language and Cognition (CLCG),
Rijksuniversiteit Groningen
gerlof [at] let.rug.nl, h.c.hopp [at] rug.nl

Abstract

This paper presents novel psycholinguistic evidence on the factors governing pronoun resolution in German. To determine the relative influence of linear order versus grammatical function of potential referents in the German Mittelfeld on pronoun interpretation, an interpretation-preference task was run on 23 German natives. Subjects chose the preferred referent for fully ambiguous personal pronouns. The results across different verb types show that grammatical function, yet not linear order, predicts pronoun resolution in German.

Introduction

Anaphors constitute a prime example of ambiguity in language. Many factors have been proposed that influence the interpretation of anaphora, and pronouns in particular. We can distinguish three general types of factors: (a) world knowledge considerations such as the plausibility of the resulting reading; (b) linguistic constraints like agreement or binding; and (c) salience or activation of available referents. With respect to (c), researchers have proposed NP-form, distance to the anaphor, position in the sentence and grammatical function as form aspects that influence salience in different languages. One particular issue that has received attention is the relative impact of linear order (LO) and grammatical function (GF).

For instance, Rambow (1993) claims that in the German *Mittelfeld* – the topological field between the finite verb in second position, and the sentence final verb cluster – linear order determines which NP is most likely to be the antecedent of a pronoun in subsequent discourse. In particular, Rambow claims that in (1) and (2) the predictive role of LO is borne out, outweighing any potential effects of GF:

(1) Q Glauben sie, dass...
Do you think that...
eine solche Maßnahme$_1$ der russischen Wirtschaft$_2$ helfen kann?
a such measure.F.NOM the Russian economy.F.DAT help can
such a measure could help the Russian economy?
A Nein, sie$_1$ ist viel zu primitiv.
no it.F is much too primitive
No, the measure is much too primitive.

(2) Q Glauben sie, dass der russischen Wirtschaft$_2$ eine solche Maßnahme$_1$ helfen kann?
A Nein, sie$_2$ ist viel zu primitiv.
No, the economy is much too primitive.

The grammatical functions, subject and indirect (dative) object, do not change between (1) and (2). However, according to Rambow, the interpretation of the pronoun in these answers is always the leftmost Mittelfeld NP of the preceding sentence, which does change between the examples.

Note that the pronoun *sie* in (1) and (2) remains ambiguous: interpretation sie_2 is readily available in (1) as is sie_1 in (2). In other words, preferences for either interpretation are not categorical; rather, they reflect tendencies potentially based on factors like GF and LO. It is notoriously difficult to capture these tendencies using isolated examples. Yet, previous experimental and corpus research has found statistical evidence of the reality of such factors.

This paper adds to this body of research by presenting new empirical evidence on the relative influences of LO and GF on pronoun resolution in relation to the German Mittelfeld. We carried out an interpretation-preference task to investigate the relative impact of either factor. This paper is structured as follows: We briefly review some previous corpus and experimental research on pronoun resolution preferences across languages (Section 1). Then we present the experiment and its findings (Section 2). In Section 3, we discuss the results and offer suggestions for future research.

1 Previous results on LO vs GF

Much work on salience is related to, or based on Centering Theory, a theory of local discourse coherence (Grosz, Joshi, and Weinstein, 1995, CT). Originally, CT proposed GF as the main determinant of salience: subjects are more salient than objects, which are more salient than other elements in an utterance. Although CT was not intended as a theory of anaphora resolution, it has inspired many such theories and systems. A well-known early example is Brennan, Friedman, and Pollard (1987, BFP), who implemented CT for pronoun resolution, and operationalized GF by ranking entities according to a refined obliqueness scale.

Gordon, Grosz, and Gillom (1993) found increased reading times for repeated full NP versus pronoun subjects if the referent was the subject or the first mentioned NP of the previous utterance. They conclude that LO is a salience factor in English. Moreover, they speculate that any GF effect might be due to subjects often being topics.

Rambow (1993) claims LO determines salience in the German Mittelfeld. Also arguing for the importance of LO, Strube and Hahn (1999) propose a reformulation of Centering Theory based on Prince's familiarity categorization of entities (1981). The system does not use GF information. Empirically, their model outperforms BFP, both on a German and an English corpus.

Of course, the comparatively fixed word order of English leads to a close correspondence of linear order and grammatical function. In a corpus based comparison of CT variants, Poesio et al. (2004) find that substituting GF information by LO has hardly any effect. Thus, free word order languages are better suited to provide evidence that differentiates between LO or GF. Contrastive research on pronoun resolution in such languages seems to point at GF as a determinant of pronoun resolution, or at best to a mixed GF & LO model.

Järvikivi et al. (2005) present eye tracking data for Finnish. Both subjecthood (GF) and first mention (LO) influence the order and number of fixations on a picture of the stories being read. However, the data suggest that subjecthood is the stronger factor. In an interpretation study, Kaiser (2003) concludes that Finnish personal pronouns refer to the subjects, initial or not, unless the object was both initial and pronominal. For Turkish, Hoffman (1998) and Turan (1998) claim subjecthood rather than LO determines salience, based on the introspective

interpretation and acceptability of null pronouns. For Hindi, Prasad and Strube (2000) argue for GF and against LO on the basis of individual examples.

Against the background of these inconclusive and conflicting findings, this study presents psycholinguistic evidence on pronoun resolution in relation to the German Mittelfeld. By using scrambling in the Mittelfeld rather than movement to sentence initial position, this study avoids potential confounds of first mention effects and/or the particular information-structural effects of topicalization on pronoun resolution.

2 The Present Study

German, an SOV language with verb-second order in main clauses, allows comparatively free NP order. Apart from argument *topicalization* to the front of the main clause, German allows for subject-object and object-subject orders in the Mittelfeld (e.g. (1) and (2) above), so-called *scrambling*. Scrambling is constrained by various factors, such as definiteness, animacy, and information structure (Müller, 1999). Constituents scramble felicitously only when they are defocussed, i.e. when they constitute given information (Haider and Rosengren, 1998; Lenerz, 1977). In other contexts, scrambling of objects across subjects is distinctly marked.

2.1 Materials

In order to test whether GF or LO in the German Mittelfeld determine pronoun resolution, we designed an interpretation-preference task. In the task, subjects indicated the preferred referent of a pronoun in relation to a preceding sentence. For this sentence, the task manipulated the linear order and the grammatical function of NPs. The factor LO (left-right) was crossed with GF in three conditions: subject-direct object (3), subject-indirect object (4), and direct object-indirect object (5). Note that none of the relevant NPs is first mentioned.

(3) subject-direct object (su-do):

a. Die Hoffnung war, dass [der Beschluss]$_{su}$ [den Plan]$_{do}$ beeinflussen würde.
The hope was that the decision.NOM the plan.ACC affect would

b. Die Hoffnung war, dass [den Plan]$_{do}$ [der Beschluss]$_{su}$ beeinflussen würde.

(4) subject-indirect object (su-io):

a. Alle dachten, dass [der Sohn]$_{su}$ [dem Vater]$_{io}$ ähnelte.
All thought that the son.NOM the father.DAT resembled

b. Alle dachten, dass [dem Vater]$_{io}$ [der Sohn]$_{su}$ ähnelte.

(5) direct object-indirect object (do-io):

a. Die Professorin stellte [dem Kollegen]$_{io}$ [den Studenten]$_{do}$ vor.
The professor introduced the colleague.DAT the student.ACC part.

b. Die Professorin stellte [den Studenten]$_{do}$ [dem Kollegen]$_{io}$ vor.

There were six items per condition, 18 items in total. All verbs were of the canonical su-obj order; the ditransitives were verbs of the su-io-do class (Haider, 1993). In the design of the items care was taken to avoid potential interfering effects on pronoun interpretation: All items were controlled and matched for animacy, definiteness, number and gender, and they all displayed unambiguous case-marking. Further, the items were preceded by a context sentence that introduced both NPs. This context sentence neutralized information-structural differences between the NPs and, by making the NPs given information, established the contextual requirements for felicitous scrambling. The items and their preceding contexts were followed

by a sentence with a referentially fully ambiguous pronoun and a question to be answered by the subjects. For (3) this would be:

(6) Aber er war zu unstrukturiert.
But it was too unstructured.

(7) Was war zu unstrukturiert? (a) Der Plan (b) Der Beschluss (c) Etwas anderes
What was too unstructured? (a) The plan (b) The decision (c) Something else

To control for plausibility of interpretation and to test for lexical biases in pronoun preferences, all items were tested in a plausibility-rating study with a separate group of 12 German natives. In the plausibility test, the pronoun in sentence (6) was replaced by the full noun phrases. Subjects were presented with both options (as in (8) for the sentence in (3)) and they had to indicate which sentence, if any, seemed more plausible.

(8) a. Aber der Plan war zu unstrukturiert.
But the plan was too unstructured.
b. Aber der Beschluss war zu unstrukturiert.
But the decision was too unstructured.

Due to a bias towards one of the readings, four items were excluded from further analysis. For the remaining 14 items, order and syntactic function of the NPs constituted the only differences for pronoun resolution.

2.2 Subjects and Procedure

Twenty-three German adult native speakers took part in the experiment. The subjects were recruited individually by e-mail. The task was untimed and conducted via the Internet. Subjects logged on to the experimental web site and filled in some biographical information. Subjects were told that they would read ambiguous sentences and were asked to give their preferred interpretation by choosing an answering to a question about the items. A complete experimental example as seen by the subjects for item (3), looked like the following (but without the italicized gloss):

(9) Das Gremium fasste einen Beschluss, der den Plan zur Umsatzsteigerung ändern sollte.
The board made a decision that should change the plan for increasing turnover.

Die Hoffnung war, dass der Beschluss den Plan beeinflussen würde.
It was hoped that the decision would affect the plan (canonical word order).

Aber er war zu unstrukturiert.
But it was too unstructured

Was war zu unstrukturiert?"
What was too unstructured?

a. Der Beschluss
The decision
b. Der Plan
The plan
c. Etwas anderes
Something else

The experimental items were interspersed with 18 filler items of various other structures involving ambiguous and unambiguous pronouns, and pseudo-randomized. Two lists were used which counterbalanced the order of items, the order of answering options and scrambling.

2.3 Results

The results are given in Tables 1 and 2. In these tables, we refer to grammatical function in terms of obliqueness, with the subject being least oblique and the indirect object most oblique. Table 1 shows that for the su-do condition, the subject was chosen as the preferred antecedent of the fully ambiguous pronoun roughly two-thirds of the time, whereas the leftmost NP was chosen at levels indistinguishable from chance. In the su-io condition, the subject was also preferred as the pronoun antecedent; again, the leftmost NP was chosen at chance levels. The same finding obtains if we collapse the su-do and the su-io conditions to see the effects of subjecthood vs objecthood (see column *su-X*). For ditransitive verbs, the rightmost column in Table 1 demonstrates that neither obliqueness nor leftmost NP accounts for more than 50% of pronoun resolution. In subsequent statistical analyses, three pairwise Wilcoxon Signed-Rank tests were run for each condition: Subject vs Leftmost NP, Subject vs Object (for do-io: Direct Object vs Leftmost NP and Direct Object vs Indirect Object, respectively) and Leftmost NP vs Rightmost NP. For the su-do condition, only the Subject vs Object comparison showed a significant difference ($z = -2.425$, $p = 0.015$). For the su-io condition, the Subject vs Leftmost NP ($z = -2.517$, $p = 0.012$) and the Subject vs Object comparison ($z = -3.823$, $p < 0.001$) reached significance. For the combined subject-object conditions (*su-X*) , both Subject vs Leftmost NP ($z = 2.862$, $p = 0.004$) and the Subject vs Object ($z = -4.509$, $p < 0.001$) comparisons showed significant differences. Across conditions, none of the Leftmost vs Rightmost NP comparisons reached significance. In the do-io condition, no comparison yielded significant differences. In sum, subjecthood turns out to be a better predictor of anaphoric reference than linear order, irrespective of the obliqueness of the object. For object antecedents, neither grammatical function nor linear order seems to govern pronoun resolution.

To show the lack of relation between anaphoric reference and word order in more detail, Table 2 presents a breakdown of the results by condition and word order. Table 2 shows that the subject preference does not interact with reordering, i.e. even when the more oblique noun phrase shifts leftwards, anaphoric preferences do not shift leftwards. Instead, anaphor resolution preferences remain oriented to the subject in the su-do and su-io conditions; in the do-io condition, the preferences remain indeterminate. This further demonstrates that linear order does not underlie pronoun resolution in relation to NPs in the German Mittelfeld.

		Condition			
Resolved to		su-do	su-io	*su-X*	do-io
least oblique		44	78	122	68
	%	64.7	67.8	66.7	49.3
leftmost		36	59	95	69
	%	52.9	51.3	51.9	50.0
#Items		68	115	183	138

Table 1: Comparison of obliqueness & leftmost NP preferences

		Least Oblique	
Condition	Word order	#yes	#no
su-do	su left	22	13
	do left	22	11
su-io	su left	41	17
	io left	37	20
do-io	do left	33	36
	io left	35	34

Table 2: Interaction between LO & GF

3 Discussion and Conclusion

The results across different types of NP arguments in the German Mittelfeld indicate that GF, or, more specifically, subjecthood is a predictor of pronoun resolution. Even though referentially fully ambiguous, pronouns were found to refer at levels above chance to the subject of a preceding embedded clause that offered multiple potential antecedents. The finding that subjects are chosen as antecedents roughly two-thirds of the time shows that there is a robust resolution preference that nevertheless is not categorical. GF turned out not to have an effect on pronoun resolution for object-object ambiguities. Counter to the claims by Rambow (1993), LO was not found to determine pronoun resolution in relation to the German Mittelfeld. In fact, in the present study, there is no evidence to suggest that LO affects resolution at all.

With respect to the role of subjecthood, these results are in line with the findings by Järvikivi et al. (2005), Kaiser (2003) for Finnish and Hoffman (1998) and Turan (1998) for Turkish. However, in their eye-tracking experiments, Järvikivi et al. (2005) tested for differences between subjecthood and first-mention effects, and not linear order. In a similar vein, the effects of LO, first-mention and topicalization potentially overlap and interact in the other studies. The present study avoids the interference of first-mention or topicality effects with LO, and the lack of LO effects found in the present study might suggest that LO effects previously reported in the literature could be due to these interfering factors. Future research on e.g. topicalization in German etc., will show whether this tentative suggestion is on the right track and holds cross-linguistically.

However, testing anaphoric preferences with scrambling in the German Mittelfeld also faces several potential problems. First, scrambling in German is marked and infrequent. Secondly, scrambling is restricted to specific discourse contexts. In particular scrambling of objects across subjects is a marked and infrequent reordering option in German. Hence, it could be that linear order effects in this study are affected or even masked by the frequency bias against scrambling. While it is difficult to isolate the effects of frequency, there is some evidence that frequency does not modulate linear order preferences in pronoun resolution with respect to the German Mittelfeld. In a corpus study on NP order using the NEGRA II corpus that consists of about 20.000 written sentences, Kempen and Harbusch (2003) report that there are huge frequency differences between accusative-initial and dative-initial orders compared to nominative-initial orders of full NPs in the German Mittelfeld. Compared to 513 nominative-accusative (i.e. su-do) orders, there is only one case of an accusative-nominative (do-su) order, for dative-marked indirect objects, there are 20 cases of dative-nominative (io-

su) orders compared to 43 nominative-dative (su-io) orders. This corresponds to a ratio of roughly 1 to 500 for do-su orders and roughly 1 to 2 for io-su orders. If frequency differences of scrambling affected linear order effects in pronoun resolution, we would thus expect to see a concomitant difference in anaphoric preferences between the su-do and the su-io conditions. Yet, the figures in Tables 1 & 2 demonstrate that there is no such difference between conditions. Moreover, Kempen and Harbusch (2003) note that, for ditransitive verbs, io-do orders by far outnumber do-io orders in the corpus (14 to 3). However, anaphoric preferences do not shift depending on NP ordering in the do-io condition (see Table 2). Consequently, we conclude that frequency differences of NP order in the German Mittelfeld do not modulate pronoun resolution preferences and thus cannot account for the lack of LO effects attested in this study.

As for the discourse requirements, scrambling in German is felicitous only in particular discourse contexts, namely those in which the scrambled constituent denotes given information. This requirement on scrambling was met in the present study in that all relevant NPs were given in preceding discourse contexts as in previous studies (Scheepers, Hemforth, and Konieczny, 2000). This way, information-structural differences between these NPs were neutralized, so that the effects of word order could be isolated. The prototypical case of Mittelfeld scrambling, however, is arguably when a given object NP fronts across an information-structurally new (and focussed) constituent (Müller, 1999). In future research, it would be interesting to vary the information-structural contexts of ambiguous pronouns systematically to test for potential interactions between word order and information structure in anaphor resolution.

Ideally, experimental psycholinguistic studies on pronoun resolution as this one should be supplemented by corpus research to determine whether the same factors mandate anaphoric preferences in comprehension and naturalistic production. However, the multitude of potential relevant variables that need to be controlled for in conjunction with the rarity of scrambling in German will most likely lead to problems of data sparseness. Worse still, identifying the exact properties of contexts in a natural language corpus is difficult and subject to inter-annotator variation. In future research, we hope that these problems can be overcome so that we gain a more comprehensive understanding of pronoun interpretation.

Acknowledgments

As with the antecedents of pronouns, linear order of the authors' names does not correlate with their respective contributions to the paper. We would like to thank three anonymous reviewers for helpful comments on an earlier version of this paper. The first author gratefully acknowledges financial support by the Netherlands Organization for Scientific Research (NWO), grant number 051.02.071, as part of the NWO Cognition Programme.

References

Brennan, Susan E., M.W. Friedman, and Carl J. Pollard. 1987. A centering approach to pronouns. In *Proceedings of the 25th Annual Meeting of the ACL*, pages 155–162.

Gordon, Peter, Barbara Grosz, and Laura Gillom. 1993. Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, 17(3):311–347.

Grosz, Barbara, Aravind Joshi, and Scott Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):203–225.

Haider, Hubert. 1993. *Deutsche Syntax, generativ*. Günter Narr, Tübingen.

Haider, Hubert and Inger Rosengren. 1998. *Scrambling*, volume 49 of *Sprache und Grammatik*. Lund University.

Hoffman, Beryl. 1998. Word order, information structure, and centering in Turkish. In M. Walker, A. Joshi, and E. Prince, editors, *Centering Theory in Discourse*. Clarendon University Press, Oxford, pages 251–273.

Järvikivi, Juhani, Roger van Gompel, Jukka Hyönä, and Raymond Bertram. 2005. Ambiguous pronoun resolution. contrasting the first-mention and subject-preference accounts. *Psychological Science*, 16(4).

Kaiser, Elsi. 2003. Word order, grammatical function, and referential form: On the patterns of anaphoric reference in Finnish. *Nordlyd*, 31(1):245–260.

Kempen, Gerard and Karin Harbusch. 2003. An artificial opposition between grammaticality and frequency: comment on Bornkessel, Schlesewsky & Friederici (2002). *Cognition*, 90:205–210.

Lenerz, Jürgen. 1977. *Zur Abfolge Nominaler Satzglieder im Deutschen*. Günter Narr, Tübingen.

Müller, Gereon. 1999. Optimality, markedness and word order in German. *Linguistics*, 37:777–818.

Poesio, Massimo, Rosemary Stevenson, Barbara di Eugenio, and Janet Hitzeman. 2004. Centering: A parametric theory and its instantiations. *Computational Linguistics*, 30(3):309–363.

Prasad, Rashmi and Michael Strube. 2000. Discourse salience and pronoun resolution in Hindi. In *Penn Working Papers in Linguistics 6.3*. University of Pennsylvania, pages 189–208.

Prince, Ellen. 1981. Toward a taxonomy of given-new information. In P. Cole, editor, *Radical Pragmatics*. Academic Press, New York, pages 223–256.

Rambow, Owen. 1993. Pragmatic aspects of scrambling and topicalization in German: A Centering approach. MS, University of Pennsylvania.

Scheepers, Christoph, Barbara Hemforth, and Lars Konieczny. 2000. Linking syntactic functions with thematic roles: psych-verbs and the resolution of subject-object ambiguity. In B. Hemforth and L. Konieczny, editors, *German sentence processing*. Kluwer, Dordrecht, pages 95–135.

Strube, Michael and Udo Hahn. 1999. Functional Centering. grounding referential coherence in information structure. *Computational Linguistics*, 25(3):309–344.

Turan, Ümit Deniz. 1998. Ranking forward-looking centers in Turkish: Universal and language-specific properties. In A Joshi M. Walker and E. Prince, editors, *Centering Theory in Discourse*. Clarendon University Press, Oxford, pages 139–161.

Ambiguity in Anaphora Workshop Proceedings, pages 13–21
ESSLLI 2006, Málaga, Spain, 7–11 August 2006
Edited by Ron Artstein and Massimo Poesio
http://cswww.essex.ac.uk/anaphora/

On the (Ir)relevance of Psycholinguistics for Anaphora Resolution

Lucas Champollion
champoll [at] ling.upenn.edu

Abstract

Psycholinguistic experiments show that pronouns tend to be resolved differently depending on whether they occur in main or subordinate clauses. If a pronoun in a subordinate clause has more than one potential antecedent in the main clause, then the pronoun tends to refer to the antecedent which has a certain thematic role (depending on the verb and on the subordinating conjunction). In contrast, pronouns in main clauses tend to refer back to the subject of the previous main clause, and this tendency is not affected by any verbs or conjunctions. In natural language processing, these findings have recently led to a proposal that pronoun resolution systems should have a split architecture, i.e. that they should use different mechanisms for pronoun resolution in the two cases.

With the help of two parsed and coreference-annotated corpora, this paper estimates the impact of the split-architecture proposal. The findings of this work are as follows: (1) Subject pronouns in authentic texts behave the same way in main and subordinate clauses. (2) The number of sentences in which a split architecture would behave differently than a system that treats both cases the same way is close to zero. Therefore, a separate treatment of resolution within and across units is unlikely to improve the performance of any system. This result casts a doubt on the split-architecture proposal, and more generally on approaches that directly incorporate psycholinguistic results into performance-oriented algorithms for anaphora resolution without assessing the relative importance of the phenomena that underlie them.

1 Introduction

Both in natural language processing and in psycholinguistics, the resolution of pronouns has long been a center of attention.

Computational approaches have ranged from purely syntax-oriented treatments (Hobbs, 1978) to work in the framework of centering theory (Joshi and Kuhn (1979); Kehler (1997); Joshi et al. (to appear)) to analyses based on statistical methods (e.g. Ge, Hale, and Charniak, 1998) and genetic programming (Orasan et al., 2000).

Psycholinguists have studied the processes involved in human anaphora resolution. Seemingly contradictory results were obtained in experiments by Stevenson et al. (2000) and Hudson-D'Zmura and Tanenhaus (1998). Stevenson et al. (2000) reports that the choice of verbs in a clause affect the interpretation of pronouns in subsequent subordinate clauses. Certain arguments of the verb (depending on the verb and on the conjunction between the clauses) are more likely to act as antecedents than others. This can be seen in the following minimal pair:

(1) a. Ken$_i$ admired Geoff so he$_i$...
b. Ken impressed Geoff$_j$ so he$_j$...

Although the pronoun *he* is ambiguous in these sentence fragments, subjects of a sentence completion experiment preferred to resolve it in both cases as indicated, that is, they coindexed it with the experiencer of the verb. Thus, this experiment seems to show that antecedents are preferred based on their thematic roles and not on their subjecthood.

Hudson-D'Zmura and Tanenhaus (1998) report experiments that seem at first sight to contradict this finding. When participants were presented with the following sentences (without indexings) and asked to judge the continuations for naturalness, they strongly preferred the subject interpretation shown in (2a).

(2) Max_i despises Ross_j.

a. He_i always gives Ross_j a hard time.

b. He_j always gives Max_i a hard time.

Crucially, this tendency was unaffected by the thematic role of the subject. In the previous example, the subject was the experiencer, while in the following example, it is the agent. Yet participants still favored the subject interpretation, shown in (3a).

(3) Jack_i apologized profusely to Josh_j.

a. He_i had been rude to Josh_j yesterday.

b. He_j had been offended by Jack_i's comment.

To the extent that this experiment is comparable with the previous one, it seems to show exactly the opposite tendency: that antecedents at least of subjects are preferred based on their subjecthood (a preference known as subject-to-subject parallelism) and not on their thematic roles.

One way of making sense of this contradiction is to assume that there is a distinction between pronoun resolution within a main clause and its subordinate clauses on the one hand and pronoun resolution across main clauses on the other hand. A sentence-completion experiment by Miltsakaki (2002), henceforth Miltsakaki, confirms this hypothesis (see also Miltsakaki, 2003). She reports a strong preference for the following coindexings:

(4) a. The groom_i hit the best man_j violently. However, he_i...

b. The groom_i hit the best man_j violently although he_j...

This minimal pair exhibits subject-to-subject parallelism in the case of two main clauses (4a), but not in the case of a main and a subordinate clause (4b). In the latter case, the main clause verb *hit* seems to focus its experiencer *the best man* and thereby to make it more likely to be an antecedent. (From here on, following Miltsakaki, I will refer to pronouns which are located in a dependent (subordinate) clause with respect to the clause of their antecedent as *intrasentential*. Pronouns whose antecedent is located in a different main clause will be called *intersentential*. This case includes clausal conjunction. For example, in the sentence "John loves Mary and she loves him", both pronouns are intersentential, because their antecedent is located in a different main clause.)

2 Miltsakaki's model

Miltsakaki's anaphora resolution architecture models this split behavior. In her model, entities inside main clauses are ranked in two different ways: according to grammatical function for the

purposes of resolution across main clauses, and according to semantic focusing preferences for resolution within a main clause and its subordinate clauses. These preferences are determined lexically by the main verb and by discourse connectives (i.e. subordinating conjunctions). Whenever a pronoun occurs inside a subordinate clause and there is no compatible potential antecedent inside that clause (for example because the pronoun is a subject), the pronoun is resolved to the most highly ranked available candidate, if any, inside the main clause as defined by semantic focusing preferences. In a second step, pronouns that could not be resolved so far are matched against candidates from the previous discourse unit, this time ranked according to grammatical function. Thus, the preferred reading in (2) as well as in (4a) will be obtained because pronouns will first be resolved to the subject (highest ranked in terms of grammatical function), whereas in the case of (4b), the pronoun is resolved to the expression *the best man* (highest ranked in terms of semantic focus).

Both the tendency of certain verbs to promote one of their arguments as a potential antecedent and the tendency for subject pronouns to resolve to subject antecedents have been previously observed in the literature on anaphora resolution.

- In the natural language processing (NLP) literature, Mitkov (1997) observes the tendency of certain verbs to promote their objects, which he calls “verb preference”. In contrast to Miltsakaki, however, he regards this as only one of several preference factors influencing pronoun resolution. — Psycholinguistic studies of activation of antecedents by *implicit causality verbs*, a closely related phenomenon, are cited in McDonald and MacWhinney (1995).

- The subject-to-subject parallelism is explicitly modeled in several centering-based algorithms, such as Left-to-Right Centering (Tetreault, 1999) or the RAFT/RAPR algorithm (Suri and McCoy, 1994). Again, Mitkov (1997) cites this as just one of several factors. — In psycholinguistics, the tendency of the first element of a sentence (which, in English, generally coincides with the subject) to influence subsequent pronoun resolution has been called the "advantage of first mention" (see McDonald and MacWhinney (1995) for references).

However, Miltsakaki is the first to suggest, based on the dichotomy in the findings described above, that a separate treatment of intra- and intersentential resolution is appropriate for the purpose of anaphora resolution.

While the results of the psycholinguistic experiments described above are beyond doubt, it is less clear how, if at all, the performance of a pronoun resolution algorithm would be improved by redesigning it along the lines of Miltsakaki’s proposal. In other words, the reported effects are undoubtedly true, but are they relevant to natural language processing? More generally, how can and should results obtained in psycholinguistics be incorporated in the development of applications whose goal is to improve system performance?

In the remainder of this paper, I investigate the relevance of Miltsakaki’s two-way model in two corpus-based experiments. It is shown that pronouns found in corpora do not exhibit the kind of split behavior you would expect from them based on Miltsakaki’s model. Furthermore, the corpora are analyzed in detail to show that the number of sentences in which semantic focusing properties can apply is very low and thus Miltsakaki’s algorithm is not going to increase the performance of anaphora resolution algorithms more than marginally. I conclude by a brief discussion of the role of psycholinguistic results for anaphora resolution in natural language processing.

3 Experiments

In order to assess the extent to which overall pronoun resolution could be improved by Miltsakaki's proposal, two parsed and manually coreference-annotated corpora from different domains were used. Both corpora are subsets of the Penn Treebank (Marcus et al., 1993). The first corpus is a collection of the Wall Street Journal articles 1 to 199 (about 94,100 words in total); the second corpus consists of three fictional texts taken from the Brown corpus section of the Penn Treebank (about 8,700 words). From the standpoint of anaphora resolution, the major difference between the two corpora is the abundance of *he* and *she* tokens in the fictional corpus. By contrast, *it* is the most frequently mentioned pronoun in the Wall Street Journal corpus.

A list of all pronoun-antecedent pairs from the corpora was automatically extracted. In case the pronoun had multiple antecedents (i.e. when the corresponding discourse entity had been mentioned several times), only the last one (the closest to the pronoun) was recorded as its antecedent. Each pronoun or antecedent extracted from the Wall Street Journal subcorpus was identified as subject or nonsubject. (Due to poor quality of annotation, subjects could not be identified in the Brown subcorpus.) Finally, for each pronoun-antecedent pair it was recorded whether it was located in the same clause, in two different clauses but in the same sentence, or in two different sentences.

(*Technical note.* Two coindexed expressions occurring in the same sentence were counted as interclausal if and only if there existed an SBAR node that dominated one but not both of them. In the Penn Treebank annotation, SBAR nodes identify most finite subordinate clauses. I did not consider other subordinate clauses. This is in part guided by Miltsakaki's working assumption (p.c.) that only the boundaries of *finite* subordinate clauses should be relevant for her algorithm. Moreover, among those subordinate clauses that are not identified by SBAR nodes, many are instances of inverted direct speech (identifiable by SINV nodes), and like many others, Miltsakaki's algorithm does not make provisions for resolution in connection with quoted texts.)

3.1 Experiment 1

In a first experiment, I determined how often a subject pronoun refers to an antecedent that is also a subject, depending on whether they are inter- or intrasentential. (More precisely, I determined how often the nearest clause that mentions the referent of a pronoun mentions it at least in the subject. This formulation filters out irrelevant factors such as parentheticals: For example, we regard the sentence "[John Doe]$_i$, [vice chairman of XYZ Co.]$_i$, announced that he$_i$ would resign" as a case of (intrasentential) subject-to-subject parallelism even though the subject of the main clause is strictly speaking not the *closest* antecedent of the pronoun.)

Since Miltsakaki suggests a separate treatment for intra- and intersentential pronoun resolution, a difference in behavior would indicate that her separate treatment is on the right track. More precisely, since the intersentential component of Miltsakaki's algorithm tries to resolve subjects to subjects but the intrasentential component does not have that preference, her approach would be validated if intersentential pronouns in subject position resolved more often to subjects than intrasentential pronouns did.

For this experiment, the set of pronouns was restricted to *he, she, it* and *they*, since these are the only pronouns that can occur as subjects of finite clauses and this experiment dealt specifically at subject pronouns. (In addition, *I, we, you* and possessive pronouns were excluded because it is not clear whether Miltsakaki intends these pronouns to fall into the scope

of her algorithm.) The results of this experiment are shown in Table 1. Again, note that the Brown subcorpus could not be included in this experiment.

Intersentential subject pronouns	
Antecedent is a subject:	682 instances (85%)
Antecedent is not a subject:	123 instances (15%)
Intrasentential subject pronouns	
Antecedent is a subject:	179 instances (87%)
Antecedent is not a subject:	26 instances (13%)

Table 1: *A comparison of pronouns which are subjects with respect to subjecthood of their antecedents.*

As can be seen, subject pronouns uniformly tend to resolve to the subject of the closest clause that mention their discourse referent at least once. There is no significant difference between intra- and intersentential pronouns. This result is in contrast with Miltsakaki's split model of pronoun resolution, and it suggests that at least for applications dealing with newspaper-style texts, it may not be necessary to treat intra- and intersentential pronouns differently as Miltsakaki suggests, since most of them tend to be resolved to subject position anyway.

In this experiment, *all* subject pronouns in the Wall Street Journal corpus have been considered, as opposed to considering only pronouns in contexts that are comparable to the psycholinguistic experiments described above. Therefore, while the result does show a tendency to foreground certain discourse referents by placing them in subject position again and again, it neither confirms nor disconfirms a possible tendency of subject pronouns to refer back to subjects as opposed to other potential antecedents.

For example, it may be the case that most of the intersentential pronouns resolve trivially to the subject of the main clause, as there is simply no other grammatically compatible candidate available in the clause. Under this hypothesis, semantic focusing preferences would hardly ever have to be applied, and their application would be undetectable from the previous experiment. Another possibility is that semantic focusing preferences do apply often, but they happen to rank the subject highest in most cases. Since semantic focusing preferences are determined by lexical properties of individual verbs, it would then be conceivable that in other text genres more verbs are used which tend to rank the subject lower. Thus, the question arises how often the application of semantic focusing preferences would really make a difference, as opposed to a straightforward resolution procedure operating without focusing preferences.

3.2 Experiment 2

To answer this question, the following experiment was carried out. For each of the two corpora, those anaphoric pronouns were isolated for which semantic focusing properties actually had a chance of picking out one potential antecedent over another. Miltsakaki claims that this is only the case for pronouns in subordinate clauses whose closest antecedent occurred in the clause on which that clause was dependent. The corpora were filtered accordingly.

For this experiment, nonsubject pronouns were included, since Miltsakaki intends the split resolution algorithm to apply to those as well. Again, since it is not clear if *I, we, you* and possessive pronouns are supposed to fall in the scope of her algorithm, they were excluded. This means that the following pronouns were considered: *he, she, it, him* and objective case *her* (as opposed to possessive case *her*, which is exemplified by *Mary is looking for her purse*.)

All sentences in which there was only one grammatically compatible potential antecedent

in the relevant clause (and so semantic focusing preferences could trivially not apply) were eliminated. Furthermore, all sentences were removed in which there was more than one compatible potential antecedent but they were not assigned different thematic roles by the verb and so its focusing preferences could not have been used to disambiguate. This was the case when two potential antecedents were nested, for example. The sorting and removal had to be performed semi-automatically, as person, number, and binding constraints are not annotated in the corpora.

In this way, only those sentences were retained in which Miltsakaki's algorithm would actually apply semantic focusing preferences during resolution. The result is as follows. Seven sentences in total, containing eight relevant pronouns, were found. All of them were found in the Wall Street Journal corpus. By comparison, this corpus contains 846 instances of the pronouns under consideration. This means that on all except seven sentences (99% of the corpus), Miltsakaki's split-architecture resolution algorithm will yield exactly the same results as an otherwise equal algorithm that ignores semantic focusing preferences. In other words, if the corpus is representative, then the maximum improvement we can expect from switching to an architecture that takes semantic focusing preferences into account is around one percent. Since focusing preferences are likely to often coincide with subject-to-subject parallelism, the real number is likely to be even lower.

(Two reviewers are worried that we might be drawing conclusions based on a very small amount of evidence, i.e. seven examples, and that more examples should first be collected. But the very fact that almost no examples could be found is the main result of this experiment.)

A note of caution: The corpora are annotated for coreference, but not for potential antecedents. If a discourse referent is referred to at least twice (for example, by two definite descriptions, or by a definite description and a pronoun), then the two referents are marked as coreferential and can be assumed to be potential antecedents to other pronouns. But discourse referents which are only mentioned once are not marked as potential antecedents. They are merely marked as NP (noun phrase). However, it is not possible to simply assume that the set of NPs and the set of potential antecedents are equal, because not all entities which the corpus annotation considers NPs are potential antecedents:

(5) John does not see a car$_i$. #It$_i$ is blue. (von Heusinger, 2000)

(6) #Today$_i$'s notes will be posted online after it$_i$ is over.

For this reason, only those NPs which are annotated as coreferential could safely be considered potential antecedents for the purposes of this experiment. Therefore, an unknown number of "critical sentences" (sentences that contain a pronoun for which semantic focusing preferences will have the chance to apply) may have been missed. Identifying all potential antecedents in the corpus by hand is beyond the scope of this paper. Unfortunately, I am not aware of any parsed corpora in which all potential antecedents, as opposed to just the coreferential noun phrases, have been annotated and on which the experiment could be therefore carried out with greater accuracy.

An upper bound on this error has been estimated by eyeballing the data. An informal count performed on a 500-word sample, taken at random from one Wall-Street-Journal file, identified 114 potential antecedents, of which 81 (71%) were annotated as coreferential with some other noun phrase and were therefore visible to this experiment. If we assume the number of "critical sentences" to be roughly proportional to the number of potential antecedents, it can therefore be cautiously estimated that the number of "critical sentences" that were missed by the above

procedure is less than one third of the actual number. In other words, if the estimate is correct, there might perhaps be around 11 or 12 such sentences in the corpus, but not a lot more.

Taken together with the results of the previous experiment, this suggests that least in the newspaper and fictional genres, focusing preferences are of limited relevance as a factor in pronoun resolution.

4 Conclusion

While psycholinguistic experiments may seem to suggest that two different mechanisms are responsible for pronoun resolution within and across sentences, these experiments have considered a type of sentence that appears to occur very rarely in actual texts, if at all. The algorithm described in Miltsakaki (2002) may model human preferences in anaphora resolution acccurately when running on a certain set of restricted cases, but its most important feature – a separate treatment of intra- and intersentential pronouns – is unlikely to result in a significant improvement in performance.

It has to be stressed that these results are preliminary and a larger corpus study would be necessary to confirm them, preferably one that involves corpora from more varied domains than those used here. Note, too, that not all potential antecedents have been annotated in the corpora used here. Therefore, as explained above, the number of sentences in which semantic focusing preferences apply may be larger than what is reported here. However, if it were much larger, we would expect this to result in a marked difference in the frequency of subject-to-subject parallelism within as opposed to across clauses. This has been shown not to hold for the present corpus.

The identification of (abstract) thematic roles is a hard task. It is notoriously difficult to assign thematic roles consistently even by hand, let alone to build a system that identifies them (see e.g. Thompson et al., 2003). The focusing preferences of verbs could perhaps be restated in terms of more easily identifiable features such as grammatical function. However, no wide-coverage investigation has yet been carried out on whether focusing preferences are predictable from more easily obtainable features. For this reason, it is a welcome result that a unified treatment of pronouns within and across sentences seems possible. The simple heuristics of subject-to-subject resolution suggested by Miltsakaki (2002) and others for intersentential pronouns can likely be applied to intrasentential cases without any significant loss of accuracy.

The present study can of course not answer the general question of psycholinguistic results for preferences in anaphora resolution should be handled in performance-oriented algorithms. While psycholinguistics can bring attention to hitherto unknown anaphora resolution *factors* in the sense of Mitkov (1997), evaluating actual resolution systems is the only way to know how the best use can be made of these factors. As an example, see Mitkov (1997) for a comparative evaluation of two approaches based on the same set of factors.

Nevertheless, the present work has shown one way of how psycholinguistic results, specifically the ones cited, should *not* be handled: Directly implementing the semantic focusing preferences of verbs, as proposed by Miltsakaki, can require resources that are difficult to obtain (such as automatic semantic role labeling), while it seems unlikely that the overall performance will be affected at all. Speaking generally, corpus studies like the present one are a convenient tool of estimating the impact that a new anaphora resolution factor is likely to have on overall performance while avoiding the need and cost of implementing and evaluating the factor in an actual system.

Finally, the results described here make it necessary to rethink what we believe to be the *function* of phenomena we model by abstractions like subject preference, centering rules,

or semantic focusing of thematic roles. They are sometimes (see e.g. Hudson-D'Zmura and Tanenhaus, 1998, for centering rules) seen as strategies that readers or listeners unconsciously apply in order to constrain the interpretation process and in this way control inferential complexity. That is, since the strategies highlight certain entities as more likely antecedents, pronoun resolution is made easier. But if the cases in which this can happen are in practice very rare, then the overall reduction in processing load is very small, and phenomena like semantic focusing would be very inefficient strategies. They should then perhaps better be assumed to be epiphenomena of some more general, unknown processes, and their relative importance should be reassessed (contra e.g. Stevenson et al., 2000, p. 226: "Pronoun resolution is primarily determined by focusing..."). This would essentially mean that why we are so good at real-time pronoun resolution, and how we manage to reduce our inferential load most of the time when we do it, becomes an open question again.

5 Acknowledgements

I would like to thank Eleni Miltsakaki for her extensive support of this project and for her patience in explaining the details of her proposal to me. This project would not have seen the light without her initiative. I am also indebted to her for suggesting the first experiment to me. I wish to thank John Trueswell for helpful discussion, and ESSLLI reviewers for their valuable comments on this paper.

For access to the two corpora that were used in the experiments, I am grateful to Tom Morton (WSJ corpus) and Joel Tetreault (fictional corpus). (Unfortunately, the corpora are not publicly available at the time of writing. Requests for any of the corpora may be sent to the author of this paper and will be forwarded accordingly.)

References

Niyu Ge, John Hale, and Eugene Charniak. 1998. A statistical approach to anaphora resolution. In *Proceedings of the Sixth Workshop on Very Large Corpora* 161–170.

Klaus von Heusinger. 2000. Anaphora, antecedents, and accessibility. In *Theoretical Linguistics* 26:75-93.

Jerry R. Hobbs. 1978. Resolving pronoun references. In *Lingua* 44: 311–338.

Susan Hudson-D'Zmura and Michael Tanenhaus. 1998. Assigning antecedents to ambiguous pronouns: The role of the center of attentian as a default assignment. In M. Walker, A. Joshi, and E. Prince, editors, *Centering Theory in Discourse*, 273–291. Clarendon Press, Oxford.

Aravind K. Joshi and Steven Kuhn. 1979. Centered logic: The role of entity centered sentence representation in natural language inferencing. In *Sixth International Joint Conference on Artifical Intelligence* 435–439. Tokyo.

Aravind K. Joshi, Rashmi Prasad, and Eleni Miltsakaki. To appear. Anaphora resolution: a centering approach. In *Encyclopedia of Language and Linguistics, 2nd edition.* Elsevier.

Andrew Kehler. 1997. Current theories of centering for pronoun interpretation: A critical evaluation. In *Computational Linguistics*, 23(3):467–475.

Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn Treebank. In *Computational Linguistics*, 19(2):313–330.

Janet L. McDonald and Brian MacWhinney. 1995. The time course of anaphor resolution: Effects of implicit verb causality and gender. Journal of Memory and Language, 34, 543ś566

Eleni Miltsakaki. 2002. Toward an aposynthesis of topic continuity and intrasentential anaphora. In *Computational Linguistics*, 28(3):319–355.

Eleni Miltsakaki. 2003. The syntax-discourse interface: Effects of the main-subordinate distinction on attention structure. Doctoral dissertation, University of Pennsylvania.

Ruslan Mitkov. 1997. Factors in anaphora resolution: they are not the only things that matter. A case study based on two different approaches. In *Proceedings of the ACL '07/EACL '97 Workshop on Operational Factors in Practical, Robust Anaphora Resolution.* 14–21. Madrid, Spain.

Constantin Orasan, Richard Evans, and Ruslan Mitkov. 2000. Enhancing preference-based anaphora resolution with genetic algorithms. In *Proceedings of NLP '2000*, 185–195. Patras, Greece.

Rosemary Stevenson, Alistair Knott, Jon Oberlander, and Sharon McDonald. 2000. Interpreting pronouns and connectives: Interactions among focusing, thematic roles and coherence relations. *Language and Cognitive Processes* 15(3):225–262.

Linda Z. Suri and Kathleen F. McCoy. 1994. RAFT/RAPR and centering: A comparison and discussion of problems related to processing complex sentences. In *Computational Linguistics*, 20(2):301–317.

Joel R. Tetreault. 1999. Analysis of syntax-based pronoun resolution methods. In *Proceedings of the 37th Annual Meeting*, 602–605. University of Maryland, June. Association for Computational Linguistics.

Cynthia A. Thompson, Roger Levy, and Christopher D. Manning. 2003. A generative model for semantic role labeling. In: *Proceedings of the European Conference on Machine Learning (ECML)*, 397-408.

Ambiguity in Anaphora Workshop Proceedings, pages 23–30
ESSLLI 2006, Málaga, Spain, 7–11 August 2006
Edited by Ron Artstein and Massimo Poesio
http://cswww.essex.ac.uk/anaphora/

EFFECT OF RELATIVE PRONOUN TYPE ON RELATIVE CLAUSE ATTACHMENT

Claire Delle Luche[1], Roger P. G. van Gompel[2], Frédérique Gayraud[1], & Bruno Martinie[1]

1 – Laboratoire Dynamique du Langage, Lyon, France.

2 – University of Dundee, United Kingdom.

ABSTRACT

Accessibility hierarchies (Ariel, 1990, 2001; Givón, 1992; Gundel, Hedberg, & Zacharski, 1993) assume that the form of anaphoric expressions signals the relative saliency of the antecedent. We argue that the form of relative pronouns in relative clauses has a similar function and therefore influences attachment preferences. We conducted two questionnaire experiments in which we investigated whether attachment preferences for ambiguous relative clauses are affected by the type of relative pronoun that is used. Experiment 1 showed a difference in attachment preference between *qui* and *lequel*, indicating that the form of the relative pronoun affects attachment preferences. Experiment 2 demonstrated that the difference observed in Experiment 1 is not due to differences in informativity between *qui* and *lequel*, suggesting that instead, it is due to a difference in markedness (*qui* is more frequent and shorter).

INTRODUCTION

Many functional linguistic theories assume that the form of anaphoric expressions signals how accessible their antecedent is (e.g., Ariel, 1990, 2001; Givón, 1992; Gundel, Hedberg, & Zacharski, 1993). These accessibility theories claim that anaphoric expressions can be ranked according to the accessibility of the antecedent that they tend to refer to. For example, pronouns signal that the antecedent is highly accessible, whereas noun phrases and names signal that the antecedent is relatively inaccessible. Hence, pronouns are ranked higher on the accessibility hierarchy than noun phrases and names. More generally, short anaphoric expressions that provide little semantic and syntactic information tend to be high on the accessibility hierarchy, whereas longer expressions that contain more information about their antecedent tend to be low on the accessibility hierarchy.

Most evidence for accessibility hierarchies comes from corpus studies, which show that in conversations and texts, people tend to use anaphoric expressions that are high on the accessibility hierarchy when referring to very accessible antecedents, but expressions low on the hierarchy when referring to antecedents that are inaccessible (Givón, 1992; Gundel et al., 1993). For example, Gundel et al. (1993) analysed the distribution of different anaphoric expressions and observed that pronouns were more frequently used when the antecedent was in focus and therefore highly accessible, whereas definite noun phrases were mostly used to refer to uniquely identifiable antecedents that were not in focus.

Accessibility hierarchies also receive support from reading studies. In a series of experiments, Gordon and colleagues (Gordon, Grosz, & Gilliom, 1993; Gordon & Chan, 1995) have shown that names are harder to process than pronouns when they refer back to

an antecedent name that is the subject of the preceding sentence. This effect has been dubbed the *repeated name penalty*. The repeated name penalty is affected by the syntactic role of the antecedent noun phrase: When the antecedent name is a direct object, the penalty is much reduced. The repeated name penalty effect has been taken to support *centering theory*, a computational theory accounting for coherence in texts (Grosz, Joshi, & Weinstein, 1983, 1995). More generally, it provides support for the idea that the preference for a particular anaphoric expression is affected by the saliency of the antecedent. When the antecedent has the syntactically highly salient role of subject, an anaphor that is high on the accessibility hierarchy such as a pronoun is easier to process than a name, which is low on the hierarchy. But when the antecedent is a direct object and therefore less salient, the processing advantage for pronouns relative to names disappears.

An interesting question is whether relative pronouns can also be ranked on the accessibility hierarchy. As suggested by the term relative *pronoun*, they can be considered to be a type of anaphor. This idea is consistent with Hemforth, Konieczny, and Scheepers (2000), who argued that the processing of relative clauses (RCs henceforth) involves both syntactic attachment of the RC into the preceding tree structure and anaphoric binding of the relative pronoun. If this is true and relative pronouns indeed behave similarly to personal pronouns, we expect that relative pronouns also signal how accessible their antecedent is. This should have an effect on how people process ambiguous RCs such as (1), which have been investigated in much psycholinguistic research.

(1) *The journalist interviewed the daughter of the colonel who had had the accident*.

A number of studies have shown that in English, the RC *who had had the accident* in (1) is preferentially interpreted as modifying the second noun phrase (NP2 henceforth) *the colonel* rather than *the daughter* (NP1) (e.g., Carreiras & Clifton, 1999; Cuetos & Mitchell, 1988), while other studies suggest that there is no strong preference for either analysis (Carreiras & Clifton, 1993; Traxler, Pickering, & Clifton, 1998). This is consistent with the late closure principle (Frazier, 1979, 1987), which claims that the ambiguous RC should be attached as low as possible into the preceding tree structure, and is also compatible with a recency principle, which predicts that it should be attached to the most recent phrase (e.g., Gibson, Pearlmutter, Canseco-Gonzalez, & Hickok, 1996; Stevenson, 1994). By contrast, in languages such as French, Spanish, German, and Dutch, there is a preference for attachment to NP1. There has been much debate about the reasons behind the NP1 attachment preference in these languages and the reasons behind the cross-linguistic differences. One possibility, suggested by Mitchell, Cuetos, Corley, and Brysbaert (1995) is that NP2 attachment is more frequent in English, whereas NP1 attachment occurs more frequently in languages such as Spanish and Dutch. By contrast, Frazier and Clifton (1996) argued that pragmatic principles affect RC attachment differently in different languages, while Gibson et al. (1996) argued that recency, which favours attachment to the most recent noun phrase (NP2 in 1) and predicate proximity, which favours attachment to the head of the predicate (NP1) have different weights in different languages. Finally, Hemforth et al. (2000) claimed that in languages like German and Spanish, relative pronouns are processed like personal pronouns and are therefore preferentially interpreted as coreferent with the most salient NP, that is, NP1. By contrast, in English, relative pronouns are often omitted or generalised complementisers (*that*), so the parser relies more on syntactic processing strategies favouring NP2 attachment.

The current study did not aim to distinguish between the different theories of RC attachment, but instead, it aimed to investigate whether the form of the relative pronoun affects RC attachment preferences. Until now, none of the theories has considered the

possibility that the type of relative pronoun might affect RC processing, and no studies have addressed this question. However, if accessibility hierarchies generalise to relative pronouns, we expect that attachment preferences for ambiguous RCs should be affected by the type of relative pronoun: RCs with relative pronouns that signal a highly accessible antecedent should preferentially attach to the most accessible NP, whereas this preference should be less strong for RCs with relative pronouns that signal a less accessible antecedent.

In order to investigate the influence of relative pronouns on RC attachment preferences, we will compare two types of relative pronouns in French. It has been demonstrated (Zagar, Pynte, & Rativeau, 1997) that in sentences like (2) containing a *qui* RC, readers of French prefer to attach the RC *qui semblait plus confiant* to *l'avocat* (NP1 attachment) rather than to *la chanteuse* (NP2 attachment).

(2) *Un journaliste aborda l'avocat de la chanteuse qui semblait plus confiant. (A journalist approached the barrister$_{MASC}$ of the singer$_{FEM}$ who seemed more confident$_{MASC}$.)*

However, French also has a different type of relative pronoun, namely *lequel* or *laquelle*. *Lequel* and *laquelle* may be lower on the accessibility hierarchy than *qui* because *lequel* and *laquelle* are more marked, that is, they are phonologically longer and less frequent than *qui*. Furthermore, *lequel* and *laquelle* are marked for gender and number, so they are also more informative than *qui*, which does not have gender and number marking. Both factors may affect the position of the relative pronoun on the accessibility hierarchy (Ariel, 1990, 2001). Therefore, RCs with *lequel* or *laquelle* should be attached to less accessible NPs. Therefore, RCs with *lequel* or *laquelle* should be attached to less salient NPs. Assuming that NP1 is most salient, this predicts that in a sentence like (3), readers should make a local attachment to NP2 more frequently, resulting in a less strong NP1 attachment preference than in *qui* RCs.

(3) *Un journaliste aborda l'avocate de la chanteuse, laquelle semblait plus confiante.* (*A journalist approached the barrister$_{FEM}$ of the singer$_{FEM}$ who seemed more confident$_{FEM}$.*)

EXPERIMENT 1: QUI VS. LEQUEL/LAQUELLE

The purpose of Experiment 1 was to test whether the type of relative pronoun affects RC attachment in French. We compared *qui* (1) and *lequel/laquelle* (2) in non-restrictive RCs following a 'NP1 of the NP2' structure. The accessibility account predicts that *qui* RCs should preferentially attach to the most accessible NP (presumably NP1, e.g., Zagar, Pynte, & Rativeau, 1997), but this preference should be weaker for *lequel/laquelle* RCs. This is because *qui* is less marked and less informative than *lequel/laquelle*, so *qui* should refer to more accessible antecedents (Ariel, 1990).

METHOD

Participants Fifty-six undergraduates at Lyon 2 Lumière University participated in the experiment as part of their courses. All were native speakers of French.

Materials and procedure Thirty-two sets of experimental sentences were constructed in two versions, one with *qui* RCs (4), and one with the relative pronoun *lequel* or *laquelle* (5).

(4) *Je connais le père du maçon, qui est amusant*. (*I know the father*$_{NP1}$ *of the mason*$_{NP2}$ *who is funny*)

(5) *Je connais le père du maçon, lequel est amusant*. (identical meaning)

Each experimental trial consisted of a sentence such as (3) and (4) followed by two statements, one consistent with the NP1 attachment interpretation (*The father is funny*) and one consistent with NP2 attachment (*The mason is funny*). Participants were instructed to tick the option that was 'most correct'. NP1 and NP2 were matched for gender, length and number of syllables.

Prior to the experiment, we conducted a pretest to ensure that the RC was not biased toward NP1 or NP2. Twenty participants, none of whom took part in Experiment 1, were asked to rate on a seven-point scale the plausibility of statements consistent with NP1 attachment and NP2 attachment. For the 32 sentences that we selected, there was no overall preference for either NP1 (5.83) or NP2 (5.95) attachment.

Two lists were constructed using a between subjects design: one list presented *qui* RCs, the other *lequel/laquelle* RCs. The experiment lasted about 20 minutes. The order of statements was counterbalanced for the two lists. The questionnaires were run in large groups of participants.

RESULTS AND DISCUSSION

We conducted two ANOVAs on the percentages of NP1 attachment responses, one with subjects (*F1*) and one with items (*F2*) as the random variable. The ANOVAs contained condition (*qui* vs. *lequel/laquelle*) as a between subjects and within items variable. Figure 1 presents the mean percentage of NP1 attachment by condition. The results showed that for *qui* RCs, participants strongly preferred NP1 (87.2% of trials), but this preference was much weaker (70.6%) with *lequel/laquelle* RCs. The percentage of NP1 attachments differed from chance in both the *qui* RCs ($F1(1,27) = 17.70$; $p < .01$; $F2(1,31) = 26.51$; $p < .01$) and the *lequel*/*laquelle* RCs ($F1(1,27) = 4.40$; $p < .01$; $F2(1,31) = 12.64$; $p < .01$). Most important, the difference between conditions was significant both by subjects ($F1(1,54) = 9.01$; $p < .01$) and items ($F2(1,31) = 4.73$; $p = .04$).

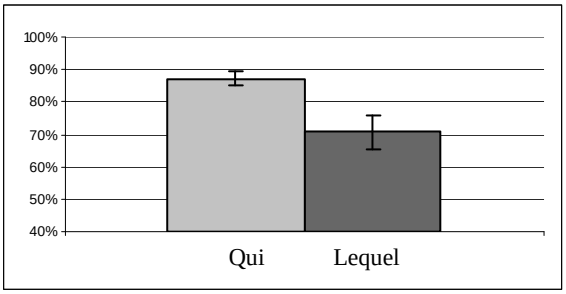


Figure 1: Mean NP1 attachments (%)

EXPERIMENT 2: A QUI VS. AUQUEL

Qui may be higher on the accessibility hierarchy than *lequel*/*laquelle* because (1) *qui* is less marked than *lequel*/*laquelle*, that is, *qui* is phonologically shorter and more frequent than *lequel*/*laquelle* and (2) because *qui* is less informative than *lequel*/*laquelle*, that is, *lequel*/*laquelle* are marked for gender and number, whereas *qui* is not. Ariel (1990, 2001) argued that both factors may affect an anaphor's position on the accessibility hierarchy, so both factors may affect RC attachment preferences.

The goal of Experiment 2 was to test whether the difference in attachment preferences we found in Experiment 1 were due to a difference in markedness between *qui* and *lequel*/*laquelle* or due to a difference in informativity between *qui* and *lequel*/*laquelle*. In the current experiment, we controlled the relative pronouns for phonological length by comparing dative RCs containing either *à qui* or *auquel*. These relative pronouns are also more similar in frequency than *qui* and *lequel*/*laquelle* in Experiment 1. However, they differ in informativity: *auquel* is marked for gender (masculine) and number (singular), whereas *à qui* is not. Hence, if *à qui* and *auquel* RCs have different attachment preferences, this must be due to the difference in informativity. By contrast, if they have the same attachment preferences, this suggests that the difference in attachment preference in Experiment 1 was due to a difference in markedness (length and frequency differences between *qui* and *lequel*/*laquelle*).

METHOD

Participants Twenty-four participants from the same population took part in this experiment. None had participated in Experiment 1 or the pretests.

Materials and procedure The method and design were similar to those in Experiment 1. On the basis of norms from a plausibility pretest (20 participants, same procedure as the pretest in Experiment 1), we selected 32 sentences that did not differ in plausibility between conditions (NP1, 4.87; NP2, 4.91). Sentences contained either *à qui* (6) or *auquel* (7).

(6) *Je connais le collègue de l'étudiant, à qui la bibliothécaire apporte le livre. (I know the colleague of the student, to whom the librarian is giving the book)*

(7) *Je connais le collègue de l'étudiant, auquel la bibliothécaire apporte le livre.* (identical meaning)

We used the same fillers as in Experiment 1. Two lists were constructed using a between subjects design.

RESULTS AND DISCUSSION

As in Experiment 1, we conducted analyses by subjects and items on the percentages of NP1 attachment responses with the variable condition as a between subjects and within items variable. Figure 2 presents the mean attachment preference by condition. The mean percentages of NP1 attachment were high (Figure 2) both with *à qui* (87.6%) and *auquel* (85.3%). The percentage of NP1 attachments differed from chance for *qui* RCs ($F1$ (1, 11) = 10.05; $p < .01$; $F2(1,31) = 10.52$; $p < .01$) as well as *auquel* RCs ($F1(1,11) = 4.56$; $p < .01$; $F2(1,31) = 22.32$; $p < .01$). No significant difference between the two conditions was found ($Fs < 1$): Participants chose NP1 equally often with *auquel* RCs as with *à qui* RCs.

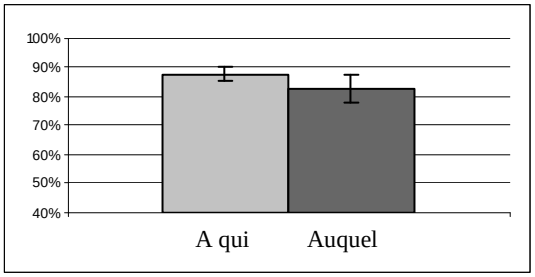


Figure 2: Mean NP1 attachments (%)

The results of Experiment 2 demonstrate that the difference in informativity between *à qui* and *auquel* did not contribute to differences in attachment preferences. This suggests that the difference observed in Experiment 1 must be due to a difference in markedness. The position of the relative pronoun on the accessibility hierarchy is thus affected by markedness rather than informativity.

GENERAL DISCUSSION

Two experiments investigated whether and how RC attachment preferences are affected by the form of the relative pronoun. Experiment 1 showed a strong NP1 attachment preference when the relative pronoun was *qui*, but this preference was much reduced when it was *lequel* or *laquelle*. This result provides support for the idea that relative pronouns can be ranked on the accessibility hierarchy. Relative pronouns such as *qui* signal highly accessible antecedents, and are therefore preferentially interpreted as referring to NP1, whereas relative pronouns such as *lequel* and *laquelle* signal less accessible antecedents, so the preference for NP1 attachment is less strong.

An important question is what factors determine a relative pronoun's position on the accessibility hierarchy. *Lequel* and *laquelle* are more marked than *qui*, because they are longer and less frequent than *qui*. But in addition, they are also more informative than *qui*, because they contain gender and number marking. In Experiment 2, we controlled for markedness of the relative pronoun by contrasting *à qui* and *auquel* (which are similar in length and frequency), while manipulating informativity (gender and number marking). The experiment demonstrated that *auquel* and *à qui* had similarly strong NP1 attachment preferences despite the fact that they differ in informativity. Therefore, in Experiment 1, it was markedness rather than informativity that contributed to the difference in attachment preferences. Hence, our experiments suggest that markedness of the relative pronoun (resulting from a low frequency of the relative pronoun and its length) affects the position of the relative pronoun on the accessibility hierarchy, whereas informativity does not.

Our results have implications for both theories of anaphors and sentence processing theories. Relative pronouns appear to have properties that are similar to those of other anaphors such as personal pronouns. Similar to other anaphors, relative pronouns refer back to an earlier introduced entity in the discourse. And like other anaphors, their form influences how they are preferentially interpreted. When a relative pronoun is unmarked, that is, it is short and frequent, it signals a highly accessible antecedent. But when it has a marked form, it signals a less accessible antecedent. Therefore, unmarked relative pronouns are preferentially interpreted as referring to the most salient antecedent NP, whereas this preference is less strong for marked relative pronouns. Of course, this is not to say that relative pronouns are similar to personal pronouns in all respects. Clearly, syntactic constraints on relative pronouns are different from constraints on pronouns. For example, unlike personal pronouns, relative pronouns must be in the same sentence as their antecedent, and unlike personal pronouns, they cannot precede their antecedent in English. However, this is not surprising: It is well-known that syntactic constraints on personal pronouns and noun phrase anaphors are also different (e.g., Chomsky, 1981, Reinhart, 1983). What we would like to argue is that there are good arguments to believe that relative pronouns belong to the class of anaphoric expressions and that their processing is affected by similar factors.

Our results have important implications for sentence processing theories too, because they showed that parsing preferences are affected by the form of the relative pronoun. This is difficult to reconcile with many sentence processing theories, because they do not assign a role to the form of the relative pronoun. For example, the garden-path theory (Frazier, 1979, 1987) predicts a preference for NP2 attachment due to the application of the late closure

strategy. This does not explain why French exhibits a NP1 attachment preference, and furthermore, given that late closure is a purely syntactic strategy, it does not explain why attachment preferences are affected by the type of relative pronoun. Gibson et al. (1996) argued that NP1 or NP2 attachment is preferred depending on whether predicate proximity favouring NP1 attachment or recency favouring NP2 attachment is the strongest parsing constraint. This would explain the current results if it is assumed that recency is a stronger constraint for *lequel/laquelle* than for *qui*. However, there does not seem to be a principled reason why this should be the case. Frequency-based accounts (e.g., Desmet, De Baecke, Drieghe, Brysbaert, & Vonk, in press; Mitchell et al., 1995) fare slightly better, because it seems likely that *qui* is used more often to refer to NP1 than *lequel/laquelle*. But this raises the question: Why do these production preferences occur? Accessibility theories have an answer to this. Language producers signal that the antecedent of a relative pronoun is highly accessible by using an unmarked relative pronoun that is high on the accessibility hierarchy, whereas they signal that the antecedent is relatively inaccessible by using a marked relative pronoun that is low on the hierarchy. The results from our two experiments show that comprehenders use these accessibility cues, and therefore, the form of the relative pronoun affects the comprehension of RCs.

REFERENCES

Almor, A. (1999). Noun-phrase anaphora and focus: The informational load hypothesis. *Psychological Review, 106,* 748-765.

Ariel, M. (1990). *Accessing noun-phrase antecedents*. London: Croom Helm.

Ariel, M. (2001). Accessibility theory: An overview. In T. J. M. Sanders (Ed.), *Text representation: Linguistic and psycholinguistic aspects.* Philadelphia, PA, USA: John Benjamins.

Carreiras, M., & Clifton, C. (1993). Relative clause interpretation preferences in Spanish and English. *Language and Speech, 36,* 353-372.

Cuetos, F., & Mitchell, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition, 30,* 73-105.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

Desmet, T., Brysbaert, M., & De Baecke, C. (2002). The correspondence between sentence production and corpus frequencies in modifier attachment. *Quarterly Journal of Experimental Psychology, 55A,* 879-896.

Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies.* Ph.D. Dissertation. Indiana University Linguistics Club. University of Connecticut.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 559-586). Hillsdale, NJ: Erlbaum.

Frazier, L., & Clifton, C., Jr. (1996). *Construal*. Cambridge, MA: MIT Press.

Gibson, E., Pearlmutter, N., Canseco Gonzalez, E., & Hickok, G. (1996). Recency preference in the human sentence processing mechanism. *Cognition, 59,* 23-59.

Givón, T. (1992). The grammar of referential coherence as mental processing instructions. *Linguistics, 30,* 5-55.

Gordon, P. C., & Chan, D. (1995). Pronouns, passives, and discourse coherence. *Journal of Memory and Language, 34,* 216-231.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language, 69,* 274-307.

Gordon, P. C., Grosz, B., & Gilliom, L. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science, 3*, 311-347.

Gosz, B. J., Joshi, A. K., & Weinstein, S. (1983). Providing a unified account of definite noun phrases in discourse. *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics*, Cambridge, MA.

Gosz, B. J., Joshi, A. K., & Weinstein, S. (1995). Centering: A framework for modelling the local coherence of discourse. *Computational Linguistics*, *21*, 203-225.

Hemforth, B., Konieczny, L., & Scheepers, C. (2000). Syntactic attachment and anaphor resolution: The two sides of relative clause attachment. In M. W. Crocker, M. J. Pickering, & C. E. Clifton, Jr. (Eds.), *Architectures and mechanisms for language processing* (pp. 259-282). Cambridge, UK: Cambridge University Press.

Mitchell, D.C., Cuetos, F., Corley, M.M.B., & Brysbaert, M. (1995). Exposure-based models of human parsing: Evidence for the use of coarse-grained (nonlexical) statistical records. *Journal of Psycholinguistic Research, 24*, 469-488.

Reinhart, T. (1983). *Anaphora and Semantic Interpretation*. Chicago: University of Chicago Press.

Stevenson, S. (1994). Competition and recency in a hybrid network model of syntactic disambiguation. *Journal of Psycholinguistic Research, 23*, 295-322.

Traxler, M.J., Pickering, M.J., & Clifton, C. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language, 39*, 558-592.

Zagar, D., Pynte, J., & Rativeau, S. (1997). Evidence for early-closure attachment on first-pass reading times in French. *The Quarterly Journal of Experimental Psychology, 50A*, 421-438.

Ambiguity in Anaphora Workshop Proceedings, pages 31–39
ESSLLI 2006, Málaga, Spain, 7–11 August 2006
Edited by Ron Artstein and Massimo Poesio
http://cswww.essex.ac.uk/anaphora/

A Model of Grouping for Plural and Ordinal References

Alexandre Denis, Guillaume Pitel, Matthieu Quignard

LORIA, BP239 F-54206 VANDOEUVRE-LES-NANCY

Contact author: Guillaume.Pitel [at] gmail.com

We present a model for the resolution of plural references on groupings based on Reference Domains Theory. While the original theory does not take plural reference into account, this paper shows how several entities can be grouped together by building a new domain and how they can be accessed later on. We introduce the notion of super-domain representing the access structure to all the plural referents of a given type.

Introduction

In the course of a discourse or a dialogue, referents introduced separately could be referenced with a single plural expression (pronoun, demonstrative, etc.). The grouping of these referents may depend on many factors: it may be explicit if they were syntactically coordinated or juxtaposed, or implicit if they only share common semantic features (Eschenbach *et al.*, 89). Time is also an important factor because it may be difficult to group old mentioned referents with new ones. Because of this multiplicity of factors, choosing the right discursive grouping for a referential plural expression is ambiguous, and this ambiguity needs to be explicitly described.

We present a model of grouping based on Reference Domains Theory (Salmon-Alt, 01) which considers that a reference operation consists of extracting a referent in a domain. However the original theory barely takes into account plural reference. This paper shows how several entities can be grouped together by building a new domain and how they can be accessed later on. It introduces also the notion of super-domain D^+ that represents the access structure to all the plural referents of type D. This work is being implemented and evaluated in the MEDIA/EVALDA project (Devillers, 04).

The goal of this research is both to find a practical solution to deal with the kind of situations we met in the corpus of the MEDIA/EVALDA project, and to improve the coverage of the Reference Domain Theory, which is a representational theory of reference that focuses on describing the selection preferences from several ambiguous candidates.

1. Groupings and plural anaphora

Several kinds of clues can specify that referents should be grouped together, or at least could be grouped together. These clues may occur at several language levels, from the noun phrase level to the rhetorical structure level. We have not explored in detail the different ways of groupings entities together in a discourse or dialogue. We just describe here some of the phenomena we were confronted with while developing a reference resolution module for a dialogue understanding system.

- **Explicit Coordination** - The most basic way to explicitly express the grouping of two or more referents is using a connector such as *and, or, as well as*, etc.
"*Good afternoon, I would like to book a single room **and** a double room*"
- **Implicit Coordination** - An implicit coordination occurs when two or more referents of the same kind are present in one sentence, without explicit connector between them.
"*Does the hotel de la gare have a restaurant, like the Holiday Inn?"*
- **Repetitions/Specifications** – In some particular cases, groupings are explicitly described by the enumeration of their referents. For instance "*Two rooms. A single room, a double room*".
- **Inter-Sentential** – In the course of a dialogue, referential expressions can be grouped together, depending on several factors (common type, common predicate, semantic link).

Most of these different situations have already been thoroughly investigated in previous work. However, these methods are, from our point of view, unable to fulfill the needs we met in the particular task of the MEDIA/EVALDA project, especially with plurals, otherness, and ordinals being very frequent in our corpus.

In the standard model of plurals in the DRT (Kamp & Reyle, 93), discourse referents are grouped and assigned to a plural discourse referent (this is represented using the $\oplus$ operator), but no information can be assigned to the relative role of the individual referents within the group (which is necessary for the resolution of ordinal anaphora). Moreover, without the presence of specific markers or constructions, it seems difficult to allow the emergence of several groupings from a single list of referring expressions (for instance in the case of a co-occurence of several referents - X,Y,Z - in the same predicate, while several others - Y,Z,W - share a common type). Other approaches deal with referring expressions sharing the same type for making a group (Eschenbach *et al.*, 89), which is not sufficient for our problems, since sharing a common type is only one of the enablers of grouping.

2. Reference Domains Theory

The Reference Domains Theory (Salmon-Alt, 01) supposes that every act of reference is related to a certain domain of interpretation, in that it both describes how to extract a referent and which set of elements to extract it from. In the reference domains theory, an act of reference also modifies the structure of the reference domains of the discourse, in term of focus and partitions.

A reference domain is composed of any group of entities in the hearer's memory (discursive referents, visual objects, or concepts) and describes how each entity could be addressed through a referential expression. The theory has been developed in order to represent the diversity of access modes to the referents. The claim is that every referential expression has a different behavior which depends on the vericonditional description its referent must satisfy and on its conditions of use (the actual structure of context).

The theory considers the referring process as a dynamic extraction of a referent in a domain instead of a binding between two entities (Salmon-Alt, 00). Hence doing a reference act consists of isolating a particular entity from other rejected candidates (Olson, 70), amongst all the accessible entities composing the domain. This dynamic discrimination relies on projecting an access structure focusing the referent in the domain and facilitates further access: any extraction in a domain increases its salience, thus it is preferred for the next

interpretations.

The preferences for choosing a suitable domain are inspired from the Relevance theory (Sperber & Wilson, 86) taking into account such focalization and salience. F. Landragin & L. Romary (03) have also studied the usage of reference domains in order to model a visual scene.

2.1. Basic type

A reference domain is a structure which can reference entities by differentiating them. It is modelled as a set of **entities** (ground) and a set of **partitions** of these entities (equivalency classes or alternatively differentiation functions as in Pitel, 03). Each class groups the elements which are accessible by the same referential expression *i.e.* which can be viewed as the same object from a certain point of view. Identifying completely a referent requires to find a domain where a partition gives an unambiguous access to this entity. Using a referential expression presupposes then (optimistically) that this condition is respected for all the speakers.

2.2. Access structures

We suppose that any distinction between the referents from the excluded alternatives requires highlighting the discrimination criterion opposing them. This criterion, given by the referential expression and its context of use behaves like a partition of the accessible entities, grouping them together according to their similarities and their differences. A partition may have one of its parts **focused** (or profiled). There are, at least, three kinds of discrimination criteria:

- **discrimination on description**. Entities can be discriminated by their type, their properties, or by the relations they have with other entities. For example the name of the hotels is a discrimination criterion in "*the Ibis hotel and the hotel Lafayette*".
- **discrimination on focus**. Entities can also be discriminated by the focus they have when they are mentioned in the discourse or designated by a gesture. For example, "*these rooms*" would select focused referents in a domain, whereas "*the other room*" would select a non-focused one.
- **discrimination on time of occurrence**. Entities can also be discriminated by their occurrence in the discourse. For example "*the second hotel*" would discriminate this hotel by its rank in the domain.

Every referential expression aims to distinguish referents and exhibit a differentiation criterion; however the referents are sometimes not distinguished intentionally. For example, the indefinite plural "*two hotels*" will introduce two hotels that cannot be differentiated. But even in this case it is possible to project a differentiation structure *a posteriori* for example by next saying "*the first one*". Figure 1 shows the state of a domain after the sequence "*two hotels*", "*the first one*" (the ground is represented here as a concept in description logics, each subdomain is then a lower concept in the hierarchy). The domain H of the two hotels contains a partition which focalizes the first of its elements, the subdomain H_1.

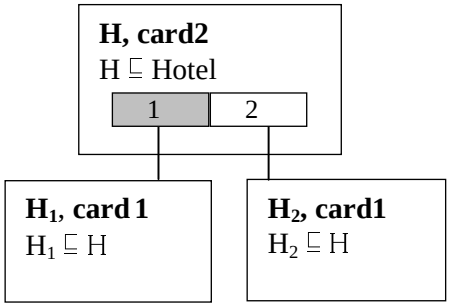


Figure 1: A domain containing two subdomains

2.3. Classical resolution algorithm

Each activated domain belongs to a list of domains ordered according to their recentness (the referential space). The resolution algorithm consists of two phases:

1. It searches a suitable, preferred domain in the referential space when interpreting a referring expression. The suitability is defined by the minimal conditions the domain has to conform to in order to be the base of an interpretation (particular description, or presence of a particular access structure). The general preference factor is the minimization of the access cost (recentness, salience or focalization).
2. It extractes a referent and restructurates the referential space, taking into account this extraction. It not only focuses the referent in its domain, but also increases the salience of the domain itself which will be preferred for further extractions.

According to the determination and description of the referential expression or to the gesture made to access to the referent, this generic scheme will be instantiated in different ways. For example a definite "*the N*" will search for a domain in which a particular entity can be solely discriminated by its type N, and the restructuring consists of focalizing the found referent in this domain. A demonstrative "*this N*" behaves differently in that it tries directly to access to the referent without imposing a strong discrimination criterion on the type, *i.e.* it finds a focalized referent in a domain which could be cast into a "N" during the restructuring phase. See (Landragin & Romary, 2003) for a classification of the different access modes.

The algorithm highlights two types of ambiguities, domain ambiguities when there are many preferred domains of interpretation with no mean to choose during the first phase, and the referent ambiguities when many referents are found without preferences. Of course a domain ambiguity implies a referent ambiguity. In a dialogue system, it is not the role of the reference module to disambiguate completely referents, but instead to propagate the ambiguity to next modules which, for instance, will solve it by asking clarifications to the user.

3. Super-domains

In order to take groupings into account in the Reference Domains Theory, we introduce two constructs in our formal toolbox. Indeed, having only one kind of domain construct doesn't allow for a correct distinction between different referent statuses.

First we distinguish plural and simple domains. The simple domains *D* serve as bases for profiling a **subpart**, or **related part** of a simple referent. For instance, if *D* = *Room*, then

one can profile a *Price* from *D*. The plural domains D^* serve either as a **generic base** or as a **plural representative** for profiling a simple domain *D*. A generic base is mandatory in our model to support the insertion of new extra-linguistic referents evoked with an indefinite construct (for instance "*I saw a black bird on the roof*"), while plural representatives are used for explicit groupings. A domain D^*_1 can also be profiled from a D^*_0, provided D^*_1 profiles a subset of the elements of D^*_0.

Second, we introduce the notion of **super-domain** D^+, from which a D^* can be profiled. The relations allowed between domains are represented on figure 2. A super-domain D^+ is the domain of all groupings D^*, including a special D^*_{all} grouping which is the representative of all evoked instances of a given category. This configuration is not intended to deal with long dialogues where several, trans-sentential groupings occur, and where older groupings may become out of access. Doing this would require a rhetorically driven structuring of the D^*_{all}.

As Reference Domain Theory is primarily targeted toward extra-linguistic referents occurring in practical dialogue, the construction of domain trees representing the supposed structuring of referent accessibility is based on an ontology. As a consequence, for each "natural" type and each subtype (for instance *Room*∧*Single*), a domain tree is potentially created (actually, one can easily imagine how this creation may be driven 'on-demand').

Another evolution from the initial Reference Domain Theory is the possibility to focalize several items of a partition. Indeed, since the resolution algorithm can focalize a whole plural domain, all elements of this domain must be focalized in all the plural domains they occur in.

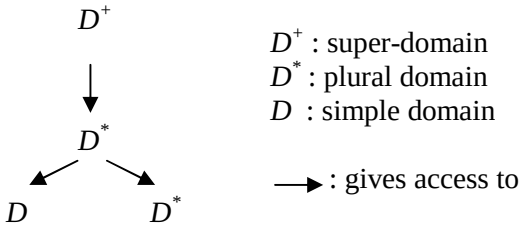


Figure 2: Access tree of Reference Domains

When new extra-linguistic referents are evoked, they are individually profiled under the D^*_{all} corresponding to their types (that is, their "natural" type, and all the subtypes they are eligible to). When some sentence-level grouping occurs or when a plural extra-linguistic referent is evoked, a D^* is created, with each of its components as children, when possible (that is, when each component is described). Figure 3 illustrates the state of the *Hotel* domain tree after a scenario with at least two dialogue acts, the first one introducing $Hotel_1$, the second one inserting a grouping of $Hotel_2$ and $Hotel_3$ (due to their co-occurence in the same utterance). One can see that all referents introduced are accessible through the special $Hotel^*_{all}$ domain.

In short:

- All new referents (singular or plural) become subdomains of D^*_{all}
- All new plural referents build up a subdomain of D^+

When a referring expression occurs, one performs the resolution through the following algorithm:

- If the referring expression is singular, performs the classical resolution algorithm in the plural domains D^* (including D^*_{all})
- If the referring expression is plural, performs the classical resolution algorithm with D^+ as the base

U: The Ibis Hotel (*Hotel*$_1$) is too expensive
S: How about the Hotel Lafayette (*Hotel*$_2$) or the Hotel de la cloche (*Hotel*$_3$)

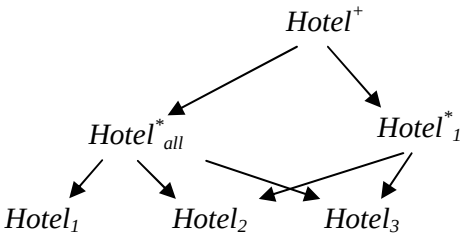


Figure 3: A domain tree built from a the scenario on the left, containing a grouping

4. Examples

A sample dialogue (figure 4) is analyzed through the algorithm presented above. This example shows how the referents introduced in an explicit coordination could be referenced as a whole "*the two hotels*", or extracted discriminately by an ordinal "*the second one*" or by an otherness expression "*the other one*". All the subdomains of H^+ (*i.e.* the plural domains of hotels) are indicated *after* each interpretation using a simplified notation. Only the ordered list of accessible entities and their focalization (bold) are noted for each subdomain (only one access structure is represented for each domain). For instance $\mathbf{H^*_{all}}$= (h_1, h_2, $\mathbf{h_3}$) means that the domain H^*_{all} is focalized in H^+, and that h_3 is focalized in H^*_{all}.

Dialogue	$\mathbf{H^+}$
U: *Is there a bathroom at the Ibis hotel* (h_1) *and the hotel Lafayette* (h_2)?	$\mathbf{H^*_0 = (h_1, h_2)}$ $H^*_{all} = \mathbf{(h_1, h_2)}$
S: *No they don't have bathrooms*	$\mathbf{H^*_0 = (h_1, h_2)}$ $H^*_{all} = \mathbf{(h_1, h_2)}$
S: *But I propose you the Campanile hotel* (h_3)	$H^*_0 = (h_1, h_2)$ $\mathbf{H^*_{all} = (h_1, h_2, h_3)}$
U: *Hmm no, how much were the two hotels?*	$\mathbf{H^*_0 = (h_1, h_2)}$ $H^*_{all} = \mathbf{(h_1, h_2}, h_3)$
S: The hotel Lafayette is 100 euros, the Ibis hotel is 75 euros	$\mathbf{H^*_1 = (h_2, h_1)}$ $H^*_0 = \mathbf{(h_1, h_2)}$ $H^*_{all} = \mathbf{(h_1, h_2}, h_3)$
$\mathbf{U_1}$: Ok, I'll take the second one	$\mathbf{H^*_1 = (h_2, h_1)}$ $H^*_0 = \mathbf{(h_1}, h_2)$ $H^*_{all} = \mathbf{(h_1}, h_2, h_3)$
$\mathbf{U_2}$: Ok, I'll take the third one $\mathbf{U_3}$: OK, I'll take the other one	$H^*_1 = (h_2, h_1)$ $H^*_0 = (h_1, h_2)$ $\mathbf{H^*_{all}} = (h_1, h_2, \mathbf{h_3})$

Figure 4: Example of dialogue (focused domains and referents are in **bold**)

In order to interpret U_1, U_2 and U_3 one needs to rely on the previous structuring of H^+. In U_1, the previously focalized domain H^*_1 is preferred to be the base for interpreting "*the second one*" because of the order discrimination. This leads to extracting h_1 hence focalizing it in H^*_1 but also in H^*_0 and in H^*_{all}. In U_2, H^*_1 cannot be the base for interpreting "*the third one*" because no entity could be discriminated this way. Therefore the only suitable domain is H^*_{all}. It is also impossible to interpret "*the other one*" in H^*_1 because of the lack of a focus discrimination between h_1 and h_2. It is however possible to choose H^*_{all} for the domain of interpretation: the excluded referents h_1 and h_2 are unfocused while h_3 gains focus.

Another example (figure 5) shows that keeping the way the referents are accessed is important in order to have a reliable state of the referential space. Compare the sequences $S_0U_0S_1U_1$ and $S_0U_0S_2U_2$. In the first one the system does not distinguish the referents from each other, and the referential expression "*the second one*" address the hotel Lafayette. In the second one the system answers the question by mentioning the prices of each hotel separately and "*the second one*" address the Campanile hotel. A reason for such phenomenon seems that it is difficult to corefer to the same referent by two different ordinal expressions successively: the extraction of h_3 instead of h_2 in U_1 would sound strange. On the contrary in U_2, "*the second one*" could refer to h_3 because of the new domain which differenciates the hotels by their prices. Actually the model could predict such behavior by the access structure of H^*_1 introduced in U_0 specifying an ordinal discrimination criterion (noted by a "o:"): if the structure does not change each hotel h_1 or h_2 could be accessed by the ordinal expression they were introduced with. The pronoun "*They*" in S_1 does not change this structure in the same way as S_2 does, that is by increasing the salience of the referents accessed by their names. This way we can constrain the interpretation of ordinals.

Dialogue	$\mathbf{H^+}$
$\mathbf{S_0}$: I propose you the Ibis hotel (h_1), the hotel Lafayette (h_2) and the Campanile hotel (h_3).	$\mathbf{H^*_0 = (h_1, h_2, h_3)}$ $H^*_{all} = (\mathbf{h_1, h_2, h_3})$
$\mathbf{U_0}$: How many are the first and the third hotel ?	$\mathbf{H^*_1 = o:(h_1, h_3)}$ $H^*_0 = (\mathbf{h_1}, h_2, \mathbf{h_3})$ $H^*_{all} = (\mathbf{h_1}, h_2, \mathbf{h_3})$
$\mathbf{S_1}$: They are expensive.	$\mathbf{H^*_1 = o:(h_1, h_3)}$ $H^*_0 = (\mathbf{h_1}, h_2, \mathbf{h_3})$ $H^*_{all} = (\mathbf{h_1}, h_2, \mathbf{h_3})$
$\mathbf{U_1}$: OK, I'll take the second one.	$H^*_1 = o:(h_1, h_3)$ $\mathbf{H^*_0 = (h_1, h_2}, h_3)$ $H^*_{all} = (h_1, \mathbf{h_2}, h_3)$
$\mathbf{S_2}$: The Ibis hotel is 100 euros and the Campanile hotel is 50 euros.	$\mathbf{H^*_2 = (h_1, h_3)}$ $H^*_1 = o:(\mathbf{h_1, h_3})$ $H^*_0 = (\mathbf{h_1}, h_2, \mathbf{h_3})$ $H^*_{all} = (\mathbf{h_1}, h_2, \mathbf{h_3})$
$\mathbf{U_2}$: OK, I'll take the second one.	$\mathbf{H^*_2 = (}h_1, \mathbf{h_3)}$ $H^*_1 = o:(h_1, \mathbf{h_3})$ $H^*_0 = (h_1, h_2, \mathbf{h_3})$ $H^*_{all} = (h_1, h_2, \mathbf{h_3})$

Figure 5: Example of cascading ordinals

5. Discussion

The scope of the groupings considered by this extension to the Reference Domains Theory is still limited. First the trans-sentential groupings are not fully studied yet. We guess that such groupings would need a rhetorical description of the discourse à la SDRT (Asher, 93). Second it considers only extra-linguistic referents, *i.e.* those having an existence outside discourse. When trying to solve references in a dialogue, one should also take into account the domains of interpretation of each speaker. Consider "U: I want an hotel in Paris", "S: I propose you two hotels". The first hotel is interpreted in the domain of mental representations of U while the two hotels proposed by S are assumed to exist outside the discourse. They can hardly be grouped together by a referential expression because they belong to different levels of reality. We guess that this kind of phenomenon could be rendered by defining accurate differentiation criteria.

Conclusion

We presented a model of grouping in the Reference Domain Theory. This theory considers that reference resolution is a matter of extracting the referent in an accurate reference domain. It suits well our needs: the groups are considered as reference domains, where any further reference (ordinals or other) can be interpreted. We introduced a particular type of domain, the superdomain, which references all the plural domains constructed at a certain time. Given this domain, the model can render dynamic effects like ordinals or otherness in plural contexts. The conditions for grouping are not examined in detail, however such domain architectures could be the backbone for modelling more complex reference effects using more precise differentiation criteria: time between two evocations of referent, rhetorical or dialogical structure or different mental spaces. This would be future work but in the meantime the model and algorithms are currently evaluated in the MEDIA/EVALDA project which aims to compare the semantic and pragmatic understanding of dialogue systems.

References

Devillers, L., Maynard, H., Rosset, S., Paroubek, P., McTait, K., Mostefa, D., Choukri, K., Bousquet, C., Charnay, L., Vigouroux, N., Béchet, F., Romary, L., Antoine, J.-Y., Villaneau, J., Vergnes, M., and Goulian, J. (2004). The French MEDIA/EVALDA Project: the Evaluation of the Understanding Capability of Spoken Language Dialog System. In *Proceedings of LREC 2004*, Lisbon, Portugal.

Eschenbach, C., Habel, C., Herweg, M., Rehkämper, K., (1989). Remarks on plural anaphora. In *Proc. Fourth Conference of the European Chapter of the Association for Computational Linguistics.*

Kamp, H. and Reyle, U. (1993). *From Discourse to Logic: Introduction to Model-theoretic Semantics of Natural Languge. Formal Logic and Discourse Representation Theory.* Kluwer Academic Publisher.

Landragin, F. and Romary, L. (2003) Referring to Objects Through Sub-Contexts in

Multimodal Human-Computer Interaction. In *Proc. Seventh Workshop on the Semantics and Pragmatics of Dialogue* (DiaBruck'03), Saarbrücken, Germany, 2003, pp. 67-74.

Olson D. (1970). Language and Thought: Aspects of a Cognitive Theory of Semantics. *Psychological Review*, 77/4, 257-273.

Pitel G., Sansonnet J-P. (2003) A Differential Representation of Predicates for Extensional Reference Resolution, In *Proc. of the 2003 International Symposium on Reference Resolution and its Application to Question Answering and Summarization*, Venice, Italia, june 2003

Salmon-alt, S. (2000) Interpreting referring expressions by restructuring context. *Proc. ESSLLI 2000, Student Session*, Birmingham, UK, August 2000.

Salmon-Alt, S. (2001) Reference Resolution within the Framework of Cognitive Grammar. *Proc. International Colloquium on Cognitive Science*, San Sebastian, Spain

Sperber, D. and Wilson, D. (1986) *Relevance, Communication and Cognition*. Basil Blackwell, Oxford.

Ambiguity in Anaphora Workshop Proceedings, pages 41–48
ESSLLI 2006, Málaga, Spain, 7–11 August 2006
Edited by Ron Artstein and Massimo Poesio
http://cswww.essex.ac.uk/anaphora/

THE ROLE OF INFORMATION STRUCTURE IN INTERPRETIVE ASYMMETRIES

Maia Duguine
EHU-U. Basque Country & U. Nantes-Naoned

Abstract[1]

This paper analyses one aspect of the so-called *two-pronoun puzzle* (or *Dahl's puzzle*). I argue that the two-pronoun puzzle hides a double puzzle; I call them Puzzle A and Puzzle B. Puzzle A is why one logically possible interpretation disappears in certain cases of ellipsis of pronominal dependencies; and the other way around, Puzzle B is why this reading is still possible when ellipsis doesn't apply. Puzzle A can be accounted for in terms of conditions on parallelism and semantic binding (cf. Fox 2000, Reinhart 2000, Büring 2005). Puzzle B has not yet been resolved in the same terms. In this paper I propose an explanation that is compatible with the analysis in Fox (2000). First, based on Basque data I show that the two-pronoun puzzle phenomenon is not restricted to VP ellipsis: it also concerns silent arguments. Then, I show how this Basque data help understanding better what is under Puzzle B: the facts appear to be less uniform once information structure is taken into account, and focus is shown to play a central role in the derivation that allows the crucial data.

1. The interpretive asymmetries of the two-pronoun puzzle

The two-pronoun puzzle, or Dahl's puzzle, is the name given to the interpretive asymmetries displayed in cases of ellipsis of pronominal dependencies (cf. Dahl 1974, Fiengo & May 1994, Fox 2000 a.o.).

In the conversation in (1)-(2) for instance, the sentence in (1), with all the pronouns with value 'Max' (indices are used just as shorthand to indicate the intended reading), can serve as an antecedent for the optional VP ellipsis in (2b) (the elided constituent is represented between brackets). In principle, the 'answers' to (1) in (2a) and (2b) are equivalent; they differ in that VP ellipsis is applied in the latter but not in the former. Depending on the reading each pronoun can get (strict 'Max' or sloppy 'Oscar'), the four interpretations in (3) are logically possible for the sentences in (2). However, since the first description of the data in Dahl (1974), it is considered that whereas (2a) allows these four interpretations, (2b) only allows (3a-b-c), excluding the interpretation in (3d), the '*strict + sloppy*' (henceforth ST-SL) reading. This asymmetry is the essence of the so-called two-pronoun puzzle.

(1) Max_i said he_i saw his_i mother.

[1] This research was supported by the grants BFF2002-04238-C02-01 of MCYT-FEDER, UPV-EHU 9 UPV 00114.130-160.09-2004 U of EHU-U. Basque Country and a Ph.D. research grant by EHU-U. Basque Country. I want to thank Aritz Irurtzun for comments and helpful discussion, and also Brendan Costello, Scott Fults and Heather Taylor, my informants of English.

(2) a. Oscar said he saw his mother, too.
b. Oscar did <*say he saw his mother*>, too.

(3) a. *Oscar said Max saw Max's mother*. [strict + strict]
b. *Oscar said Oscar saw Oscar's mother*. [sloppy + sloppy]
c. *Oscar said Oscar saw Max's mother*. [sloppy + strict]
d. *Oscar said Max saw Oscar's mother*. [strict + sloppy (ST-SL)]

What I would like to focus on in this paper is that the two-pronoun puzzle is in fact a double puzzle. On the one hand, if the logic gives four possible interpretations for a sentence, the question is why in some cases one of them is impossible. But on the other hand, if the impossibility of the ST-SL reading is accounted for in a principled manner, the problem gets reversed: if we continue to assume that (2a) and (2b) are equivalent and that they differ minimally in their phonological (non)realization, it is not clear why the sentence where ellipsis does not apply, i.e. (2a), allows the otherwise unavailable interpretation. These two puzzles are spelled-out in (4).

(4) Puzzle A: Why is the ST-SL reading impossible in VP ellipsis (cf. 2b)?
Puzzle B: Why is the ST-SL reading still possible in non-ellipsis (cf. 2a)?

Below we will see that despite its apparent redundancy, this dichotomy appears to be relevant. In this paper, I will take Fox's (2000) account of Puzzle A as a point of departure (section 2), and I will show that it does not give a straightforward way of answering to the second question in (4), Puzzle B (section 3). Then I will develop an explanation in terms of Fox's system, proposing that it is the absence of application of the condition on Parallelism that allows the ST-SL reading. First, based on Basque data I will show that the two-pronoun puzzle phenomenon is not restricted to VP ellipsis: it also concerns Basque silent arguments (section 4). Then, I will show that this Basque data can help us understanding better what is under Puzzle B: there are in fact two different structure under the allegedly uniform (2a), and the structure permitting the ST-SL reading is different from the one permitting the other readings (sections 5 and 6). Section 7 is the conclusion.

2. Puzzle A

Puzzle A can be explained as the effect of the interaction of two constraints (cf. Reinhart 2000, Fox 2000, Büring 2005). On the one hand the parallelism condition on ellipsis in (5) ensures that an elided constituent receives an interpretation parallel to that of an antecedent (cf. Chomsky & Lasnik 1993, Fiengo & May 1994, Fox 2000). An intuitive characterization of Parallelism will be sufficient for our purposes here (cf. Dalrymple et al. 1991, Fiengo & May 1994, Hardt 1999, Merchant 2001 a.o. for a discussion). What is important is that semantic binding relations are part of this characterization: for a constituent to be elided, the binding relations in it will have to be the same as those in its antecedent (cf. Fox 2000), under the assumption that ellipsis sites have full-fledged internal structure (cf. Hankamer & Sag 1976, Fiengo & May 1994, Lasnik 1995, Merchant 2001, a.o.).

(5) *Parallelism condition on ellipsis*
A constituent can be elided only if it has a parallel antecedent.

On the other hand a constraint on the semantic binding of pronouns gives the possible representations of sentences (cf. Reinhart 1983, 2000, Heim 1993, Fox 2000). In this paper I

will assume that the constraint on binding is an economy rule defining the possible antecedents for a pronoun; (6) is Fox's (2000) Rule H (cf. also Büring 2005):

(6) *Rule H*
A pronoun α can be bound by an antecedent β only if there is no closer antecedent γ such that it is possible to bind α by γ and *get the same semantic interpretation*.

Let us see how the interaction between Parallelism and Rule H can account for Puzzle A in Fox (2000). First, Rule H will give the actual representation of the pronominal dependencies in (1). Two representations are logically possible: (7) and (8). However, contrarily to (7), the binding of the second pronoun in (8) is made across another possible antecedent *he*, and it does not give a different semantic interpretation: both representations give the interpretation where 'Max said Max saw Max's mother'. Thus by Rule H, the actual representation of the binding relations in (1) can only be the one in (7), the representation in (8) being uneconomical.

(7) Max said he saw his mother.

(8) Max said he saw his mother.

Then, by Parallelism, the semantic binding relation of the possessive pronoun in the elided VP in (2b) will be the same as that in its parallel VP antecedent in (1), that in (7); hence the possessive pronoun will be able to be bound by the embedded subject pronoun, but, crucially, not by the main subject. This only possible binding relation is represented in (9):

(9) Oscar said he saw his mother

Now, if we look at the representation necessary for the ST-SL interpretation of (2b), illustrated in (10), the possessive pronoun has to be bound by the subject in the main clause.

(10) Oscar said he saw his mother [*Oscar said Max saw Oscar's mother*]

The representation of the binding of *his* in (10) is parallel to that of (8), but we have seen that by Rule H the representation in (8) is ungrammatical. In consequence, the demands of Rule H and Parallelism cannot converge in order to allow the ST-SL interpretation of the elided VP in (2b).

3. Puzzle B

Fox's (2000) system presented above gives an explanation for what we called Puzzle A. In the remaining part of this paper, I will be concerned with the second part of the two-pronoun puzzle, the Puzzle B.

Let us first spell out Puzzle B in the light of the analysis of Puzzle A given above. Under the assumption that ellipsis is an optional operation and that the ellipsis site has a full syntactic structure, (2b) has to have a counterpart in which the VP is overt and that has the same interpretation (that is, a deaccented counterpart; cf. Tancredi 1992, Rooth 1992,

Chomsky & Lasnik 1993). At first sight (2a) seems to be that counterpart, because the VP is intuitively 'parallel' to that of (1); but this is impossible, because the interpretation of (2a) and (2b) differ in that the former but not the latter allows for the ST-SL reading. However, all the things being equal, following the system in Fox (2000), Parallelism and Rule H would be predicted to apply equally, be the VP phonologically realized or not. Hence there is no apparent reason for (2a) to allow for the ST-SL reading. This 'paradox' is the essence of Puzzle B.

We have to conclude that it is only apparently that the sentence in (2a) is the overt equivalent of the sentence in (2b) (at least in the case in which we get the ST-SL interpretation). What is the difference between them? We saw in section 2 that it is the combination of the particular effects of Rule H and Parallelism that blocks the ST-SL interpretation in (2b); thus it means that the same combination is not met in the case where (2a) has ST-SL reading, that is, the effect of either Rule H or Parallelism is different. First, obviously, it cannot be that the effect of Rule H is different; it cannot be that Rule H gives a different representation for (1) when (2a) has ST-SL interpretation. Rule H applies automatically whenever there is a pronoun and its possible antecedents in a derivation; as seen above, its result is unequivocal in (1), and the only possible representation is the one in (7). Hence, the difference between (2a) and (2b) has to lie in the effect of Parallelism. The hypothesis is the following: Parallelism does not apply between (1) and (2a); in consequence the interpretation of the pronominal dependency in the latter is not dependent on that in the former. More precisely, by Rule H (1) will have the representation in (7) and (2a) will have the representation in (10); the difference between both binding relations is not blocked by Parallelism. In the following section I show that Basque gives evidence in favor of an analysis of Puzzle B in terms of a non-application of Parallelism.

4. The two-pronoun puzzle in Basque

Basque shows the same interpretive effects as English (1)-(2), but in an environment different from VP ellipsis. This language, which allows its subjects, direct objects and indirect objects to be phonologically empty (cf. Goenaga 1980, Ortiz de Urbina 1989), displays the interpretive effects of the two-pronoun puzzle when the pronouns in embedded contexts are phonologically silent (phonologically empty arguments are indicated by [e]). More specifically, in the conversation in (11)-(12), when the second, possessive pronoun is realized (cf. (12a)), the ST-SL reading in (13d) is available; but when it is phonologically empty (cf. (12b)), it is not available. The (13a-b-c) readings are available for both (12a) and (12b); and the (non-)realization of the first pronoun has no effect on the possible readings. As in the two-pronoun puzzle in English, phonological emptiness in Basque (in this case, phonological emptiness of the second, possessive pronoun) corresponds to the case where the ST-SL reading 'disappears': the second pronoun must be overt in order to display the ST-SL reading.

(11) Peiok$_i$ esan digu [berak$_i$ bere$_i$ ama ikusi duela]
Peio say Aux s/he his mother see Aux.that
'Peio told us that he saw his mother.'

(12) a. Jonek esan digu [berak/[e] bere ama ikusi duela]. [√ ST-SL]
Jon say Aux s/he his mother see Aux.that
'Jon told us that (he) saw his mother.'

b. Jonek esan digu [berak/[e] [e] ikusi duela]. [* ST-SL]
Jon say Aux s/he see Aux.that
'Jon told us that (he) saw [e].'

(13) a. *Jon told us that Peio saw Peio's mother*. [strict + strict]
b. *Jon told us that Jon saw Jon's mother*. [sloppy + sloppy]
c. *Jon told us that Jon saw Peio's mother*. [sloppy + strict]
d. *Jon told us that Peio saw Jon's mother*. [ST-SL]

If we take the two-pronoun puzzle as a property of ellipsis (cf. Fiengo & May 1994), the natural conclusion is that the phonologically empty subjects in (12) are indeed elided constituents (cf. Duguine 2006). In consequence, their behavior will be representative of the phenomenon of ellipsis.

5. Accounting for the Basque data

In this section, I show that the Basque data in (11)-(13) can be accounted for by the interaction between the properties of ellipsis and information structure.

Crucially, Basque speakers agree in that the second pronoun in (12a) has to be 'emphasized' in order to get the ST-SL reading. That is, it has to be focused. Basque focused constituents are subject to syntactic movement to the left periphery followed by verb-movement, resulting in an obligatory focus-to-verb adjacency, the focused phrase ending in the immediate preverbal position (cf. Ortiz de Urbina 1989). This formal characteristic of Basque makes a prediction. Whenever a pronoun is in postverbal position, it is not focused; hence if the second pronoun in (12a) appears in a postverbal position, the ST-SL interpretation should not be available. This prediction is actually borne out: the sentence in (14) below, as an answer to (11), cannot be interpreted as *Jon said Peio saw Jon's mother* (the readings in (13a-b-c) are however available)

(14) Jonek esan digu [berak/[e] ikusi duela bere ama].
Jon say Aux s/he see Aux.that his mother
'Jon told us that he saw his mother.'

In consequence, what the minimal pair (12a)-(14) shows is that the second pronoun has to be fronted, that is, focalized, in order for the sentence to display the ST-SL reading.

Until this point I have shown that (i) the empty arguments in (12) are elided constituents, (ii) in sentences displaying the ST-SL reading the second pronoun is necessarily overt, and (iii) in sentences displaying the ST-SL reading the second pronoun is necessarily focalized. From (ii) and (iii), it results that focalized pronouns are necessarily overt. Then, if, as stated in (i), the arguments of the verb are subject to ellipsis in Basque, it is natural to conclude that focused pronouns cannot be subject to ellipsis: focalization blocks ellipsis (cf. Merchant 2001, Han & Romero 2004 a.o. for similar ideas). The generalization is thus the following:

(15) *Generalization on focalization:*
A focused constituent cannot be subject to ellipsis.

Summarizing, pronouns are subject to ellipsis in Basque, but when they are focused, they cannot be elided. This explains why if there is a certain information structure configuration that allows a given interpretation (in this case a ST-SL reading), the elements participating in this configuration cannot be phonologically empty; it is only when the second pronoun is overt and focused that the ST-SL reading will be displayed.

6. An explanation for Puzzle B

The Basque data just presented show that the ST-SL reading is to be related to the focalization of the second pronoun. Let us return now to the Puzzle B of English described above. With regards to the sentence in (2a), speakers of English esteem that the ST-SL reading is more clearly available when the second pronoun is phonologically emphasized than when it is not. It is interesting to see that the intuition on the second pronoun being emphasized is present in the literature, but is not taken as being relevant. In Reinhart (2000) the equivalent of (2a) is described as "funny (as long as the destressing and intonation pattern required by *too* is kept)". Also, in Fiengo & May (1994: 130) it is explained in a footnote that "a speaker may place emphasis on one or the other of the pronouns in the string in order to bring out which of the [interpretations] the string corresponds to". However this is not considered as relevant: "doing so … does not transform [(2a)] into a different string in terms of anaphoric possibilities".

The intuitions of the English speakers and the notes in Reinhart (2000) and Fiengo & May (1994) point towards a unification of the phenomena in Basque and English. Obviously, the emphatization alluded to is focalization; hence in English as in Basque, the ST-SL reading is permitted only when the second pronoun is focused. If (2b) is derived from VP ellipsis, this means that the second pronoun is elided as part of the main VP. However, if this pronoun is focused, and under the assumption that the generalization in (14) holds in English too (cf. Han & Romero 2004), the impossibility of eliding the pronoun will also block the ellipsis of the bigger VP constituent. In consequence, it is only the version of the sentence in (2) where the second pronoun is focused that will be able to display the ST-SL interpretation, that is, the one where the VP is overt, (2a).

Summarizing, (2b) does not allow the ST-SL reading because it does not permit the possessive pronoun to be interpreted as focused. Hence, in fact (2a) as such does not allow for the ST-SL interpretation: there is a (2c) sentence that differs minimally from (2a) by having a F(ocus) feature on the pronoun:

(1) Max_i said he_i saw his_i mother.

(2) a. Oscar said he saw his mother, too.
b. Oscar did <*say he saw his mother*>, too.
c. Oscar said he likes $\text{his}_{[F]}$ mother, too.

Going back to the question of Puzzle B, we can see that the fact that the second pronoun is focused explains why Rule H can give the representation in (10) for (2c). The antecedent sentence in (1) has a non-focused second pronoun whereas the one in (2c) has a focused second pronoun. If this is so, the VPs are not parallel; focusing one constituent in the second VP violates Parallelism. But now, if Parallelism does not hold, the representation in (2c) is not dependent on that of the preceding sentence, and in consequence the representation of the binding relation does not have to be parallel to that of (1). In (2c), the first pronoun gets its assignment from the context ('Max'); by Rule H then, the closest possible antecedent for the second pronoun (with the referential value 'Oscar') is the subject of the main clause, as shown by the representation in (16) (capital letters indicate focusing).

(16) Oscar_j said he_i saw HIS_j mother, too.

In consequence, once the focusing effects identified, the system proposed in Fox (2000) can also account for the Puzzle B part of the two-pronoun puzzle. The 'puzzle'

derived in fact from a minimization of the role that focus can play with regards to Parallelism.

But now that we know that in a sense it is the focus on the possessive pronoun that allows the ST-SL reading, one question arises: is this the only effect of focus here? Focus is supposed to be subject to felicity conditions that have nothing to do with anaphoric dependencies and (most of the time) Parallelism. The question in fact has to be asked in the following manner: is the use of focus in (2c) felicitous? A priori it should be the case, because under the assumption that focus has uniformly a pragmatic and semantic effect (on the truth-conditions of sentences, in the sense of Jackendoff (1972)), violation of the conditions regulating these effects should lead to non-felicity. A brief look at the basis of the phenomenon of focus, 'contrastivity', used here in the broad sense, will show that focus is legitimate with regards to the constraint regulating its use. For the sake of simplicity, I will assume a definition of focus based on contrastivity: focus on a constituent makes the sentence contrast with a discursive antecedent, be it a question or a statement (cf. Rooth 1992, Roberts 1996/1998, Schwarzschild 1997, Kadmon 2001 for versions of this idea). (17) is Schwarzschild's (1994) *Constraint on Contrast in Discourse* as given in Kadmon (2001), and largely inspired by the *Contrasting Phrase Constraint* in Rooth (1992). It is based on the alternative semantics theory (Rooth 1992).

(17) *Constraint on Contrast in Discourse*
An utterance B is felicitously contrasted with another utterance A only if
$[[A]]^o \neq [[B]]^o$ and $[[A]]^o$ is a member or a subset of $[[B]]^f$

(18) $[[Y]]^o$ is the *ordinary semantic value* of Y.

(19) $[[Y]]^f$ is the *focus semantic value* of Y,
where $[[Y]]^f$ is the set of the propositions obtainable by replacing the focus with an alternative of the same type.

Let us take the plausible conversation in (20), where the ST-SL reading can be derived 'naturally'.

(20) A: Max_i said he_i saw his_i mother.
B: (No, this cannot be true:) $Oscar_j$ said he_i saw HIS_j mother. [ST-SL]

Here, we have (replacing pronouns by proper names for ease of presentation):

(21) a. $[[(20A)]]^o$ = 'Max said Max saw Max's mother'
b. $[[(20B)]]^o$ = 'Oscar said Max saw Oscar's mother'
c. $[[(20B)]]^f$ = {'Oscar said Max saw Max's mother', 'Oscar said Max saw Oscar's mother'}

$[[(20A)]]^o$ and $[[(20B)]]^o$ differ, and $[[(20A)]]^o$ is a member of $[[(20B)]]^f$, thus (20A) and (20B) are felicitously contrasted. In consequence, not only does focus on the second pronoun in (20B)/(2c) 'ensure non-parallelism', but its use is also felicitous in its discursive context.

8. Conclusion

In this paper, in the light of some recent proposals in the semantic binding literature, especially the one in Fox (2000), I showed that half of the two-pronoun puzzle remained

unaccounted for. I proposed an explanation for the fact that the sentence where the VP is overt does not behave like the sentence where it is elided, despite they are apparently equal and they are both subject to Rule H and Parallelism. Data from Basque showed that what was considered a uniform phenomenon (phonological realization vs. non-realization) is not so uniform, given that information structure comes to play a crucial role, changing the effects of Parallelism and Rule H. In consequence, the present work gives further evidence in favor of the proposal in Fox (2000): once the differences between the actual possible derivations are identified, the system based on Parallelism and Rule H accounts for Puzzle B, and hence is able to give a complete explanation for the two-pronoun puzzle.

References

Büring, D. 2005. "Bound to bind", *Linguistic Inquiry* 36: 259-274.

Chomsky, N. and Lasnik, H. 1993. "The theory of principles and parameters", in *Syntax: An International Handbook of Contemporary Research*, Vol. 1., J. Jacobs, A. von Stechow, & T. Vennema (eds), 506-569. Berlin and New York: Gruyter.

Dahl, Ö. 1974, "How to open a sentence. Abstraction in natural language", in *Logical Grammar reports* 12, University of Götenberg.

Dalrymple, M., Shieber S., and Pereira F. 1991. "Ellipsis and higher-order unification", *Linguistics and Philosophy*, 14: 399-452.

Duguine, M. 2006. "Silent arguments without *pro*: the case of Basque", ms. EHU-U. Basque Country.

Fiengo, R. and May, R. 1994. *Indices and Identity*, Cambridge, MA: MIT Press.

Fox, D. 2000. *Economy and semantic interpretation*, Cambridge, MA: MIT Press.

Goenaga, P. 1980. *Gramatika bideetan* (2nd ed.). Donostia: Erein.

Han, C. & Romero, M. 2004. "Disjunction, focus, and scope", *Linguistic Inquiry* 35: 179-217.

Hankamer, J. & Sag, I. 1976. "Deep and Surface Anaphora", *Linguistic Inquiry* 7: 391-428.

Hardt, D. 1999. "Dynamic interpretation of verb phrase ellipsis", *Linguistics and Philosophy* 22: 185-219.

Heim, I. 1993. "Anaphora and Semantic Interpretation: a Reinterpretation of Reinhart's Approach", SfS-Report-07-93, University of Tubingen, Tubingen.

Kadmon, N., 2001. *Formal Pragmatics*, Oxford: Blackwell.

Lasnik, H. 1995. "A note on pseudogapping", *Papers on Minimalist Syntax, MIT Working Papers in Linguistics* 27: 143-163.

Mechant, J. 2001. *The Syntax of Silence*, Oxford: Oxford University Press.

Ortiz de Urbina, J. 1989. *Parameters in the Grammar of Basque: A GB approach to Basque Syntax*. Dordrecht: Foris.

Reinhart, T. 1983. *Anaphora and semantic interpretation*, London: Croom Helm.

Reinhart, T. 2000. "Strategies of Anaphora Resolution", in H. Bennis, M. Everaert and E. Reuland (eds) *Interface Strategies North Holland Amsterdam*, 295-324.

Roberts, C. 1996/1998. "Information Structure in Discourse: Towards an Integrated Formal Theory of Pragmatics", ms., Semantics Archive (www.semanticsarchive.net). Originally published in *Ohio State University Working Papers in Linguistics*, 49: 91-136.

Rooth, M. 1992. "A theory of focus interpretation", *Natural Language Semantics* 1: 75-116.

Schwarzschild, R. 1997. "Why some foci must associate", ms., Rutgers University.

Tancredi, C. 1992, *Deletion, Deaccenting and Presupposition*, Ph.D. dissertation, MIT.

Ambiguity in Anaphora Workshop Proceedings, pages 49–56
ESSLLI 2006, Málaga, Spain, 7–11 August 2006
Edited by Ron Artstein and Massimo Poesio
http://cswww.essex.ac.uk/anaphora/

Evaluating a Coherence-Based Model of Pronoun Interpretation

Laura Kertz, Andrew Kehler, Jeffrey L. Elman

Abstract

We describe two pronoun interpretation experiments in which a Coherence Hypothesis is tested against preference-based accounts. The Coherence Hypothesis holds that apparent preferences in antecedent selection are actually byproducts of the inferencing processes used to establish different types of coherence. In Experiment 1 we show that preferences can be systematically disrupted through the manipulation of coherence and that when the relevant factors are balanced, preferences disappear. In Experiment 2 we show that the coherence effect is not disrupted by voice alternations (active/passive), providing evidence for a strong semantic model of coherence-driven interpretation. We speculate on the adequacy of this strong semantic model and propose additional online experiments to examine the interaction between propositional content and information structure influences on pronoun interpretation.

1. Introduction

The pronoun interpretation literature over the last three decades has followed two main lines of investigation. One approach casts pronoun interpretation as a matching process guided by 'heuristics', that is, broadly-applicable strategies for matching pronouns to their antecedents. Of particular relevance to our purposes is a body of work developed in the 1990's that examined two competing preferences: the parallel function preference (Smyth 1994, Chambers & Smyth 1998, inter alia) and the subject assignment preference (Crawley & Stevenson 1990, Crawley et al. 1990, inter alia). Centering Theory (Grosz et al. 1995 [1986], Brennan, Friedman, and Pollard 1987) incorporates neither of these preferences directly, but shares the property that pronoun interpretation is driven in large part by relationships based on grammatical roles.

In this paper, we describe two psycholinguistic experiments that support a model in which the interpretation heuristics posited in the literature are epiphenomena of processes that hearers use to establish discourse coherence. In Experiment 1, we demonstrate (following Kehler 2002) that such preferences -- in particular, the parallel function preference and the subject assignment preference -- are not reliable indicators of pronoun interpretation. In Experiment 2, we test the interaction of coherence with referent accessibility using a voice (active/passive) manipulation. We find that in an off-line task, coherence trumps accessibility. We conclude with a discussion of the possible limitations of these findings, and suggest further research using online measures.

2. Preferences and Coherence

Heuristics-based models of pronoun interpretation predict that morpho-syntactic cues can generally be used to identify a single preferred antecedent. These heuristics, however, are often in conflict. For example, the subject assignment preference identifies the preferred antecedent as the subject of the preceding clause, a condition which holds in (1a) and (1c), but not in (1b) and (1d). The parallel function preference, on the other hand, predicts that an antecedent will appear in the same argument position as the pronoun, a claim supported by (1a) and (1b), but not in (1c) or (1d). A modified version of the parallel function preference that requires full syntactic parallelism between the clauses (Smyth 1994) suggests that the presence/absence of a modifier, as in (1a-d), may shift these preferences.

(1)	Samuel threatened Justin with a knife, and		*reference*	*coherence*
a.	he blindfolded Erin (with a scarf).	(=Samuel)	parallel	parallel
b.	Erin blindfolded him (with a scarf).	(=Justin)	parallel	parallel
c.	Erin stopped him (with pepper spray).	(=Samuel)	non-parallel	cause/effect
d.	he alerted security (with a shout).	(=Justin)	non-parallel	cause/effect

The preference accounts stand in contrast to a coherence-driven model, as originally proposed by Hobbs (1979), which suggests that pronoun interpretation is not driven by morpho-syntactic cues but rather is the by-product of larger inferencing processes that support the establishment of discourse coherence. Kehler (2002) extends Hobbs's proposal, arguing that interpretation 'preferences' are actually epiphenomena of the manner in which different types of coherence are established. Preliminary support for this model was reported by Wolf et al. (2004), which we extend in several ways here. Specifically, we test Kehler's (2002) predictions in behavioral experiments, focusing on two coherence types: parallel coherence, which obtains between clauses with similar propositional content, and cause/effect coherence, which obtains between clauses denoting events that are causally linked.

3. Experiment 1

Kehler (2002) suggests that the conflicting preferences reported in the psycholinguistics literature often result from a failure to control for coherence across stimuli. To address this, we constructed stimulus sets which exhibit parallel coherence in half the cases and cause/effect coherence in the other half. As in Wolf et al, the cause/effect cases incorporated a semantic bias toward the non-parallel referent; the parallel cases incorporated no bias. Coherence was assessed during a prior norming study. Note that we used only *and* as a connective to guard against potential 'focusing' effects of connectives like *and similarly* or *and so* as claimed by Stevenson et al. (2000). We also varied the choice of verb class to guard against the possibility of a thematic role bias (Stevenson et al. 1994, Stevenson et al. 2000).[1]

To test the parallel function preference, we also varied pronoun position. Recall that while the subject assignment strategy predicts an across-the-board preference for subject

[1] Verbs in the introductory clause were drawn from one of four classes: physical action (threaten, punch, check, injure); social action (deceive, taunt, harass, salute); mental state (demoralize, admire, irritate, intimidate); and speaking verbs (bait, defend, suggest, ridicule). See Rohde et al (2006) for studies examining the interaction of thematic roles, event structure, and coherence.

antecedents, the parallel function preference predicts a subject preference for subject pronouns, and an object preference for object pronouns (a main effect of pronoun position). The modified parallel function preference suggests that this effect may be mitigated when the two clause structures are not perfectly parallel (an interaction between pronoun position and syntactic structure). To test this, we varied syntactic structure, creating a perfectly parallel condition in which an adverbial modifier in clause 2 matches the modifier in clause 1 (i.e., the parentheticals in 1a-d are included), and a partially parallel condition in which no modifier appears in clause 2 (i.e., the parentheticals in 1a-d are not included). The Coherence Hypothesis predicts parallel coreference in the parallel coherence condition, and non-parallel reference in the cause/effect coherence condition (i.e. an interaction between coherence and pronoun position).

3.1 Method

- Participants were 32 undergraduates who received extra credit for participation. All were self-reported mono-lingual native speakers of English.

- In a 2x2x2 design, we constructed 16 stimulus sets, each with 8 variants, for a total of 128 stimuli, as described above. Verb class and modifier type were also balanced across stimuli. A sample stimulus set is presented above in (1).

- In a repeated measure design, participants were tested on two instances of each stimulus type, but no participant saw more than one variant from the same set. The 16 experimental stimuli were interleaved with 24 distracters (16 stimuli from experiment 2; and 8 fillers with unambiguous pronouns), and the replications were block randomized.

- Participants were presented with a paper and pencil task in which the stimulus was followed by a question requiring that the pronoun be interpreted, as in (2). For answers matching the subject of the preceding clause, a score of 1 was assessed. For answers matching the object of the preceding clause, a score of 0 was assessed.

(2) Samuel threatened Justin with a knife, and he blindfolded Erin with a scarf. *Who blindfolded Erin?*

3.2 Results

The data showed strong support for the Coherence Hypothesis, confirming the predicted interaction between pronoun position and coherence frame. The data were not consistent with any of the other hypotheses. Gross percentages across stimuli show that the 'preferences' operate at roughly chance levels when the relevant factors (including coherence) have been balanced.

Table 1 Experiment 1

Hypothesis	*Condition*	*Antecedent* *Subj*	*Obj*	*n*
Subject Assignment	All pronouns	.52	.48	512
Parallel Function	Subject pronouns	.51	.49	256
	Object pronouns	.52	.48	256
Modified Parallel Function Coherence Hypothesis	Subject pronoun: fully parallel structure	.52	.48	128
	Object pronoun: fully parallel structure	.50	.50	128
	Subject pronoun: Parallel coherence	.98	.02	128
	Subject pronoun: Result coherence	.05	.95	128
	Object pronoun: Parallel coherence	.10	.90	128
	Object pronoun: Result coherence	.94	.06	128

Statistical analysis of these data confirms that the interaction between coherence frame and pronoun position, predicted by the Coherence Hypothesis, is significant [$F1(1,31) = 1379.23$, $p < .0001$; $F2(1,15) = 2016.158$, $p < .0001$]. A second, smaller effect, which we did not predict, was found for coherence alone [$F1(1,31) = 4.429$, $p = .044$; $F2(1,15) = 7.105$, $p = .018$].

4. Experiment 2

The results of Experiment 1 support the Coherence Hypothesis. They show that pronoun interpretation does not depend exclusively on pronoun position or syntactic structure, both of which were balanced across stimuli. Nor was the preferred antecedent signaled by discourse connectives like *and similarly* or *and so* (as suggested by Stevenson et al. 2000). Rather, pronoun interpretation depended primarily on the semantic content of the propositions and whether the relationship between them is based on similarity or causality.

While the results of Experiment 1 appear to support the type of 'strong' coherence-based model we have described, they do not rule out the possibility of a 'weaker' coherence model, in which coherence interacts with other factors such as information structure. To differentiate these possibilities, Experiment 2 tested participants on passages in which the voice of the context sentence was manipulated. Because the passive marks the surface subject as topical (and likewise reduces the topicality of the logical subject), we might expect support for pronominal reference to the logical subject via coherence to be compromised in such constructions.

Thus, the strong semantic model predicts the replication of the Experiment 1 results in both the active and the passive conditions: parallel co-reference with parallel coherence; non-parallel co-reference with cause/effect coherence. (Note, importantly, that parallelism now refers to thematic role, not argument position, i.e. semantically parallel, not syntactically). Thus, when we exclude pronoun position as a factor (using only object pronouns), the choice of antecedent (subject/non-subject) will depend on an interaction between coherence and voice. The weak model also predicts an interaction between coherence and voice, but it predicts a different pattern, one in which the clean results seen in Experiment 1 are

replicated in the active voice, but where the interaction is weakened in the passive, with voice playing a greater role.

4.1 Method

- Participants were the same as in Experiment 1.

- In a 2x2 design, we adapted 16 stimulus sets from Wolf et al. (2004), constructing 4 variants for each set, for a total of 64 stimuli. Stimuli contained all object pronouns, but varied in coherence frame (parallel vs. cause/effect; based on verb semantics) and voice (active/passive). A sample stimulus set is presented in (3).

- In a repeated measure design, participants were tested on four instances of each stimulus type. All remaining aspects of the design, including task, were identical to Experiment 1.

(3) a.	James complimented Craig, and Fiona congratulated him after the match.	parallel coherence active voice
b.	James defeated Craig, and Fiona congratulated him after the match.	cause/effect coherence active voice
c.	James was complimented by Craig, and Fiona congratulated him after the match.	parallel coherence passive voice
d.	James was defeated by Craig, and Fiona congratulated him after the match.	cause/effect coherence passive voice

4.2 Results

The data show a clear interaction between coherence and voice, predicted by both models. The pattern in the data, however, is symmetrical for active and passive, supporting the strong coherence model. Passivization does not appear to weaken the interaction between coherence and voice: we observe parallel reference in the parallel coherence frame, and non-parallel reference in the cause/effect frame. (The co-reference pattern reverses, however; in the passive, the patient role is realized in subject position.)

Table 2 Experiment 2

Condition		***Antecedent Subj***	***Non-Subj***	***n***
parallel coherence:	active voice	.22	.78	128
cause/effect coherence:	active voice	.84	.16	128
parallel coherence:	passive voice	.77	.23	128
cause/effect coherence:	passive voice	.23	.77	128

Statistical analysis of these data confirms that the interaction between coherence frame and voice is significant [$F1(1,31) = 278.52$, $p < .0001$; $F2(1,15) = 118.64$, $p < .0001$]. In the item analysis, an additional interaction was detected among item, voice, and coherence [$F2(1,15) = 2.712$, $p = .0005$], indicating some degree of variability in the stimulus set, which we are currently investigating.

5. Discussion

The results of Experiment 1 confirm the prediction in Kehler (2002) that pronoun interpretation preferences can best be understood as epiphenomena of coherence-based inferencing processes. Parallel coherence results in parallel co-reference, whereas causal inference determines co-reference in cause/effect coherence frames. The Coherence Hypothesis makes the correct predictions in cases in which the preference accounts conflict, and also predicts the subject-pronoun-to-object-antecedent pattern in the cause-effect passages that contradict both preference hypotheses. (See Table 3.) Further, the results of Experiment 2 suggest that this strong coherence-driven model of pronoun interpretation is resilient to the passive/active voice distinction.

Table 3 Predicted Resolution Patterns (Pronoun to Antecedent)

Hypothesis	***subject to subject***	***subject to object***	***object to object***	***object to subject***
Subject Preference:	yes	no	no	yes
Parallel Preferences:	yes	no	yes	no
Coherence:	yes	yes	yes	yes

Whereas the effects appear to be robust in parallel and cause/effect coherence frames, at this point we cannot be certain that it will generalize to other coherence types. It may be the case that parallel and cause/effect coherence relations rely more heavily on semantic content than do other forms of coherence, for example, narrative coherence (e.g., Hobbs's Occasion relation).

A variety of effects associated with the subject preference, topic preference, and first mention effects (among others reported in the literature), appear to be attention-based (cf. various forms of Centering Theory). Kehler (2002) suggests that attentional effects may be most prominent in Occasion frames, in which a series of events connected in time and/or space are described. For example, in a stimulus like (4), the most likely coherence frame in the final sentence is Occasion.

(4) Betty and Sally had been pestering the firm for months to buy a new photocopier. Duncan was hoping that he would be able to use it too. Betty demonstrated the new machine to Sally and Duncan asked her about it. (Crawley et al. 1990)

While it is not impossible to draw either a causal or similarity-based connection between the demonstration and the questioning, such inferences are much weaker than the sort drawn in the parallel and cause/effect variants of example (1). It is possible that Occasion is intertwined with notions such as topichood and salience to a greater degree than are the other relations. Experiment 2 did not show an interplay between attention and coherence, but if such an effect can be demonstrated, this would support a richer model of coherence recognition and pronoun interpretation than previously described (for example, in Hobbs 1979).

6. Future Work

While our off-line experiments showed that participants identify the referents that are supported by coherence establishment, they provide no measure of the difficulty with which those identifications were made. As such, experiments currently in preparation will use manipulations similar to those described here, but embedded within a self-paced reading task. We hope that this online measure will allow a fuller view of the processes involved in interpretation than the offline disambiguation task. In particular, we will examine cases like (1a) and (1d), repeated here as (5) and (6), in which the semantic information supporting coherence appears after the pronoun. We are particularly interested in whether any biases introduced by the subject assignment and/or parallel function preferences cause examples like (5) to be easier to process than those like (6), in which both heuristics point to the semantically dispreferred referent.

(5) Samuel threatened Justin with a knife, and he blindfolded Erin (with a scarf).
(6) Samuel threatened Justin with a knife, and he alerted security (with a shout).

We will also look for reading time delays in cases like (3d) from Experiment 2, repeated here as (7), for which we predicted a clash between information structure (supporting the subject antecedent) and coherence (supporting a non-subject antecedent).

(7) James was defeated by Craig, and Fiona congratulated him after the match.

Finally, the norming phase of Experiment 1 turned up a variety of interesting data relevant to the question of coherence recognition. These data suggest various ways that aspects of linguistic form and structure might affect the coherence relations that hearers identify, beyond the effects of semantics proper. We are currently developing experiments that control for these effects and initiating an investigation into the processes that drive coherence recognition.

References

Brennan, S. E., Friedman, M. W., & Pollard, C. J. (1987). A Centering Approach to Pronouns. Proceedings of the 25th Meeting of the Association for Computational Linguistics, 155-162.

Chambers, C. C., & Smyth, R. (1998). Structural Parallelism and Discourse Coherence: A Test of Centering Theory. Journal of Memory and Language, 39, 593-608.

Crawley, R. A., & Stevenson, R. J. (1990). Reference in Single Sentences and in Texts. Journal of Psycholinguistic Research, 19, 191-210.

Crawley, R. A., Stevenson, R. J., & Kleinman, D. (1990). The Use of Heuristic Strategies in the Interpretation of Pronouns, Journal of Psycholinguistic Research 19, 245-264.

Grosz, Barbara J., Joshi, A. K., & Weinstein, S. (1995 [1986]). Centering: A Framework for Modelling the Local Coherence of Discourse, Computational Linguistics 21:2, 203-225.

Hobbs, J. R. (1979). Coherence and Coreference, Cognitive Science 3, 67-90.

Kehler, A. (2002). Coherence, Reference, and the Theory of Grammar. Stanford, CA: CSLI Publications.

Rohde, H., Kehler, A., and Elman, J. L. (2006). Event Structure and Discourse Coherence Biases in Pronoun Interpretation. To appear in the Proceedings of the 28th Annual Conference of the Cognitive Science Society, Vancouver, Canada.

Smyth, R. (1994). Grammatical Determinants of Ambiguous Pronoun Resolution. Journal of Psycholinguistic Research, 23, 197–229.

Stevenson, R. J., Crawley R., & Kleinman D. (1994). Thematic Roles, Focusing and the Representation of Events. Language and Cognitive Processes 9, 519–548.

Stevenson, R. J., Knott, A., Oberlander, J., & McDonald, S. (2000). Interpreting Pronouns and Connectives: Interactions among Focusing, Thematic Roles, and Coherence Relations, Language and Cognitive Processes 15:3, 225-262.

Wolf, F., Gibson, E. & Desmet, T. (2004). Discourse Coherence and Pronoun Resolution. Language and Cognitive Processes, 19:6, 665–675.

Laura Kertz
Department of Linguistics, 9500 Gilman Drive
La Jolla, CA 92093-0108 USA
kertz [at] ling.ucsd.edu

Andrew Kehler
Department of Linguistics, 9500 Gilman Drive
La Jolla, CA 92093-0108 USA
kehler [at] ling.ucsd.edu

Jeffrey L. Elman
Center for Research in Language, 9500 Gilman Drive
La Jolla, CA 92093-0515 USA
elman [at] crl.ucsd.edu

Ambiguity in Anaphora Workshop Proceedings, pages 57–64
ESSLLI 2006, Málaga, Spain, 7–11 August 2006
Edited by Ron Artstein and Massimo Poesio
http://cswww.essex.ac.uk/anaphora/

The German Temporal Anaphor *danach* – Ambiguity in Interpretation and Annotation

MAREILE KNEES
Institut für Germanistische Sprachwissenschaft
Friedrich-Schiller-Universität Jena
mareile.knees [at] uni-jena.de

In this paper I present different types of ambiguity that occur in annotating and resolving anaphoric adverbials. My analysis concentrates on the German temporal adverbial *danach* ("after that" or "thereafter"). However, issues discussed here go beyond the analysis of this particular German word. I assume that the types of ambiguity presented in this paper occur not only with other anaphoric adverbials like *therefore* or *besides* but also with other types of anaphoric expressions (e.g. personal pronouns). At least, the "Justified Sloppiness Hypothesis" proposed by Poesio et al. (2005) leads to this conclusion. First, I present an account of the meaning and anaphoric character of the German pronominal adverb *danach*. By means of a pilot study, I then show that the resolution of *danach* involves several types of ambiguity, namely structural and referential ambiguity. Regarding referential ambiguity, we can distinguish the following sub-types: 1. competition between temporal and situation referents, 2. competition between simple and structured situation referents and 3. competition between different situation referents. Finally, I make some suggestions for handling these types of ambiguity.

1 The German Pronominal Adverb *danach*

Following Webber et al. (2003) and Miltsakaki et al. (2004), adverbial connectives like *however*, *therefore* etc. are assumed to express (as subordinate and coordinate conjunctions) binary predicate-argument relations. Nonetheless, they only get one of their two arguments structurally, namely the matrix-clause. The other argument has to be derived anaphorically from the discourse context. In German there are so-called pronominal adverbs like *danach* and *dadurch* (*thereby*) which consist of an anaphoric element (e.g. *da-*) and a relational element (e.g. *-nach* or *-durch*) (cf. Rüttenauer 1978, Fraurud 1992, Pasch et al. 2003). Due to their anaphoric element these pronominal adverbs refer to a referent previously introduced into the discourse. By means of their relational element they can also establish a temporal, causal or any other discourse relation between the referent of the antecedent (e.g. a situation or a time-entiy) and the situation-referent of the matrix-clause of the anaphor (s. example (1)). Thus, they can function as relational anaphora and adverbial connectives.

(1) **[***Eine Abordnung des Münchner Polizeipräsidiums legte an der Gedenktafel einen Kranz für den 1972 getöteten Polizeibeamten Anton Fliegerbauer nieder.***]$_1$ Danach** begaben sich die Mitglieder der Deutsch-Israelischen Gesellschaft und der anderen Gruppen zu Fuß ins 17 Kilometer entfernte Dachau.

[*A delegation of the police headquarters of Munich put down a wreath at the commemorative plaque of the police officier Anton Fliegenbauer killed in 1972.***]$_1$ After that** the members of the German-Israelic Society and other groups walked to Dachau which was 17 kilometers away.

In order to find out which factors influence the resolution of the relational anaphor *danach*, I selected and classified a variety of occurrences of *danach* from a corpus study. In the TIGER corpus, only 45 of approximately 100 occurrences of *danach* (used anaphorically) express temporal succession. The remaining ones mean "according to" and introduce the

content of a previously mentioned 'documentation'-noun like *study*, *agreement* or *report*.[1] In this paper I focus on the temporal incidences of *danach*. In general, temporal *danach* expresses the temporal succession of two temporal entities which have been mentioned in the text and which are usually conceptually related to each other. Only situations (i.e. events, processes, states) and time-entities (given by a temporal expression like *2006* or *New Year*) have a temporal dimension and can be temporally located. The situation introduced in the anaphor-matrix-clause is temporally located with respect to some reference time in the previous discourse given by a temporal expression or by an expression referring to a situation. More precisely, *danach* locates the matrix-situation in the post-phase of the reference time (s. figure 1).[2]

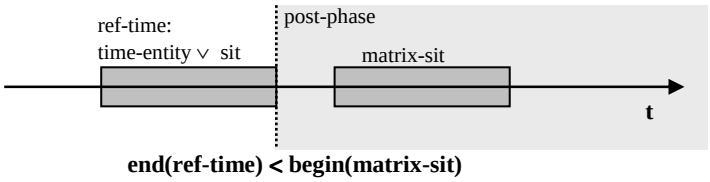


Figure 1: Temporal interpretation of *danach*

Syntactically, the antecedent of *danach* is either an S-based phrase (a sentence, a clause, a VP etc.) referring to a situation or an NP-based phrase (an NP or PP) referring to a situation, a time-entity or an object (coerced into a situation). Considering the ontological status of the referents of the antecedent and of the anaphor, we can distinguish between **situational** reference (the referent is a situation) and **temporal** reference (the referent is a time-entity).

1.1 The Pilot Study

As I assumed that the *danach*-occurrences of my corpus study might be ambiguous and therefore difficult to annotate, I did a pilot study with 6 test persons (all of them linguists). I selected 24 examples (ranging from easy to tricky) and asked the subjects not only to mark the antecedent but also to name the referent of *danach.* As I annotated the data (i.e. marked the antecedent) I largely followed the annotation convention for marking arguments of discourse connectives proposed in the Penn Discourse Treebank annotation manual (cf. Prasad et al. 2006). Thus, only S- or NP-based phrases with all their attributes were possible antecedents of *danach*. To include all attributes has two advantages. First the antecedent is not fragmented too much, so the annotation is clearer and easier. Second the annotator does not have to decide which attributes are absolutely necessary and which are not – generally a hard decision, s. example (2). Here the modification *within the forthcoming 24 hours* is not crucial for the interpretation of *danach* since it focusses on the potential occurring of the event rather than on a specific moment. Nonetheless, the modification does not spoil the interpretation. Therefore, all attributes of the antecedents are contained in the annotations.

(2) Zu spät, wie amnesty international fürchtet, **[***weil die Todesurteile binnen der kommenden 24 Stunden vollstreckt werden könnten***]$_1$**. **Danach**, so steht zu vermuten, werden wir uns den Namen Ken Saro-Wiwa endlich gemerkt haben.

Too late, as amnesty international fears, [*since the death sentences could be executed within the next 24 hours*]$_1$. It could be supposed that **thereafter** we will keep in mind the name Ken Saro-Wiwa at last.

[1] As one of the reviewers pointed out Dutch *daarna* appears to behave like German *danach*, except that it does not have a content reading (meaning *according to*) but only a temporal one.

[2] My analysis of the temporal meaning of *danach* is influenced by work on temporal connectors by Herweg (1992), Schilder (2004) and Habel & Knees (2004).

The disadvantage of keeping all (even unnecessary) attributes is that the antecedent often only partly refers to the referent of the anaphor. This also happens regardless of its unnecessary attributes. Thus, in some cases, the referent of the anaphor has to be derived from the antecedent or better from the textual anchor. Another annotation convention is to choose as antecedent only the textually closest and shortest referential expression from which the referent of *danach* can be derived.[3] In the PDTB annotation manual this convention is called "Minimality Principle" (cf. Prasad et al. 2006:13).

The subjects of the pilot study did not know about the conventions.[4] I wanted to check whether these conventions are really necessary as they are quite strict and for reasons just illustrated not completely sound. Moreover, I wanted to know whether and how my annotations differ from those of the subjects.

In general, the subjects had less problems to name the referent than to mark the antecedent. There are two possible explanations for this observation. First, as the subjects did not know about the annotation conventions, i.e. as they had no systematic method for marking the antecedent (e.g. to choose only referring expressions like an NP, PP or S and including all their attributes) they had trouble identifying the borders of the textual material to be marked. Thus, to hand out the conventions to the subject in advance might help to amend the reliability of the annotated data. Second, the referent was often only mentioned implicitly, i.e. it had to be derived from the text. Then the subjects were uncertain which and how much of the textual material was necessary in order to specify the referent. This difficulty can be explained as follows: Text understanding requires the activation of lexical and conceptual knowledge and many inferences are drawn in order to build up a coherent conceptual representation of the text. The conceptual representation of the text is much richer than the text itself (cf. e.g. Schwarz 2001, Graesser et al. 2001). As text understanding also relies on implicit knowledge which is difficult to determine explicitly, people are at a loss when they are asked to precisely determine the antecedent. They have the impression that the textual material contains less or more information than needed.

The aim of the paper is to show that annotating occurrences of *danach* and representing their meaning is a quite difficult task since there is a lot of implicit and explicit ambiguity[5] involved. I will now present different types of ambiguity found in the data. Then, I make some suggestions how annotating the examples can be facilitated (e.g. by the annotation conventions) and why we should and how we could account for the different types of ambiguity e.g. by using underspecified representations.

2 Different Types of Ambiguity

I distinguish two main types of ambiguity: structural and referential ambiguity. When the subjects specified the same referent for *danach* but marked different antecedents as in example (3), I call this structural ambiguity. For the interpretation of the example it does not make a difference whether one assumes only the NP or the whole S as antecedent.

[3] This definition of the antecedent differs from a text-linguistic one where the antecedent is the first (not the last) expression referring to the referent of the anaphor.

[4] The instructions given in the pilot study were: 1) For the antecedent: Please mark the text span that you think *danach* relates to. and 2) For the referent: Please name the entity (in the world), i.e. an event, a time-entity or an object to which *danach* refers to. 3) Please indicate the degree of difficulty for each example as easy, medium or tricky. 4) Please comment on the example, if you want to.

[5] The distinction between implicit and explicit ambiguity is taken from Poesio et al. (2005). Implicit ambiguity means that subjects do not notice the ambiguity but as a group specify different antecedents and referents. Explicit ambiguity means that the subjects notice and specify the ambiguity.

(3) **[[***Die erste Explosion***]$_1$** *ereignete sich gegen 9 Uhr in einem Pulver- und Munitionsdepot.***]$_2$** Stundenlang kam es **danach** im Abstand von 10 bis 15 Minuten zu weiteren Detonationen.

[*At 9 o'clock* [*the first explosion*]$_1$ *took place in a depot for gunpowder and ammunition.*]$_2$ For hours **after that** further detonations took place at intervals of 10 to 15 minutes.

In contrast, in referentially ambiguous cases there are several plausible referents and antecedents for *danach*. Thus, we get distinctive readings when different textual spans referring to different referents are marked as antecedents (s. example (4)).

(4) **[***In* **[***dem Schauprozess 1981***]$_1$** *wird er zu* **[***20 Jahren Haft***]$_2$** *verurteilt.***]$_3$** **Danach** lebte er als Rentner in seiner Heimatstadt.

[*In* [*the show trial 1981*]$_3$ *he was sentenced to* [*20 years in prison*]$_1$.]$_2$ **After that** he lived as pensioner in his hometown.

In (4) choosing antecedent$_1$ or antecedent$_3$ makes no difference with respect to the interpretation of the temporal constellation of the two situations related by *danach*. The end of the trial and the end of the proclamation of sentence are temporally identical. However, choosing antecedent$_2$ which is preferred due to world knowledge yields another temporal constellation between different situtations.

2.1 Annotating Structural Ambiguity

Example (3) demonstrates only one possible variant of structural ambiguity. Structural ambiguity can also be observed in anaphoric chains. In example (5) the subjects marked different antecedents while specifying the same referent, namely the bombing.

(5) **[[***Ein Mann sei herausgesprungen und habe eine Handgranate gegen das Tor und die Wachposten geschleudert, hieß es. Daraufhin sei der zweite Täter durch das Tor gerast und habe sich vor dem Botschaftseingang samt der Bombe in die Luft gesprengt.***]$_1$ [[***Die Tat***]$_2$** *geschah gegen 9.30 Uhr Ortszeit,***]$_3$** *als in der Botschaft normaler Alltagsbetrieb herrschte.***]$_4$** Teilweise standen **danach** nur noch Gerippe des Gebäudes.

[[*It is said that a man jumped out and threw a grenade against the gate and the sentinels. Thereupon, the second committer ran amuck through the gate and blew himself and the bomb up in front of entrance of the embassy.*]$_1$ [[*The action*]$_2$ *happened about 9.30am at local time,*]$_3$ *when in the embassy everything was at normal course of life.*]$_4$ **Thereafter**, in part only the framework of the building was there.

Due to the complex anaphor[6] *die Tat* referring to the propositionally structured referent (the BOMBING) mentioned in the first two sentences either antecedent$_1$ or the complex anaphor$_2$ is marked as antecedent. In the third sentence we again have the structural choice between the NP-based-antecedent$_2$ and the S-based-antecedent$_3$. One subject marked all sentences as antecedent$_4$. I will turn to this issue in section 2.2.2.

I assume that structural ambiguity as in such cases can be handled consistently by the annotation conventions proposed in section 1.1.[7] So according to the conventions antecedent$_1$ in example (3) and antecedent$_2$ in example (5) should be marked. In section 2.2.2, I will discuss why this only partly resolves the ambiguity in example (5).

2.2 Annotating Referential Ambiguity

In this section I will present three types of referential ambiguity: 1. competition between temporal and situation referents, 2. competition between simple and structured situation referents and 3. competition between several different situation referents.

[6] S. Consten & Knees (forthc.) for a detailed analysis of complex anaphora.

[7] Another study is planned in order to find out whether the inter-annotator agreement of marking the antecedent raises when the subjects get the annotation conventions beforehand.

2.2.1 Competition between temporal and situation referents

In some examples there is an ambiguity between a temporal referent (usually a time span) and a situation referent which is temporally located in that time span.

(6) Und der im Grunde hypersensible Pollock scheint von solcher Männlichkeit zutiefst beeindruckt gewesen zu sein, damals **[***im Jahre 1936***]$_1$**, als **[***er in Siqueiros New Yorker Laboratorium zur Erprobung moderner Techniken in der Kunst mitarbeitete***]$_2$**. ... So erstaunlich wie unstrittig ist nämlich die Tatsache, **[***dass Pollock dort jene Experimente bereits kennengelernt haben muss***]$_3$**, die er selbst erst **zehn Jahre danach** anwandte: die Arbeit mit Industrielacken und das Verträufeln und Verschütteln von Farbe direkt auf dem Malgrund oder mit der Spritzpistole, in einem Akt des „kontrollierten Zufalls".

And Pollock, virtually hyper-sensitive, seemed to be deeply impressed by such a masculinity, [[*at that time in 1936*]$_1$*, when he worked in Siqueiro's New Yorker laboratory in order to try out modern techniques in art.*]$_2$ *... Thus, the fact* [*that there he must have got to know these experiments*]$_3$ which he used himself **ten years after that**, is amazing and indisputable at the same time: his works with industrial gloss paint and his spilling and bloting with paint directly on the subfont or with the spray guns, in an act of „controlled chance".

In example (6) Pollock's cooperation with Siqueiros (sit_A) referred to by antecedent$_2$ and his getting in touch with these experiments (sit_B) referred to by antecedent$_3$ are interpreted as taking place at some time interval within the year 1936 referred to by antecedent$_1$. Thus, all referents selected by the subjects are at least temporally related to each other. The temporal referent provides a time frame for the situation referents (i.e. $t(sit_A) \wedge t(sit_B) \subseteq t(1936)$)[8]. Thus, *danach* anchors either at t(1936), $t(sit_A)$ or $t(sit_B)$ where the latter ones can be merged to $t_x \subseteq t(1936)$ (i.e. some time interval in 1936). So in order to account for the ambiguity, I propose the following underspecified representation: [**end(**$t(1936)) \vee (t_x \subseteq t(1936))$] < 10-years **begin**(t(using modern art techniques)) which means that the matrix-situation took place either ten years after the end of 1936 or ten years after some point in time within 1936.[9]

2.2.2 Competition between single and structured situation referents

This type of ambiguity is discussed as 'circumscription' by Webber (1987), or 'reference to structured entities' by Fraurud (1992) and Poesio et al. (2005). In example (5) (repeated here as (7)) all subjects named the bombing as referent, but one subject conceptualised it as a structured entity described by antecedent$_4$.

(7) **[[***Ein Mann sei herausgesprungen und habe eine Handgranate gegen das Tor und die Wachposten geschleudert, hieß es. Daraufhin sei der zweite Täter durch das Tor gerast und habe sich vor dem Botschaftseingang samt der Bombe in die Luft gesprengt.***]$_1$** **[[***Die Tat***]$_2$** *geschah gegen 9.30 Uhr Ortszeit,*]$_3$ *als in der Botschaft normaler Alltagsbetrieb herrschte.***]$_4$** Teilweise standen **danach** nur noch Gerippe des Gebäudes.

[[*It is said that a man jumped out and threw a grenade against the gate and the sentinels. Thereupon, the second committer ran amuck through the gate and blew himself and the bomb up in front of entrance of the embassy.*]$_1$ [[*The action*]$_2$ *happened about 9.30am at local time,*]$_3$ *when in the embassy everything was at normal course of life.*]$_4$ **Thereafter**, in part only the framework of the building was there.

I consider this example to be similar to examples discussed in Poesio et al. (2005). They assume that people do not perceive this type of ambiguity as infelicitous, if the following conditions hold ("Justified Sloppiness Hypothesis"):

1) "Both explicitly mentioned potential antecedents *x* and *y* are elements of an underlying mereological structure with summum $\sigma = x \oplus y$ which has been explicitly constructed

[8] t stands for the ‚life' or ‚run'-time of the situations and the time-entity.

[9] There are incidences where *danach* definitely locates the matrix-situation within and not after the time frame given by the temporal referent, s. the following example: **[***During the Weimar Republic the interventionism was modern.***]$_1$** *What happened* ***thereafter*** *in the Weimar Republic, has a great deal to do with the fact that this concept did not work as a problemsolver.*

(and made salient) in the dialogue". This is only partly true for this example, as antecedent$_1$ describes several sub-events of the summum bombing and not only a single element.

2) "[I]t is possible to construct a p-underspecified interpretation" which means that the anaphor can refer to each of the elements or to the summum. Again, this condition holds only partly (for the same reason as before). The anaphor can refer to the summum (*die Tat*$_2$), to antecedent$_1$ (describing several sub-events) or it refers to all of them (antecedent$_4$).
3) "All possible interpretations […] are equivalent for the purpose of the plan." This condition also needs a slight modification in order to hold: All possible interpretations in this example are equivalent with respect to the conceptualisation of the structured referent, as all subjects named the bombing as referent. Thus, examples like (7) can be explained by an expanded version of the "Justified Sloppiness Hypothesis".

(8) **[***In den 80er Jahren trieben Gerüchte, dass die FDA Virazole als Mittel gegen Aids zulassen werden, die Aktienkurse von ICN und ihren beiden Tochtergesellschaften scharf in die Höhe.***]$_1$ [***Panic nahm die Gelegenheit wahr, um eigene Aktien im Wert von 13 Millionen Dollar abzugeben.***]$_2$ [***Dies brachte ihm zahlreiche Aktionärsklagen ein, die zum Teil heute noch anhängig sind.***]$_3$]$_4$ **Kurz danach** setzte die FDA das Unternehmen unter Druck, das im Zusammenhang mit der ursprünglich begrenzten Genehmigung verbreitete Werbematerial für Virazole zurückzurufen.

[*In the eighties rumours that FDA would admit Virazole as drug against Aids boosted the stock price of ICN and its two affiliates.*]$_1$ [*Panic availed himself the opportunity of selling his shares amounting to 13 million dollars.*]$_2$ [*This brought him in numerous accusations of shareholders, of which some are still pending.*]$_3$]$_4$ **Shortly thereafter** the FDA enforced his company to call back the advertisement for Virazole which had originally been approbated for limited use.

In example (8) a structured entity can be constructed, namely the chain of cause and effect resulting in the FDA'S ENFORCEMENT in the anaphor-matrix-clause. But only some of the subjects named this referent, while others preferred as referent the raising of the share price referred to by antecedent$_1$ or Panic's selling of his shares referred to antecedent$_2$. So this example differs from the previous one in two aspects: first, there is no complex anaphor explicitly introducing the structured referent and second, the single situations in antecedent$_1$, antecedent$_2$ and antecedent$_3$ are causally connected to each other. This justifies their grouping but as the results of the pilot study show the situations do not have to be clustered. As far as I am aware there is no accomplished solution for handling this kind of ambiguity (cf. Fraurud 1992, Poesio et al. 2005). Following the PDTB annotation manual (Prasad et al. 2006) cases like this are handled in accordance with the "Minimality Principle". It states that only textual material that is "*minimally required* and *sufficient* for the interpretation of the relation" is marked as argument of a discourse connective whereas "other span of text that is perceived to be relevent (but not necessary) in some way to the interpretation of arguments is annotated as *supplement information*" (s. Prasad et al. (2006:13); highlighted text as in original).[10] Thus, according to the PDTB manual only one – probably antecedent$_2$ – would be marked as antecedent. The other ones would get the status of a supplement. But this is only a convention and not a proper solution for the ambiguity of this example.

Webber (1987, 1991) and Fraurud (1992) suggest that discourse structure guides the interpretation of anaphoric expressions emphasising the role of the right frontier constraint. I think we have to accept that some anaphoric references are just ambiguous and that Poesio et al.'s (2005) p-underspecification provides a good means for representing this type of ambiguity.

[10] The antecedent of the complex anaphor *dies* ("this") in the third sentence would also be marked as supplement.

2.2.3 Competition between several situation referents

The third type of ambiguity concerns cases involving several plausible situation referents. In example (9) half of the subjects named Jacksons demanding as referent for *danach* whereas the other half specified the Nuremberg trials as referent where the trials are only inferred.[11]

(9) Es sollte nach Robert Jacksons Willen kein Einzelfall bleiben; **[***er forderte schon damals die Einrichtung eines Internationale Strafgerichtshof***]$_1$**. ... Gewiß, die Straßburger Konvention für Menschenrechte, die Völkermord-Konvention der Vereinten Nationen entstanden unter dem Eindruck von **[***Nürnberg***]$_2$**, Angriffskrieg gilt nicht länger als Recht souveräner Staaten. ... Ein Internationaler Strafgerichtshof aber nimmt erst heute, **50 Jahre danach**, allmählich Gestalt an: in Den Haag, wo es um die Kriegsverbrechen im ehemaligen Jugoslawien geht.

According to Robert Jackson's volition, this should not remain an individual case; [*he has already demanded at that time that an International Criminal Division should be established.*]$_1$ ... Surely, the Convention of Human Rights from Strasbourg, the Convention of Genocide of the UN emerged from the impression of [*Nuremberg*]$_2$; war of aggression is not longer a right of sovereign states. ... But an International Criminal Division starts to gradually take shape today, **50 years after that**: in Den Haag where the war crimes of the former Yugoslavia are dealt with.

Both referents are plausible. The former Jacksons demanding as the International criminal division mentioned in antecedent$_1$ is described again in the anaphor-matrix-clause, so both the antecedent-clause and the anaphor-clause are thematically connected by *danach*. Nonetheless, the latter referent derived from antecedent$_2$ is also plausible not only as it is textually closer to the anaphor, it also gets causally related to the establishing of the International criminal division (and maybe even to Jacksons Demanding). Moreover, the Nuremberg trials are triggered by textual anchors like *Völkermord* ("genocide"). Nonetheless, Jacksons demanding does not have to be inferred. Thus, choosing antecedent$_1$ reflects staying closer to the explicit textual material while choosing antecedent$_2$ suggests a more conceptually guided decision. Still, both referents are somehow conceptually related. Thus, this example may also be only another case of ambiguity by structured enities. But we cannot formulate a preference for one referent, so I conclude that we have to allow for some instances of 'proper' ambiguity.

3 Summary and Conclusions

In the paper I first presented an account of the meaning and anaphoric character of the German pronominal adverb *danach*. By means of a pilot study, I have shown that the resolution of *danach* involves several types of ambiguity, namely structural and referential ambiguity which have to be dealt with in the annotation of *danach*-occurrences. In the case of referential ambiguity we can distinguish the following sub-types: 1. competition between temporal and situation referents, 2. competition between simple and structured situation referents and 3. competition between several different situation referents. In conlusion, I argue that referential ambiguity has to be accounted for by underspecified representations, while structural ambiguity (e.g. anaphoric chains, competing NP- and S-antecedents) can be dealt with by the annotation conventions proposed. We should keep in mind that annotation conventions are a good means to standardise annotations but they are sometimes only partial solution for problems you have to deal with annotating and interpreting anaphoric references.

Finally, we have to consider that the subjects often chose one candidate from a set of plausible referents not noticing an ambiguity (implicit ambiguity). Thus, as *danach* seems to allow for some ambiguity, recipients are not expected to be able to precisely determine its antecedent and/or its referent. So we better account for some types of ambiguity, e.g. by

[11] The use of *Nuremberg* to refer to the TRIALS is an example of the kind of metonymy where a location name stands for something that happened there (s. Markert / Nissim's (2003: 9) on metonyny and reference resolution).

underspecified representations, since humans seem to tolerate them as Poesio et al.'s (2005) "Justified Sloppiness Hypothesis" suggests.

Acknowledgements I would like to thank Manfred Consten (Jena), Mascha Averintseva-Klisch (Tübingen), Yannick Versley (Tübingen) and the anonymous reviewers for helpful comments on this paper. This paper has been written within the context of the research project "KomplexTex", granted by the Deutsche Forschungsgemeinschaft (SCHW 509/6-2).

References

Consten, M. & M. Knees (forthc.). Complex Anaphors in Discourse. In A. Benz and P. Kühnlein (eds.). *Constraints in Discourse*.

Fraurud, K. (1992). Situation Reference. What does ‚it' refer to?. In K. Fraurud. *Processing Noun Phrases in Natural Discourse*. PhD thesis. Depart. of Linguistics, Stockholm University.

Graesser, A., P. Wiemer-Hastings & K. Wiemer-Hastings (2001). Construction Inferences and Relations during Text Comprehension. In: T. Sanders, J. Schilperoord, & W. Spooren (eds.). *Text representation: Linguistic and psycholinguistic aspects*, Amsterdam, 249-271.

Habel, Ch. & M. Knees (2004). On Generating Verbal Descriptions of Temporal Successsion. In E. Buchberger (ed.). *Proceedings of KONVENS 2004*, Wien, 53-60.

Herweg, M. (1992). Aspectual requirements of temporal connectives: Evidence for a two-level approach to semantics. In J. Pustejovsky & S. Bergler (eds.). *Lexical Semantics and Knowledge Representation*, 185–200.

Markert, K. / Nissim, M. (2003). Corpus-Based Metonymy Analysis. In *Metaphor and Symbol* 18:3.

Miltsakaki, E.; R. Prasad; A. Joshi & B. Webber (2004). Annotating discourse connectives and their arguments. *Proceedings of the HLT/NAACL Workshop on Frontiers in Corpus Annotation*. Boston, MA. 2004.

Pasch, R. / Braußе, U. / Breindl, E. / Waßner, U. H. (2003). *Handbuch der deutschen Konnektoren. Linguistische Grundlagen der Beschreibung und syntaktische Merkmale der deutschen Satzverknüpfer* (Konjunktionen, Satzadverbien und Partikeln). Berlin, New York: de Gruyter.

Poesio, M., P. Sturt, R. Artstein & R. Filik (2005). *Underspecification and Anaphora: Theoretical Issues and Preliminary Evidence*. Technical report CSM-438, University of Essex Department of Computer Science, October 2005.

Prasad, R. / Miltsakaki, E. / Dinesh, N./ Lee, A. / Joshi, A. / Webber, B. (2006). *The Penn Discourse TreeBank 1.0. Annotation Manual*. IRCS Technical Report IRCS-06-01. Institute for Research in Cognitive Science, University of Pennsylvania. March 2006.

Rüttenauer, M. (1978). *Vorkommen und Verwendung der adverbialen Proformen im Deutschen*. Hamburg: Buske-Verlag.

Schilder, Frank (2004). Temporale Konnektoren im Diskurs. In H. Blühdorn, E. Breindl & U. H. Waßner, (eds.). *Brücken schlagen. Grundlagen der Konnektorensemantik*, 161-184.

Schwarz, M. (2001). Establishing Coherence in Text. Conceptual Continuity and Text-world Models. *Logos and Language. Vol. II, No. 1*, 15-24.

Webber, B. (1987). *Two Steps Closer to Event Reference*. MS-CIS-86-74. INC LAB 42. Dept. of Computer and Information Science, University of Pennsylvania.

Webber, B. (1991). Structure and ostension in the interpretation of discourse deixis. *Language and Cognitive Processes 6*. 107-135.

Webber, B., A. Joshi, M. Stone, and A. Knott (2003). Anaphora and Discourse Structure. *Computational Linguistics, 29(4)*, 545-587.

Ambiguity in Anaphora Workshop Proceedings, pages 65–72
ESSLLI 2006, Málaga, Spain, 7–11 August 2006
Edited by Ron Artstein and Massimo Poesio
http://cswww.essex.ac.uk/anaphora/

TOWARDS A MODULAR APPROACH TO ANAPHOR RESOLUTION

Arnout W. Koornneef, Frank Wijnen & Eric Reuland
Utrecht Institute of Linguistics OTS
Utrecht University
e-mail: arnout.koornneef [at] let.uu.nl

Abstract

Recent accounts of anaphor resolution propose a two-route architecture for the interpretation of pronominals (e.g., Grodzinsky & Reinhart, 1993; Reuland, 2001). A pronoun can either be resolved by a grammatical operation in *logical syntax* (i.e., variable binding) or through value assignment in *discourse* (i.e., co-reference). Reuland (2001) proposes that an interpretation through variable binding requires less processing resources and is therefore preferred over a co-reference interpretation. Rule I compares variable binding with co-reference interpretations to decide whether a co-reference dependency is allowed (Grodzinsky & Reinhart, 1993; Reinhart 2000). This rule prevents discourse processes from by-passing logical syntax where the latter rules out an interpretation as ungrammatical. The question is whether Rule I always compares variable binding with co-reference interpretations, or is only executed if a dependency is initially ruled out by logical syntax. In an eye-tracking experiment we manipulated the interpretation of ambiguous and unambiguous pronouns in an attempt to explore how variable binding, co-reference and Rule I influence the way readers resolve pronouns. The results show that if a pronoun was ambiguous between a variable binding and co-reference antecedent, the variable binding antecedent was initially preferred even if discourse information clearly favoured the co-reference antecedent. Therefore, we argue that logical syntactic processes function independently from discourse processes, indicating that it may be warranted to consider logical syntax and discourse as distinct modules of the language system. Furthermore, the language system does not seem to apply Rule I in cases where both variable binding and co-reference lead to the same grammatical interpretation suggesting that Rule I is only relevant in cases that are potentially ungrammatical.

Background

Experimental studies have identified a range of factors that influence the resolution process for pronouns. For instance, subjecthood or first mention (e.g., McDonald & McWhinney, 1995), gender information, (e.g., Arnold, Eisenband, Brown-Schmidt & Trueswell, 2000), recency or distance (e.g., Clark & Sengul, 1979), and implicit causality information of interpersonal verbs (e.g., Caramazza, Grober, Garvey & Yates, 1977; Koornneef & Van Berkum, 2006) all seem to affect the search for a referent. Within psycholinguistics the current debate mainly focuses on the question *when* during comprehension these factors influence the pronoun resolution process. For instance, some argue that gender information becomes available and is put to use before other factors can have an affect (e.g., Ehrlich, 1980), while others claim that gender is only used during special, strategic or later processing (e.g., Greene, McKoon & Ratcliff, 1992). Still others propose fully dynamic

accounts where multiple sources of information are used in parallel to guide the resolution process (e.g., Arnold et al.). However, these different views have in common that they all implicitly assume that the pronoun is connected to a referent by accessing the discourse representation of the preceding written or spoken text. More specifically, they presume a one-route architecture to resolve a pronoun.

On the other hand, linguistic accounts of pronoun resolution (e.g., Reinhart, 1983) argue that the language system has two ways by which a pronoun is connected to an antecedent; (i) the pronoun behaves as a variable and is *bound* by its antecedent, or (ii) the pronoun receives a value from the discourse storage through co-reference. Consequently, sentence (1) has actually two possible representations, one in which '*he*' is bound by '*the clown'* (ex. 1a) and one in which the value of '*he'* can be freely chosen from the discourse. If in the latter '*the clown'* is picked as the proper antecedent for the pronoun the two derivations have the same interpretation (ex. 1b).

(1) The clown$_i$ thinks that he$_i$ is funny
  (a) The clown λx (x thinks x is funny)
  (b) The clown λx (x thinks a is funny) and $a = x$

Variable binding is only possible if the antecedent c-commands the pronoun. Informally, a c-commanding antecedent can be characterized as 'being higher in the syntactic tree' of a sentence than the pronominal.[1] As a result, variable binding is licensed in example (1) but not in example (2). In the former '*the clown'* is higher in the syntactic tree than '*he'*, but in the latter there is no c-command relation because the antecedent and pronoun appear in different sentences. Hence, in structures like (2) only a co-referential dependency can be constructed. On the other hand, if the antecedent contains a quantifier such as '*every'*, co-reference is ruled out (see ex. 3) and variable binding is in fact the only option for establishing an anaphoric dependency (e.g., ex. 4).

(2) The clown$_i$ knows it for sure. He$_i$ is funny.
(3) *Every clown$_i$ knows it for sure. He$_i$ is funny.
(4) Every clown$_i$ thinks that he$_i$ is funny

The two-route architecture of the system comes with one important implication illustrated in (5). Logical syntactic constraints (i.e., Principle B) rule out a variable binding relation between the pronoun and the antecedent, as the pronoun is structurally too close to the antecedent (i.e., not 'free' in its binding domain). However, now we still have to explain why the availability of a co-referential interpretation does not systematically bypass the effect of Principle B.

(5) *The clown$_i$ hates him$_i$.

[1] A widely accepted formal definition of c-command is as follows: phrase α c-commands phrase β if and only if phrase α does not contain phrase β and the first branching node dominating phrase α also dominates phrase β (see Reinhart, 1983, for discussion).

A solution is provided by Rule I.[2] The original idea was that Rule I always compares the variable binding interpretation to the co-reference interpretation and opts for the former unless the two potential dependencies yield different interpretations (Grodzinsky & Reinhart, 1993). By applying this rule to (5), both variable binding and co-reference are ruled out; variable binding is blocked through Principle B and co-reference is therefore blocked as well because the two interpretations would be identical (i.e., in both cases '*he = the clown'*). However, on this construal Rule I does not only bear on ill-formed dependencies like (5), but on all dependencies that have an antecedent in a potential binding position, and a co-reference solution that points to that same antecedent. Thus, in grammatical example (1), where both variable binding and co-reference are possible in principle, the costly application of Rule I is predicted as well. Recently, however, Reinhart (2000) has proposed that Rule I only prevents co-reference from bypassing variable binding when the grammar rules out a variable binding interpretation and, hence, is only executed in potential ungrammatical cases like (4) and not in cases like (1).

According to Reuland (2001), variable binding and co-reference result from distinct processes in two functionally independent modules. Furthermore, it is proposed that a general economy principle governs the division of labour between the modules, based on the number of cross modular steps. That is, variable binding is thought to be more economic than co-reference because it requires fewer steps. By incorporating the economy principle, the theory assumes a processing hierarchy in which variable binding has precedence over co-reference. Hence, if a pronoun is ambiguous between a variable binding and co-reference antecedent the pronoun is predicted to be (initially) interpreted as referring to the former and not the latter. Given the logic of the theory this preference may stem from two sources; (i) differences in processing cost intrinsic to the processes themselves, or (ii) a time course effect – the variable binding option becomes available before the co-reference option and, hence, is initially taken. A similar proposal was made by Frazier and Clifton (2000) who hypothesize that 'a bound-variable interpretation is preferred because the perceiver need only consult the Logical Form representation (not the discourse representation) in order to identify the bound-variable analysis of the sentence'. In a series of reading time and questionnaire experiments they examined this hypothesis and reported mixed results. It seemed that consistent results could only be found in VP-ellipsis contexts and did not generalize to other (quantificational) contexts, which led them to reject their 'LF-only' hypothesis.

In this study, we addressed the hypothesis that the parser prefers a variable binding over a co-reference dependency using structures very different from VP-ellipses. In addition, we studied whether the preference as implied by Reuland's model might be attributed to an intrinsic difference in the associated processing costs or is rather a time

[2] A simplified version of original Rule I is: a pronoun cannot co-refer with a particular antecedent if the co-reference interpretation is indistinguishable from what would be obtained if the antecedent binds the pronoun.

A simplified version of revised Rule I is: a pronoun cannot co-refer with a particular antecedent if (i) the antecedent is in a potential binding position, yet cannot bind the pronoun (due to Principle B), and (ii) the co-reference interpretation is indistinguishable from what would be obtained if the antecedent binds the pronoun.

It is important to note that the original and revised Rule I are formulated in a way that correctly allows co-reference to by-pass Principle B in some specific situations. For instance, the dependency 'him = Bill' is allowed in the following example: "It is clear what Bill and Mary have in common, she adores him, and $Bill_i$ adores him_i too".

course effect. Finally, we examined whether the revised version of Rule I should be preferred over the original version, or alternatively, that we should maintain the latter.

An eye-tracking experiment

In an eye-tracking experiment 36 Dutch participants read a series of short stories containing 36 experimental items (latin square design). In the experiment we manipulated the resolution of ambiguous and unambiguous pronouns to examine the issues discussed above.

The hypothesis that logical syntactic operations are preferred over discourse operations predicts that if a pronoun is ambiguous between a variable binding and co-reference antecedent the language system initially constructs a variable binding dependency, even if the context strongly supports a co-reference dependency. To test this prediction we presented stories like (6) and (7). In both stories the critical sentence contains the ambiguous pronoun '*he'* with two potential antecedents (i.e., '*every worker'* and '*Paul'*). In story (6) the context preceding the pronoun supports a reading in which *he* refers to the variable binding antecedent (i.e., '*he = every worker'*). In story (7), on the other hand, the preceding context strongly supports a co-reference reading (i.e., '*he = Paul'*)[3]. Reuland's model predicts that in story (7) initially the 'wrong' (variable binding) antecedent is chosen, because this dependency is initially preferred over the co-reference dependency. As a result, in (7) readers have to re-analyze their initial interpretation to construct the more suitable co-reference dependency.

(6) <u>*Biased towards variable binding antecedent (VB-bias)*</u>
A working day in the factory is always very tough. Especially today a lot of workers, among them the very old man Paul, could barely cope. ***Every worker*$_{\mathbf{i}}$** *who just like Paul was running out of energy, thought it was very nice that* ***he*$_{\mathbf{i}}$** *could go home early this afternoon.* After a hot shower things would probably look better.

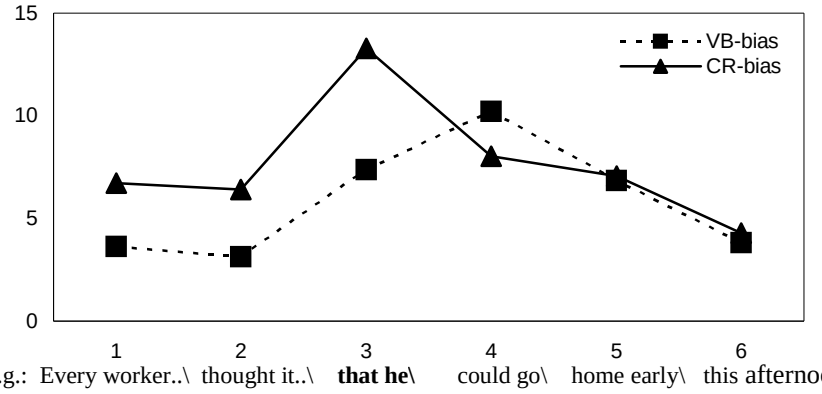
(7) <u>*Biased towards coreference antecedent (CR-bias)*</u>
A working day in the factory is always very tough. Especially today the very old man Paul could barely cope. *Every worker who knew that* ***Paul*$_{\mathbf{i}}$** *was running out of energy, thought it was very nice that* ***he*$_{\mathbf{i}}$** *could go home early this afternoon*. After a hot shower things would probably look better.

Whereas readers did not seem to have obvious processing difficulties during the initial interpretation of the CR-biased pronoun as indicated by the absence of significant first-pass differences in any of the regions, they clearly slowed down during later stages of processing, as significant longer reading times emerged in the CR-biased condition in the second-pass measure (see Figure 1). More specifically, readers re-fixated the critical region (i.e., containing the pronoun) and the preceding region longer in the CR-biased condition than in the VB-biased condition. This finding, particularly the difference in the pre-critical region, suggests that readers were engaged in re-analysis when they read a CR-biased

[3] We tested our materials in two internet-based questionnaires. In one of them we simply asked participants to indicate how they interpreted the pronoun by giving them a 2-choice option between the variable binding antecedent and the co-reference antecedent. In the other we presented the stories up to the critical pronoun and instructed the participants to finish the story with the first ending that came to mind. Both experiments revealed an equally strong bias towards the intended interpretation.

context. Hence, this result is consistent with the idea that a variable binding reading is initially preferred over a co-reference reading even if the context strongly supports the latter.

Figure 1. *Mean Second-Pass Durations in msec per character for the Variable Binding Biased (VB-bias) and Co-reference Biased condition (CR-bias).*



To test whether this preference for a variable binding interpretation occurs because logical syntactic computations are intrinsically less costly than discourse computations, we presented stories containing unambiguous pronouns like (8) and (9). In both stories the pronoun '*he'* has only one possible referent (i.e., '*every worker'* or '*Paul'*). By hypothesis, the process that assigns the referent to the pronoun differs between conditions. In (8) the pronoun can only be resolved at the logical syntactic level because the antecedent is quantified, and therefore, by definition not referential. On the other hand, in (9) no c-command relation exists between '*Paul'* and '*he'* and the pronoun can only receive its value through co-reference. If co-reference is intrinsically more costly than variable binding we should expect longer reading times in (9) around the critical pronoun.

Furthermore, we included stories like (10) and (11) to check whether the original or revised version of Rule I is preferred. According to the original version of Rule I, a costly comparison between variable binding and co-reference should be made in (10) and (11) because both processes are available in principle. Consequently, longer reading times are predicted in story (10) and (11) than in (8) and (9). In contrast, revised Rule I does not predict a difference in reading times as this version suggests that the rule is only executed if an antecedent in a potential binding position is ruled out by the grammar, which is not the case in any of the stories.

(8) <u>*Variable binding only (VB-only)*</u>
A working day in the factory is always very tough. Especially today a lot of workers could barely cope. ***Every worker*$_{\mathbf{i}}$** *who was running out of energy, thought it was very nice that* ***he*$_{\mathbf{i}}$** *could go home early this afternoon.* After a hot shower things would probably look better.

(9) <u>*Co-reference only (CR-only)*</u>
A working day in the factory is always very tough. Especially today the very old man Paul could barely cope. ***Paul*$_{\mathbf{i}}$** *was running out of energy. It was very nice that* ***he*$_{\mathbf{i}}$** *could go home early this afternoon.* After a hot shower things would probably look better.

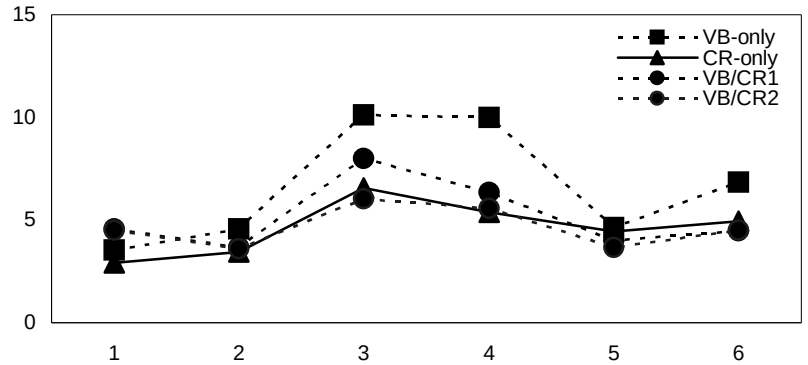
(10) *Variable binding/co-reference 1 (VB/CR1)*
A working day in the factory is always very tough. Especially today a lot of workers could barely cope. ***All workers$_i$*** *who were running out of energy, thought it was very nice that* ***they$_i$*** *could go home early this afternoon.* After a hot shower things would probably look better.

(11) *Variable binding/co-reference 2 (VB/CR2)*
A working day in the factory is always very tough. Especially today the very old man Paul could barely cope. ***Paul$_i$****, who was running out of energy, thought it was very nice that* ***he$_i$*** *could go home early this afternoon.* After a hot shower things would probably look better.

We observed no significant differences in first-pass, regression path and total reading durations in any of the regions. The analysis of the second-pass durations, on the other hand, revealed an unexpected effect. In the Variable Binding-only condition longer second pass reading times emerged in the region that directly followed the critical pronoun (region 4, see Figure 2). The comparisons between the other conditions revealed no significant differences. As such, this finding provides no support for the hypothesis that discourse operations require more processing resources than logical syntactic operations.

To examine the predictions of the different versions of Rule I, we compared the mean reading times collapsed over the conditions in which only one route was available (the *only* conditions) to the mean reading times collapsed over the conditions in which both routes were available (the *VB/CR* conditions). The analyses revealed no significant difference in any of the regions. These results are consistent with the predictions of revised Rule I, and not with original Rule I.

Figure 2. *Mean Second-Pass Durations in msec per character for the Variable Binding Only (VB-only), Co-reference Only (CR-only) and Variable Binding/Co-reference Conditions (VB/CR1 and VB/CR2).*



Conclusions

In an eye-tracking experiment we manipulated the interpretation of ambiguous and unambiguous pronouns in an attempt to explore how variable binding, co-reference and Rule I influence the way readers resolve pronouns. The results for unambiguous pronouns

revealed that no processing differences existed for pronouns that could be connected to their antecedents by only one route or, alternatively, by both routes. This result seems inconsistent with the original idea that Rule I always compares the variable binding interpretation to the co-reference interpretation and opts for the former unless the two potential dependencies yield different interpretations (Grodzinsky & Reinhart, 1993). The absence of a costly (and intuitively redundant) comparison between the two derivations in sentences with grammatical variable binding dependencies is, on the other hand, completely consistent with the predictions of revised Rule I, in which it is assumed that the rule is only relevant under specific circumstances (Reinhart, 2000). That is, the parser does not seem to apply Rule I blindly, but instead only uses it for structures similar to *the clown$_i$ hates him$_i$*. These structures never allow a variable binding interpretation as the pronoun is structurally too close to the antecedent (i.e., a violation of Principle B), yet presented within the right context a co-referential dependency can be established, provided that revised Rule I allows it (see Footnote 2 for an example).

Consistent with Reuland's (2001) processing hierarchy for anaphor resolution we found that ambiguous pronouns are preferably connected to a variable binding antecedent even if discourse information clearly favoured a competing co-reference antecedent. This suggests that the preference for variable binding dependencies is not restricted to VP-ellipsis, but is a general tendency of the human parser during anaphor resolution. Furthermore, as discourse constraints seem to be unable to immediately influence the preference of the parser, we propose that the variable binding interpretation is established independently from the discourse storage, which is consistent with a two-route architecture for pronominal resolution (Reinhart, 1983; Grodzinsky & Reinhart, 1993; Reuland, 2001). Moreover, our results indicate that it may be warranted to consider logical syntax and discourse as distinct modules of the language system (Reuland, 2001).

As noted, if the construction of a dependency in the logical syntactic module has precedence over the construction of a dependency in the discourse module, this can either indicate that the former is intrinsically more economic than the latter, or it can be a time course effect, in the sense that the logical syntax module is accessed earlier. For the latter issue it may then be relevant that, in the case of unambiguous pronouns we found that a pronoun that could only be resolved through variable binding required *more* processing resources than a pronoun that could only be resolved trough the co-reference route. This result appears to go against an approach based on intrinsic economy. It could be caused by the specific antecedents we used in our stories.[4] Burkhardt (2004), for instance, suggests that quantified antecedents like '*every worker*' are more complex than non-quantified antecedents like '*the worker*' or '*Paul*' and, as a result, increase the processing load during pronoun resolution. In her account the most economic dependency between a pronoun and antecedent is established with so-called 'light' quantifiers like '*everyone*'.[5] If it is indeed more costly for the parser to connect a pronoun to an antecedent like '*every worker*' than to an antecedent like '*Paul*', the preference for the quantified antecedents in our ambiguous materials cannot be exclusively attributed to a difference in processing costs for variable

[4] As one anonymous reviewer pointed out, the critical pronouns in the CR-only and VB-only conditions appeared in different structural contexts, which may have caused the unexpected effect. However, this would left unexplained why the reading times for the CR-only condition did not differ from the VB/CR conditions as in the latter conditions the structural context was exactly the same as in the VB-only condition.

[5] Currently we are conducting an eye-tracking experiment in Dutch in which we compare the interpretation of pronouns with light-quantified, quantified and non-quantified antecedents.

binding and co-reference. This might suggest that the language system prefers a binding interpretation, not because it is by definition more economic (that depends on the type of the antecedent), but due to the time course of processing. That is, the parser initially considers c-commanding antecedents by consulting the logical syntactic module and only later has access to the discourse module after which other (sometimes more suitable) antecedents are evaluated. At this point, however, this explanation is rather speculative and needs to be examined more explicitly in future experiments.

References

Arnold, J., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J.C. (2000). The rapid use of gender information: Evidence of the time course of pronoun resolution from eye tracking. *Cognition, 76,* B13-B26.

Burkhardt, P. (2004). *Representation and Interpretation at the Syntax-Discourse Interface: Establishing Dependency*. Ph.D. Dissertation, Yale University, New Haven, USA.

Caramazza, A., Grober, E., Garvey, C., & Yates, J. (1977). Comprehension of anaphoric pronouns. *Journal of Verbal Learning and Verbal Behavior, 16,* 601-609.

Clark, H. H., and Sengul, C. J. (1979). In search of referents for nouns and pronouns. *Memory and Cognition, 7,* 35-41.

Ehrlich, K. (1980). Comprehension of pronouns. *The Quarterly Journal of Experimental Psychology, 32,* 247-255.

Frazier, L., & Clifton, C. Jr. (2000). On bound variable interpretations: the LF-Only hypothesis. *Journal of Psycholinguistic Research 29,* 125-139.

Greene, S. B., McKoon, G., & Ratcliff, R. (1992). Pronoun resolution and discourse models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 182,* 266-282.

Grodzinsky, Y., & and Reinhart, T. (1993). The innateness of binding and coreference. *Linguistic Inquiry, 24,* 69-101.

Koornneef, A.W. & Van Berkum, J.J.A. (2006) On the use of verb-based implicit causality in sentence comprehension: Evidence from self-paced reading and eye tracking. *Journal of Memory and Language, 54(4),* 445-465.

McDonald, J. L. & MacWhinney, B. (1995). The time course of anaphor resolution: Effects of implicit verb causality and gender. *Journal of Memory and Language, 34,* 543-566.

Reinhart, T. (1983). *Anaphora and Semantic Interpretation*. University of Chicago Press, USA.

Reinhart, T. (2000). Strategies of anaphora resolution. In *Interface Strategies,* ed. Bennis, Everaert, & Reuland, 295-325. Amsterdam: Royal Academy of Arts and Sciences.

Reuland, E (2001). Primitives of binding. *Linguistic Inquiry 32,* 439-492.

Ambiguity in Anaphora Workshop Proceedings, pages 73–82
ESSLLI 2006, Málaga, Spain, 7–11 August 2006
Edited by Ron Artstein and Massimo Poesio
http://cswww.essex.ac.uk/anaphora/

On pronouns in Catalan and game theory[1]

Laia Mayol
University of Pennsylvania
laia [at] babel.ling.upenn.edu

The goal of this paper is to explore whether game theory can account for the distribution of pronouns in Catalan, a null-subject language. I extend the game-theoretical account proposed by Clark and Parikh (2006), which uses games of partial information to account for anaphora production and resolution in English. Their proposal needs to be extended to account for the fact that null-subject languages make extensive use of null pronouns in subject position, apart from overt pronouns, proper names and definite descriptions. The main idea is that speaker and hearer choose the most efficient strategy with the highest payoff, given the type of entity they want to refer to (more or less salient) and the kind of information they want to convey. Game theory provides an elegant framework to capture anaphora production and resolution, by giving payoffs to each option and probabilities to information states. However, anaphora production and resolution in Catalan constitute a challenge for this type of approach. While this approach elegantly captures the use of an overt pronoun to mark a special information structure, its use in intrasentential anaphora situations is more problematic to model.

1. Introduction

In Catalan, a pronoun in subject position can be either overt or phonologically null. The goal of this paper is to model the distribution of overt and null pronouns within a game-theoretical framework, extending the account proposed for English by Clark and Parikh (2006) and discussing whether game theory makes the right predictions.

This paper is structured as follows. Section 2 gives an overview of game theory and its applications to linguistics. Section 3 briefly reviews Clark and Parikh's proposal. Section 4 introduces the pronominal system of null-subject languages. Section 5 analyzes some cases of intrasentential anaphora in connection with the choice and interpretation of referring expression. Section 6 deals with the situations in which overt pronouns are mandatory as markers of a particular information structure. Section 7 concludes.

2. Game theory and linguistics

Game theory (GT) is the study of mathematical models of conflict and cooperation between intelligent rational decision-makers (Myerson, 1991). In linguistics, GT has mainly been used in semantics and pragmatics, since it provides a good framework to explain why speakers and hearers (that is, rational agents) choose a certain action, i.e. why they utter a

[1] Thanks to Robin Clark, Prashant Parikh and Josh Tauberer for many helpful comments and to Lucas Champollion for many interesting observations and for letting me use his ideas of how to draw a game tree.

sentence or interpret a sentence with a particular meaning in a particular context. Specifically, game theory has been applied to derive the semantics of questions, quantifiers scope, discourse anaphora (Clark and Parikh, in preparation) and implicatures (Parikh, 2000).

The basic idea of GT is that an agent must often decide among several possible actions; in such cases, he will obviously choose the one he prefers. It is possible to translate this preference by giving a numerical value (or payoff) to each option. The option with the highest number will be the preferred action. However, sometimes payoffs are uncertain, so that every possible outcome has a certain probability associated with it. In that case, the agent might choose the action with the highest expected payoff or he might prefer the action with the minimum risk. When there is more than one agent making decisions, the action one agent decides to do might affect the other agents' payoffs and, thus, the other agents' decisions. That is, one agent needs to consider the other agents' actions and payoffs in order to choose his best option. In this sense, there is a strategic interaction among all rational agents. When no agent has an incentive to change his action (given all others agents' actions), an equilibrium (called Nash equilibrium) is reached and the game is solved. In a given game, there may be several Nash equilibria; the one(s) with the highest payoffs for both agents is called Pareto-Nash equilibrium.

A game of partial information is that in which, at some point, one of the agents does not know which state he is in. The usefulness of games of partial information for linguistics is straightforward (Parikh, 2000). Speaker and hearer are rational agents; the speaker is trying to convey some information by uttering a proposition (among the several possible propositions she could utter), and the hearer is trying to correctly interpret this proposition (among several interpretations, given the fact that language is ambiguous). Both agents are trying to minimize production and processing costs (by, for example, avoiding unambiguous but extremely long sentences), while communicating successfully.

3. Clark and Parikh's approach

Consider the simple text in (1):

1. A cop saw a hoodlum. He yawned.

There are several issues regarding the choice of referring expression in this small text, both from the speaker and hearer point of view. Namely, how does the hearer (B) know who *he* refers to? Why does the speaker (A) choose to utter *he* instead of a definite description (DD)? Clark and Parikh (2006) view this problem as a game of partial information in which A and B share some knowledge and in which they try to find the most efficient strategy to solve the game of communicating the utterance. On the one hand, A uses particular discourse anaphors when she expects B to be able to correctly identify the referent. On the other hand, B chooses antecedents to discourse entities based on how he expects A to refer to each entity. As both agents, speaker and hearer, are aware of this fact (they both know they are playing this game), they can find a maximally efficient solution, that is they can compute a Pareto-Nash equilibrium as a solution for the game. This equilibrium is maximally efficient in the sense that for A it is the best way to encode the meaning that she wants convey and, for B, given the form he has heard, it is the best way to interpret the referring expression.

The game tree in Figure 1 shows the moves of A and B and the payoffs they get in each situation for each option. There are two trees for each information state s and s', with probabilities p and p'. The tree rooted in s is the one in which A intends to refer to the cop

(Subj, henceforth), and s' is the one in which A intends to refer to the hoodlum (Obj, henceforth). For each tree, A may use a pronoun or a DD. If a pronoun is used, B may resolve the anaphor correctly or may make a mistake. This is precisely what makes this game a game of partial information: when hearing a pronoun, B does not know whether he is in t or in t' (this is indicated by circling both states).

In s, if A uses a DD, B will surely resolve the anaphor correctly. However, the payoffs will not be very high due to production and processing costs (DDs are longer and syntactically more complex than pronouns) and due to the assumption that referring to a prominent element (the subject) with a full description rather than a pronoun entails some cost. In the example, these assumptions are modeled by payoffs of (6)[2]. If A uses a pronoun in s and B correctly resolves the anaphor, the payoffs are higher (10), since the costs are much less. However, if B interprets Obj instead of Subj, the payoffs would be negative (-10) and would lead to an undesirable situation. In s', the situation is very similar. However, if a DD is used, the payoffs are (7) (and not (6) as in s), which represents the fact that using a DD for a less prominent entity (the object) is assumed to be less costly. Also, if A chooses a pronoun and B correctly chooses Obj, the payoffs are (8) and not (10), because it is less efficient to pronominalize a less prominent element. If A chooses a pronoun and B incorrectly chooses Subj, the payoffs are again negative (-10).

To summarize, payoff are assigned following this two principles:
1. Longer expressions are most costly, while pronouns are less expensive
2. It is cheaper to refer to more salient entities with pronouns, and to less salient entities with DDs.

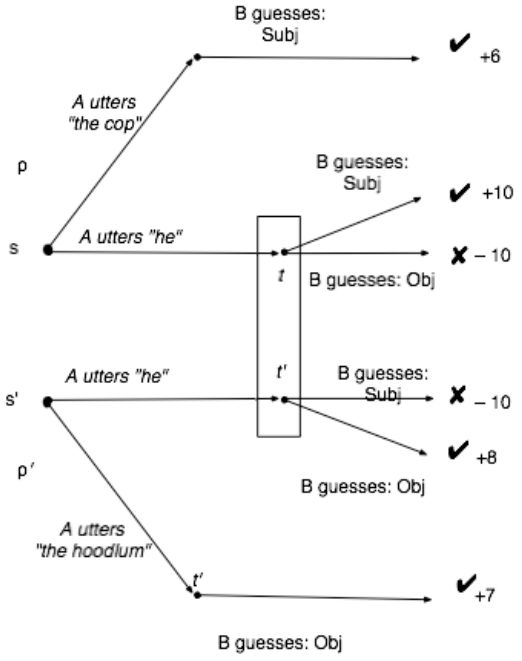


Figure 1

[2] They assume payoffs are equal for speaker and hearer.

In absence of further information, they assume that the two information states are equally likely, that is p = p' = 0.5. In that case, there are two Nash equilibria. One corresponds to A uttering *he* in s, uttering *the hoodlum* in s' and B choosing Subj whenever he's in {t, t'}. This equilibrium has an expected payoff of p·10 + p'·7 = 8.5. The other equilibrium corresponds to A uttering *the cop* in s, *he* in s' and B choosing Obj. This equilibrium has an expected payoff of p·6 + p'·8 = 7. Since the payoff of the first equilibrium is higher, it is the unique Pareto-Nash equilibrium and it is the strategy A and B will follow.

Clark and Parikh also show how this account can deal with apparent counterexamples. The basic idea is that several factors can influence the probabilities, so that one of the information states becomes more likely. For instance, note the following contrast:

2. a. John called Bill a Republican. Then he insulted him.
   b. John called Bill a Republican. Then HE insulted him.

The partial game just presented correctly predicts that the pronoun in (a) should refer to John. The game for (b) should be identical to the game for (a). However, the contrastive stress in the pronoun has the effect of affecting the probabilities, so that p' > p; that is, it becomes more likely that the speaker wants to refer to the object of the previous sentence.

4. The pronominal system of null subject languages

Catalan, as other Romance languages including Italian or Spanish, is a null-subject language and has a double system of pronouns: (a) strong lexical pronouns and (b) clitic and empty pronouns.

3	a. Ell t' estima a tu	b. [∅] T' estima
	he obj-cl loves to you	obj-cl loves
	`He loves you'	`He loves you'

In subject position, there is an alternation between overt and null pronouns (*ell* in 3a, null in 3b). From a pragmatic point of view, weak pronouns are the unmarked form in Catalan (Vallduví, 2002). It has been suggested that overt pronouns are used when, for some reason, it is difficult to access the referent they denote. For instance, one such case would be when there are two possible antecedents for the pronoun. The use of overt pronoun could help to select the less accessible antecedent. We will explore some of these cases in section 5. However, Vallduví (2002) also notes that in some cases the use of overt pronouns is mandatory. Constructions in which a stressed form is needed (clefts, answers to a wh-question, comparative constructions, focus constructions, constructions with an elliptical verb) will require the use of an overt pronoun, since a null pronoun cannot, of course, be stressed. Such cases will be examined in section 6.

5. Intrasentential anaphor in Catalan

Consider the following mini-discourses:

4. a. El Joan va pegar el Pere. [∅] Està enfadat
      The Joan hit-PAST the Peter. Is angry
   b. El Joan va pegar el Pere. Ell està enfadat.
      The Joan hit- PAST the Peter. He is angry
   c. El Joan va pegar el Pere. El Pere/?? El Joan està enfadat
      The Joan hit- PAST the Peter. The John/The Peter is angry

Alonso-Ovalle et. al. (2002) used the Spanish version of 4a and 4b in an experimental setting. They asked subjects the question "Who is angry?". When presented with the text in (4a), the subject of the first sentence was chosen as antecedent for the null pronoun in 73.2%

of the answers. When presented with (4b), the percentage of responses choosing the subject of the first sentences dropped to 50.2%, while the percentage of responses choosing the object of the first sentence increased considerably.

I reproduced the same experiment with thirteen Catalan speakers and the results were fairly similar to the Spanish experiment. For the null pronoun, in 68.75% of the sentences, the speakers said the pronoun referred to the subject, in 18.75%, to the object and in 12.5% they judged the pronoun to be ambiguous. In contrast, with the overt pronoun, 50% of the time the speakers said that the pronoun was referring to the subject and 50% to the object.

From the Spanish and Catalan experiments, we see the following generalizations:

- The null subject of (4a) is mostly interpreted as unambiguously referring to the subject of the previous sentence, Joan. Being interpreted as the object of the previous sentence is not impossible, but much less likely.
- The facts of (4a) sharply contrast with the situation in (4b). Experimental data from Spanish and Catalan indicates that both the subject and the object of the sentence are equally likely.

As for (4c), the felicity of using a proper noun (or a DD) depends on the grammatical function of the phrase referring to this entity in the previous sentence. Using a proper noun to refer to a previous subject (the most salient entity of the sentence) is highly infelicitous. Using a proper noun to refer to the object of the previous sentence (which is less salient) is perfectly felicitous.

Modeling these facts with game theory is a challenging task. The game between hearer and speaker to resolve the anaphor, shown in Figure 2, looks similar to the game for the English discourse in Figure 1, but the complexity has increased because A now has three choices, instead of two: she can use an overt pronoun, a null pronoun or a proper noun. B has to decide whether A wants to refer to Joan, Subj, or to Pere, Obj.

Let's first assume that A wants to refer to Subj (the actual information state is s). In that case, if A uses a proper noun, B will surely interpret it correctly, but the payoffs will not be high because of the production and processing costs and because it is not efficient to use a proper noun to refer to the subject. We model this with payoffs of (5). If A uses an overt pronoun and B correctly interprets Subj as the referent, the payoffs will be higher than in the case of the proper noun, but not maximally high (not (10)) because there is a better strategy to refer to a maximally salient entity. This strategy is the use of a null pronoun: it is the most efficient and less costly strategy to refer to maximally salient entities and, thus, the payoffs are (10). When using a pronoun (overt or null), there is the chance that B makes a mistake and incorrectly interprets Obj instead of Subj. In such cases, the payoffs are negative (-10).

Let's now assume A wants to refer to Obj. If B incorrectly interprets Subj in the case of both pronouns, the payoffs are again negative (-10). If A uses a null pronoun, the payoffs will be (9), and not (10) as in s, since it is not as efficient the use a null pronoun to refer to a non-maximally salient entity as to refer to a maximally salient one. The use of a proper noun yields a better payoff than in s, (7) versus (5), because we are referring to a less salient entity.

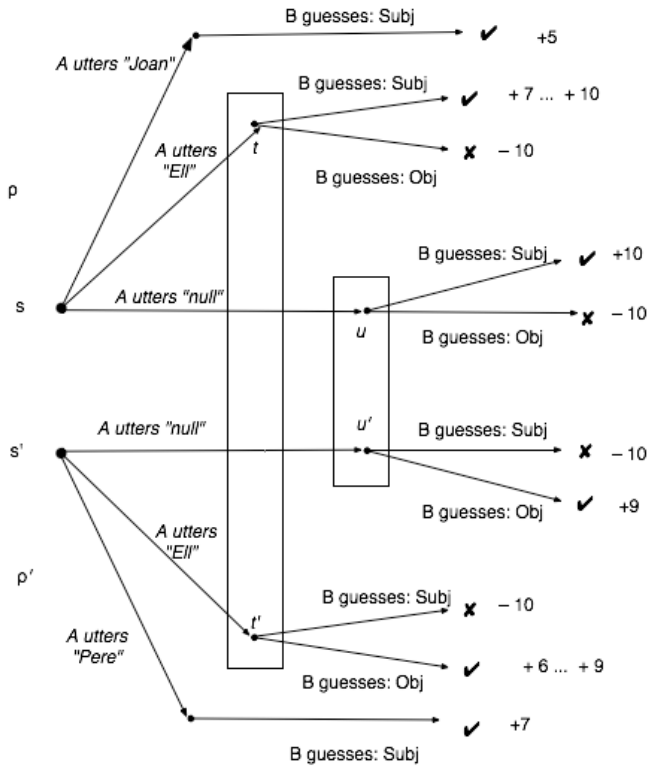


Figure 2

The difficult task is to assign a payoff to the overt pronoun. As mentioned, the use of an overt pronoun is considered ambiguous. This suggests that the payoffs of using an overt pronoun should be the same in s and s'. However, no payoff for the overt pronoun will fully capture the data we want to model:

- If the payoffs consist of any fixed value greater than 7, the Pareto-Nash equilibrium will predict that A should choose the null pronoun in s, utter the overt pronoun in s' and B choose Subj whenever he's in {t, t'} and the Obj consistently whenever he's in {u, u'}. The intuitive reasoning behind this Pareto-Nash equilibrium is the following: the hearer knows that the speaker will always use a null pronoun to refer to the subject and, therefore, is able to conclude that the speaker will only use an overt pronoun to refer to the object.
- If the payoffs consist of any fixed value smaller than 7, B is predicted to choose a proper noun to refer to Obj, since it will have a higher payoff than the overt pronoun.

Therefore, neither option is consistent with the data I want to model. This suggests that in order to model the ambiguity of the overt pronoun we need its payoffs to be indeterminate and to fall within a range. To maintain the Subj-Obj asymmetry, the payoff for the subject should be in the range between 7 and 10 and the payoff for the object between 6 and 9. Modeling the overt pronoun's payoff as a range and not a fixed value can capture a number of things:

1. When hearing an overt pronoun, B will not be able to choose a pure strategy, because, whenever the payoff for the overt pronoun is smaller than 7, there are two

Pareto-Nash equilibria: one in which B has to interpret the pronoun referring to the Subj and the other to the Obj. Therefore he will have to choose randomly between the two options.

2. Referring to the Subj with a proper noun is highly infelicitous. Our model predicts exactly this, since there are at least two more efficient alternative strategies: the most efficient and unambiguous use of a null pronoun and the ambiguous use of an overt pronoun. In contrast, the use of a proper noun to refer to the Obj of the previous sentence is fully acceptable. Given that the overt pronoun strategy was ambiguous and its payoffs indeterminate, the proper noun strategy is the most efficient unambiguous strategy.

3. A change of probabilities, by biasing towards one of the information states, should be able to affect the choice of pronouns. Namely, when s' is more likely than s, the speaker will be able to use the null pronoun to refer to Obj. This biasing can be caused by lexical semantics, by choice of discourse markers, etc. For example, in (5) the null pronoun unambiguously refers to the previous object. The biasing comes from the lexical semantics of the verbs (if X hits Y, Y is likely to spend time in hospital and not the other way round) and from the consequence connective "per això" '*therefore'*.

   5. El Pere va apallissar el Joan. Per això, va haver-se d' estar-se un mes a l' hospital
      The Peter beat-pst the John. So, had-pst of stay-pst one month in the hospital
      "Peter beat John. So, he had to stay one month in the hospital"

An obvious question to ask is why should the overt pronoun be associated with a range in its payoffs. Although I don't have a conclusive answer, I would like to suggest that it is related to the fact that the overt pronoun covers a middle ground in a scale of efficiency. The null pronoun is clearly the most efficient way of referring to maximally salient entity and a proper noun is the less efficient, but more safe way. The overt pronoun falls in the middle and it has a mixed status: it is a pronoun, but, unlike most pronouns, is not the most economical way of referring to some entity. A related question would be how this mixed status of the overt pronoun affects the choice of a speaker. The data I have been examining has to do with interpretation and not with production. A plausible consequence of the indeterminacy of overt pronouns would be that speakers will avoid them unless they have an extra reason to use them. Section 6 deals with what these extra reasons may be.

To conclude this section, the Catalan data presented is a challenging case for game theory:

- The use of a null pronoun unambiguously refers to the subject of the previous sentence and these facts are translated in high payoffs for this strategy. Reference to the previous object only arises in case of biasing.
- The use of a proper noun is unacceptable for the subject of the previous sentence and acceptable for the object. Both facts are predicted by our model
- The use of an overt pronoun seems to be ambiguous for the hearer as for the referent choice. We have modeled this by proposing that the payoffs for the overt pronoun are not fixed, but fall within a range. The hearer when trying to interpret an overt pronoun will not be able to choose a pure strategy and will have to randomly choose between Obj and Subj.

6. Overt pronouns as markers of a particular information structure.

In this section, I discuss situations in which an overt pronoun is obligatory in subject position in Catalan, in spite of being the marked option and of there being always the hypothetical alternative of using a null pronoun. Overt pronouns in Catalan can be mandatory even when there is no ambiguity as to which entity is its referent. Catalan is a language with rich verbal inflection that indicates the grammatical person of the subject of the sentence. For instance, in (6) the verb (*nego*, 'deny') is unambiguously first person singular. Therefore, the pronoun does not play any disambiguation role. However, using a null pronoun is not an option and the discourse would be unfelicitous if such a pronoun was used. Why is it so? All examples in which this is the case have a special informational structure, which involve focusing or contrast and which require a segment that may be stressed. Obviously, a null pronoun cannot be stressed and, therefore, an overt pronoun must be used.

> 6. "Hom ha arribat a afirmar que Bernat Mies és el conegut pintor berguedà Joseph Maria de Martín. **Jo** ho nego rodonament [..]"
> "One has even claimed that Bernat Mies is the famous painter Joseph Maria de Martin. **I** deny it vigorously

In (6), there is a clear contrast between what people claim and what the speaker claims. Note that in English, a special focus accent would be used on the pronoun. In Catalan, actually pronouncing the pronoun is enough to convey the contrast.

We can model this situation using partial games[3]. For B, the ambiguity is not between two possible antecedents for the pronoun, but between two possible interpretations: the literal interpretation (*l*), in which the pronoun just refers to some discourse antecedent, and the informational-structure interpretation (*i.s.*), in which apart from the referential use of the pronoun it also conveys some other information (contrast, focusing, etc.). That interpretation *i.s.* contains more information that *l*, and this fact will be encode in the payoffs.

Figure 3 captures all the relevant facts. The two initial nodes s and s' represent two different situations in which A wants to convey either *i.s.* is *l* respectively. B, when hearing the pronoun (overt or null), will not be sure in which state he is in: t or t' in the case of an overt pronoun, q or q' in the case of a null pronoun. In addition, we also need to consider other propositions that A might have said but chose not to. In s, A could have said a proposition *ex* in which the information structure is explicit (for instance, "In contrast, I"). In s', where the speaker does not want to convey any extra information structure, A could have chosen to remain silent, proposition υ.

As for the payoffs, in s, if A uses an expression that explicitly conveys the information structure, B will correctly interpret it, but the payoffs will not be very high due to processing and production costs. If A uses a null pronoun in s, the payoffs will be negative (-10), since a null pronoun cannot express this information structure meaning and the hearer will surely misinterpret. If A uses an overt pronoun, B can interpret *l* (and, therefore, the payoffs will also be negative) or can correctly interpret *i.s.*. In that, case the payoffs will be (8), since there are some processing costs but less than in the case of the explicit proposition.

[3] My analysis follows very closely Parikh's (2000) analysis of implicatures.

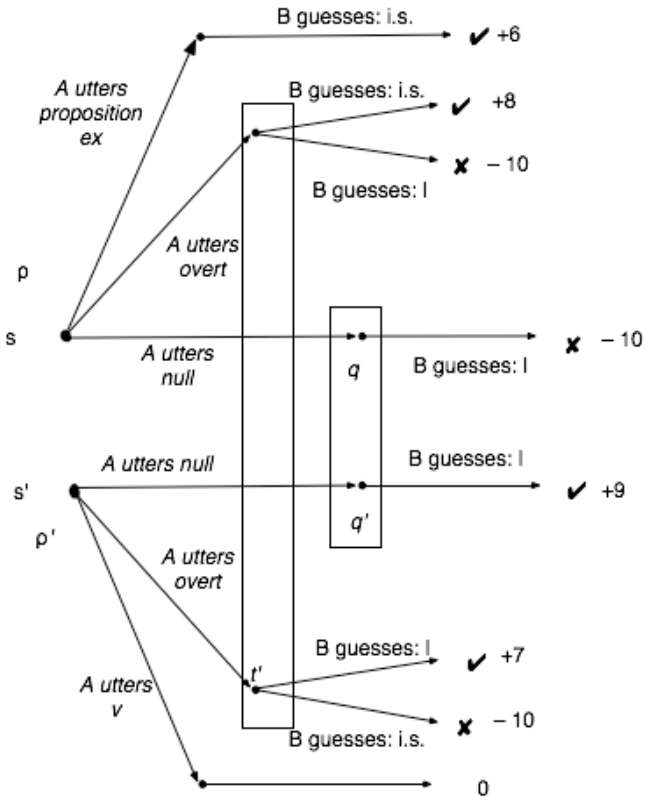


Figure 3

In s', if the null pronoun is used, B will correctly interpret *l* and the payoffs will be positive. If an overt pronoun is used, B may incorrectly interpret *i.s.* and the payoffs will be negative or the hearer may successfully interpret *p*, in which case the payoffs will be positive. However, the payoffs will be lower than the same situation in s, since in s' the information conveyed is less and, therefore, has less value. Thus, we model this by assigning payoffs of (7). Finally, probabilities will be assigned to the initial nodes, s and s'. In this case, we can assume both information states have equal probabilities.

There is a unique Pareto-Nash equilibrium with an expected payoff of 8.5 ($p \cdot 8 + p' \cdot 9$), which is the following: {(s, overt), (s', null); ({t, t'}, i.s.), ({q, q'}, l)}. In information state s, A should utter an overt pronoun; in information state s', a null pronoun. In the information set {t, t'}, A should interpret *i.s.* (the literal interpretation of a pronoun plus the extra information structure meaning) and in the information set {q, q'}, B should interpret just the literal meaning *l*. That is, when a null pronoun is used, no information structure meaning is conveyed. In contrast, using an overt pronoun is a possible way to convey such information structure. Therefore, even if the absence of ambiguity or biasing, an overt pronoun will be used if the speaker needs to convey meaning related to contrast or focusing.

In this case, the game-theoretical approach provides an elegant account of why a speaker should use an overt pronoun in absence of ambiguity if she wants to convey a special information structure.

7. Conclusion

In this paper, I have examined whether the game-theoretical approach to pronoun choice and anaphora resolution that Clark and Parikh (2006) have proposed for English can be extended to a null-subject language like Catalan.

Catalan data for intrasentential anaphora pose a challenge for this game-theoretical account specially because of the behavior of an overt pronoun, which speakers judged could equally refer to the previous subject or the previous object. This cannot be modeled be assigning a fixed payoff to the overt pronoun. In contrast, assigning an indeterminate payoff within a range could account for the empirical data. In contrast, game theory can very easily and elegantly account of why overt pronouns are mandatory in cases where the speakers want to convey a particular information structure.

8. References

Alonso-Ovalle, Luis; Charles Clifton, Lyn Frazier and Susana Fernández Solera (2002). Null vs. Overt Pronouns and the Topic-Focus Articulation in Spanish. *Journal of Italian Linguistics*, 14:2.

Clark, Robin and Prashant Parikh (2006). "Game Theory and Discourse Anaphora". Submitted to the *Journal of Logic, Language and Information*.

Myerson, Robin (1991). *Game Theory. Analysis of Conflict.* Cambridge: Harvard University Press.

Prashant Parikh (2000). *Language in Use*. CSLI Publications, Standford, CA.

Sally, David (2003). "Risky speech: behavioral game theory and pragmatics" *Journal of Pragmatics*, Volume 35, Number 8, pp. 1223-1245(23).

Vallduví, Enric (2002). "L'oració com a unitat informativa". Dins: Solà, Joan et al. (ed.). *Gramàtica del català contemporani*. Barcelona: Empúries. Cap. 4.

Ambiguity in Anaphora Workshop Proceedings, pages 83–89
ESSLLI 2006, Málaga, Spain, 7–11 August 2006
Edited by Ron Artstein and Massimo Poesio
http://cswww.essex.ac.uk/anaphora/

Disagreement Dissected: Vagueness as a Source of Ambiguity in Nominal (Co-)Reference

Yannick Versley
Seminar für Sprachwissenschaft
Universität Tübingen
E-mail: versley [at] sfs.uni-tuebingen.de

Using a qualitative analysis of disagreements from a referentially annotated newspaper corpus, we show that, in coreference annotation, vague referents are prone to greater disagreement. We show how potentially problematic cases can be dealt with in a way that is practical even for larger-scale annotation, considering a real-world example from newspaper text.

1 Introduction

Since the early investigations by Hirschman et al. (1997) and the critique of the MUC-7 annotation scheme put forward by van Deemter and Kibble (2000), several large corpora have been annotated with coreference relations, with refinements in terms of annotation schemes (Poesio, 2004), as well as in terms of support by the annotation tools.

After van Deemter and Kibble and their critique of coreference in the case of bound anaphora, critique of the concept of coreference itself came only from Poesio and Reyle (2001), who argue that in the case of mereologically structured entities (both physical objects and abstract objects like plans), it is possible that underspecified references occur without any loss of understandability on the part of the reader/listener, and propose an underspecified DRT representation to cope with these cases.

Poesio and Artstein (2005) argue that coreference annotation can (and possibly should) use underspecified representations to cope with these ambiguous cases. In a study with 18 subjects in an annotation setting, they found that, in the text used, 42.6% of the markables were at least implicitly ambiguous (in the sense that at least two antecedents were chosen by more than two coders each), of which a little more than one third were marked explicitly by annotators when they had been told to do so.

These results are also highly relevant for large-scale annotation efforts like ACE[1], the dutch KNACK-2002 corpus (Hoste and Daelemans, 2004), or the ongoing effort to add coreference annotation to the text of the German TüBa-D/Z treebank (Hinrichs et al., 2005), since these ambiguities may well occur not only in spoken dialogues but also in edited newspaper texts.

In the remaining part of this paper, we will provide a qualitative analysis of disagreements in the TüBa-D/Z corpus, to show that a certain class of cases, namely those involving vague objects, are prone to genuine ambiguities and lead to a decrease in annotator agreement where they occur. This class includes, but is not limited to, the cases that Poesio and Reyle call the 'meronymy pattern'.

[1] http://projects.ldc.upenn.edu/ace/

We contend that ambiguities are interesting, even given that we do not wish to annotate these ambiguities explicitly, since we can raise the annotator's awareness of these ambiguities and propose a resolution of these ambiguities based on independent principles[2].

2 Data examined

In the referentially annotated TüBa-D/Z corpus of written German, we examined the nominal coreference annotations for the 60 articles that had been annotated by two annotators. Inter-annotator agreement (as indicated by F-measure following Vilain et al. (1995)'s scoring scheme) is at 0.83 for all mentions[3]. After normalizing differing spans by mapping them to nodes in the treebank using fuzzy matching, and projecting every markable to the span where it should be following the annotation guide, the inter-annotator agreement improves to F=0.85, which is a visible improvement but less than what Hirschman et al. (1997) found in their study when they let annotators discuss and agree on markables and their boundaries. For full NP mentions only (excluding NPs in predicative positions, as in 'Peter is [the greatest fool of all]'), the agreement is at F=0.78.

We classified every full NP mention that any one of the two annotators had annotated as being coreferent with another mention (including pronominal mentions) with a semantic class label using the following categorization[4]:

- Persons (PER) are natural persons, including plural person NPs used metonymously to denote some organization (the conservatives, the policemen).
- Organizations (ORG) are formal groupings of persons that are seen as a single actor (e.g. political parties, sports clubs, research institutes)
- Events (EVT) are abstract objects that have a (more or less well-defined) temporal boundary and often result in a change in the state of affairs (e.g. bombings, financial mergers, strikes).
- Locations (LOC) are all geopolitical entities (countries, cities etc.) as well as geographical and physical features.
- Objects (Obj) are things that can be possessed and used and which are generally not seen as being able to perform actions of their own. They may or may not have a material form (as in bank accounts, or electronic books).
- Temporal entities (TMP) are regions of time that are referred to explicitly (e.g. the next week, the eighth day of the strike, Christmas 2006).
- All the rest (Other) is a disconcerting hodge-podge comprising propositions, organizational roles (as opposed to the person filling that role), concepts, legal rights, plans etc. that we did not want to distinguish further.

[2] As an example for such independent principles, consider ambiguous modifier attachment in syntax annotation, where ambiguities are usually solved by attaching to the higher candidate.

[3] Hirschman et al. (1997) also give an agreement figure of F=0.83, but they counted the elements of appositional constructions as two markables linked by a coreference relation while we count appositional constructions as a single markable. Because these additional links between appositions are trivial to annotate, the agreement on the remaining relations is probably slightly better in TüBa-D/Z than in MUC-7.

[4] The annotation of semantic classes was performed by the author of this paper. Zaenen et al. (2004), who did a study with 3 annotators for a slightly finer coding scheme, found that the agreement they got for this task was quite good ($\kappa = 0.92$).

	(total)			(disagree)			error rate		
	pl	sg	all	pl	sg	all	pl	sg	all
PER	156	297	453	33	25	58	0.21	0.08	0.13
ORG	38	310	348	8	39	47	0.21	0.13	0.14
LOC	12	204	217	1	41	42	0.08	0.20	0.19
EVT	31	165	196	5	42	47	0.16	0.26	0.24
Obj	29	80	109	11	11	22	0.38	0.14	0.20
TMP	—	14	14	—	5	5	—	0.36	0.36
Other	16	95	111	1	18	19	0.06	0.19	0.17

Table 1: Disagreements by number (singular/plural)

Looking at the disagreement shown in table 1, we can see that there is significant interaction between disagreements and semantic classes ($\chi^2 = 20.77$, $p < 0.01$), and between disagreements and number ($\chi^2 = 4.76$, $p < 0.05$). Single persons, organizations and objects have the lowest error rates[5], whereas plural objects and temporal entities (which only occured with singular number) exhibit an unusually high error rate.

Several error types contribute to these discrepancies. If we distinguish between 'hard' disagreements, where both annotators' versions can be seen as equally valid, and soft errors, where it is obvious that one of the annotators just overlooked a possible antecedent, we find that many of the errors for single locations and all of the errors for temporal mentions are indeed soft errors and would possibly profit highly from better annotation tools: in these cases, the location or the temporal region is uniquely (and thus unambiguously) specified, but since they are always uniquely specified (and never anaphoric in the sense that context information from a specific antecedent was needed for the interpretation). Additionally, keeping track of the temporal and spatial locations in a story is usually not required, while keeping track of the protagonists of a story (usually persons and/or organizations) is required for its understanding. In three of the five erronous coreference decisions regarding temporal mentions, we found that, to realize the coreference relations between the mentions, it would be necessary to make certain inferences that a cursory reader will almost certainly not make.

For plural objects, another source of disagreement is overrepresented, the ambiguity whether a given mention is used in a specific or in a generic sense, typically when a class of objects is denoted. As a simplified example, the sentence "John threw out the red shirts" can have a specific reading (where John acts on a well-defined set of shirts, moving them from his closet to the dustbin) and a generic one (where John expresses his attitude toward the class of red shirts, and he's less likely to buy one again).

The 'hard' disagreements in the PER class mostly involve groups of persons, which are *not* generic, but the actual set of persons that they denote is vague, and annotators decided differently on the question whether two vague objects corefer. Problems with vague reference are usually suspected with event coreference, which is why general event coreference is usually excluded from annotation schemes that are geared towards the reliable annotation of large text quantities, but the presence of this problem for groups of persons (and organizations) suggests that a principled treatment of vague reference would benefit not only the coreference annotation for nominalized events, but also that for groups of persons, which are as important (disagreement-wise) as the former.

[5]We defined the error rate as the ratio between the number of disagreements and the number of markables that were coreferent to another markable in at least one annotator's version.

3 Coreference of Vague Objects

In order to be able to notice, discuss, and possibly resolve ambiguities (or equivalently, argue that a certain annotation is right, wrong, or left ambiguous by the annotation guidelines), we need to complement our naïve understanding with a (semi-)formal description of what coreference is; it is mostly uncontroversial that we build some kind of model from the text, with mentions referring to entities that either have been mentioned in the text or can be accommodated in the model. In terms of the scene that the text describes, it is unlikely that several blatantly dissimilar interpretations arise for the kind of text that we are investigating. Thus, ambiguities must be attributed to the reference relation between mentions on one hand and pieces of modeled reality on the other hand, and identity conditions between these pieces, which are both non-issues with concrete referents, or some vague referents like mountains that can be individuated by their peak. Without an obvious individuation criterion, coreference decisions can become difficult.

Consider the following sentences, taken from the TüBa-D/Z corpus[6]:

(1) a. Für ein "barrierefreies Bremen" gingen deshalb gestern [1 mehrere hundert behinderte Menschen] auf die Straße – und demonstrierten für "Gleichstellung statt Barrieren".

b. "Warum immer wir?" fragten [2 die Versammelten] deshalb auf Plakaten.

It is intuitively clear that the person groups from mentions 1 and 2 have a large overlap, but, seen in isolation, the real-world extensions of the two mentions do not seem to be identical, as not every demonstrator had disabilities, and neither did every one of them carry a poster with the indicated question. On the other hand, saying that 1 and 2 denote different entities would miss the point, since the author meant to talk about the group of demonstrators and not several largely overlapping subsets of it.

We would like to treat the demonstrators as one entity that is described by several predications and not several distinct entities, just as we would not want to talk about multiple clouds when there is just one cloud in the sky to which several predicates apply differently on different parts.

If we treat the conditions of being disabled and of carrying posters as incidental and instead use the demonstrating as the defining property of the crowd of mentions 1 and 2, we can coerce the individual predicates of being disabled, and wanting to push for a "barrier-free Bremen", to (vague) predicates of groups by taking a majority view.

That is, the article talks about a crowd of demonstrators that

- wanted to push for a "barrier-free Bremen"
- comprised (about) several hundred people
- consisted (in a significant proportion) of disabled people
- had some posters asking "Why always us?" (cumulative reading)

Intuitively, this is much closer to the intended interpretation than talking about several overlapping but not identical crowds. But we have to make sure that we will not run into problems this way, or at least that we know it when we do — if we point to a crowd, it is unclear whether we mean this set of persons or another one that differs in only one person belonging or not to the set, giving us multiple equally possible extensions for that crowd.

[6]Translation: (a) For a "barrier-free Bremen", [1 several hundred disabled people] went onto the streets yesterday — and demonstrated for "Equality, not Barriers". (b) "Why always us?" [2 the congregated] asked on the posters.

The problem of vagueness in reference has been studied extensively (see Weatherson, 2005), and we will use Smith and Brogaard (2001)'s superevaluationist account of reference to vague objects and predications of these objects.

Smith and Brogaard posit that you can, for a vague object, give multiple precisifications relevant to a certain context – for a cloud, several cloud-shaped sets of water molecules, for a crowd, multiple sets of persons, or, for a lorry loaded with oranges, the lorry with or without the oranges.

A statement is then judgeable and true (supertrue) iff we can instantiate every singular term with a corresponding family of aggregates and that, however we select a single possibility from the family of aggregates, the statement is true.

If we construct a logical form out of the sentences from the crowd example and model all possible crowd extensions, the conclusion ("the crowd from sentence (1) is the same as the crowd from sentence (2)") will obviously *not* be supertrue since we could always choose two different extensions for the two crowds. Saying that two crowds are the same when they have a large overlap would partially solve the problem, but leads us to Sorites-style paradoxes where the crowd of demonstrators is the same crowd as another crowd etc., where the last crowd of the chain is the same as a crowd totally different from the first one. We can posit identity independently of the extension for objects where we have an individuation criterion, for example the peak for a mountain, or for humans, but not for crowds.

This is where the idea of a (cognitive) model comes in, since we can conceptually separate the process of model construction from the process of model verification (i.e. seeing if the model fits the real world, or asking questions about possible conclusions). This is also done in SDRT (Asher and Lascarides, 2003), where the construction of the discourse model is done using a quantifier-free default logic, whereas the semantics of the model itself uses a monotonic logic without quantifiers. In a similar fashion, we want to handle vague predications in the semantics of the discourse model itself, but not in the construction of the model.

For our crowd example, we could construct a discourse model with a single referent for both mentions and supplement it with possible extensions of the crowd, of consisting of several hundred people etc. such that the predications of the text (in the form of the discourse model, with the given sets of possible extensions) are supertrue. But we could also construct a discourse model with two separate referents for the two mentions. What keeps us from positing another model with two (or even more) overlapping but unrelated crowds?

We could argue that there is nothing that keeps us from positing a model with two overlapping crowds that are both mentioned in the text, and that the distinction between the one-crowd and the two-crowd model is best left underspecified. But we would like to be able to choose as the preferred interpretation one of the two possible discourse models (which we posit can both be constructed from the text and are both supertrue when evaluated in conjunction with plausible predicate extensions).

In terms of the number of entities involved, a model with multiple overlapping crowds would be larger, and by positing more identity relations we decrease the number of entities in the model. We can say that we only want to consider *minimal* models, as proposed by Gardent and Webber (2001), more specifically what they call locally minimal.

This leaves open the question what to do when there are multiple minimal models that all make the statements of the discourse judgeable and true, and it can be argued that an approach using underspecification like the one of Poesio and Reyle (2001) is still needed for these cases. Poesio and Reyle's example of "hooking up the engine to the boxcar and sending *it* to London" is not disambiguated by our criteria since we cannot distinguish between the interpretation where the train is referenced as a (vague) extension of the boxcar and that where

it is referenced as a (vague) extension of the engine, at least not using domain-independent principles. But our account correctly predicts that "stirring up the butter and the sugar and cooking *it* on a stove" is awkward, since the mixed substance cannot be seen as a vague extension of either of the two.

We also hope that using underspecification, or ad-hoc ambiguity resolution, is needed for fewer cases and can be used with greater confidence, allowing for a better compromise between simplicity and annotation quality than relying on multiple annotators to make consistent ad-hoc judements.

4 Conclusion

We have provided a quantitative analysis of disagreements in a referentially annotated corpus, the TüBa-D/Z corpus of written German, and shown that entities with vague extensions like groups of persons are subject to greater-than-average disagreement among annotators. We proposed a theoretical framework based on Smith and Brogaard (2001)'s superevaluationist account of reference to vague objects and minimal models that can help to better understand and resolve difficult cases of coreference, complementing Poesio and Reyle (2001)'s approach by stating further conditions on when underspecification is really necessary. A further study is needed to show if and by how much the improved theoretical framework leads to better agreement among annotators and, generally, better annotation quality, as we hope and at least van Deemter and Kibble (2000) seem to imply in their article.

Acknowledgements I would like to thank Heike Zinsmeister, Mareile Knees, Sandra Kübler and Piklu Gupta as well as three anonymous reviewers for helpful comments on a draft of this paper. The author's work was supported as part of the DFG collaborative research centre (Sonderforschungsbereich) "SFB 441: Linguistische Datenstrukturen".

References

Asher, N. and Lascarides, A. (2003). *Logics of Conversation*. Cambridge University Press.

Gardent, C. and Webber, B. (2001). Towards the use of automated reasoning in discourse disambiguation. *Journal of Logic, Language and Information*, 10:487–509.

Hinrichs, E., Kübler, S., and Naumann, K. (2005). A unified representation for morphological, syntactic, semantic and referential annotations. In *ACL Workshop on Frontiers in Corpus Annotation II: Pie in the Sky*, Ann Arbor.

Hirschman, L., Robinson, P., Burger, J., and Vilain, M. (1997). Automating coreference: The role of automated training data. In *Proc. of AAAI Spring Symposium on Applying Machine Learning to Discourse Processing*.

Hoste, V. and Daelemans, W. (2004). Learning Dutch coreference resolution. In *Fifteenth Computational Linguistics in the Netherlands Meeting (CLIN 2004)*.

Poesio, M. (2004). The MATE/GNOME scheme for anaphoric annotation, revisited. In *Proc. of SIGDIAL'04*, Boston.

Poesio, M. and Artstein, R. (2005). Annotating (anaphoric) ambiguity. In *Corpus Linguistics 2005*, Birmingham.

Poesio, M. and Reyle, U. (2001). Underspecification in anaphoric reference. In *Fourth International Workshop on Computational Semantics (IWCS-4)*.

Smith, B. and Brogaard, B. (2001). A unified theory of truth and reference. *Logique et Analyse*, 43(169-170):49–93.

van Deemter, K. and Kibble, R. (2000). On coreferring: Coreference in MUC and related annotation schemes. *Computational Linguistics*, 26(4):629–637.

Vilain, M., Burger, J., Aberdeen, J., Connolly, D., and Hirschman, L. (1995). A model-theoretic coreference scoring scheme. In *Proceedings of the 6th Message Understanding Conference*. Morgan Kaufmann.

Weatherson, B. (2005). The problem of the many. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. The Metaphysics Research Lab, Stanford University.

Zaenen, A., Carletta, J., Garretson, G., Bresnan, J., Koontz-Garboden, A., Nikitana, T., O'Connor, M. C., and Wasow, T. (2004). Animacy encoding in english: why and how. In *ACL 2004 Workshop on Discourse Annotation*.

Research on Language and Computation

Ambiguity and semantic judgments

Special issue, edited by Massimo Poesio and Ron Artstein

We invite articles for a special issue on ambiguity and semantic judgments from a computational, theoretical and psychological perspective. Much research in computational linguistics assumes that tasks have a single answer: word sense disambiguation looks for an unambiguous sense in context, anaphora resolution algorithms look for a unique antecedent, question-answering systems look for the best answer, semantic role labeling identifies the most appropriate role, and so on. Yet theoretical and psychological evidence show that ambiguity is abundant, and semantic annotation tasks often display disagreements between coders which are the result of genuine ambiguity rather than annotation error.

We are interested in ambiguity, broadly defined. On the one hand, there are cases where ambiguities constitute clearly distinct interpretations, preserved despite the context. On the other hand, there are instances of underspecification which may or may not be construed as ambiguous given a context. And in between there may be cases where different modes of processing give rise to differences of emphasis which may or may not warrant classifying as ambiguities. All these shades of variation, and the disputes they give rise to, call for more empirical study of matters of ambiguity, especially as they pertain to semantic judgments used in corpus annotation and computational implementation.

For this special issue we are looking for high-quality, original, full-length journal articles on any aspect pertaining to ambiguity and semantic judgment. We especially welcome articles on the following topics:

— Computational implementations which take ambiguity into account

— Empirical research on ambiguity and annotator agreement

— Psychologically motivated research on semantic ambiguity

Deadline for submissions: 15 October 2006. Late submissions will only be considered if time and space allow. It would be helpful if authors who intend to submit an article could let us know by 1 August 2006, or as soon as possible thereafter.

Length: There is no formal length restriction, but please try to keep the length of the articles moderate (around 25–30 pages). If an article is so long as to exclude other articles from the issue, we may ask the authors to shorten it.

Blind review: Please do not include any information identifying the author in the manuscript submitted for review.

Submission method: For review purposes, please submit your article as a PDF attachment to **ambiguity@essex.ac.uk**. Include contact information in the body of the email.

Further information: http://cswww.essex.ac.uk/ambiguity/

Ambiguity in Anaphora

Workshop Program

Monday, 7 August 2006

14:00–14:10 Introduction and welcome
Massimo Poesio and Ron Artstein

14:10–14:50 The German temporal anaphor *danach* – Ambiguity in interpretation and annotation
Mareile Knees

14:50–15:30 Disagreement dissected: Vagueness as a source of ambiguity in nominal (co-)reference
Yannick Versley

Tuesday, 8 August 2006

14:00–14:45 Evaluating a coherence-based model of pronoun interpretation
Laura Kertz, Andrew Kehler, and Jeffrey L. Elman

14:45–15:30 Towards a modular approach to anaphor resolution
Arnout W. Koornneef, Frank Wijnen, and Eric Reuland

Wednesday, 9 August 2006

14:00–14:45 Effects of word order and grammatical function on pronoun resolution in German
Gerlof Bouma and Holger Hopp

14:45–15:30 Effect of relative pronoun type on relative clause attachment
Claire Delle Luche, Roger P. G. van Gompel, Frédérique Gayraud, and Bruno Martinie

Thursday, 10 August 2006

14:00–14:45 On pronouns in Catalan and game theory
Laia Mayol

14:45–15:30 The role of information structure in interpretive asymmetries
Maia Duguine

Friday, 11 August 2006

14:00–14:45 On the (ir)relevance of psycholinguistics for anaphora resolution
Lucas Champollion

14:45–15:30 A model of grouping for plural and ordinal references
Alexandre Denis, Guillaume Pitel, and Matthieu Quignard