Doing Dutch Pronouns Automatically in Optimality Theory

Gerlof Bouma
Center for Language and Cognition (CLCG), Rijks*universiteit* Groningen
Postbus 716, 9700 AS Groningen, The Netherlands
gerlof@let.rug.nl

Abstract

Pronoun resolution algorithms often use elaborate and complicated rules and weighted factors. In this paper I will use the framework of Optimality Theory to implement an automatic pronoun resolution system for Dutch. By ranking constraints, Optimality Theory can be used to model complex behaviour and preferences, whilst keeping the constraints clear and simple. The system is developed by quantative evaluation of different configurations on a small corpus of newspaper articles, containing 222 pronouns. Even though, from a practical perspective, the performance is wanting, some linguistically interesting results are given.

1 Introduction

In this paper I will present some results of the development of an automatic pronoun resolution system for Dutch, i.e. a system where there is no intervention or correction in the stages from flat text to resolved pronouns. The system uses full grammatical information, which it receives from Alpino (Bouma et al., 2001), a wide coverage grammar for Dutch, by means of dependency structures. This syntactic information is needed at several points in the resolution process, from identifying the pronouns to determining agreement values and salience rankings.

In pronoun resolution there is typically a division in the factors between rules and preferences. The divide between the two is not clear cut. Some preferences are strongly preferred over others, wereas some rules maybe should be allowed to be broken every once and a while. In Optimality Theory (OT, Prince and Smolensky (2002)) the total interaction of rules and preferences is modeled by assuming soft-constraints that are ranked. The strict ranking, as opposed to, say, a weighting of constraints, means the system stays conceptually clear.

First I will give a short description of the pronouns that are resolved. Then after giving a description of the algorithm and introducing the constraints that are used, I will discuss the results of testing several configurations on a small corpus.

2 Target Pronouns

The pronouns to be resolved by the algorithm are the third person personal and possessive pronouns, as listed in table 1. These pronouns were selected because they are expected to behave as topic-sensitive pronouns. The reflexive *zichzelf* was added to be able to see the effects of the binding theory. [1]

In the first and second row in table 1 are the common-gender pronouns, used for male and female referents and common nouns, the neuter pronoun is used for neuter nouns. Dutch plural is unspecified for gender.

The system will not resolve *het* ('it'). Apart

[1]The *self*-less reflexive *zich* is mostly encountered in inherently reflexive contexts and was therefore omitted.

	singular			
	c/m	*c/f*	*n*	*plural*
subject	hij	zij, ze	(het)	zij, ze
non-subject	hem	haar	(het)	hun, hen, ze
possessive	zijn	haar	zijn	hun
reflexive		$\cdots$ zichzelf $\cdots$		

Table 1: The target pronouns

from the obvious non-anaphoric *het*, like in weather predicates, extrapositions and clefts, most occurences of *het* in the corpus had non-nominal antecedents.

3 Algorithm

3.1 General Outline

After tokenisation and parsing, the text is traversed by the algorithm. On encountering a possible antecedent (currently most NP's), a feature structure is created. The FS has a semantic part with natural agreement values, humanhood and an index, and a syntactic part with grammatical agreement, syntactic function and category. The FS is put on a stack A of possible antecedents.

On encountering a pronoun P from the target category, the OT system is triggered, with P as the input. The intricacies of the OT system are given in the subsection below.

The output of the OT system is the *antecedent* for P. A full FS for P is constructed by combining the pronoun's syntactic information with the semantic features of the antecedent. The result is put on A, to serve as a possible antecedent itself.

3.2 Implementing Optimality Theory

Abstractly, running an OT model could be seen as going through the following three steps:

1. Create for a given input *I* a set of possible outputs. This set is called Gen(*I*), its members are referred to as *candidates*

2. Check for each candidate which constraints it satisfies, and which not.

3. Calculate the *harmony ranking* of candidates, i.e. how well each candidate satisfies the constraints, taking into account that satisfying certain constraints is more important than satisfying others. The most harmonic candidate (the *optimal* one), is selected as the output.

Several computational problems arise for OT, mainly concerning the construction and the size of Gen(I). There are various ingenious solutions to these and other problems (Tesar, 1995; Gerdemann and van Noord, 2000). However, the application of these solutions to the type of system I am proposing here, is not very straightforward. Rather, by making some limiting assumptions, the abstract OT system above can be implemented almost directly:

ad 1 The input of the system is a partial FS containing syntactical information on the pronoun to be resolved *P*. The output is an antecedent for P and Gen(P) therefore is the stack with FS's of possible antecedents, *A*, which is composed incrementally. The stacksize is (somewhat arbitrarily) set to 20. Increasing the size has no influence on resolution accuracy, the effects of reducing it are only noticable when going down to 10 or below.

ad 2 The constraints are evualuated by checking feature values of the candidate, or by comparing feature values of the input and the candidate, depending on the constraint.

ad 3 Instead of creating a ranking, Gen(P) is systematically searched for just the optimal candidate.[2]

4 Resolution Constraints

The constraints that I propose are traditional resolution factors (see e.g. Mitkov (1999); Hirst (1981) for an overview) that in some cases have been adapted for Dutch. Also they are cast in OT-style formulations, which typically means no (or fewer) conditionals.

While introducing the constraints I will give a few partial rankings. In the experiment possible total rankings are compared.

[2]The concept of a harmony ranking was introduced here because it is a useful tool for reasoning about the behaviour of an OT system. In subsection 4.3, I will come to use it.

4.1 Agreement

For describing pronoun-antecedent agreement behaviour (but also other types of agreement), a distinction is often made between agreeing *ad sensum*: with properties of the referent, or *ad formam*: with properties of the word form (Corbett, to appear).

Dutch pronoun-antecedent agreement is a mixed phenomenon. Dutch NP syntax has two genders: common (or: *de*), and neuter (*het*) gender. For inanimate referents, agreement is *ad formam*. Generally *het*-words get *het* and *de*-words get either *hij* or *zij*. Human referents take their pronoun *ad sensum*. In case of other animate referents the choice is free (Geerts et al., 1984).[3]

Number agreement is generally *ad formam*. The standard exceptions to this rule are group nouns, which allow for both singular and plural agreement.

I propose two constraints to model agreement:

RESPECT HUMAN SEX**:** use the appropriate pronoun for a human referent.

AGREE**:** agree semantically or syntactically.

Syntactic agreement values are filled in in a candidate's FS. Candidates with human referents are also given semantic agreement values. Group nouns get plural semantic agreement.

One constraint, that is to be ranked higher than RHS, will take care of a special case: when the antecedent is a pronoun itself, only syntactic agreement is allowed. The '*' is read as 'don't':

*CHANGE YOUR MIND**:** If the antecedent is a pronoun, there has to be syntactic agreement.

4.2 Binding Restrictions

A relatively simple binding theory can be formulated using versions of the well-known binding principles.

PRINCIPLE A**:** *Zichzelf* takes a co-argument as an antecedent

DISJOINT REFERENCE PRINCIPLE**:** Never take a co-argument as an antecedent

Notice that the DRP is stated without hedges, the fact that the conflicting PRINCIPLE A is ranked over it makes that the binding behaviour is modeled correctly (cf. Beaver (to appear)).

4.3 Topichood

Normally pronouns refer to salient entities, or even: to the topic. How salience should be measured and how to determine what the topic is, is a central issue in many approaches to pronoun resolution of realworld texts.

Centering Theory (CT, Grosz et al. (1995)) links specific definitions of topic ('backward-looking center') and salience ('forward-looking center list'), to model coherence. Utterances yield salience rankings of entities after grammatical function. The topic of an utterance is the most salient entity of the previous utterance that is realised in the current one. Moreover, CT says that if there are pronouns in a sentence, one of them should realize the topic.

If we regard the salience ranking in CT — *subject* $\succ$ *objects* $\succ$ *other* — as a harmony ranking, we can formulate it in constraints. The constraint ranking SUBJECT $\gg$ OBJECT realizes this. The two constraints can be characterized as:

SUBJECT/OBJECT**:** The entity is realized in subject/object position.

An incremental, CT based algorithm is Left-Right Centering (LRC, Tetreault (2001)). On encountering a pronoun, the LRC algorithm looks through the list of discourse entities it has come across in the current sentence. The first entity that meets agreement and binding constraints is selected. If no such is found, it starts looking at the salience ranking from the previous utterance, etc.

Assuming that the same criteria apply to the partial salience ranking of the current sentence, as do to the ranking of previous ones, an implementation of LRC in OT — LRCOT for short — can be given.

The two function constraints are ranked below:

SENTENTIAL PROXIMITY**:** Don't have a sentence boundary between the pronoun and the

[3] This description applies to s.c. Northern Dutch. Speakers of Southern Dutch (mainly in Belgium) allow for syntactic agreement with a human referent. Also they may use a three-gender system, where the common gender is still divided between feminin and masculin.

		SP	SU	OB
1	Su(U_0)			*
2	Ob(U_0)		*	
3	Xx(U_0)		*	*
4	Su(U_{-1})	*		*
5	Ob(U_{-1})	*	*	
...	Su(U_{-2})	**		*

Table 2: Harmony ranking of SP ≫ SU ≫ OB

antecedent.[4]

This is a constraint of the integer type: a mark is given for each sentence boundary. Table 2 shows part of the harmony ranking for SEN PROX ≫ SUBJECT ≫ OBJECT.

Adding the agreement and binding constraints yields the complete ranking for LRCOT. The procederal precedence of the salience rankings over agreement and binding in LRC is not reflected in LRCOT. Binding and agreement are ranked over the other constraints because it is more important to obey them.

Lets look at an example. The LRCOT tableaux for resolving (3) in contexts (2) and (2′) are given in (4), (5). [5]

(1) Jan$_j$ houdt van dansen
jan loves dancing

(2) Jans$_j$ vader$_v$ is dansleraar
jan's father is dancing teacher

(2′) Zijn$_j$ vader$_v$ is dansleraar
his father is dancing teacher

(3) Hij? danst al zijn hele leven
he danced PART his whole life

(4)

	1–2–3: *Hij*	SP	SU	OB
	Jans$_j$	*	*!	*
☞	*Jans$_j$ vader$_v$*	*		*
	Jan$_j$	**!		*

(5)

	1–2′–3: *Hij*	SP	SU	OB
	Zijn$_j$	*	*!	*
☞	*Zijn$_j$ vader$_v$*	*		*
	Jan$_j$	**!		*

Both times the pronoun is resolved to Jan's father. The third candidate in (4) and(5) is ruled out because it is the only candidate to violate SP twice. The competition between the first and second candidate therefore is decided by the SUBJECT constraint.

However, when informally consulting several native speakers, there did seem to be a difference in interpretation. With 1-2′-3, *Hij* = Jan is an option, perhaps even the preferred one.

The fact that this difference does not turn up in Standard CT, is because CT relates pronouns to topics, but not vice versa. Intuitively: the failure to pronominalize the topic Jan, somehow signals a topic shift.[6] Or, turning it around, pronominalizing a topic keeps it available for further mentioning, and it should therefore be highly salient.

One way of dealing with this is to take word form directly into account when determining salience. For instance using the following constraint:

PRONOUN: The antecedent is a pronoun.

Effectively this constraint says that a previous topic should be picked up. Ranking it above SUBJECT gives alternation between the two contexts. The interpretation of 1–2–3 does not change, but *hij* in 1–2′–3 is now resolved to Jan, as shown in (6).

(6)

	1-2′-3: *Hij*	SP	PR	SU	OB
☞	*Zijn$_j$*	*		*	*
	Zijn$_j$ vader$_v$	*	*!		*
	Jan$_j$	**!	*		*

Tetreault (o.c.) also discusses a version of his algorithm, LRC-P, that ranks pronouns over all other things. It outperforms the default variant of LRC.

[4]By sentence I mean what the tokenizer takes to be a sentence.

[5]For space I have omitted the agreement and binding constraints and reduced Gen(*Hij*) to the relevant candidates. The '☞' indicates the antecedent, '!' a violation that takes the candidate out of the competition.

[6]Beaver (to appear) can show the difference between the two contexts because of his constraint: *the topic is pronominalized*, but without effect.

4.4 Parallelism

Smyth (1994) showed that in very restricted cases it pays off to use the strategy *pick a parallel function*, in stead of *pick a subject*. Since Alpino gives us information about grammatical rôles, we can formulate a parallelism constraint like the following:

PARALLELISM**:** The pronoun and the antecedent are in the same argument position.

If the pronoun is not in argument position, this is by definition violated by every candidate, and the constraint is rendered invisible.

5 Results

Testing was carried out on a small corpus of articles from the art section of the Dutch *Volkskrant* newspaper. The pronouns from the target group were hand-annotated by the author for antecedents. The head of the last non-pronominal mention of the entity a pronoun referred to, was taken to be the antecedent. In this manner many occurences of what normally be regarded as cataphors were 'annotated away'. Also the previous mention might be an NP-anaphor, or a used proper name. Neither the annotator nor the algorithm has any information about paragraphs or other document structure apart from sentence boundaries.

There were 222 pronouns from the target group in the corpus. It consisted of 9 articles with a total of about 5000 words in 304 sentences. Of the 222, 13 were lost because of parsing errors from Alpino. A score of 209/222 (94.1%) would therefore be the maximum.

For evaluating the output of the algorithm, two kinds of errors were discerned. A pronoun might be considered correctly resolved if the algorithm coindexes it with an antecedent that, in the annotation, bears the same index as the pronoun. But if this antecedent is a pronoun itself, an error in the resolution may have occurred earlier. This would mean that the error is percolated up through the chain: although the right *coindexation* is given, the pronoun at the end of the chain receives the wrong *anchor*, and the wrong interpretation.

In table 3 results for both ways of counting are given. Which number is interesting depends on the motivation of the research. For more linguistically motivated projects the percolation of errors may be not so interesting, for applications on the other hand it is important pronouns get the right interpretation (see e.g. Stuckhard (2001), for a more elaborate discussion of counting).

To avoid ambiguities, a final constraint DISTANCE was added at the lower end of every hierarchy. DISTANCE gives a mark for every NP between the pronoun and the antecedent. From two equally good candidates the nearest is alway chosen.

There are several baselines in table 3: always taking the closest candidate; taking the closest fitting candidate and taking the closest fitting subject.

Looking at the upper part of the table, it is striking that the subject baseline scores as well as it does, confirming the observation made many times before that subjecthood is a good indicator for topichood. LRCOT, however performs poorly: it does not meet the subject baseline. This is mainly due to the fact that SENTENCE PROXIMITY is too restrictive.

Adding PRONOUN gives a sharp increase in coindexation score. The staying behind of the anchor score for LRCOTP can be ascribed to two things: firstly preferring pronouns means that changes in chains are not picked up, although this effect is being balanced by the strict SENTENCE PROXIMITY, secondly selecting a pronoun more often makes the chains longer, thus increasing the chance of ending up with a wrong interpretation at the end of the chain.

SENTENCE PROXIMITY has been replaced from 6 onwards, by a boolean constraint MAX S$-$1, that requires the antecedent to be in the current or previous sentence. This places antecedents in the current and previous sentence on equal footing, but penalizes coindexing with a NP farther back.

In 6, I also introduce the operator constraint addition, which is not part of standard OT.[7] The marks of two constraints are added to function as

[7] Albro (1998) uses the term *mutually unranked constraints*, but this is slightly misleading.

No.		Constraint ranking	CC[a]	AC[b]
1	*base*	DST	57	36
2	*base*	BA[c] ≫ DST	88	51
3	*lrcot*	BA ≫ SP ≫ SU ≫ OB ≫ DST	112	79
4	*base*	BA ≫ SU ≫ DST	122	95
5	*lrcotp*	BA ≫ SP ≫ PR ≫ SU ≫ OBJ ≫ DST	132	75
6		BA ≫ MS1 ≫ PA + PR ≫ SU ≫ OB ≫ DST	141	87
7		BA ≫ MS1 ≫ HU ≫ PA + PR ≫ SU ≫ OB ≫ DST	143	97
8		BA ≫ MS1 ≫ PH ≫ PA + PR ≫ SU ≫ OB ≫ DST	147	95
9		BA ≫ MS1 ≫ PH ≫ HU ≫ PA + PR ≫ SU ≫ OB ≫ DST	153	105

[a]Number of correct coindexations
[b]Number of correct anchors
[c]The ranked Binding and Agreement constraints

Table 3: Summary of testresults

one integer type constraint. The harmony ranking for PRONOUN + PARALLEL is: *a parallel pronoun* ≻ {*a pronoun, a parallel item*} ≻ *others*.

In the final rankings two more constraints are introduced, that have been formulated on the basis of error analysis:

HUMAN: The antecedent has a human referent.

POSSESSIVES HACK: Possessives take their antecedent in the current sentence.

As can been seen in 7 and 8 in table 3 the two constraints each only introduce slight improvement, but adding them both pushes performance to the maximum of these experiments.

The effect of HUMAN should probably be explained by the nature of the corpus. The POSS HACK is of an *ad hoc* character. It was inserted because the error analysis showed that many of the wrongly resolved possessives had taken an antecedent too far away.

This constraint is far from satisfactory, there are examples in which the constraint is too strict, as well as cases in which it is too loose. An interesting approach would be to assume two kinds of possessives: those that are topic sensitive, and those that are proximity sensitive. It is hoped that a deeper analysis of the data will show whether there is indeed evidence for such a distinction and whether an implementation of the disctinction is feasible in the current framework. Kameyama (1998, p103) mentions a separate treatment of all possessives based on proximity. But performance decreased when using a formulation of the POSS HACK that forced possessives to take the nearest antecedent.

The best configuration in this experiment is the one in 9. Considering the fact that the resolution is fully automatic,[8] the coindexing score (153/222 = 68%) is promising. The anchor score (105/222 = 47,3%) is however too low for the algorithm to be of practical use yet.

6 Comparison

6.1 Other automatic resolution methods

Comparison with other automatic pronoun resolution approaches is difficult. The impact of the use of different resources, having different target classes, training and testing on different corpora and, most of all, working with a different language makes comparison practically impossible.

Still I would like to shortly mention two other algorithms, to give a frame of reference.

Firstly, Mitkov et al. (2002) develop a fully automatic version of an existing knowledge-poor algorithm. The algorithm is developed and tested on a set of English computer manuals, and, although not relying on large knowledge bases and in-depth analysis of the input, it uses 14 factors ranging from syntactic information to genre-specific collocation patterns. These factors are weighted and

[8]About 15 coindexation errors are directly due to Alpino, but it is hard to predict what the impact would be of using fully corrected parses

applied to a candidate set consisting of agreeing and accessible possible antecedents. Also, the system automatically filters non-nominally anaphoric and non-anaphoric occurances of *it*. They achieve an *anchor* score of around 60% on 2263 pronouns.

Secondly, op den Akker et al. (2002) give results on a rule-based, automatic resolution algorithm for Dutch, meant for use in a text summerization task. They report a recall of 69,3%, and a precision of 73,4% for their system which uses weighted indicating factors and shallow syntactic information. This amounts to a *coindexation* score of 50,9%, on a corpus of 440 pronouns. Significantly, this number includes the attempts at resolution of non-nominally anaphoric *het*.

6.2 Other anafora-related OT proposals

Theoretically, the system I have proposed is OT Semantics (Hendriks and de Hoop, 2001), this means that the hearer perspective is taken: the input is a surface form (a pronoun), the output a meaning (an antecedent). Two recent proposals regarding OT and anaphora have taken two quite different approaches.

In his article, Beaver (to appear) gives an OT reformulation of the BFP algorithm (COT). He then extends it to a more comprehensive formalisation of discourse coherence and anaphora resolution. COT can be used both for production and comprehension of discourse.

There are seperate constraints in Beaver's system to model salience, link salience to topicality, and link topics to pronouns, whereas in my proposal all these things are implicit in the harmony ranking of candidates. Also there are constraints to regulate the behaviour of other anaphoric nominals. This results in an elegant and more extensive theory of anaphora, but at considerable costs regarding the character and size of Gen(I). A naive implementation of the model would be computationally very heavy.

Recoverability Optimality Theory (ROT, Buchwald et al. (2002)) is a type of Bidirectional OT. Optimizing production includes optimizing for comprehension. The surface form is then a realisation of the intended meaning that balances speaker economy and hearer recoverability. Input, or intended meaning has the form: intended salience ranking and intended logical form. The constraints mainly link salience directly to recoverability and penalize the use of unreduced forms. Buchwald et al. exploit the possibility of reranking certain constraints, to give a typology of anaphora systems.

The three proposals share a common background (viz. CT) but differ in goals and abstractness. The current state of the rather abstract ROT is not such that one could start analysing complex sentences and discourse with it. And, although Beaver's model can be used for this, its complexity prevents us from using it on a larger scale. My system, finally, can easily be used on a large corpus, but "analysis" is restricted to the resolution of pronouns; it does not offer the explanations or even the level of description, the former two do.

7 Conclusions

I have presented an implementation of an automatic pronoun resolution system for Dutch. Using Optimality Theory, the system remains clear and intuitive, whilst allowing me to start modelling the complex behaviour of pronouns.

On the application front, results are still bleak. On the linguistic side, however, developing the programme and error analysis have already given interesting hints for further research: the use of *het* in Dutch, the influence of *humanhood* on salience, and the peculiar behaviour of possessives.

In the experiments presented here, the basic salience determiner was obliqueness. Strube and Hahn (1999) have argued that information status is a more appropriate measure for salience, especially for free-word-order languages like German. It would be interesting to compare the predictions made by standard Centering and s.c. Functional Centering for Dutch, but even though an OT formulation of Strube's incremental Functional CT algorithm (Strube, 1998) is straightforward to give,[9] information status of an entity is complicated to compute. It *could* be approximated by using the typical NP-forms instead.

Other research directions include: analysis of plural pronouns and plural antecedents — there is

[9] BA $\gg$ MS1 $\gg$ OLD $\gg$ MED $\gg$ SP $\gg$ LEFT ALIGNMENT

no solution for conjunctions and split antecedents at the moment — and reference to implicitly realized entities. In connection to the latter, the Functional Centering framework is interesting, too, since it provides us with a principled way of dealing with indirect reference.

Finally, it would be interesting to see if a hybrid between my system and Beaver's model could be made: an incremental algorithm with a limited backtracking ability. The amount of backtracking necessary could then itself be a measure of the coherence of a discourse.

Acknowledgements

I would like to thank Gosse Bouma and Gertjan van Noord for guiding me in this research and three anonymous reviewers for helpful comments. The work is supported in the framework of the NWO Cognition Program with financial aid from the Netherlands Organization for Scientific Research (NWO), grant number 051.02.071.

References

D. M. Albro. 1998. Three formal extensions to primitive optimality theory. In *Proc. of the 17th International Conference Computational Linguistics*.

D. Beaver. to appear. The optimization of discourse anaphora. *Linguistics and Philosophy*.

G. Bouma, G. van Noord, and R. Malouf. 2001. Alpino: Wide-coverage computational analysis of dutch. In *Computational Linguistics in The Netherlands 2000*.

A. Buchwald, O. Schwartz, A. Seidl, and P. Smolensky. 2002. Recoverability optimality theory, discourse anaphora in a bidirectional framework. In Bos, Foster, and Matheson, editors, *Proceedings of EDILOG 2002*, pages 37–44, Edinburgh.

G. G. Corbett. to appear. Agreement: Terms and boundaries. In William Griffin, editor, *The Role of Agreement in Natural Language: Proc. 2001 Texas Linguistic Society Conference*.

G. Geerts, W. Haeseryn, J. de Rooij, and M. C. van den Toorn, editors. 1984. *Algemeen Nederlandse Spraakkunst*. Wolters-Noordhoff, Groningen.

D. Gerdemann and G. van Noord. 2000. Approximation and exactness in finite state optimality theory. In J. Eisner, L. Karttunen, and A. Thériault, editors, *Finite-State Phonology. Proc of the 5th Workshop of the ACL Special Interest Group in Comp. Phon.*, pages 34–45, Luxembourg.

B. Grosz, K. Joshi, and S Weinstein. 1995. Centering: A framework for modelling the local coherence of discourse. *Computational Linguistics*, 21(2):203–225.

P. Hendriks and H. de Hoop. 2001. Optimality theoretic semantics. *Linguistics and Philosophy*, 24(1):1–32.

G. Hirst. 1981. *Anaphora in Natural Language Understanding: A Survey*. Springer-Verlag, Berlin.

M. Kameyama. 1998. Intrasentential centering: A case study. In M.A. Walker, A.K. Joshi, and E.F. Prince, editors, *Centering Theory in Discourse*, pages 89 – 112. Clarendon Press, Oxford.

R. Mitkov, R. Evans, and C. Orasan. 2002. A new, fully automatic version of mitkov's knowledge-poor pronoun resolution method. In *Proceedings of CICLing-2000*, Mexico City.

R. Mitkov. 1999. Anaphora resolution: the state of the art. Working paper based on the COLING'98/ACL'98 tutorial on anaphora resolution.

R. op den Akker, M. Hospers, , E. Kroezen, A. Nijholt, and D. Lie. 2002. A rule-based reference resolution method for dutch discourse. In S. Harabagiu and A. Ferrandez, editors, *Proc. 2002 Symposium on Ref. Res. in NLP*, pages 59–66, Alicante.

A. Prince and P Smolensky. 2002. ROA 537-0802 - Optimality theory: Constraint interaction in generative grammar. Technical report.

R. Smyth. 1994. Grammatical determinants of ambiguous pronoun resolution. *Journal of Psycholinguistic Research*, 23(3):197–229.

M. Strube and U. Hahn. 1999. Functional centering: Grounding referential coherence in information structure. *Computational Linguistics*, 25(3):309–344.

M. Strube. 1998. Never look back: An alternative to centering. In *COLING-ACL '98*, pages 1251–1257.

R. Stuckhard. 2001. Design and enhanced evaluation of a robust anaphor resolution algorithm. *Computational Linguistics*, 27(4):479–507.

B. Tesar. 1995. *Computational Optimality Theory*. Ph.D. thesis, University of Colorado at Boulder.

J. R. Tetreault. 2001. A corpus-based evaluation of centering and pronoun resolution. *Computational Linguistics*, 27(4):507–521.