An integrated framework for text planning and pronominalisation

Rodger Kibble and Richard Power
ITRI
University of Brighton
Brighton BN2 4GJ
UK.

Email: {Rodger.Kibble | Richard.Power}@itri.brighton.ac.uk

Abstract

This paper describes an implemented system which uses centering theory for planning of coherent texts and choice of referring expressions. We argue that text and sentence planning need to be driven in part by the goal of maintaining referential continuity and thereby facilitating pronoun resolution: obtaining a favourable ordering of clauses, and of arguments within clauses, is likely to increase opportunities for non-ambiguous pronoun use. Centering theory provides the basis for such an integrated approach. Generating coherent texts according to centering theory is treated as a constraint satisfaction problem.

1 Introduction

1.1 Issues in pronoun generation

The appropriate realisation of anaphoric expressions is a long-standing problem in NLG research. However, as McCoy and Strube (1999) observe, few researchers have developed sophisticated algorithms for pronoun generation. A typical approach, exemplified by Dale (1993), Reiter and Dale (1997) is to pronominalise some distinguished referent which was mentioned in the previous sentence according to a domain-dependent criterion of prominence or salience. McCoy and Strube (op cit.) offer a more complex algorithm based on the notion of "discourse threads", for which they report an accuracy of 85% when tested against a corpus of naturally-occurring texts. Their approach makes some fundamental assumptions about discourse structure which appear to be beyond the capabilities of current text and sentence planners and are incompatible with the widely-accepted notion of discourse structure as a tree with non-crossing branches (e.g., Mann and Thompson 1987).

We argue for an approach which integrates the tasks of text planning and choice of referring expression on the following grounds:

- *portability*: this approach should be compatible with any system that employs hierarchical text planning and certain basic grammatical categories;
- *coherence*: we claim that text planning needs to be driven in part by the goal of maintaining referential continuity: obtaining a favourable ordering of clauses, and of arguments within clauses, is likely to increase opportunities for non-ambiguous pronoun use.

The latter claim is not new, but underlies the Centering Theory (CT) of Grosz, Joshi and Weinstein (1995, hereafter "GJW").

1.2 Issues in Text Planning

Text Planning is one of the distinct tasks identified in Reiter's "consensus" architecture for Natural Language Generation (Reiter 1994, Reiter and Dale 1997):

Text Planning - deciding the content of a message, and organising the component propositions into a text tree;

Sentence Planning - aggregating propositions into clausal units and choosing lexical items corresponding to concepts in the knowledge base;

Linguistic realisation - surface details such as agreement, orthography etc.

Following Scott and de Souza (1990), we assume that the component propositions to be realised in a text are organised in a tree structure in which terminal nodes are elementary propositions and non-terminal nodes represent discourse relations as defined by e.g., Rhetorical Structure Theory (RST, Mann and Thompson 1987). This structure only partially constrains the linear order in which the propositions will be realised — in other words, any RST structure specifies a range of possible text plans. We propose as an additional constraint that the generator should seek to maximise *continuity of reference* as determined by centering theory, and we argue that this enables us to select the most cohesive variants from a set of text plans. The RST tree itself is produced by an interactive knowledge base editor which allows a user to control both semantic content and rhetorical structure via a sequence of choices guided by a natural language interface.

2 Reconstructing Centering for NLG

2.1 Centering in a nutshell

The main assumptions of Centering theory are:

1. For each utterance in a discourse there is precisely one entity which is the centre of attention or *center*. The center in an utterance U_n is the most grammatically salient entity realised in U_{n-1} which is also realised in U_n. This is also referred to as the *backward-looking center* or *Cb*. The notion of "salience" for the purposes of centering theory is most commonly defined according to a hierarchy of grammatical roles: SUBJECT > DIRECT OBJECT > INDIRECT OBJECT > OTHERS (see e.g., Brennan et al 1987). For alternative approaches see e.g., (Strube and Hahn 1999), (Walker et al 1994).

2. There is a preference for consecutive utterances within a discourse segment to keep the same entity as the center, and for the center to be realised as Subject or *preferred center* (*Cp*). Kibble (1999) dubbed these principles **cohesion** and **salience** respectively. Pairs of successive utterances $\langle U_n, U_{n+1} \rangle$ are classified into the transition types shown in Fig. 1, in the order of preference specified by Grosz et al's "Rule 2".

3. The center is the entity which is most likely to be pronominalised: Grosz et al's "Rule 1" in its weakest form states that if any entity is referred to by a pronoun, the *Cb* must be.

CONTINUE: **cohesion** and **salience** both hold; same center (**or** $Cb(U_n)$ undefined), realised as Subject in U_{n+1};

RETAIN: **cohesion** only; i.e. center remains the same but is not realised as Subject in U_{n+1};

SMOOTH SHIFT: **salience** only; center of U_{n+1} realised as Subject but not equal to $Cb(U_n)$;

ROUGH SHIFT: neither **cohesion** nor **salience** holds.

NO CB: this transition is used by some researchers but is not discussed by GJW.

Figure 1: Centering Transitions

2.2 Pronominalisation

Text genres are known to vary in the extent to which pronouns are used. The CT-based framework allows us to experiment with different strategies for choosing when to use a pronoun, including:

1. Never use anaphoric pronouns — for instance, in certain legal documents or technical manuals where there must be no possibility of ambiguity.
2. Always pronominalise the *Cb*.
3. Use a pronoun for non-*Cb*s only if the *Cb* is pronominalised.
4. Pronominalise the *Cb* only after a Continue transition.

Strategy 3 is favoured by (GJW) who cite psychological evidence that "the *Cb* is preferentially realised by a pronoun in English and by equivalent forms (i.e., zero pronouns) in other languages" (op cit., p. 214). However, in the implementation reported in section 3 we opt for the more restrictive strategy 4. The generation approach outlined below enables us to experiment with different strategies and compare the resulting texts.

2.3 Centering and discourse structure

The canonical formulation of CT is concerned with local cohesion, specifying fluent transitions between consecutive sentences in a discourse segment and favouring sequences which maintain the same center. Our implementation incorporates two extensions which have implications for more structured discourse: Strube and Hahn's (1999) "cheapness" principle, which favours transitions that introduce a new topic in a salient position, and Cristea's Veins Theory (Cristea et al 1998) which redefines transitions in terms of rhetorical hierarchy rather than linear sequence of clauses (see section 3.3 for discussion).

"Cheapness" is satisfied by a transition pair $\langle\langle U_{n-1}, U_n\rangle, \langle U_n, U_{n+1}\rangle\rangle$ if the preferred center of U_n is the Cb of U_{n+1}. For example, this test is satisfied by a RETAIN-SHIFT sequence but not by CONTINUE-SHIFT, so it is predicted that the former pattern will be used to introduce a new center. (This claim is consistent with the findings of Brennan 1998, Brennan et al 1987.) If we consider examples 1a-e below, the sequence c-d′-e′, including a RETAIN-SHIFT sequence, reads more fluently than c-d-e even though the latter scores better according to the canonical ranking.

1. a. John has had trouble arranging his vacation.
   b. He cannot find anyone to take over his responsibilities.
   c. He called up Mike yesterday to work out a plan. CONTINUE
   d. He has been pretty annoyed with Mike recently. CONTINUE
   e. Mike cancelled last week's project meeting at short notice. EXP-SMOOTH SHIFT
   d′. Mike has annoyed him a lot recently. RETAIN
   e′. He cancelled last week's project meeting at short notice. SMOOTH SHIFT

The "cheapness" principle illustrates the need for global optimisation. We noted above that there is evidence that a RETAIN-SHIFT sequence is the preferred way of introducing a new center after a sequence of CONTINUE. However, in a sequence CONTINUE-RETAIN-SHIFT the SHIFT is predicted in its local context but the RETAIN is not; whenever RETAIN is a cheap transition following CONTINUE, another CONTINUE would be cheap as well. The RETAIN is motivated as it enables a "cheap" SMOOTH SHIFT, and so we need a way of evaluating the whole sequence CONTINUE-RETAIN-SHIFT versus CONTINUE-CONTINUE-SHIFT.

2.4 Centering in NLG

CT has developed primarily in the context of natural language interpretation, focussing on anaphora resolution (see e.g., Brennan et al 1987). Curiously, NLG researchers have tended to overlook GJW's proposal that

> Rule 2 provides a constraint on speakers, and on natural-language generation systems ... To empirically test the claim made by Rule 2 requires examination of differences in inference load of alternative multi-utterance sequences that differentially realize the same content.
>
> GJW, p. 215.

With a few exceptions (e.g., Mittal et al 1998, Kibble 1999, Kibble and Power 1999, Cheng 2000) NLG researchers have interpreted CT as a theory of pronominalisation only (e.g., Dale 1992). In this paper we concentrate on *planning*, aiming to determine whether the principles underlying the constraints and rules of the theory can be "turned round" and used as planning operators for generating coherent text. Text planning in conformity with CT will need to follow the following set of heuristics:

1. Plan the order of clauses so that adjacent clauses have at least one referent in common.
2. **Cohesion**: Prefer orderings which maintain the same Cb in successive clauses.
3. **Salience**: Realise as Subject of U_n the most grammatically salient entity in U_{n-1} which is mentioned in U_n (the Cb).
4. **Cheapness**: Realise as Subject of U_n an entity which is mentioned in U_{n+1} (and will therefore be Cb of U_{n+1}).

Breaking down the problem like this reveals that there are various ways the distinct tasks can be slotted into an NLG system. **Cohesion** naturally comes under Text Planning: ordering a sequence of utterances to maintain the same entity as the center, within constraints on ordering determined by discourse relations. However, identifying the center depends on grammatical salience, which is normally determined by the Sentence Planner - for example, choice of active or passive voice. Three possibilities are:

- "Incremental" sentence-by-sentence generation, where the syntactic structure of U_n is determined before the semantic content of U_{n+1} is planned. That is, the Text Planner would plan the content of U_{n+1} by aiming to realise a proposition in the knowledge base which mentions an entity which is salient in U_n. We are not aware of any system which performs all stages of generation in a sentence-by-sentence way, and in any case this type of architecture would not allow the cheapness principle to be implemented as it would not support the necessary forward planning.
- A pipelined system where the "topic" or "theme" of a sentence is designated independently as part of the semantic input, and centering rules reflect the information structure of a discourse. This approach was suggested by Kibble (1999), proposing that text and sentence planning should be driven by the goal of realising the designated topic in positions where it will be interpreted as the Cb. However, this is not really a solution so much as a refinement of the problem, since it simply shifts the problem of identifying the topic. Prince (1999) notes that definitions of "topic" in the literature do not provide objective tests for topichood and proposes that the topic should be identified with the centre of attention as defined by CT; however, what would be needed here would be a more fundamental definition which would account for a particular entity being chosen to be the centre of attention.
- The solution we adopt is to treat the task of identifying Cbs as an optimisation problem, giving different weightings to the different transitions. We assume that certain options for syntactic realisation can be predicted on the basis of the argument structure of predicates, which means that centering transitions can be calculated as part of Text Planning.

3 Implemented prototype

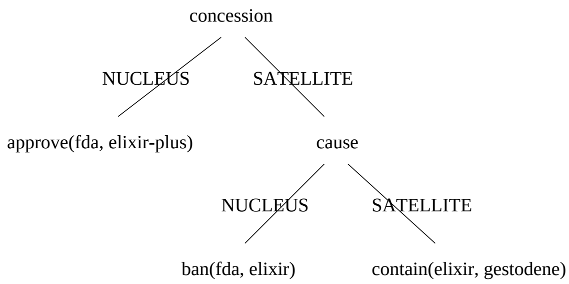

Figure 2: Rhetorical structure

The text planner has been developed within ICONOCLAST, a project which investigates applications of constraint-based reasoning in Natural Language Generation using as subject-matter the domain of medical information leaflets. Following (Scott and de Souza 1990), we represent rhetorical structure by graphs like figure 2, in which non-terminal nodes represent RST relations, terminal nodes represent propositions, and linear order is unspecified. The task of the text planner is to realize the rhetorical structure as a *text structure* in which propositions are ordered, assigned to textual units (e.g., sentences, paragraphs, vertical lists), and linked where appropriate by discourse connectives (e.g., 'since', 'however').

Even for a simple rhetorical input like figure 2 many reasonable text structures can be generated. Since there are two nucleus-satellite relations, the elementary propositions can be ordered in four ways; several discourse connectives can be employed to realize each rhetorical relation (e.g. *concession* can be realized by 'although', 'but' and 'however'); at one extreme, the text can be spread out over several paragraphs, while at the other extreme it can be squeezed into a single sentence. With fairly restrictive constraint settings, the system generates 24 text-structure patterns for figure 2, including the following (shown schematically):

A. Since $contain(elixir, gestodene)$, $ban(fda, elixir)$. However, $approve(fda, elixirplus)$.

B. $approve(fda, elixirplus)$, although since $contain(elixir, gestodene)$, $ban(fda, elixir)$.

The final output texts will depend on how the propositions are realized syntactically; among other things this will depend on centering choices within each proposition.

In outline, the procedure that we propose is as follows:

1. Enumerate all text structures that are acceptable realizations of the rhetorical structure.

2. For each text structure, enumerate all permissible choices for the Cb and Cp of each proposition.

3. Evaluate the solutions, taking account of referential coherence among other considerations, and choose the best.

For the example in figure 2, centers can be assigned in four ways for each text-structure pattern, making a total of 96 solutions.

As will probably be obvious, such a procedure could not be applied for rhetorical structures with many propositions. For examples of this kind, based on the relations 'cause' and 'concession' (each of which can be marked by several different connectives), we find that the total number of text-structures is approximately 5^{N-1} for N propositions. Hence with $N = 4$ we would expect around 600 text structures; with perhaps 5-10 ways of assigning centers to each text structure, the total number of solutions would approximate to 5000. Global optimization of the solution therefore becomes impracticable for texts longer than about five propositions; we address this problem by a technique of *partial optimization* in which a macro-planner fixes the large-scale structure of the text, thus defining a set of micro-planning problems each small enough to be tackled by the methods described here.

Stage 1 of the planning procedure is described elsewhere (Power, 2000); we focus here on stages 2 and 3, in which the text planner enumerates the possible assignments of centers and evaluates which is the best.

3.1 Choosing centers

Given a text structure, we enumerate all permissible centering assignments as follows:

1. Determine the predecessor U_{n-1} (if any) of each proposition U_n.

2. List the potential Cbs and Cps of each proposition, henceforth denoted by ΣCb and ΣCp.

3. Compute all combinations from ΣCb and ΣCp that respect the fundamental centering constraint that $Cb(U_n)$ should be the most salient candidate in U_{n-1}.

Two criteria for determining the predecessor have been implemented; the user can select one or other criterion, thus using the NLG system to test different approaches. Following a *linear* criterion, the predecessor is simply the proposition that precedes the current proposition in the text, regardless of structural considerations. Following a *hierarchical* criterion, the predecessor is the most accessible previous proposition, in the sense defined by Veins Theory (Cristea et al 1998). We will return to this issue later; for now we assume the criterion is linear.

$\Sigma Cb(U_n)$ (potential Cbs of proposition U_n) is given by the intersection between $Cf(U_n)$ and $Cf(U_{n-1})$ — i.e., all the referents they have in common. The potential Cps are those referents in the current proposition that can be realized as most salient. Obviously this should depend on the linguistic resources available to the generator; the system actually uses a simpler rule based on case roles within the proposition. Figure 3 shows the potential Cbs and Cps for the proposition sequence in solution A.

Our treatment of salience here simplifies in two ways. First, we assume that syntactic realization serves only to distinguish the Cp from all other referents, which are ranked on the same level: thus effectively SUBJECT > OTHERS. Secondly, we assume that the system already knows, from the event class of the proposition, which of its case roles can occur in subject position: thus for *ban*, both arguments are potential Cps because active and passive realizations are both allowed; for *contain*, only *elixir* is a potential Cp because we disallow 'Gestodene is contained by Elixir'.

U	**Proposition**	$\Sigma Cb(U)$	$\Sigma Cp(U)$
U_1	contain(elixir, gestodene)	[ ]	[elixir]
U_2	ban(fda, elixir)	[elixir]	[fda, elixir]
U_3	approve(fda, elixir-plus)	[fda]	[fda, elixir-plus]

Figure 3: *Cb*s and *Cp*s for solution A.

With these simplifications, the enumeration of centering assignments is straightforward; in the above example, four combinations are possible, since there are two choices each for $Cp(U_2)$ and $Cp(U_3)$, none of which leads to any violation of the basic centering constraint. This constraint only comes into play if there are several choices for $Cb(U_n)$, one of which coincides with $Cp(U_{n-1})$.

3.2 Evaluating solutions

Various metrics could be used in order to evaluate centering choices. One possibility, for example, would be to associate a cost with each transition, so that perhaps 'Continue' (the best transition) has zero cost, while 'No *Cb*' (the worst transtion) has the highest cost. However, we have preferred the approach mentioned earlier in which cohesion and salience are evaluated separately; this allows us to include the further criterion of cheapness.

Although this paper focusses on centering issues, it is important to remember that other aspects of text quality are evaluated at the same time: the aim is to compute a global measure so that disadvantages in one factor can be weighed against advantages in another. For instance, text pattern B is bound to yield poor continuity of reference because it orders the propositions so that U_1 and U_2 have no referents in common. Text pattern A avoids this defect, but this does not necessarily mean that A is better than B overall; there may be other reasons, unconnected with centering, for preferring B to A.

The system evaluates candidate solutions by applying a battery of tests to each node of the text plan. Each test identifies whether the node suffers from a particular defect. For instance, one stylistic defect (at least for the rhetorical relations occurring in figure 2) is that of placing nucleus before satellite; in general, the text reads better if important material is placed at the end. For each type of defect, we specify a weight indicating its importance: in evaluating continuity of reference, for example, the defect 'No *Cb*' might be regarded as more significant than other defects. Summing the weighted costs for all defects, we obtain a total cost for the solution; our aim is to find the solution with the lowest total cost.

Regarding centering, the tests currently applied are as follows.

Salience violation
: A proposition U_n violates salience if $Cb(U_n) \neq Cp(U_n)$. This defect is assessed only on propositions that have a backward-looking center.

Coherence violation
: A proposition U_n violates cohesion if $Cb(U_n) \neq Cb(U_{n-1})$. Again, this defect is not recorded when either U_n or U_{n-1} has no Cb.

Cheapness violation
: A proposition U_n violates cheapness if $Cb(U_n) \neq Cp(U_{n-1})$.

No backward-looking center
: This defect is recorded for any proposition with no Cb, except the first proposition in the sequence (which by definition cannot have a Cb).

Applied to the four solutions to text structure A, with all weights equal to 1, these definitions yield costs shown in Figure 4. According to our metric, solutions A1 and A2 should be preferred to A3 and A4 because they incur less cost. This result can be assessed by comparing the following output texts, in which the generator has followed the policy of pronominalizing the Cb only after a 'Continue' transition:

A1. Since Elixir contains gestodene, the FDA bans Elixir. However, it approves ElixirPlus.

Solution	U	$Cb(U)$	$Cp(U)$	**Defects**
A1	U_1	$\emptyset$	elixir	none
	U_2	elixir	fda	salience
	U_3	fda	fda	cohesion
			TOTAL	2
A2	U_1	$\emptyset$	elixir	none
	U_2	elixir	elixir	none
	U_3	fda	fda	cohesion, cheapness
			TOTAL	2
A3	U_1	$\emptyset$	elixir	none
	U_2	elixir	fda	salience
	U_3	fda	elixir-plus	salience, cohesion
			TOTAL	3
A4	U_1	$\emptyset$	elixir	none
	U_2	elixir	elixir	none
	U_3	fda	elixir-plus	salience, cohesion, cheapness
			TOTAL	3

Figure 4: Costs of solutions A1 - A4.

A2. Since Elixir contains gestodene, it is banned by the FDA. However, the FDA approves Elixir-Plus.

A3. Since Elixir contains gestodene, the FDA bans Elixir. However, ElixirPlus is approved by the FDA.

A4. Since Elixir contains gestodene, it is banned by the FDA. However, ElixirPlus is approved by the FDA.

Of course we are not satisfied that this metric is the best; an advantage of the generation approach is that different evaluation methods can easily be compared.

3.3 Hierarchical centering

The linear approach, illustrated above, assigns centers on the basis of a proposition sequence, flattening the original hierarchy and ignoring nucleus-satellite relations. This means, for example, that in a text of two paragraphs, proposition $U_{2.1}$ (the first proposition in the second paragraph) has to be treated as the successor to $U_{1.N}$ (the final proposition of the first paragraph), even if $U_{1.N}$ is relatively insignificant (the satellite of a satellite, perhaps). One's intuition in such cases is that some more significant proposition in the first paragraph should become the focus against which continuity of reference in the second paragraph is judged.

Veins Theory (Cristea et al 1998) provides a possible formalization of this intuition, in which some earlier propositions become inaccessible as a rhetorical boundary is crossed. The theory could be applied to centering in various ways; we have implemented perhaps the simplest approach, in which centering transitions are assessed in relation to the nearest *accessible* predecessor. In many cases the linear and hierarchical definitions give the same result, but sometimes they diverge, as in the following alternative to solutions A and B:

C. $ban(fda, elixir)$ since $contain(elixir, gestodene)$.
However, $approve(fda, elixirplus)$.

Following Veins Theory, the predecessor of $approve(fda, elixirplus)$ is $ban(fda, elixir)$; its linear predecessor $contain(elixir, gestodene)$ (an embedded satellite) is inaccessible. This makes a considerable difference: under a hierarchical approach, fda can be the Cb of the final proposition; under a linear approach, this proposition has no Cb.

4 Conclusion

This paper has highlighted some implications of Centering Theory for planning coherent text. We show that by making some assumptions about which entities are potential Cps, we can determine Cbs, Cps, and hence transitions, in the text planning stage. This allows the text planner to select the proposition sequence that yields the best continuity of reference, or to balance the goal of referential continuity against other factors. For instance, there may be a preference for Satellite to follow Nucleus for some discourse relations, even if this results in a greater number of defects according to centering considerations. There are difficulties in evaluating algorithms for specific tasks which are embedded in a generation system, since the quality of the output is limited by the functionalities of the system as a whole. In particular, the task of generating appropriate referring expressions cannot be tackled in isolation from other tasks which contribute to the coherence of a text.

The implementation of Centering reported here is a special case of text planning by constraint satisfaction, where the user has control over the different constraints, and this approach means that different strategies for e.g. clause ordering and pronominalisation can easily be compared by inspecting the resulting texts. The evaluation metrics we have presented here are provisional and are a matter for further detailed research, which our approach to text generation will facilitate.

Acknowledgements

This work was supported by the UK EPSRC under grant references L51126, L77102 (Kibble) and M36960 (Power).

References

S Brennan 1998. Centering as a Psychological Resource for Achieving Joint Reference in Spontaneous Discourse. In Walker, Joshi and Prince (eds), *Centering Theory in Discourse*, Oxford.

S Brennan, M Walker Friedman and C Pollard 1987. A Centering Approach to Pronouns. In *Proc. 25th ACL* .

H Cheng 2000. Experimenting with the Interaction between Aggregation and Text Planning, Proceedings of ANLP-NAACL 2000.

D Cristea, N Ide and L Romary, 1998. Veins Theory: A model of global discourse cohesion and coherence. In Proc COLING/ACL'98, pp 281-285, Montreal.

R Dale 1992, *Generating Referring Expressions*, MIT Press.

B Grosz, A Joshi and S Weinstein 1995, Centering: a framework for modelling the local coherence of discourse. *Computational Linguistics.*

R Kibble 1999, Cb or not Cb? Centering theory applied to NLG. *ACL workshop on Discourse and Reference Structure.*

R Kibble and R Power,1999. Using centering theory to plan coherent texts, Proceedings of the 12th Amsterdam Colloquium.

W Mann and S Thompson 1987, Rhetorical Structure Theory: A Theory of Text Organisation. In L Polanyi (ed.), *The Structure of Discourse.*

K McCoy and M Strube, 1999. Generating Anaphoric Expressions: Pronoun or Definite Description? *ACL workshop on Discourse and Reference Structure.*

V Mittal, J Moore, G Carenini and S Roth 1998, Describing Complex Charts in Natural Language: A Caption Generation System. *Computational Linguistics.*

R Power 2000. Planning Texts by Constraint Satisfaction, to appear in Proceedings of COLING 2000.

E Prince 1999. How not to mark topics: "Topicalization" in English and Yiddish. Ms, Linguistics Department, University of Pennsylvania.

E Reiter 1994. Has a consensus NL generation architecture appeared, and is it psycholinguistically plausible? In *Proc. INLG 7.*

E Reiter and R Dale 1997, Building Applied Natural-Language Generation Systems. *Journal of Natural-Language Engineering*

D Scott and C de Souza, 1990. Getting the message across in RST-based text generation. In Dale, Mellish and Zock (eds), *Current Research in Natural Language Generation*, Academic Press.

M Strube and U Hahn 1999, Functional Centering - Grounding Referential Coherence in Information Structure. *Computational Linguistics.*