

Centering and the Optimization of Discourse*

David Beaver, Stanford University

July 28th 2000

Abstract

In this paper the Centering model of anaphora resolution and discourse coherence (Grosz, Joshi and Weinstein, 1983) is reformulated in terms of Optimality Theory (OT) (Prince and Smolensky 1993). A first version of this reformulated model is proven to be descriptively equivalent to an earlier algorithmic statement of Centering due to Brennan, Friedman and Pollard (1987). However, the new model is stated declaratively, and makes clearer the status of the various constraints used in the theory. In the second part of the paper, the model is extended in various ways, demonstrating the advantages of the OT reformulation. First to be considered are alternative versions of the constraints on topic-hood (how the “backward-looking center” is identified) and salience (the definition of the “forward-looking center list”). Then, after relating the model to recent proposals in OT semantics and pragmatics (Blutner 2000, de Hoop and Hendriks 2000), three new applications are described. It is shown how the theory can be applied to natural language generation, to the evaluation/optimization of complete texts, and to the interpretation of accented pronouns.

*The central idea in this paper (of re-interpreting the Centering transition classification schemas as ranked OT constraints) was first presented at the Tenth European Summer School in Logic, Language and Information at Saarbrücken in August 1998, and was also presented in talks at the Stanford CSLI Workshop in Logic, Language and Computation and the Amsterdam Colloquium in 1999. The meat of the current paper was first presented at the Stanford University Semantics Fest, in March 2000. I am grateful for feedback from those present at all occasions of presentation, and to Brady Clark, Edward Flemming, Barbara Grosz, Beth Levin, Peter Sells, Maria Wolters and Henk Zeevat.

Overview

In the last twenty years, the fields of formal semantics and pragmatics have seen a great deal of research on the interpretation of anaphora. Of particular note is the development of dynamic approaches to meaning [KR93, Hei82, GS91]. Yet there has been a curious near absence of work within this tradition on anaphora *resolution*: models have tended to concentrate on absolute semantic constraints on what can be anaphoric to what, rather than to build up detailed pictures of which discourse entities are salient, and hence likely to be referred to, at which times.

There is a separate strong tradition of work on anaphora resolution [GJW83, Sid83, GS86, Gun98]. This work is similarly dynamic, in that the core of these models is an account of the impact of an utterance on the information state of conversational participants. With one or two notable exceptions (e.g. [Rob98]), work on anaphora resolution has received far more attention in the Natural Language Processing and psycholinguistics communities than from formal semanticists and pragmaticists.

The current paper is in part of an effort to bridge the gap between these separate communities. Ideally, a psychologically and computationally motivated theory of anaphora resolution should be developed that is empirically motivated, formally precise, makes clear the relationship between anaphora resolution and other linguistic phenomenon, and which draws on tools from mainstream linguistic theory. This is an ambitious goal, and the current paper is restricted to practical sub-goals. I concentrate on reformulating an existing theory of anaphora resolution, Centering, and then demonstrating some of the benefits of the reformulation.

Centering [GJW83, GJW95] is intended to model discourse coherence, inference by conversational participants, and anaphora resolution. A quite separate line of linguistic research has produced Optimality Theory (OT), a formal framework for reasoning about combinations of linearly ranked linguistic constraints [PS93]. OT has been extraordinarily influential in phonology, has made a significant impact on some areas of syntax (e.g. formal models of typology), and, of particular relevance to the current enterprise, has recently started to make inroads into syntax, semantics and pragmatics.¹ I will present a restatement and development of the Centering model in Optimality Theory.

Part I of the paper “Centering in OT” begins with a detailed statement of

¹For OT work on semantics and pragmatics, see [Blu00a, BJ99, DvR, HdHar, dH00, dHdS98, vdDdH98, Zee99, Zee00, Lan00].

a standard variant of Centering. It is not my purpose to motivate Centering here, for which readers are referred to the original papers. Having stated the archetype theory, I describe the OT reformulation, and then demonstrate the application of the resulting system, COT, with respect to examples. This part of the paper ends with a section making formally precise the sense in which COT is a reformulation rather than a descriptively new theory.

While part I is largely conservative with respect to the existing theory of Centering, part II, “New Directions in Discourse Optimization” suggests innovations to the model, and ways in which it might be integrated within a wider model of discourse processing and conversational inference. Relevant ideas drawn from other recent OT-based theories of the semantics-pragmatics interface as well as from other areas of linguistic theory, are described, and compared with COT, and on this basis several extensions to COT are proposed. These include applications to the evaluation of complete texts, to text generation, and to the interpretation of stressed pronouns. The paper ends with suggestions for further research, and discussion of how the developments in the paper relate to the original goals of Centering theory.

Part I

Centering in OT

1 Introduction to Centering Theory

The early articles in [WJP98] provide a good introduction to the theory of Centering. The theory resulted from the fusion of two lines of thought. On the maternal side, it incorporates ideas from Grosz and Sidner’s work on anaphora resolution and discourse coherence, work which has appeared in their models of local and global discourse structure [Sid83, GS86]. The paternal line, from which the framework’s name descends, includes work on inference in discourse by Joshi and associates [JK79, JW81]. The first published paper drawing together these lines of thought was [GJW83], and a more extended presentation of the framework did not appear in print until [GJW95].

The original architects of the theory stated it at an abstract and general level. Presumably this reflected both the breadth of application intended for the theory, and the fact that it drew together considerably different lines of thought. However, this abstractness, and the fact that the definitive [GJW95] did not appear for so long, has allowed considerable room for

interpretation by other authors, and in turn this has led to occasional unclarity about both the form and content of the theory. However, we may sum up the main themes of Centering as follows:

1. The attentional state of language users evolves dynamically through production or comprehension of a discourse, on a sentence by sentence basis.
2. Attentional state is related to ease of inference: certain inferences associated with salient entities are made more easily than comparable inferences unrelated to salient entities.
3. The way in which the attentional state changes may be classified into a small number of transition types.
4. Coherence of a discourse is dependent on the attentional transitions made in processing that discourse. In particular, the most coherent discourses will involve a steady evolution of participant's attentional state, rather than rapid change.
5. One crucial aspect of the attentional state is the discourse entities under discussion, or *centers* of attention.
6. By considering all ways in which linguistic form may relate to the centers of attention, and trying to maximize coherence of the discourse, we may make predictions about when anaphoric expressions should be used, and how anaphora is resolved.

The model presented in [BFP87] (BFP) cashes out some of these themes with sufficient precision to produce a predictive model of anaphora resolution. The model provides much of the groundwork for the reformulation I will propose, although it differs in a number of crucial respects. First, BFP, as with all other existing presentations of Centering, is intrinsically procedural. The model I will propose is stated declaratively, although it has a decision procedure. The declarative statement of the theory does not deny the dynamic nature of Centering, but abstracts away from any particular algorithm or heuristics that might be used in an implementation. Second, the models differ in the degree to which linguistic generalizations about anaphora resolution are integrated into a single level of description. Whereas the model I will propose is in this sense highly integrated, in BFP the generalizations of centering are stated at a number of levels, some as absolute constraints on reference, some as transition specifications, and some

as preferences between transitions. In the model I will propose transitions are no longer a core part of the theory, although they may still be identified epiphenomenally.

In [Kam98], the late Megumi Kameyama also suggested a system of defeasible constraints that would make the transitions epiphenomenal. Her work, like BFP, is an inspiration behind the current one. I will incorporate her suggestion, but (a) use a different set of constraints and (b) make the model more predictive by stating a constraint ranking. The resulting system allows easy calculation of anaphoric resolution preferences, thus greatly clarifying the work which she began. Despite the relevance of Kameyama’s model, it is the more formal BFP model that provides the basis of COT, and the remainder of this section will be taken up with a description of BFP.

In the Centering model, a sentence provides a mapping from an input information state to an output state. However, the output state, at least those aspects of it described by Centering, does not capture the meaning of the sentence. Rather, the state represents the sentence’s anaphoric potential, and, in particular, captures the relative salience of various discourse entities. A state is a simple data structure comprising: a *backward-looking center* (name of a single discourse referent), and a *forward-looking center list* (list of referent names). The backward center is a link with the previous sentence: it is the most significant discourse entity under discussion in both the current and previous sentences. C_B^n is the backward-looking center of the n-th sentence of a given discourse. The forward-looking center list, notated C_F^n for the n-th sentence, is a list of all the discourse entities in a sentence. In the BFP model, this list is ordered according to argument role, using the standard hierarchy, sometimes referred to as *grammatical obliqueness*. The *subject* is the least oblique argument, and becomes the first element of the forward-looking center list. By virtue of this privileged position, it is also termed the *preferred center*. The remainder of the forward-looking center list consists of the *direct object*, then *indirect objects*, and then *adjuncts*.

In standard centering there are three transition types, *continue*, *retain* and *shift*, and in BFP the latter is itself broken down into two subtypes.

Continuing is when the backward-looking center is unchanged ($C_B^{n-1} = C_B^n$), and is also the preferred center of the new sentence ($C_B^n = C_P^n$).

Retaining means the backward-looking center is unchanged ($C_B^{n-1} = C_B^n$), but is no longer in preferred position ($C_B^n \neq C_P^n$), signaling that a shift is likely to occur in the following sentence.

Shifting is what happens when the new backward-looking center is different

from the old ($C_B^{n-1} \neq C_B^n$). If the backward-looking center is the same as the preferred center ($C_B^n = C_P^n$), the transition is known as a **smooth** shift. If the backward-looking center is different from the preferred center ($C_B^n \neq C_P^n$), what results is a **rough** shift.²

When resolving anaphora, different analyses may correspond to different transition types. A ranking is given over the different transitions: analysis involving least change (and the least hint of coming change) is preferred. Thus, continuing is favored over retaining, which is preferred over a smooth shift, which in turn is preferred over a rough shift.

The process of anaphora resolution is based on a four stage algorithm which involves firstly generating alternative resolutions, then pruning out those resolutions that conflict with certain absolute constraints, and then applying the transition ranking. In more detail, although still glossing over some of the details, the BFP algorithm runs as follows:

Construct The alternative possibilities for anaphoric resolution are constructed. Each possibility maps all the pronouns in the sentence to discourse entities in such a way as to respect agreement features. For each possibility, C_F^n consists of all the referents of NPs in the sentence, and C_B^n is chosen from C_F^{n-1} , or is chosen to be NIL. A NIL backward-looking center means that there is no link to a previous sentence, as, for example, in an initial sentence of a discourse.

Filter Possibilities are discarded unless all of the following conditions are met:

1. If there are pronouns in the current sentence, then one of them refers to the backward-looking center of the current sentence;
2. The backward-looking center is mapped onto the entity mentioned in the current sentence which is highest ranked in the previous sentence's forward-looking center list;
3. Syntactic coreference constraints are upheld.

The first of these is what is known in the Centering literature, following [GJW95], as *Rule 1*.³

²BFP use the terminology *shift* for a smooth shift and *shift-1* for a rough shift, but the texture-based terminology is now more common.

³Rule 2 in [GJW95] is the preference relation over transitions.

Classify Classify each possibility as one of the four transition types using the criteria above.

Select Choose the best possibility, using the ranking over transition types.

Consider the treatment of the third sentence in the following example, where the second sentence is assumed to be already interpreted with the resolution indicated using indices:

- (1) a. Jane_{*i*} likes Mary_{*j*}.
 b. She_{*i*} often brings her_{*j*} flowers.
 c. She chats to the young woman for ages.

For this example the forward-looking center list from the second sentence C_F^2 will be $\langle \text{Jane, Mary, flowers} \rangle$.

Construct Agreement facts prohibit “she” or “the young woman” referring to flowers, so the only possibilities constructed involve each of these expressions referring to Jane or Mary. This results in the 16 possibilities for the pair $\langle C_B^3, C_F^3 \rangle$ shown in table (1).⁴

Filter Items 5–8 and 13–16 fail the first filter, since the backward-looking center is not pronominalized despite the presence of a pronoun. Items 3–8, 10, 11 and 13–16 fail the second filter, since C_B^n is not the most salient item mentioned, and items 9–16 fail the third, which prohibits argument coreference. This leaves us with items 1 and 2.

Classify Item 1 is classified as *continuing*, since $C_B^{n-1} = C_B^n = C_P^n$. Item 2 is classified as *retaining* since $C_B^{n-1} = C_B^n \neq C_P^n$.

Select Candidate 1 wins out, since continuations are ranked higher than retentions. So it is predicted that “she” refers to Jane, and “the young woman” to Mary.

Before moving on to consider the COT reformulation of centering, I would like to point out some peculiarities and shortcomings of BFP.

First, the algorithm takes a decidedly parsing/interpretation oriented perspective. Although application of BFP to generation is discussed briefly

⁴I ignore the referent for “ages”. For convenience, in table (1) the (new) preferred center is underlined, and the final column of the table indicates which filters are broken by each possibility.

	C_B^{1c}	C_F^{1c}	Filters
1	Jane	$\langle \underline{\text{Jane}}, \text{Mary} \rangle$	
2	Jane	$\langle \underline{\text{Mary}}, \text{Jane} \rangle$	
3	Mary	$\langle \underline{\text{Jane}}, \text{Mary} \rangle$	2
4	Mary	$\langle \underline{\text{Mary}}, \text{Jane} \rangle$	2
5	NIL	$\langle \underline{\text{Jane}}, \text{Mary} \rangle$	1,2
6	NIL	$\langle \underline{\text{Mary}}, \text{Jane} \rangle$	1,2
7	flowers	$\langle \underline{\text{Jane}}, \text{Mary} \rangle$	1,2
8	flowers	$\langle \underline{\text{Mary}}, \text{Jane} \rangle$	1,2
9	Jane	$\langle \underline{\text{Jane}}, \text{Jane} \rangle$	3
10	Jane	$\langle \underline{\text{Mary}}, \text{Mary} \rangle$	3
11	Mary	$\langle \underline{\text{Jane}}, \text{Jane} \rangle$	2,3
12	Mary	$\langle \underline{\text{Mary}}, \text{Mary} \rangle$	2,3
13	NIL	$\langle \underline{\text{Jane}}, \text{Jane} \rangle$	1,2,3
14	NIL	$\langle \underline{\text{Mary}}, \text{Mary} \rangle$	1,2,3
15	flowers	$\langle \underline{\text{Jane}}, \text{Jane} \rangle$	1,2,3
16	flowers	$\langle \underline{\text{Mary}}, \text{Mary} \rangle$	1,2,3

Table 1: Possible resolutions of example (1)c

in the original paper, the algorithm as described above is not reversible in any obvious way.⁵

Second, the algorithm seems to need two sentences to ‘warm up’, before it really gets into gear for the remainder of the discourse. For any discourse, C_B^1 will be NIL. Since NIL is neither the backward-looking center of any previous sentence, nor the referent of the subject of the first sentence, presumably the first sentence will be classified as a rough shift. Normally, C_B^2 will be the referent of some anaphor in the second sentence. Since this anaphor will pick out something referred to in the first sentence, and not NIL, it follows that $C_B^1 \neq C_B^2$. Thus the second sentence will also typically be a shift of some sort.⁶ For instance, the first two sentences of the following example are analyzed as rough shifts:

- (2) a. Mary is happy.
 $C_B^1 = \text{NIL}$, $C_F^1 = \langle \text{Mary} \rangle$, *rough shift*.
- b. Jane just gave her a book.
 $C_B^2 = \text{Mary}$, $C_F^2 = \langle \text{Jane}, \text{Mary} \rangle$, *rough shift*.
- c. She loves to read.
 $C_B^3 = \text{Mary}$, $C_F^3 = \langle \text{Mary} \rangle$, *continue*

Given that the second sentence is a shift, and that, apart from syntactic agreement, the main mechanism for choosing anaphoric antecedents in BFP is the preference for avoiding shifting, it follows that BFP will make relatively weak predictions about anaphora resolution in the second sentence of a discourse. Some examples may clarify:⁷

⁵But see [Kib00] for an attempt at basing a generation component on the BFP model.

⁶Note that if there were a choice of anaphoric and non-anaphoric readings, a possibility not made explicit in BFP, then the preference for maintaining a constant C_B might actually cause non-anaphoric readings to be preferred over anaphoric ones, to ensure that for each sentence $C_B = \text{NIL}$, an unwelcome consequence.

⁷In examples (3)–(6), gender is varied simply as a control, to show that gender is not a decisive constraint on resolution. Note also that an alternative choice of main verb in the first sentence of each pair can produce different resolution preferences. In particular, as Beth Levin has pointed out to me, choosing “see” rather than “argue” reverses the preference. This is an interesting observation. It is arguable that in the discourse “A motorist saw a police officer. She was gesticulating wildly.”, the second sentence can be taken as being implicitly from the (visual) perspective of the motorist, thus explaining the preference for “she” to be identified with the object of the previous sentence. The argument I present in the main text goes through independently of these complications.

- (3) a. A motorist was arguing with a police officer.
 $C_B^1 = \text{NIL}$, $C_F^1 = \langle \text{the motorist, the officer} \rangle$, *rough shift*.
- b. She was gesticulating wildly.
 Either (i) $C_B^2 = \text{the motorist}$, $C_F^2 = \langle \text{the motorist} \rangle$, *smooth shift*,
 or (ii) $C_B^2 = \text{the officer}$, $C_F^2 = \langle \text{the officer} \rangle$, *smooth shift*.
- (4) a. A motorist was arguing with a police officer.
 b. He was gesticulating wildly.
- (5) a. A police officer was arguing with a motorist.
 b. She was gesticulating wildly.
- (6) a. A police officer was arguing with a motorist.
 b. He was gesticulating wildly.

Regarding all of examples (3) to (6), informants have expressed a strong preference for resolution of the pronoun in the second sentence to the subject of the first sentence. However, BFP classifies both of these readings as smooth shifts in all four cases, and thus does not rank them relative to each other. So the second sentence of each discourse is predicted to be ambiguous.⁸

It is certainly arguable that BFP are right to predict that the second sentence in (3) to (6) is ambiguous, although there appears to be a clear preference for one reading over the other. What is peculiar is that the BFP model makes qualitatively different predictions about some other two sentence discourses and about three sentence discourses in general. For instance, compare (3) to (7):

- (7) a. A motorist was arguing with a police officer.
 $C_B^1 = \text{NIL}$, $C_F^1 = \langle \text{the motorist, the officer} \rangle$, *smooth shift*.
- b. She was asking her to go away, and
 $C_B^2 = \text{the motorist}$, $C_F^2 = \langle \text{the motorist, the officer} \rangle$, *smooth shift*.

⁸To date, I have only polled two colleagues on spoken discourses like (3). Both strongly agreed with my own judgments. More extensive empirical validation is clearly in order. However, this would not be germane to my point. As far as I can see, the failure of BFP to make strong predictions about the second sentence of the discourse results from a failure to fully specify how the model should work in these cases. It does not appear to have been motivated by linguistic evidence that the second sentence in a discourse is commonly ambiguous in this way. Gathering evidence to support the unmodified BFP model might be one way to repair the failing that I have identified, although my initial empirical research seems to suggest that this would be difficult.

- c. she was gesticulating wildly.
 $C_B^3 = \text{the motorist}$, $C_F^3 = \langle \text{the motorist} \rangle$, *continue*.

With two pronouns present rather than one, BFP no longer predicts ambiguity, provided one of the two is in subject position. Analyzing the subject pronoun in (7)b as referring to the police officer would result in a rough shift. But analyzing it as referring to the motorist results in a smooth shift, and so this is preferred. Likewise, the choice is forced in (7)c: resolving the single pronoun to the police officer would produce a smooth shift, whereas resolving it to the motorist is a case of continuing, and so is preferred. Empirically, this analysis of (7)b is only half-right. The reading where “she” picks out the motorist and “her” picks out the police officer in (7)b is available, and may even be preferred for a majority of speakers, but the competing reverse reading is also available. There is a preference for an interpretation of (7)c that is parallel to (7)b, although, again, there may be ambiguity.

My point, then, is this: in all of (3)b, (7)b and (7)c there is some interpretational preference, but there is also some ambiguity. There is no evidence that these are qualitatively different cases in terms of which discourses are ambiguous and which are not. However, BFP makes qualitatively different predictions about (3)b to those it makes about (7)b and (7)c. This is a shortcoming of the theory.

Next, I turn to a major limitation of BFP: the only anaphors dealt with in the published algorithm are pronouns. This, in turn makes the status of the Rule 1 filter peculiar. Rule 1 is normally taken to mitigate against using definite descriptions for C_B , and to prevent interpretation of definite descriptions as co-referential with C_B when a pronoun is also present. However, the lack of definite descriptions in BFP means that such situations do not even arise within the theory’s application domain.

What effects, then, does Rule 1 have in BFP? One effect is to filter out pathological possibilities where C_B^n is not even mentioned in the current sentence despite the presence of anaphoric links. In example (2), above, interpretations involving “NIL” and “flowers” failed both the first filter (Rule 1) and the second filter. Such examples of filtering do not seem to correlate with anything that the original architects of Centering might have had in mind as a function for Rule 1, or anything motivated by any explicit empirical study. Besides this, it is notable that *all* the readings filtered by Rule 1 in (2) would also be filtered by the second filter, whereas the reverse is not true. If this type of filtering were the only motivation, Rule 1 would be completely superfluous.

The second effect of Rule 1 relates to its originally intended function,

and can be seen in constructed examples where a proper name happens to co-refer with the previous preferred center, and agreement prevents the only pronoun in the sentence from referring to the previous preferred center. Unfortunately, such examples do not necessarily support the BFP model. Consider the final sentence of example (8):

- (8) a. Mary likes tennis.
 $C_B^1 = \text{NIL}$, $C_F^1 = \langle \text{Mary}, \text{tennis} \rangle$, *rough shift*.
- b. She plays Jim quite often.
 $C_B^1 = \text{Mary}$, $C_F^1 = \langle \text{Mary}, \text{Jim} \rangle$, *smooth shift*.
- c. He used to be Mary’s doubles partner.
 $C_B^1 = \text{Mary}$, $C_F^1 = \langle \text{Jim}, \text{another Mary} \rangle$, *smooth shift*.

According to BFP, the reading where “He” refers to Jim and “Mary” refers to the same individual named in the first sentence is predicted not to be available. It is filtered out because Mary would be the backward-looking center but not pronominalized, even though there is a pronoun present. It is inappropriate that this reading is ruled out completely, although some speakers may find the text awkward. In fact the authors of BFP mention the possibility of making Rule 1 into a preference rather than a constraint. The treatment of (8) is returned to in section 5.

The possibility of altering the status of Rule 1 brings me onto my next point: given that linguistic generalizations can be expressed at any of the algorithm’s four stages, how are we to judge where a particular generalization belongs?

Part of Rule 1 could easily have been expressed in the construction stage. The algorithm could have been altered in such that the only interpretation candidates considered map C_B^n onto some element occurring in both C_F^n and C_F^{n-1} , and only onto NIL if there was no such element. Similarly, other filtering constraints could have been expressed as construction rules, and *vice versa*. Why should a syntactic agreement test be built into the construction phase, but a syntactic co-occurrence test be built into the filtering stage?

On the other hand, the suggestion that Rule 1 be made a default essentially amounts to moving it into the Classification and Selection phase. To make Rule 1 a default in a way consistent with the general framework, it would seem that we would have to double the number of transition types. Each of the current transitions would bifurcate into one version in which Rule 1 was followed, and one in which it was not. Having thus defined eight

transition types, a linear ordering would then be defined over them. There are, in principle, $8! = 40320$ such orderings.⁹

More generally, if we have k independent binary classification constraints, we have $2^k!$ possible orderings over transition types. Adding further defeasible classification constraints to BFP produces huge numbers of possible orderings, and there is little reason to believe that this space of orderings will be useful to the linguist or the computational linguist. Rather than discussing methodological and implementational issues which this raises, I now move on to an alternative analysis in which orderings over constraints are defined directly, instead of being defined indirectly via orderings over transition types.

2 Reformulating Centering

In OT models, there are standardly two levels of representation, an input and an output. In OT syntax it is standard to assume that the input is an LF-like structure, and the output is a string. Relative to a given fixed input, the constraints are used to find the optimal output.

In the COT model, the two levels of representation are again, roughly, form and meaning. The first is a (partially) syntactically analyzed sentence, and the second is a mapping from referring NPs in the sentence to their referents. From a parsing/interpretation perspective, we take the form as input, and calculate the optimal output. This will be the form in which COT is applied throughout Part I of this paper. In section 7, in Part II, it is shown how the system can be used ‘in reverse’ to help select alternative forms on the basis of a fixed meaning. Such a generation perspective gives the standard direction of optimization in OT-syntax, and for this reason practitioners of OT-syntax may initially find some aspects of the description below confusing.

Given some input to a OT model, the constraints provide a way to select an optimal interpretation from a set of candidates. The set of candidates is

⁹In BFP, transition classification is based on two binary constraints, $C_B^n = C_B^{n-1}$, and $C_B^n = C_P^n$. If assume that Rule 1 is more important than these two classification constraints, and that the requirement that $C_B^n = C_B^{n-1}$ continues to take precedence over the requirement that $C_B^n = C_P^n$, we would be left with just one ordering over classifications that incorporated Rule 1. Mere inclusion of Rule 1 into the transition classification schema is thus not particularly difficult, although it would lead to an ordering over eight different transition types. My point is that the use of transition classification schemes, and the intuitions behind orderings over them, tends to become rapidly less transparent as the number of classification constraints rises.

assumed to be in principle unrestricted. For instance, the output candidate set in OT syntax could be all the syntactic trees defined over some set of rewrite rules, or the set of all strings over some atomic language. In practice, a given OT paper will generally only consider a set of constraints pertinent to a small group of phenomena, and the constraints required to determine other aspects of the input-output mapping are not explicit. It is thus standard to restrict the candidate set to relevant alternatives, those assumed not to be ruled out by constraints that are unrelated to the phenomena at hand. So it will be with COT: the candidate set only partially specifies the meaning of a sentence, and the only candidates that will be considered are those that seem of interest for a theory of anaphora resolution.¹⁰

Next we come to the question of how the optimal candidate is chosen. Firstly, it should be realized that while some constraints are boolean with respect to candidates, some are not. For instance, the AGREE constraint is not boolean. It is possible for two candidates both to violate AGREE, but one to involve more violations. In this case, we count one violation for each non-agreeing anaphor. Given two candidates A and B and a constraint χ , let us say that A is at least as good as B with respect to χ ($A \geq_{\chi} B$) provided A has no more violations of χ than B. Candidate A is superior to candidate B if (i) there is some constraint χ such that for each constraint χ' higher ranked than χ ($A \geq_{\chi'} B$), and (ii) A has strictly fewer violations of χ than B.

Having outlined the basic principles of OT, I now move to the reformulation of Centering. I find it convenient to make a terminological change: *topic* instead of *backward-looking center*. This inessential modification, discussion of which is postponed until section 5, suggests interesting links with a wide literature based in quite different empirical domains within linguistics. For this and the following two sections of the paper, the *topic* of a sentence is defined to be the entity referred to in both the current and the previous sentence, such that the relevant referring expression in the previous

¹⁰One issue which is not dealt with in this paper is the nature of what in OT is called GEN, the function/algorithm that creates the candidate set. I assume that GEN creates pairs of all possible forms and meanings with no further restriction. The limited sets of constraints that are considered mean that COT is only sensitive to a few select features of the forms and meanings, such as the obliqueness of arguments and identity of referents. Some forms or meanings generated may be so unlike what we expect of forms or meanings that features like obliqueness and referent identity are undefined for them. In this case, these forms/meanings are assumed to violate all relevant constraints, thus rendering them non-optimal, and irrelevant to our considerations. For example, GEN might produce a pair consisting of a certain sentence and a peanut: the peanut will be a candidate meaning, but certainly non-optimal. I will omit peanuts and other oddities from the tableaux in this paper.

sentence was minimally oblique. If there is no such entity, the topic can be anything.¹¹ Alternatives to this definition of *topic* will be considered later.

The various generalizations on which the BFP resolution algorithm is based will now be expressed using six linearly ranked constraints. Additionally, I will require that a list is maintained of which entities were referred to in the previous sentences, what the grammatical obliqueness of each referring expression was, and which was topic. No further apparatus specific to Centering is required. In this section, I will state the constraints and specify the ordering. As will become clear, all constraints are either already present in BFP in something close to the required form, or else are uncontroversial, so little motivation or further explanation will be required.

Here are the constraints, in rank order, with the top constraint being the strongest:

AGREE Anaphoric expressions agree with their antecedents in terms of number and gender.

PRINCIPLE-B Co-arguments of a predicate are disjoint.

PRO-TOP The topic is pronominalized.

FAM-DEF Each definite NP is familiar. This means both that the referent is familiar, and that no new information about the referent is provided by the definite.

COHERE The topic of the current sentence is the topic of the previous one.

ALIGN The topic is in subject position.

The ordering of most of the constraints can be related to the BFP algorithm. Suppose that two OT constraints mirror operations taking place in the BFP algorithm, and that the operation corresponding to the first constraint takes place earlier in the algorithm than the operation corresponding to the second constraint. Then the first constraint is higher ranked than the second constraint. There are two exceptions to this principle. First, FAM-DEF, as will be discussed, does not correspond directly to a BFP constraint. Second, COHERE and ALIGN both relate to a combination of the Classify and Select stages of the BFP algorithm, and their relative ordering is

¹¹To clarify, if for a discourse initial sentence, the second clause of the definition of topic is intended to apply, so the topic can be anything. This replicates the effect of C_B being set to NIL in BFP.

not determined by temporal precedence in the algorithm. Rather, the relative ranking of these two constraints with respect to each other mirrors the ranking of transition types in BFP, in a way that should be obvious to those familiar with the Centering literature.

The top two constraints, AGREE and PRINCIPLE-B reflect ideas that are familiar from the syntactic literature. They are found in the construct and filter stages of BFP, respectively. Their relative ordering is arbitrary in the current work.

PRO-TOP has essentially the effect of Centering’s Rule 1. However, the original Rule 1 includes an if-clause; “if there are pronouns in the sentence then...”. Given that the original rule was an absolute constraint, it was essential that the if-clause restricted the rule’s application. However, in COT constraints are defeasible. If there are pronouns, then PRO-TOP will function comparably to Rule 1, providing a preference for interpretations that make the topic (i.e. C_B) into a pronoun. But if there are no pronouns, then all candidate interpretations will be equally bad as far as PRO-TOP is concerned, which means that PRO-TOP will not have any effect in the final preference over candidate interpretations. Examples in the following section should clarify.¹²

One subtle aspect of OT is that combinations of default rules can have the effect of producing absolute constraints. In this particular case, the interaction of PRO-TOP and the lower ranked FAM-DEF produces the effect of Rule 1’s infeasibility. To show this, it is important first to clarify the interpretation of FAM-DEF. Significantly, the class of definites is taken to include pronouns, definite descriptions and proper names.¹³

Suppose that there is a possible interpretation where some proper name or definite description in the current sentence refers to the sentence topic. Suppose further that there are pronouns in the current sentence which refer

¹² PRO-TOP can also be seen as an instance of a cross-linguistically more general principle that topics are reduced. Bresnan [Bre99] suggests a similar constraint to PRO-TOP: “Reduced \Leftrightarrow TOP”. More generally, a rule governing the form of the topic is just a special case of the rules relating the form of NPs to their incoming and outgoing salience. Such rules, at least in as far as they relate form to incoming salience, have been developed in work on the Givenness Hierarchy [GHZ83]. I return to these issues in section 8.

¹³Note that although proper names are preferentially familiar, this will not mean that a use of “Jane” is typically taken to be anaphoric upon a previous use of “Mary”. The notion of familiarity allows for “no new information about the referent is provided by the definite.” Thus a use of “Jane” co-referential with a previous use of “Mary” would constitute a violation of FAM-DEF, just as a non-anaphoric use of “Jane” would. In effect, the constraint will cause multiple uses of “Jane” to preferentially refer to the same individual, except where higher ranked constraints say otherwise.

to discourse entities other than the topic. Then this interpretation cannot be optimal. Why? Because this reading breaks PRO-TOP and not the lower ranked FAM-DEF. But there must be alternative interpretations which break FAM-DEF by allowing the proper name or definite to refer to a novel entity. All these alternatives are such that the topic can be identified with the referent of some pronoun, so they do not conflict with PRO-TOP. Thus they are preferred to the original interpretation which did conflict with PRO-TOP. We will see an example of this reasoning shortly.

The last two constraints, COHERE and ALIGN are just the classification constraints used in BFP to specify transition types. COHERE says that we prefer not to change topic.¹⁴ ALIGN literally requires the topic to be subject, but for canonical English sentences this is equivalent to saying that the topic is the preferred center of the current sentence. In fact, in as much as it is appropriate to identify the notions of *backward-looking center* and topic, ALIGN is independently motivated in recent literature on OT-syntax.¹⁵

The authors of BFP, although they do not share my terminology, make it clear that COHERE is more important than ALIGN. Where we differ is that in BFP the relative ranking of these constraints is stated indirectly, as an ordering over transitions, whereas in COT the ranking is stated directly. More generally, in COT all constraints are ranked directly.

If we wanted to expand BFP to include k defeasible constraints we would have to decide between $2^k!$ transition rankings. In COT, the number of rankings for a given k is $k!$, and increases much more slowly than in BFP. This, of course, is no argument for COT being *a priori* a better model than BFP, or *vice versa*.¹⁶ But it may be suggestive of why, from personal experience, I

¹⁴The presence of this constraint requires that an interlocutor's *information state* determines what the topic of the previous sentence was, but I will not explicitly define notions of information state or *update* in this paper. The dynamic model of anaphora resolution in [Bea99a] does make explicit a notion of information state — the states used there would have to be augmented with a *topic register* if an interface between COT and Dynamic Semantics were to be developed.

¹⁵For instance, see the treatment of Swedish in [Sel00]. Note that Sells uses a combination of two constraints, one to say that the topic is left aligned in the clause, and another to say that the subject is left aligned. It is only in *canonical* sentences that these produce the same effect as the single ALIGN constraint used here. My ALIGN is sufficient for demonstrating the COT framework, but further work ought to explore the use of constraint combinations to model effects of non-canonical word order on coherence and anaphora resolution. Also note that in other work [Sel99], Sells uses *prominence* relationships which mirror the use of the forward-looking center list in centering. In the basic version of COT, these prominence relationships are built into the definition of topic, but I take this to be provisional. See also Aissen's use of scales, e.g. [Ais99].

¹⁶In the absence of more restrictions on what a constraint can be like, or of how many

find direct rankings over constraints easier to work with than rankings over transitions.

3 Application of COT

In this section, COT is applied to a range of examples. Interpretations of sentences are compared using a tableau method which is standard in OT. To begin, consider the following discourse:

- (9) a. Jane_{*i*} likes Mary_{*j*}.
 b. She_{*k*} often goes around for tea_{*l*}.
 c. The woman_{*m*} is a compulsive tea drinker.

In most cases, I will not explicitly detail how the topic is calculated. But for the first sentence of (9)a, I will show the calculation. Since there is no previous sentence, by definition the topic can be anything. Intuitively, there are three relevant possibilities, that the topic, T, is the referent of “Jane_{*i*}” (notated $T = i$), that the topic is the referent of “Mary_{*j*}”, ($T = j$), or that the topic is some other entity ($T \notin \{i, j\}$). We can represent these three possibilities in the following tableau:

(10)

Example (9)a	AGREE	PRINC-B	PRO-TOP	FAM-DEF	COHERE	ALIGN
$T = i$			*	**	*	
$T = j$			*	**	*	*
$T \notin \{i, j\}$			*	**	*	*

In the tableau, the top row lists the input, which here is identified by the relevant example number, and then the constraints in rank order, with the strongest constraint on the left. Each of the following rows details the behavior of one candidate interpretation with respect to the constraints, each star marking a constraint violation. The best of the candidates under

constraints are allowed, we cannot even say that one of the two models is more expressive than the other.

consideration is found by looking at each constraint column from the left until finding a column in which one candidate has fewer violations than any other. This is then the optimal candidate, and is dignified with a “☞”.

In the case of (9)a, all three candidates trivially violate several constraints: they violate PRO-TOP since there are no pronouns, so whatever is topic it will not be pronominalized; they violate FAM-DEF twice since there is no previous discourse which could make the two proper names familiar in the required sense; and they violate COHERE, again since there is no previous sentence, and thus no commonality of topic with the previous sentence. However, two of the candidates additionally violate ALIGN. The only candidate which does not is the first, where the topic is identified with the referent of the subject NP.¹⁷ Hence, the first candidate is optimal.

For the second sentence, (9)b, simple reasoning shows that the optimal candidate must be one where the topic is the referent of the subject pronoun. Thus I will only compare candidates where this condition is met, and not explicitly indicate the referent of the topic. The three candidate resolutions for the pronoun considered will be: $k = i$, $k = j$, and $k \notin \{i, j\}$ (i.e. the pronoun resolves to Jane, Mary, and something else, respectively). As seen in the following tableau, only the first candidate, satisfies COHERE, so the pronoun resolves to Jane. Note that this is a case where BFP fails to choose between the first and second candidates, due to the fact that both would be classified as shift transitions.

(11)

		AGREE	PRINC-B	PRO-TOP	FAM-DEF	COHERE	ALIGN
☞	$k = i$						
	$k = j$					*	
	$k \notin \{i, j\}$				*	*	


The third sentence of (9) is one involving a definite description and no pronoun. It is important to realize that I have not imposed any requirement

¹⁷Note that this reasoning crucially relies on the presence of a referring expression in subject position. As the model currently stands, an expletive in grammatical subject position would lead to no preference at all regarding the topic. This may be problematic, but could be dealt with in a number of ways, for instance by modifying the ALIGN constraint to say not that the topic is the subject, but that the topic is minimally oblique.

that the definite description actually does describe what it refers to: presumably in a more complete model this would be a high ranking constraint. But the matter is more complex. A full treatment of definites would involve consideration of the extent to which the evolving common ground establishes what a description refers to, and consideration of the ease with which information not yet established can be *accommodated* (in the sense of [Lew79]). These issues, although crucial, go beyond what is standardly discussed under the rubric of Centering, and beyond what I aim to achieve in this paper. See [Blu00a, Zee99] for discussion of accommodation in an OT framework.

There are three relevant resolution possibilities for the definite NP “The woman_{*l*}” in (9)c, $m = i$, $m = l$, and finally $m \notin \{i, l\}$. PRO-TOP fails for all candidate interpretations, so this constraint does not end up affecting resolution. The $m \notin \{i, l\}$ candidate where the definite is not anaphoric fails on two additional counts, FAM-DEF and COHERE. The candidate mapping “The woman_{*l*}” onto tea violates the same COT constraints as the $m \notin \{i, l\}$ candidate, although here FAM-DEF fails not because the referent is new, but, somewhat bizarrely, because “The woman_{*l*}” is not already established to be tea.¹⁸ The $m = i$ candidate, where the definite refers to Jane, involves a familiar reference for the definite and continuity of topic, so this candidate is optimal:

(12)

	AGREE	PRINC-B	PRO-TOP	FAM-DEF	COHERE	ALIGN
Example (9)c						
 $m = i$			*			
$m = l$			*	*	*	
$m \notin \{i, l\}$			*	*	*	

(9)c illustrates the fact that PRO-TOP, although it is closely related to Rule 1, does not require the extra if-clause “if there are pronouns in the sentence then...”. In case there are no pronouns, PRO-TOP simply becomes irrelevant to the choice of candidate.¹⁹

¹⁸Arguably “The woman_{*l*}” also differs from “tea_{*l*}” in grammatical gender, so that the second candidate in (12) also violates AGREE. The question of whether non-pronominal anaphors in English have grammatical or semantic gender is not tackled in the current paper.

¹⁹Example (9)c is what Centering would classify as a *continuation*, although, from a

In various ways, the treatment of (9)a–c illustrate minor departures from BFP, concerning the topic of the first sentence of a discourse, the resolution preferences in the second sentence, and the treatment of definite descriptions. The following examples are intended to illustrate commonalities between COT and BFP, although in some cases the presence of definite descriptions does mean that the examples go beyond what is strictly dealt with by the BFP model. In each case, the first and second sentences are assumed to have been processed, resulting in the anaphoric relationships indicated by co-indexation.

Examples (13)c and (14)c both involve a pronoun in subject position that can agree with the previous subject. A theory which required that parallelism be maximized would presumably resolve the subject pronoun to the referent of the previous subject in both cases. However, parallelism is not a deciding factor *per se* in COT or BFP, and neither is subjecthood of the antecedent. It happens that according to both models, in (13)c the antecedent is the previous subject. But both models agree that this is not the case for (14)c.

- (13) a. Jane_i likes Mary_j.
 b. She_i often goes around for tea with her_j.
 c. She_k chats to the young woman_l for ages.
- (14) a. Jane_i is happy.
 b. Mary_j gave her_i a present_k.
 c. She_l smiled.

The tableau for (13)c is shown in (15). The first two candidates are the obvious two alternative resolutions, and an extra possibility has been included simply to illustrate the effect of PRINCIPLE-B. As can be seen, the high ranking of this constraint means that the third candidate in the table, a reading in which co-arguments co-refer, is far from optimal. The second candidate, in which a definite description is resolved to the previous

formal point of view, the BFP algorithm does not cover this particular case: as indicated BFP does not include any explicit treatment of definite descriptions. I have chosen to include some treatment of definites in part because it allows me to provide examples that illustrate the effects of Rule 1/PRO-TOP more transparently than do examples not involving definite descriptions.

sentence’s subject, produces multiple violations: the topic is not pronominalized, and is not aligned with the subject. In contrast, the parallel-subject reading does not violate any constraints, and is selected.

(15)

	AGREE	PRINC-B	PRO-TOP	FAM-DEF	COHERE	ALIGN
Example (13)c						
☞ k=i, l=j						
k=j, l=i			*			*
k=i, l=i		*				

In (16) a tableau for (14)c is presented. Here the two candidates included are one in which the single anaphor co-refers with the previous subject, which was not topic, and one in which the anaphor co-refers with the previous direct object, which was topic. The first of these produces a conflict with COHERE, and is ruled out.

(16)

	AGREE	PRINC-B	PRO-TOP	FAM-DEF	COHERE	ALIGN
Example (14)c						
☞ l=j					*	
☞ l=i						

As noted, BFP agrees with COT in cases like (13)c and (14)c. Modulo the absence of definite descriptions in BFP, it is clear that both examples would be classified as continuations, where the preferred reading is the one in which the topic/ C_B remains constant.

The next example, (17)c, in which the only pronoun is not in subject position, is one that BFP would classify as *retaining*. As shown in the immediately following tableau, COT duplicates this result. Any anaphoric reading, i.e. any reading in which NPs in (17)c co-refer with elements in (17)b, will conflict with ALIGN, and the preferred reading is the only one which maintains constant topic.

(17) a. Jane_i is happy.

- b. She_{*i*} was congratulated by Freda_{*j*},
- c. and Mary_{*k*} gave her_{*l*} a present_{*m*}.

(18)

		AGREE	PRINC-B	PRO-TOP	FAM-DEF	COHERE	ALIGN
Example (17)c							
☞	$l=i$						*
	$l=j$					*	*

The next two cases illustrate BFP shifts. The first, (19) involves a smooth shift in its third sentence, and the second, (21), involves a rough shift in its third sentence. In both cases COT predicts the same shift of topic/ C_B as BFP. In the first case, the topic shifts because the only interpretations in which both pronouns gain anaphoric readings involve a reference to an entity that was in subject position in the previous sentence but non-topical there. In this case, shown in the tableau in (20), ALIGN comes into play to determine the optimal choice. In the second case, (21)c, the only candidates which satisfy COHERE and ALIGN would violate other higher ranked constraints. In the tableau, (22), only the correct reading and another candidate violating AGREE are shown.²⁰

- (19)
- a. Jane_{*i*} is happy.
 - b. Mary_{*j*} gave her_{*i*} a present_{*k*}.
 - c. She_{*l*} smiled at her_{*m*}.

(20)

		AGREE	PRINC-B	PRO-TOP	FAM-DEF	COHERE	ALIGN
Example (19)c							
☞	$l=i, m=j$					*	
	$l=j, m=i$					*	*

²⁰I have implicitly restricted the candidate set to rule out interpretations where indefinite NPs are anaphoric. This could, of course, have been stated as a further ranked constraint.

- (21) a. Jane_i is happy.
 b. Mary_j gave her_i a present_k.
 c. Somebody_k unwrapped it_l.

(22)

Example (21)c	AGREE	PRINC-B	PRO-TOP	FAM-DEF	COHERE	ALIGN
l=i	*					*
☞ l=k					*	*

The final example in this section illustrates how the high ranking of PRO-TOP mirrors the centering stipulation that the backward-looking center be pronominalized if there are any pronouns at all. Consider an interpretation of (23)c in which “Fred_l” is co-referential with its use in the previous sentence, and “her_m” picked out Jane. In BFP, Fred would be C_B, a conflict with Rule 1 would arise, and this interpretation would be filtered out. In COT, such an interpretation would violate PRO-TOP, as shown in the first row in (24). On the other hand, readings of the sentence in which successive uses of “Fred” pick out different individuals violate FAM-DEF. But this constraint is lower ranked than PRO-TOP, so such readings are preferred to those in which PRO-TOP is violated. Similar argumentation would apply to the variant in (23)c’, although here the use of a definite description goes beyond the BFP fragment.

- (23) a. Jane_i is happy.
 b. Fred_j gave her_i a present_k.
 c. Fred_l amused her_m.
 c’. The young man_l amused her_m.

(24)

	AGREE	PRINC-B	PRO-TOP	FAM-DEF	COHERE	ALIGN
Example (23)c						
$l=j, m=i$			*		*	
$l \notin \{i,j,k\}, m=j$				*		*
$l=j, m \notin \{i,j,k\}$			*	*	*	
$l,m \notin \{i,j,k\}$				**	*	

In effect, what COT predicts for (23)c (or (23)c') is a peculiar case of a retaining transition. I have my doubts about whether this prediction is correct, and in section 5 I consider a modification to the model which would produce different results. Whether BFP predicts a retaining transition, or whether it simply predicts no output at all, would depend on what alternatives are allowed for the interpretation of the proper name “Fred”. If multiple alternatives are allowed (they are not explicitly disallowed in [BFP87]), then the predictions will be the same as for this first version of COT.

At this point, the reader might wonder what exactly is the relationship between COT and BFP? They are, in fact, equivalent in a strong, formal sense.

4 Equivalence of COT and BFP

It will now be shown that BFP and COT make identical predictions about anaphora resolution. The proof is restricted to what I take to be the domain of effective application of BFP. I thus exclude non-pronominal anaphora, and exclude the first two sentences of a discourse.²¹

The proof is in three parts. The first two parts concern the fact that both BFP and COT rule out *pathological* interpretations, i.e. those breaking syntactic constraints or Centering’s Rule 1. The third part concerns preferences between resolutions in non-pathological cases. Each part of the proof corresponds to the demonstration of one proposition, with a fourth proposition combining these into a more general result.

²¹The proof in fact could be extended to cover the second sentence of a discourse if one minor change were made to BFP. If C_B^1 were taken to be the referent of the subject of the first sentence, rather than NIL, then I believe the predictions of BFP would not only be improved, but also would agree with those of COT. The first sentence of a text, by an implicit but reasonable assumption in the BFP model, contains no sentence external anaphora, and so is not relevant.

Proposition 1 *Resolutions breaking syntactic constraints are never COT optimal, and never correspond to preferred BFP transitions.*²²

Suppose that a resolution R in COT conflicts with at least one of AGR or PRINCIPLE-B. In such a case there are guaranteed to be alternative candidate interpretations which satisfy both of these constraints. To see this, observe that by breaking FAM-DEF repeatedly in some other candidate interpretation, we can ensure that no definite NPs are interpreted as anaphoric and that none corefer. AGR and PRINCIPLE-B are then satisfied trivially.

In BFP, interpretations failing agreement constraints are not produced in the *construction* phase of the algorithm, and interpretations failing syntactic co-reference constraints are removed in the *filter* stage. In either case, such interpretations are ruled out before transition-based preferences between interpretations are even considered. So R will also not be the interpretation predicted by BFP. The reverse argument, that if syntactic violations rule a candidate out in BFP, it will also not be the optimal COT candidate, is similar.

Proposition 2 *If a candidate is syntactically acceptable, then violation of PRO-TOP ensures it is not COT optimal iff Rule 1 prevents it from being classified as a BFP preferred transition.*²³

If there are no pronouns in the sentence, then all candidates will fail PRO-TOP, so the other constraints will select the optimal candidate. Also, in this case Rule 1, which is conditional on the presence of pronouns, will not filter any interpretations, and other aspects of the model will select the preferred interpretation. The first and third parts of the proof determine in such cases that COT and BFP make identical predictions on any anaphoric readings of the sentence.

If there are pronouns, then we can follow similar reasoning to the first part of the proof. If some candidate R violates PRO-TOP, then we can show that it will not be COT optimal, because there will be other interpretations similar with respect to the higher ranked constraints but satisfying PRO-TOP. One way to find such a candidate is by considering non-anaphoric readings obtained by violating FAM-DEF. Recall that the definition of topic says that if there is no entity commonly referred to in both the current and

²²The impact of Proposition 1 in COT is seen in tableaux which include candidates violating syntactic constraints, (15) and (22)

²³All tableaux in the preceding section including one or more candidates violating PRO-TOP exemplifying the COT side of Proposition 2.

Interpretation	COHERE	ALIGN	BFP
i	A		continuation
	B	*	retain
ii	A		continuation
	B	*	smooth
iii	A		continuation
	B	*	rough
iv	A		retain
	B	*	smooth
v	A		retain
	B	*	rough
vi	A	*	smooth
	B	*	rough

Table 2: Equivalence between transitions and constraints

previous sentences, the topic can be anything. If there is no anaphora, we can arbitrarily choose the topic to be the (non-anaphoric) referent of some pronoun in the current sentence. In this case PRO-TOP is satisfied. So R will not be the optimal candidate.

Still under the assumption that there are pronouns, a resolution violates PRO-TOP iff it also violates Rule 1. So R will be filtered in the BFP model, and will not be the preferred interpretation. So, if there are pronouns (i) if R violates PRO-TOP then R will not be selected in either model, and (ii) if R violates Rule 1 then R will not be selected in either model.

Proposition 3 *Suppose two resolutions A and B satisfy AGREE, PRINCIPLE-B and PRO-TOP, and have identical clashes (possibly none) with FAM-DEF. Then COT ranks A above B iff BFP ranks A above B.*²⁴

To see this, consider table (2), which shows how certain interpretations A and B might fare with respect to the COT constraints COHERE and ALIGN, and what transitions these would correspond to in the BFP taxonomy.

The table lists all possible combinations of clashes with COHERE and ALIGN such that A would be COT-preferred to B. For each of these, the

²⁴The COT side of Proposition 3 is illustrated by (15) and (16) (corresponding to BFP continuing transitions), (18) (retain), (20) (smooth shift) and (22) (rough shift).

corresponding BFP transition is uniquely determined, and in each case the A transition out-ranks the B transition. This demonstrates the proposition from left to right. The reverse direction is similarly simple. The BFP column lists all the transition pairs such that A out-ranks B. For each of these the pattern of COT clashes with COHERE and ALIGN is uniquely determined, and in each case A would be the COT preferred candidate. This demonstrates the proposition from right to left.

In combination with the first two parts of the proof, a more general claim, but one in which I have made the relevant caveats explicit, follows:

Proposition 4 *Given a non-initial, non-second sentence in which the only referring expressions are proper nouns and pronouns, if either COT or BFP predicts an interpretation involving anaphoric interpretation of all pronouns, then both do, and in this case they predict the same interpretation.*

Part II

New Directions in Discourse Optimization

5 Topic and Salience in COT

So far I have attempted to remain descriptively faithful to a standard variant of Centering theory. In this section I will review how the model might be improved with respect to the definition of topic. Currently, the topic is defined as the entity referred to in both the current and the previous sentence, such that the relevant referring expression in the previous sentence was minimally oblique. This definition involves grammatical obliqueness in the previous sentence, but obliqueness is really being used as a practical operationalization of *salience*, or what psychologists might term *activation*.²⁵ Accordingly, I split the discussion below into one subsection concerning the notion of topic, and one concerning salience.

²⁵See [Arn98] for an extensive discussion of the relation between activation/salience and topic/focus.

5.1 Topic

First, let us return briefly to the terminological move from *backward-looking center* to *topic*. Consider the following quote from Katz [Kat80] (partially cited also in [BY83]): “The surface-subject position imposes the rhetorical or stylistic role of DISCOURSE TOPIC on an NP occupying it, especially one that has been moved into that position... The notion of a discourse topic is that of the common theme of the previous sentences in the discourse, the topic carried from sentence to sentence as the subject of their predications.” Provided we allow that Katz’s notion of *imposes* is the defeasible preference found in Centering theory, then it is clear that the backward-looking center is very like what Katz referred to as the discourse topic.

My use of *topic* as opposed to *discourse topic* is consistent with much contemporary use of the term in syntactic theory — see e.g. the uses of the term in [Ais92], or the even more recent discussion of Chinese in [Shi00]. I would agree with anyone who suggests that *topic* is an overloaded notion. However, my feeling is such overloading can eventually pay off more handsomely than the introduction of yet more un-sullied and connotation-free terminology. This was, in my opinion, the case with use of the similarly overloaded term “presupposition”, about which I have written at length elsewhere.²⁶

Reinhart [Rei82] is often taken to have given the definitive statement of what linguists mean by *topic*, and what they ought to mean. Reinhart argues against defining topics in terms of given material. She argues that topics are primarily what a sentence is *about*, and that givenness is neither a sufficient nor necessary condition for topicality. It is notable that her conclusions do not appear to match the use of *topic* here, and her arguments, although I will not repeat them here, are well taken.²⁷

Even if one fully accepts the points that Reinhart makes, that would not necessarily invalidate the use of *topic* in the current paper. As will be shown shortly, the definition of *topic* used above can be stated instead as a high ranking set of OT constraints which relate the topic to what is mentioned and what is salient. As a result of this move, COT is no longer restricted to any

²⁶I note that Ellen Prince has independently made the same terminological shift to *topic* for C_B , a fact I became aware of during a talk she presented at the January 2000 LSA meeting in Chicago. If the reader accepts no other grounds for using *topic*, perhaps an argument from Prince’s authority will suffice?

²⁷Reinhart’s notion of topic matches that used in some but not all contemporary syntactic theory. Aissen [Ais92] does cite the *aboutness* of the topic as a central feature, whereas, for example, Shi [Shi00] explicitly sides against any notion of topic based on aboutness.

one strict definition of *topic*. The constraints defining *topic* can be violated. In particular, they could, in principle, be violated if the alternative was violation of a higher ranking constraint embodying Reinhart’s requirement that the topic is what the sentence is about.

Moving to a constraint based notion of topic, while initially remaining faithful to standard Centering theory, is achieved by removal of the definition of topic used so far, and addition of the following constraints at the top of the COT ranking:

ONE SENTENCE WINDOW Only discourse entities mentioned in the previous sentence are salient.

ARG SALIENCE One discourse entity is more salient than another if the first was referred to in a less oblique argument position than the second in the same sentence.²⁸

UNIQUE TOPIC With respect to any sentence, there is exactly one discourse entity which is the topic of that sentence.

SALIENT TOPIC The topic of a sentence is the most salient discourse entity referred to in that sentence.

We now have a model in which topic is constrained rather than defined. This has two immediate consequences. First, the topic ranking could be ranked lower: see section 7 for a discussion of what effects this would have. Second, as already indicated, if there were a generally agreed on definition of aboutness, we could consider adding a constraint ranked higher than those above requiring that the topic is what the sentence is about. In effect this would mean that topics were primarily what a sentence was about, and only secondarily common themes between sentences. Such an analysis would meet Reinhart’s stated objections to defining topic as given material, while still preserving the insight that topics generally are just that.²⁹

²⁸We might consider adding to the definition of ARG SALIENCE that the first discourse entity will be more salient than the second if it occurred in a higher clause. Something like this is implicitly assumed in the analyses of (28) and (29), below. Also, to be more precise, the definition should account for cases where some entity is referred to multiple times in the same sentence, in which case on the current definition it might both out-rank and be out-ranked by some other referent.

²⁹The difficulty of defining aboutness (*pace* Reinhart) makes it difficult to state a constraint that the topic is what the sentence is about, which suggests to me that this latter line of research is best left for another occasion.

I have no strong allegiance to any of the above constraints replacing the topic definition, or to their ranking: the proposal is open to negotiation, both for English and more generally.

I would like to suggest one further obvious modification to the constraints on topic in COT. This concerns the positioning of the rule PRO-TOP, which is currently ranked very high, so as to mimic the effects of Rule 1 in BFP. I suggest reordering the constraints, so that PRO-TOP is ranked between FAM-DEF and COHERE. A first result of this move is that a text like (8), repeated below, is predicted to have the reading in which only one “Mary” is referred to, and “He” refers to Jim. As discussed earlier, this reading is filtered in BFP, and is sub-optimal in the earlier version of COT. If this reading is to be allowed, then how might the slight oddity of the discourse be explained? One obvious approach in COT would be to say that the reading is available because it is optimal given the text, but the text is odd because it is not the optimally generated text given the meaning. Text generation is considered in section 7.

- (8)
- a. Mary likes tennis.
 - b. She plays Jim quite often.
 - c. He used to be Mary’s doubles partner.

5.2 Salience

The question of how the topic is defined is distinct from (although closely related to) the question of what the relative salience of different discourse entities is, i.e. how the forward-looking center list is ordered. It is this latter question to which I will now turn.

There are many further questions to be asked about the notion of topic. What if the topic, in the sense I use it, is changing? Are there then two topics? Since my choice of terminology equates topic with C_B , it is clear that there will just be one topic. One case where multiple topics might occur is during switch of topic, in which case there might be both an *old* or *continuing* topic, and a *new*, *switch* or *contrastive* topic. Thus, for example, what is conventionally *wa*- marked in Japanese might then be equated with the *new* topic, not the *old*. Clearly it would be a mistake to equate Japanese *wa*- marked constituents with the C_B . Again, I leave a more detailed examination of this issue, and comparison with Kuno’s use of *topic* [Kun73], for another occasion. Cross-linguistic work dividing topics into separate categories includes [Ais92], and [VV97]. Also see the discussion of contrastive topic in [Bur99], and the excellent cross-linguistic discussion of information structure in [Lam94].

There have been several suggestions for how the forward-looking center list should be formed that differ from the model in BFP. For example, it has been suggested [Kun87, WIC94] that for Japanese, NPs marked (by the choice of main verb) as *empathetic* are highly salient, and that *wa*-marked NPs are even more salient. This result could be arrived at simply by adding two extra constraints at the very top of the ranking, in the following order:

SALIENT WA If in the previous sentence discourse entity α was realized by a *wa*-marked form, and discourse entity β was also realized in that sentence, then α is more salient than β .

SALIENT EMPATHY If in the previous sentence discourse entity α was marked as empathetic, and discourse entity β was not, then α is more salient than β .³⁰

Similarly, it has been suggested initially for German by Strube and Hahn [SH99], and then for English by Strube [Str98], that NP form in the previous sentence is a better predictor of salience than argument position. Here *form* refers to the question of whether a discourse entity was realized by a null pronoun, by a regular pronominal form, by a short description, and so on. I will return to the issue of NP form in section 8. For the moment, observe that given a suitable notion of *minimal form*, the generalization could be modeled by using the following constraint ranked above ARG SALIENCE:

SALIENT FORM If in the previous sentence discourse entity α was realized by a more minimal form than discourse entity β , then α is more salient than β .³¹

This latter line of work includes suggestions for intra-sentential anaphora, which would require removal of the constraint ONE SENTENCE WINDOW. A more general rule than this would be:

LAST S SALIENCE One discourse entity is more salient than another if the first was referred to in the previous sentence and the second was not.

³⁰For detailed discussion of what it means for an argument to be marked as *empathetic*, readers are referred once more to the works cited above, [Kun87, WIC94].

³¹This constraint is *not* to be confused with constraints like Bresnan's "Reduced \Leftrightarrow TOP", mentioned in footnote 12. SALIENT FORM concerns the effect of reducing an expression on its future salience, whereas Bresnan's constraint, like PRO-TOP, concerns the interdependence of the form of the expression on its current salience. Put in terms of standard Centering, if we are considering the form of an NP in sentence n , then SALIENT FORM concerns the interdependency of the form of the NP with C_F^n , whereas Bresnan's constraint and PRO-TOP concern interdependency of the NP form with C_F^{n-1} .

Such a rule, capturing at least partially the idea that salience declines over time, opens up the possibility not only of a treatment of intra-sentential anaphora, and perhaps of the relationship between bound and discourse anaphora, but also of longer distance anaphora, such as occurs with the global focusing mechanisms discussed by Grosz and Sidner [GS86].

There are many other respects in which the notion of salience could be changed. Paramount is the need to allow entities other than those referred to by a previous NP to be salient, as made clear in recent work of Eckert and Strube [ES00]. They show that in a sizeable corpus less than half the anaphoric links were to entities explicitly introduced by NPs, and it is clear that models of *bridging*, propositional anaphora, VP-anaphora and temporal anaphora are all dependent on a far better developed notion of salience than that found in standard centering models or COT.

6 The Directionality of Semantics

De Hoop and Hendriks [HdHar] (hence H&H) provide a perspective on the interpretation of quantification and comparatives, and the interaction of these phenomena with (intonational) focus.³² Constraints they use include the following:

Principle B If two arguments of the same semantic relation are not marked as being identical, interpret them as being distinct.

DOAP Don't Overlook Anaphoric Possibilities. Opportunities to anaphorize text must be seized.³³

Topicality As the antecedent of an anaphoric expression, choose a topic.

Parallelism As the antecedent of an anaphoric expression, choose a (logically, structurally or thematically) parallel element from the preceding clause.

In the H&H model, forms are inputs, and meanings are outputs, as the authors claim quite explicitly:

³²The papers [dH00, dHdS98, vdDdH98] are on related themes.

³³DOAP is introduced in [Wil97], although not in a OT context.

... OT syntax optimizes syntactic structure with respect to a semantic input. One might say that OT syntax takes the perspective of a speaker, therefore, who has a certain thought and who wants to express this correctly and optimally in a syntactic structure. OT semantics, on the other hand, takes the point of view of a hearer, who hears (or reads) an utterance with a certain syntactic structure and wants to interpret this structure correctly and optimally. In OT semantics, the input is a well-formed syntactic structure.... That is, in OT syntax, the candidates which are evaluated respect to the relevant constraints are syntactic structures. In OT semantics, on the other hand, candidate outputs that are subject to evaluation are interpretations.[HdHar]

The proposal that OT syntax is speaker based, but OT semantics is hearer based, is attractive. But it is unnecessarily restrictive.

It is possibly true that a slim majority of constraints that OT syntacticians have used so far refer exclusively to surface form, and it is conceivable that a majority of constraints that semanticists and pragmaticists use will refer exclusively to meanings or information content. To the extent that this is true, it may represent the sociology of linguistics, and it may represent some deeper fact about autonomy of grammar components. If a particular theory is concerned primarily with constraints that refer to only one grammar component, it is clear that the best demonstrations of that theory will involve taking that component to be the output. However, I would like to suggest that the most important challenge for both OT-syntacticians and OT-semanticists lies in stating the theory that relates these components. A theory that is rich in such relational constraints is not exclusively directional in the way that H&H suggest.

Of the four H&H constraints mentioned above, DOAP appears to refer exclusively to the output, i.e. the meaning. **Principle B** and Parallelism explicitly relate form and meaning. And, **Topicality** might also fall into this class depending on what definition of *topic* was given. Given that so many of the constraints used in H&H are relational, it seems that the theory could be applied in both directions³⁴.

In COT, all constraints are relational³⁵, and, as will be shown in the following section, the theory can be applied in either direction. If such use of

³⁴The other H&H constraints are **Emptiness** and **Avoid Contradiction**, which concern the output, and **Forward Directionality**, which arguably is relational.

³⁵Arguably AGREE is non-relational: this would depend on whether agreement is purely formal, or involves some reference to meaning. Given that AGREE is intended to con-

relational constraints is to be typical of papers on OT semantics and pragmatics, then it seems that theories of OT semantics under the H&H definition will also be theories of form. So I conclude that the H&H dichotomy is not useful as defined. There are syntactic constraints, semantic constraints and relational constraints, but interesting OT theories of language will generally not be easily labeled as OT syntax or OT semantics. One thing that OT offers us is a new way of looking at the syntax-semantics-pragmatics interface which emphasizes the significance of relational constraints, and treats purely syntactic, purely semantic and purely pragmatic constraints as parochial special cases. This puts OT grammar in line with proponent of other integrated grammar formalisms, such as HPSG [PS94] and LFG [Bre82], in that it emphasizes the integration of information from different components, and suggests that syntax, semantics and pragmatics are mutually constraining.³⁶

7 Generating with COT

Consider (23), repeated below as (25) with appropriate indexation for the NPs in the third sentence:

- (25) a. Jane_i is happy.
 b. Fred_j gave her_i a present_k.
 c. Fred_j amused her_i.

What are the alternative surface forms of the third sentence of (25) that could have been used to express the proposition that Fred, i.e. the same Fred already under discussion, amused Jane, i.e. the same Jane? Potentially, the candidate set might be large. Let us restrict ourselves to features that COT

strain the resolution not only of pronouns, but also of definite descriptions, which are not grammatically marked for gender, I take AGREE to be a relational constraint.

³⁶With regard to the modularity of language, there is a significant difference between the practice in OT syntax/semantics and that in OT phonology. In OT phonology, it is standard to take both the input and output as levels of phonological form. Perhaps this reflects a property of language, that phonology is to some degree autonomous from other components. But once again, it could also reflect the sociology of the field of linguistics. Note that in recent work Kiparsky [Kip] has suggested a OT model that integrates aspects of lexical morphology and phonology. Work of this sort must give an indication of the extent to which OT phonology can remain autonomous, and the extent to which phonological constraints interact with constraints referring to other aspects of grammar.

might constrain. I will consider only simple sentences consisting of a single clause with two NPs, and will fix the lexical choice of the main verb. Further, I will allow both active and passive forms, and allow each NP to be either a pronoun or a proper name. Taking as the input the desired meaning, in the context of the first two sentences of (25), the following tableau results:

(26)

	AGREE	PRINC-B	PRO-TOP	FAM-DEF	COHERE	ALIGN
Context: (23)a,b Meaning: <i>amuse(j,i)</i>						
☞ He amused her.					*	
Fred amused her.			*	*	*	
She was amused by him.					*	*
She was amused by Fred.			*	*	*	*
He amused Jane.				*	*	
Fred amused Jane.			*	**	*	
Jane was amused by him.				*	*	*
Jane was amused by Fred.			*	**	*	*

The prediction of COT, then, is that the third sentence of (25) should not be realized as in (25)c “Fred amused her”, but instead as “He amused her.” This is correct, in as much as (25)c does seem highly marked. But what does it mean to say that a particular form is *marked*?

A number of issues arise here. First, I observe that in the absence of any further constraints on the use of passive forms, sometimes a passive form will be preferred over a canonical use of a verb. In particular, if the desired meaning had instead been *amuse(i,j)*, then the prediction would be that “He was amused by her.” would be preferred to “She amused him.” However, this prediction relies on two assumptions that might be called into question: that there is no general preference for active verb forms, and that “He was amused by her.” conveys precisely the same content as “She amused him.”

The question of whether alternative surface forms have the same meaning is a complex one. Some, most notably Dwight Bolinger, have argued that all distinctions of surface form signal distinct information content.³⁷ This is

³⁷The idea that distinctions in form correspond to distinctions in meaning pervades all of Bolinger’s work. This passage from [Bol77] is in many ways typical: “Tell [a man in the street] that if two ways of saying something differ in their words or their arrangement

a view that is attractive within OT grammar, and perhaps even inevitable. My own view is a slightly weakened variant of Bolinger's: any distinction of form can signal a distinction of information content, but speakers do not always intend to signal such distinctions.

I would like to take a broad view of meaning, encompassing not only literal content, but also implicatures such as those arising from discourse structure. In particular, it may be that "He was amused by her." is naturally taken as an explanation of Fred's generosity, whereas "She amused him.", while it can be taken this way, can also be taken as a continuation of the narrative. The example is too artificial for judgments to be sound, but I hope is sufficient to illustrate my point: meaning goes beyond mere Logical Form. This is what explains why the non-optimal candidates in (26) are grammatical. They are grammatical, but would be used to signal something going beyond the meaning specified. For instance, one possibility for the non-optimal (25)c "Fred amused her." is that the use of the full proper name "Fred" indicates that we are supposed to be seeing Fred from a new perspective, say that of Jane. Perhaps we are supposed to imagine Jane saying to us, or to herself, "Fred amuses me". The possibilities are endless, and exploring them goes beyond what is possible here.

Ultimately, it will be important to clarify and explain the feeling that (25)c is marked. I tentatively suggest that we consider an explanation like the following: (25)c is not the simplest way to express the literal meaning arrived at by compositional semantic analysis and other constraints (e.g. those relating to Centering), and therefore (25)c must signal something else. If a hearer can identify what is being signaled on a particular occasion of utterance, then that use will be felicitous. But if hearers are unable to identify what is being signaled, they will perceive infelicity. The linguist's starring of a marked sentence indicates that there are no contexts in which that sentence could be used to signal something, or at least that the linguist has insufficient imagination to identify an appropriate pair of a context and a signal.

I would like to make one final point about an assumption made in the above analysis of (25), and in all the interpretation analyses earlier in the paper. The assumption is that optimization is performed on a sentence by sentence basis. With regard to selecting the optimal candidate for (25), I considered candidate surface forms in the light of the prior linguistic context.

they will also differ in meaning, and he will show as much surprise as if you told him that walking in the rain is conducive to getting wet. Only a scientist can wrap himself up in enough sophistication to keep dry under these circumstances."

Conceptually, it makes equal sense to optimize a sentence with respect to the following linguistic context, or with respect to a combination. Thus, we can pose the question does COT predict that a or a' is better in (27) below?

- (27) a. Jane was given a medal by the President.
 a'. The President gave a medal to Jane.
 b. She keeps it on the mantelpiece.

The first sentence of (27), optimized with respect to its (null) prior context, could be realized as either a or a'. But if we optimize with respect to the following sentence, a preference emerges for the passive a, since this enables the topic of the first sentence to be in subject position, and to be the same topic as that of the second sentence.

More generally, it is possible to apply COT to compare the felicity of arbitrarily large texts. The only obstacle to doing this is that it is necessary to decide how to count violations of constraints in different sentences. To demonstrate the possibility of optimizing entire texts, I propose that we count violations in a multi-sentence discourse in the most obvious way: we form one tableau using the standard COT constraint ranking, we enter violations of each constraint in the column corresponding to the violated constraint regardless of the sentence in which the violation occurred, and then select the optimal candidate using the standard OT method. On this basis, we can compare, for instance, the two texts of Grosz and Sidner [GS98], originally adapted from [GJW95], in (28) and (29). The boxes with which I have decorated certain phrases designate which NP would refer to C_B in BFP, or the topic in the basic COT model from section 2, and the significance of the underlining will be explained shortly.

- (28) a. John went to his favorite music store to buy a piano.
 b. He had frequented the store for many years.
 c. He was excited to be going to the store to actually buy a piano.
 d. It was the biggest music store in the area.
 e. It had just the kind of piano that he wanted.
 f. It was closing just as John arrived.

- (29)
- a. John went to his favorite music store to buy a piano.
 - b. It was a store John had frequented for many years.
 - c. He was excited to be going to the store to actually buy a piano.
 - d. It was the biggest music store in the area.
 - e. He knew that it had just the kind of piano that he wanted.
 - f. It was closing just as John arrived.

A theory of Centering should account for the fact that (28) is considerably more awkward than (29). However, Grosz and Sidner note that the sentence by sentence classifications of transitions in BFP and indeed the bulk of later Centering literature do not provide any way to evaluate the coherence of complete texts.

Applying COT to these texts rather naively produces the tableau in (30).³⁸ Rather than notating violations with the usual *, I have notated them with the letter for the line in which the violation occurs, and omitted violations in the first line, on which the two texts do not differ.

(30)

	AGREE	PRINC-B	PRO-TOP	FAM-DEF	COHERE	ALIGN
Common meaning of (28)/(29)						
☞ Text (28)					d	
Text (29)			b c f		c f	b c e f

The preference for the first text is correctly predicted, but a few comments on this example of text-optimization are in order.

First, it should be noted that BFP does differentiate between the two texts, albeit crudely. BFP would analyze the second text as involving multiple violations of Rule 1. For instance, in the second line, the use of the

³⁸I use the term “naively” because I have taken the two texts to be the only two candidates, and glossed over the fact that they do not in fact mean the same thing. I assume that all aspects of meaning where the texts differ are sufficiently insignificant that it is not important for the text to remain faithful to them. Also note that in (28)e and (29)e pronouns are used which refer to an individual (John) not mentioned in the previous line, although this is not reflected in (30). Significantly, this use of a pronoun is felicitous. Perhaps this indicates that COT should be augmented with a notion of global focus, in the manner of [GS86], or actor focus in the sense of [Sid83].

full name “John” to pick out the preferred center from the previous line, in combination with the fact that a pronoun is also present in the second line, would be such a violation. Thus the desired interpretation would be filtered out: BFP does not even allow us to interpret (29)b such that “John” refers to the same individual mentioned in the previous line. Similarly, “John” in (29)f could not refer to the same individual as that mentioned in (29)e, and allowing BFP to treat definite descriptions in the obvious way, the store mentioned in (29)c could not be the same one mentioned in (29). Whether BFP predicts (29) to be uninterpretable, or whether it predicts that the interpretation involves multiple Johns and multiple stores is not entirely clear, although I have assumed the latter in this paper. Either way, (29) is differentiated from (28), and in a way that suggests explanations for (29)’s infelicity. We may then ask whether BFP provides us with a general way of differentiating good texts from bad. The answer to this question must be negative. Violations of Rule 1 take on an enormous significance in BFP, and can allow texts to be differentiated. But preferences among transitions act at a sentence by sentence level, so that the BFP algorithm does not directly provide a metric for comparing pairs of texts that differ only in terms of which transitions occur. In contrast, COT does provide a simple way of comparing such pairs.

My second comment in regard to the COT analysis of (28) and (29) relates to the fact that the reasons why COT predicts relative infelicity of (29) are arguably different from those cited in [GS98]. Grosz and Sidner contend that the jerkiness of (29) results from repeated changes in what they term the “center”, by which I take it they mean C_B . This seems intuitively reasonable. However, according to the original definition of topic in COT (or of C_B in BFP) the topic changes only twice, and at relatively well spaced intervals, in line d and f. According to COT, the biggest problem with (29) is that it involves multiple violations of PRO-TOP. In BFP these would correspond to violations of Rule 1, so that the natural interpretation of the text would not even be available. How can this difference between explanations of infelicity itself be explained?

One possibility is that when Grosz and Sidner were analyzing the two texts, they identified C_B so as to be consistent with Rule 1 wherever possible. Thus, for example, they might have taken C_B of (29)b to be the store, whereas the definitions used in BFP and COT identify topic/ C_B as John. This effect could be mirrored in the variant of COT introduced in section 5 by lowering the ranking of the four topic/salience constraints ONE SENTENCE WINDOW, ARG SALIENCE, UNIQUE TOPIC and SALIENT TOPIC. Let us follow the earlier suggestion from section 5 that PRO-TOP is to be below

FAM-DEF. I will now consider the analysis of (28) and (29) with the four topic/salience constraints immediately below PRO-TOP.

The new ranking does not significantly affect the analysis of (28), for which, after the first sentence, there were no violations of PRO-TOP, FAM-DEF or any stronger constraint. However, it does affect (29), altering which NPs are taken to be topical. The underlining in (29) marks the principal NP referring to the topic under the new analysis, but the underlining is omitted where the analysis of topic is unchanged. The resulting tableau, from which I have omitted the AGREE and PRINCIPLE-B constraints, is presented in (31):

(31)

Common meaning of (28)/(29)	FAM-DEF	PRO-TOP	O-S-W	ARG-SAL	UNIQUE	SAL-TOP	COHERE	ALIGN
Text (28)			e				d	
Text (29)			e			b c	b c d	e

For (29), there are no longer any violations of PRO-TOP, but there are two violations of SALIENT-TOPIC, and three of COHERE. On this analysis, in the first four lines, the topic changes from John, to the store, to John, and back to the store, while in none of these cases is the topic shift signaled by violations of ALIGN. This is certainly a sequence of transitions that might justify Grosz and Sidner’s informal description. Furthermore, my intuition is that the first four sentences of (29)b are indeed more “jerky” than the last two sentences, lending some further credence to the proposed modification.

The alternative analysis of (29) demonstrates the power and flexibility of the COT framework, and the potential it has for capturing intuitions about constraints on discourse cleanly. However, it would be foolish to identify the correct ranking of constraints solely on the basis of my intuitions about Grosz and Sidner’s intuitions about a single text, and I have provided no evidence at all concerning the ranking of some constraints, such as ONE-SENTENCE-WINDOW. If ONE SENTENCE WINDOW were changed in an appropriate way, then the topic of line (e) might turn out to be John, not the store. Making such a change, and keeping PRO-TOP as the highest ranked constraint governing topic choice, would result in an analysis of the text in which every line had a different topic from the previous one, which

is perhaps the analysis that Grosz and Sidner had in mind. It is clear that more refined empirical study is needed.

8 Bidirectional OT and Switch Reference

So far, I have discussed only proposals which utilize a single tableau in one direction or the other, albeit that I have used the same constraints in both directions. There are now several proposals concerning the interaction between form and meaning that go beyond the standard unidirectional tableau, in particular one proposal due to Smolensky, and another by Blutner which has been developed further by him and others. After discussing the basic ideas behind their proposals for bidirectional OT, I will introduce a variant and show that it has the potential to vastly extend the potency and coverage of COT. In particular, I will show the efficacy of the approach for predicting the form of referring expressions.

Smolensky, in unpublished work [Smo98], defines a notion of *recoverability*: a meaning is recoverable if the optimal candidate C expressing that meaning in a meaning \rightarrow form tableau, is such when C is used as an input to a form \rightarrow meaning tableau, what results as optimal output is the original meaning.³⁹ Smolensky's purpose is to explain cases of *ineffability*, whereby a certain meaning is never realized by any linguistic form in a given language.⁴⁰ In effect, Smolensky's policy is to require that the only forms that

³⁹As pointed out in [Blu00a], Smolensky and Prince had in fact considered using OT in reverse in [PS93]. However, the notion of *lexicon optimization* they propose is applied in the OT phonology setting, whereby the input being constrained is a phonological representation. The goal of lexicon optimization appears not to be empirical: it is a process that acts to *tidy up* the set of inputs, where these inputs represent an underlying form not directly available to the linguist's scrutiny. This process is designed never to affect the set of surface forms produced. In contrast, the other uses of bidirectional OT discussed may impact which surface forms are produced, or when they are produced.

⁴⁰To many semanticists, ineffability will not sound like the sort of thing that one would want to explain. For semanticists often take it as a fact of life that language is expressive enough to convey whatever is in need of communication. However, it should be realized that Smolensky's notion of meaning, and of the meaning-form map, is much more restrictive than that used, e.g., in the current paper. For Smolensky, a meaning is something like an LF, and the relationship between a meaning and a form is something like the relationship between LF and Surface Structure in, say, Extended Standard Theory. Smolensky is presumably not taking into account the possibility of realizing a meaning involving a single main predicate at LF using multiple conjoined clauses at surface structure. The perspective offered in the current paper does not restrict meanings to resemble forms, and provides no absolute restrictions on the relationship between meaning and form. A single *infor* (the situation theorist's atomic unit of information) might require an entire scientific

may ever be realized are those that provide recoverability for some meaning.

Blutner, together with Jaeger, Zeevat and Dekker and van Rooy [Blu00a, BJ99, DvR, Zee99] have proposed explaining a wide range of phenomena using a similar approach. These phenomena include *blocking*: what happens when an apparently sure-fire candidate is beaten in a run-off with a surprise outsider. For example, the sure-fire candidate could be a form produced by regular grammatical processes, and the outsider a conventionalized, perhaps lexicalized, irregular form. Partial blocking occurs when the defeated candidate goes on to enter and win a new contest, in which it would not even have taken part had it one its first campaign. Thus a form produced by regular grammatical processes may take on a meaning that would otherwise be secondary or unavailable. Both of these processes seem well suited to an OT framework, although in fact they seem to strain even OT to its limits.

The formalizations of blocking due to Blutner and others mentioned above, as with Smolensky's application of *recoverability*, utilize meta-level mechanisms that sit outside of the standard tableau. Blutner and Jaeger utilize alternatives to the standard OT notion of how inputs are related to outputs which they refer to as *weak* and *strong* optimality. However, for current purposes, it is sufficient to illustrate using an unusual constraint within the tableau. This constraint will favor recoverability of meanings, but also the converse, which might be termed *re-generability*: if a form is optimally interpreted as having some meaning, then that meaning should optimally be realized by the original form. What we arrive at is a biconditional: a meaning should be optimally realized as a certain form if and only if that form is optimally interpreted as having that meaning.

Let us term our new constraint SYMMETRY (SYM). We must be wary of the fact that SYMMETRY makes reference to what is optimal in the system, which leads to the potential for circularity. To avoid such problems, I will define SYMMETRY as: a meaning should be optimally realized as a certain form using all the constraints except SYMMETRY if and only if that form is optimally interpreted as having that meaning, using all the constraints except SYMMETRY.

Let us add SYMMETRY to the constraint ranking and formalize it as follows. Write $A \triangleright B$ to mean that given input A, there is a unique optimal output B. This is to be calculated using the same tableau except without SYMMETRY. A and B are taken as form and meaning or *vice versa* as appropriate. Let M and F be a meaning and form being evaluated in some

article to convey it, and a single *mot juste* might convey more information than an entire scientific article. (The difficulty, of course, is in finding the right word.)

tableau. Then SYMMETRY is defined as the condition $M \triangleright F \leftrightarrow F \triangleright M$.

To see how SYMMETRY can be applied, let us firstly consider a problem in COT. The problem is that sometimes forms are generated from which the meaning is not recoverable. To exemplify this, suppose that in the context of (32)a,b, the speaker wished to say that Jim winked, which might potentially be done using any of c, c' or c".

- (32) a. Fred_i was eating.
 b. He_i saw Jim_j.
 c. He_k winked.
 c'. Jim_k winked.
 c". HE_k winked.⁴¹

Let us augment COT with Schwarzschild's constraint AvoidF, "avoid focus" [Sch99], and rank it below all other constraints. It is easy to see that (32)c will (still) be the preferred realization, as shown in (33).

(33)

	AGREE	PRINC-B	PRO-TOP	FAM-DEF	COHERE	ALIGN	AvoidF
Context: (32)a,b Meaning: <i>winked(j)</i>							
☞ He winked.					*		
HE winked.					*		*
Jim winked.			*		*		

However, (32)c, "He winked", is preferentially interpreted as meaning that Fred winked in this context:

⁴¹I use capitalization to indicate that "HE" carries focal stress. In fact, a realization of (32)c", in a context in which winking is not especially salient, would normally include stress on both "HE" and "winked", and optionally each word might be realized as an entire intonational phrase, or an intermediate phrase in the sense developed by Pierrehumbert and used in the ToBI system of intonational transcription — see e.g. [PH90]. A possible transcription in the ToBI system would involve H*L on the pronouns, and H*L-L% on the verb. However, I skate over details of the intonation for the remainder of this paper.

(34)

		AGREE	PRINC-B	PRO-TOP	FAM-DEF	COHERE	ALIGN	AvoidF
Context: (32)a,b								
Example: (32)c								
☞	k = i							
	k = j					*		

So “He winked” is here the optimal form for the meaning *winked(j)*, but *winked(j)* is not the optimal meaning for “He winked”. To deal with this problem, I propose to use the symmetry constraint. It is easiest to see how the analysis works in what I will call a *bidirectional tableau*. The essence of a bidirectional tableau is that it provides a relative evaluation for form-meaning pairs, and not just for forms given a meaning, or for meanings relative to a form.⁴² Thus we must pick both a candidate set of forms, and a candidate set of meanings. The standard constraints, are then evaluated as normal. Violations of the SYMMETRY constraint are easily evaluated:

1. If a candidate pair is optimal with respect to all other pairs that vary on either the form or the meaning dimension, but not both, mark it as a winner.
2. For each winning pair $\langle M, F \rangle$, star the SYMMETRY column for all pairs $\langle M', F \rangle$ and $\langle M, F' \rangle$.
3. If a candidate pair is optimal with respect to all other pairs that vary on either the form or the meaning dimension, but not both, mark it as a winner.

The following tableau is obtained with SYMMETRY ranked below AGREE and PRINC-B, but above everything else:⁴³

⁴²Blutner and others have used variant tableaux to capture non-standard notions of optimality. The ideal bidirectional tableau would be three dimensional, with meanings, forms and constraints each on separate axes, but so far no ideal way of representing this on paper has been found.

⁴³I use the victory symbol to mark that a candidate is both a winning form relative to the meaning, and a winning meaning relative to the form.

(35)

Context: (32)a,b Meaning	Form	AGREE	PRINC-B	SYM	PRO-TOP	FAM-DEF	COHERE	ALIGN	AvoidF
☞ <i>winked(f)</i>	He winked								
	HE winked.			*					*
	Jim winked.			*	*	*			
<i>winked(j)</i>	He winked			*			*		
	☞ HE winked.						*		*
	Jim winked.				*		*		

There is a great deal to be said about this analysis. To start with, “Jim winked” is predicted to be dis-preferred. In this regard, it should be noted that if in some context, like that of a written text, the accent on the pronoun was not readily detectable, then “Jim winked” would indeed become the preferred candidate of those listed. Thus an alternation between “HE” and “Jim” can be predicted. One way of deriving such an alternation formally would be to allow the relative ordering of AvoidF and PRO-TOP to vary. More generally, variations on the available constraints, and variations on the intended meaning, will obviously produce different realizations. Thus, for example, a meaning incorporating some implied contrast between Fred and Jim might, with appropriate constraints, be realized using a stressed proper name, as “JIM winked”.

I do not want to suggest that the analysis *solves* the problem of focused (or *strong*) pronouns. I do claim that the SYMMETRY constraint, and bidirectional OT more generally, provide an interesting perspective on the notion of markedness. The analysis of stressed pronouns is just one example of the more general insight of Blutner and others that bidirectional OT can make predictions which adhere to Horn’s *division of pragmatic labour*: “The use of a marked (relatively complex and/or prolix expression when a corresponding unmarked (simpler, less ‘effortful’) alternate expression is available tends to be interpreted as conveying a marked message (one which the unmarked alternative would not or could not have conveyed.” [Hor84][p.22]. In this case, the marked construction involves special use of accent, and the marked interpretation is one involving a topic shift.

Analyses of stressed pronouns in Centering have previously been given by Kameyama [Kam99] and Cahn [Cah95]. Cahn’s short paper provides some interesting suggestions as to how Centering theory can be combined with the Pierrehumbert and Hirschberg theory of intonational meaning [PH90]. The fact that Cahn considers an array of different accent types is a great

strength of that paper, and any future work on accenting of pronouns should clearly follow that lead. However, some of the predictions made by Cahn’s combined theory are dubious. In particular, she claims that a simple high accent on a pronoun should allow that pronoun to refer to the backward-looking center, and that only a complex low-high accent should produce the type of alternation found in (32)c’.⁴⁴

As was mentioned earlier in this paper, Kameyama’s approach to Centering has much in common with COT. And, specifically with regard to accenting of pronouns, her conclusions are also in tune with the proposal I have made. Kameyama suggests a general principle, as follows: “Complementary Preference Hypothesis: A focused pronoun takes the complementary preference of the unstressed counterpart.” It is clear that in the case of a single stressed pronoun, Kameyama’s principle may fall out from the more general SYMMETRY constraint applied here.

More generally, the analysis predicts that use of intonationally marked forms could be used to signal many departures from expectation, and topic shift is just one of them. Brown and Yule are amongst those to have observed an apparent mismatch between syntactic and intonational cues as to information status: “if syntactic and intonational forms are both regarded as criteria for ‘givenness’ [...] these forms may supply contradictory information to the hearer” [BY83][p.188], as cited also in [Nak97]. Any rigid theory of intonational meaning, for instance one encapsulating the generalization that focal stress marks new information, is immediately faced with uses of intonational marking that do not conform to that generalization. Perhaps a more general theory of markedness, like one based on SYMMETRY constraint, would resolve these apparent difficulties: it would allow intonational marking to have quite different effects in different contexts.

One of the most significant restrictions in Centering theory is that it does not provide a sufficiently general account of the form of referring expressions. On the other hand, the so-called Givenness Hierarchy [GHZ83] provides a much more general account of the form of referring expressions, but does not attain the degree of precision found in models of Centering such as BFP or

⁴⁴To my ear, use of an L+H* on “HE” would produce an additional indication of contrast, as if it had been previously been suggested that Fred performed some action after seeing Jim, when in fact it was Jim that produced the action. Informally, use of a simple H* produces the effect of switched reference, with less perceived contrast. Nakatani has done impressively thorough empirical research on this and related issues [Nak97], but the only moral I can draw is that empirical issues are vexed in this area. Maria Wolters and I have piloted speaker production experiments to determine the pitch contours used by speakers when the topic is changing. We hope to report on these at a later date.

COT. The Givenness Hierarchy organizes referring expressions as regards the extent to which they depend on linguistic context for their interpretation, and predicts that speakers always select the most contextually dependent (i.e. given) form that is interpretable by the hearer.

As Zeevat has observed [Zee99], the Givenness Hierarchy is naturally formulated in terms of Blutner-Jaeger weak optimality. I refrain from a detailed analysis here, but in essence the idea can be translated into COT terms as follows. Referring forms are progressively more marked, whether this be in terms of their length, morphological complexity, or some measure of semantic complexity. Furthermore, suppose that the least marked meaning for a sentence is a simple predication. Then a generalization of SYMMETRY might provide a way of *aligning the scales* of markedness of form and markedness of meaning.⁴⁵

For example, suppose that linguistic context makes some discourse entity mildly salient although there are many more salient entities of similar semantic category, and that a speaker must decide between a pronoun, a short definite description and a longer one. In this case, SYMMETRY rules out the pronoun because a hearer would not be able to recover the correct meaning. If both the short and long descriptions might potentially lead to recoverable meanings, then the longer can be ruled out by assumption of its inherently greater markedness. Furthermore, if a speaker chooses to use a long description where a shorter one might have done, SYMMETRY will ensure that the hearer concludes that a ‘special’ meaning is intended. This special meaning might, for example, involve the introduction of a new discourse entity (perhaps *accommodation* of a referent in the sense of [Lew79]), or it might involve breaking the assumption that the speaker only wished to convey one piece of information, the main predication. Perhaps the speaker wished to also convey certain extra information pertaining to the already mildly salient entity.

It should now be clear that the above analysis of accented pronouns is merely a special case of a wider analysis which remains to be developed in detail, one which might provide hope for a formal combination of Centering Theory and the Givenness Hierarchy. Gundel [Gun98] provides an excellent discussion of the potential benefits of such an integration.

⁴⁵My use of the terminology *aligning the scales* is intentionally suggestive of a possible link with Aissen’s typological work on the realization of arguments in OT. Some thoughts on an appropriate generalization of SYMMETRY are in [Bea00b].

9 Discussion

Until now it was not obvious how the four stage BFP algorithm could naturally be stated declaratively. COT is the first such statement. This declarativity means that COT is equally suited for generation or interpretation. In contrast, the BFP algorithm is suited for interpretation only. It could not be used to generate texts directly: at best it could be used as a filter to determine whether previously generated texts produced the intended interpretation. Generation in COT is a more subtle affair, since COT does not merely filter out texts which lack the desired interpretation. It also filters out texts which capture the the correct interpretation, but capture it sub-optimally.

Another issue which is clarified in COT is the relation between Centering’s Rule 1 and the two transition classification tests. In previous work, Rule 1 was seen as qualitatively different from the transition classification tests, despite the fact that no empirical evidence has been cited showing that they are different in kind. In COT Rule 1 is no longer *qualitatively* different from the transition classification tests. All three are stated as defeasible constraints. However, the COT ranking makes Rule 1 *quantitatively* different from the transition classification tests, i.e. stronger. Yet in COT the relative strength of constraints can be altered, and this flexibility is original to the COT framework. It applies not only to the status of Rule 1, but also to other components of the theory, such as the definition of C_B , or topic as I have termed it.

In previous work [Bea99b, Bea99a, Bea00a] I have developed a framework termed Transition Preference Pragmatics (TPP). This is a proposal for how Dynamic Semantics⁴⁶ should interact with pragmatics. One observation motivating TPP is that many proposals in Dynamic Semantics fail to take the process of anaphora resolution sufficiently seriously. What makes this situation acute is that the analysis of anaphora is one of Dynamic Semantics’ main applications. The conclusion I argue for is that interpretation of a sentence should not deterministically fix the effect of an information update. Rather, the meaning of a sentence should define a non-deterministic relation between possible incoming linguistic contexts, and possible outgoing linguistic contexts. In TPP, pragmatics provides a preference ordering over alternative incoming-outgoing context pairs. In this way, compositional

⁴⁶Here I use Dynamic Semantics in the sense of Groenendijk and Stokhof [GS91, GSV95]. This is arguably closer to Heim’s earlier work [Hei82] than Kamp’s [KR93], although the differences are not necessarily of empirical significance.

semantics may underspecify the effect of an anaphor, and the pragmatic component can resolve the underspecification. The model in [Bea99a] shows how a simple account of anaphora resolution based on parallelism can be applied in the TPP framework. The current paper develops a richer model of pragmatic interpretation preferences that could, in principle, be interfaced with the TPP semantic component. The result would be a model which incorporated a dynamic notion of meaning, and both semantic and pragmatic constraints on anaphoric linkage. This would produce a system empirically superior either to standard accounts of anaphora in Dynamic Semantics or Centering Theory.⁴⁷ I believe it will be clear to those who study TPP how it could be combined with the account developed here, but I leave this to further work.

Considering possible developments at an even more general level, the COT model I have proposed is founded in terms of the costs and benefits of various linguistic forms to the conversational participants. One of the driving forces behind early the Centering proposals of Joshi and associates [JK79, JW81] was the idea that speakers choose forms which minimize processing costs to hearers.⁴⁸ This idea is visible not only in the analysis of accented pronouns proposed above, but also in the analysis of generation and text optimization: COT models the fact that it may be cheaper in the long-run to use an form which is in the short-term relatively expensive. For instance, a speaker may choose a form in which the topic is not in subject position because it will reduce the costs incurred by a *following* sentence in which a topic shift is needed. This idea is explicit in early work on Centering, but submerged in BFP and much following literature: it is formally explicit for perhaps the

⁴⁷Note that Roberts [Rob98] has described a way in which Centering could be integrated with Kamp's DRT. Her goals are closely related to mine. Centering is an intrinsically dynamic theory. Yet those with a dynamic bent, who are reading the current paper will be acutely aware that, as noted previously, I say little explicitly about the dynamics of linguistic context, and never specify the details of incoming and outgoing contexts formally. Thus there is much work to be done. A natural way to proceed would be to follow the suggestions of Blutner [Blu00b], who defines preferences over pairs of linguistic forms and output contexts relative to a fixed input context. The proposal in TPP is interestingly related: there preferences are defined over pairs of input-output contexts relative to a fixed linguistic form. This casual comparison suggests that we might eventually consider defining preferences over triples of input contexts *and* linguistic forms *and* output contexts.

⁴⁸The clearest presentation of the relationship between costs/benefits and the Blutner-style analysis is found in the work of Dekker and van Rooy [DvR]. They show that the various non-standard notions of optimality developed by Blutner and Jaeger can be viewed in terms of game theory. In this model, relative payoffs of different actions, such as production of a particular linguistic form, are derived from the underlying OT constraint set, and optimality is reduced to special case of strategic equilibrium.

first time in COT.

Here I would note that in some recent interpretation directed work on Centering⁴⁹, there has been discussion of processing issues. In particular, Kehler [Keh97] has observed that the speaker's tendency to use computationally cheap shortcuts, like identifying the subject with the most topical discourse entity, is not in fact captured in the BFP model. Strube [Str98] has gone further, suggesting replacement of the BFP algorithm with an entirely incremental algorithm which works from left to right through a sentence interpreting each anaphoric expression as the most salient entity discourse entity possible.

One thing Strube's model has in common with COT is that it does away with Centering's transitions. Of course, Strube's model also does away with Centering's predictions, which is one respect in which it differs from COT. None the less, if Strube is right about the efficacy of incremental interpretation, then this still does not show that other facets of Centering should be dispensed with altogether. What it does show is that a theory of discourse, whether it be applied to interpretation or generation, should take account of the processing advantages of incrementally interpretable text. The framework I have described allows processing benefits for the hearer to be reflected in choices made by the speaker. My hope is that this quality will appeal both to those who stick by Centering orthodoxy, and to radicals like Strube.⁵⁰

I have presented a framework in which theories of anaphora resolution can be developed, and I have argued that the framework provides fertile ground for further theoretical exploration of this and related issues. I have mentioned some such areas, but omitted others, such as the significance of rhetorical relations in discourse structure.⁵¹

⁴⁹A generation perspective was taken in all the early papers on Centering, but the interpretation perspective is dominant in more formal work following BFP.

⁵⁰There is an extensive psychological literature on Centering — see [HD89, GGG93, Bre95] or various papers in [WJP98]. The above discussion of processing factors leads to the question of what the significance of COT is for psychological models. This should be explored in terms of two sub-questions. First, can previous work on preferences for different Centering transitions be reinterpreted in terms of underlying constraints such as I have proposed? Second, can psychologically motivated models of processing cost be used to derive constraints that should be part of a future COT model, or be used to help choose between alternative possible families of constraints? These are not questions which have easy answers.

⁵¹Grosz and Sidner's work [GS86] shows how a sophisticated theory of discourse structure is relevant to anaphora resolution, but detailed implementation of their proposals remains elusive. Comparison of COT with the wide ranging proposals of Hobbs

It is obvious that such theoretical development must be accompanied by rigorous empirical work. One approach would be to motivate, perhaps functionally, a large set of constraints, and thus obtain, by arbitrary reordering of these constraints, an even larger space of possible theories. We could then ask which of these theories best captures the anaphoric relationships present in elicited text in a production experiment, or which best captures the anaphoric relationships in a tagged corpus. In this way we might hope to learn which theoretical intuitions concerning discourse function significantly determine discourse form.

It is my hope that by enabling uniform description of both Centering and variants on it, the framework I have developed (rather than the specific theory) will facilitate the empirical research that remains to be done.

References

- [Ais92] Judith Aissen. Topic and focus in Mayan. *Language*, 68(1):43–80, 1992.
- [Ais99] Judith Aissen. Markedness and subject choice in Optimality Theory. *Natural Language and Linguistic Theory*, 17:673–711, 1999.
- [AL94] Nicholas Asher and Alex Lascarides. Intentions and information in discourse. In James Pustejovsky, editor, *Proceedings of the Thirty-Second Meeting of the Association for Computational Linguistics*, pages 35–41, San Francisco, 1994. Association for Computational Linguistics, Morgan Kaufmann.
- [Arn98] Jennifer Arnold. *Reference Form and Discourse Patterns*. PhD thesis, Stanford University, 1998.
- [Bea99a] David Beaver. The logic of anaphora resolution. In Paul Dekker, editor, *Proceedings of the 12th Amsterdam Colloquium*. University of Amsterdam, 1999.
- [Bea99b] David Beaver. Pragmatics (to a first approximation). In Jelle Gerbrandy, Maarten Marx, Maarten de Rijke, and Yde Venema, editors, *JFAK — Essays Dedicated to Johan van Benthem on the Occasion of his 50th Birthday*. Vossiuspers, Amsterdam University Press, 1999.

[Hob85, HSAM93] and Asher and Lascarides [LA, AL94] would also be valuable, although beyond the scope of the current work. The use of non-monotonic inference in these theories is reminiscent of the non-monotonicity inherent to OT. Connections with COT can probably best be seen via the intermediary of [dHdS98], which makes a comparable use of discourse structure and temporal relations. Also in this regard, note that observations of Kehler [Keh97, Keh93] probably necessitate some introduction of rhetorical relations into the COT model.

- [Bea00a] David Beaver. Pragmatics, and that's an order. In David Barker-Plummer, David Beaver, Johan van Benthem, and Patrick Scotto di Luzio, editors, *Symbols, Logic and Computation*. CSLI Press, 2000.
- [Bea00b] David Beaver. Weak optimality and supersymmetry, 2000. Manuscript, Stanford University. Available on request.
- [BFP87] Susan Brennan, Marilyn Friedman, and Carl Pollard. A centering approach to pronouns. In *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*, Cambridge, Mass., 1987. Association for Computational Linguistics.
- [BJ99] Reinhard Blutner and Gerhard Jäger. Competition and interpretation: The German adverbs of repetition. available at <http://www2.hu-berlin.de/asg/blutner/>, 1999.
- [Blu00a] Reinhard Blutner. Some Aspects of Optimality in Natural Language Interpretation. ROA-390-04100, 2000.
- [Blu00b] Reinhard Blutner. Some aspects of optimality in natural language interpretation. Technical report, Humbolt University, Berlin, 2000.
- [Bol77] Dwight Bolinger. *Meaning and Form*. Longman, New York, 1977.
- [Bre82] Joan Bresnan, editor. *The Mental Representation of Grammatical Relations*. The MIT Press, Cambridge, MA, 1982.
- [Bre95] Susan Brennan. Centering attention in discourse. *Language and Cognitive Processes*, 10:137–167, 1995.
- [Bre99] Joan Bresnan. The emergence of the unmarked pronoun. In Geraldine Legendre, Sten Vikner, and Jane Grimshaw, editors, *Optimality-theoretic Syntax*. MIT Press, 1999.
- [Bur99] Daniel Buring. On D-trees, beans, and B-accent, 1999. MS., UCSC.
- [BY83] Gillian Brown and George Yule. *Discourse Analysis*. Cambridge University Press, 1983.
- [Cah95] Janet Cahn. The effect of pitch accenting on pronoun referent. In *Proceedings of the 33rd International Joint Conference in Artificial Intelligence (Student Session)*, pages 290–2. Association for Computational Linguistics, 1995.
- [dH00] Helen de Hoop. Optional Scrambling and Interpretation. In H. Bennis, M. Everaert, and E. Reuland, editors, *Interface Strategies*, pages 153–168. KNAW, Amsterdam, 2000.
- [dHdS98] Helen de Hoop and Henriette de Swart. Temporal adjunct clauses in Optimality Theory, 1998. OTS Utrecht.

- [DvR] Paul Dekker and Robert van Rooy. Optimality Theory and Game Theory: Some Parallels. available at <http://turing.wins.uva.nl/pdekker/papers.html>.
- [ES00] Miriam Eckert and Michael Strube. Dialogue acts, synchronising units and anaphora resolution. *Journal of Semantics*, 2000. To appear.
- [GGG93] Peter Gordon, Barbara Grosz, and Laura Gilliom. Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, 17(3):311–347, 1993.
- [GHZ83] Jeanette Gundel, Nancy Hedberg, and Ron Zacharski. Cognitive status and the form of referring expressions in discourse. *Language*, 69:274–307, 1983.
- [GJW83] Barbara Grosz, Aravind Joshi, and Scott Weinstein. Providing a unified account of definite noun phrases in discourse. In *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics*, pages 44–49, Cambridge, Mass., 1983. Association for Computational Linguistics.
- [GJW95] Barbara Grosz, Aravind Joshi, and Scott Weinstein. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):203–226, 1995.
- [GS86] Barbara Grosz and Candace Sidner. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12:175–204, 1986.
- [GS91] Jeroen Groenendijk and Martin Stokhof. Dynamic predicate logic. *Linguistics and Philosophy*, 14:39–100, 1991.
- [GS98] Barbara Grosz and Candace Sidner. Lost intuitions and forgotten intentions. In Marilyn Walker, Aravind Joshi, and Ellen Prince, editors, *Centering Theory in Discourse*, pages 89–112. Clarendon Press, Oxford, 1998.
- [GSV95] Jeroen Groenendijk, Martin Stokhof, and Frank Veltman. Coreference and modality. In Shalom Lappin, editor, *Handbook of Contemporary Semantic Theory*. Blackwell, Oxford, 1995.
- [Gun98] Jeanette Gundel. Centering Theory and the Givenness Hierarchy: Towards a synthesis. In Marilyn Walker, Aravind Joshi, and Ellen Prince, editors, *Centering Theory in Discourse*, pages 359–400. Clarendon Press, Oxford, 1998.
- [HD89] Susan Hudson-DZmura. *Centering: A Framework for Modelling the Local Coherence of Discourse*. PhD thesis, University of Rochester, 1989.
- [HdHar] Petra Hendriks and Helen de Hoop. Optimality Theoretic Semantics. *Linguistics and Philosophy*, to appear.

- [Hei82] Irene Heim. *On the semantics of Definite and Indefinite Noun Phrases*. PhD thesis, Umass. Amherst, 1982.
- [Hob85] J. Hobbs. The coherence and structure of discourse technical report csli-85-37, 1985.
- [Hor84] Laurence Horn. Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In Deborah Schiffrin, editor, *Meaning, Form, and Use in Context: Linguistic Applications*, pages 11–42. Georgetown University Press, Washington, DC, 1984.
- [HSAM93] Jerry Hobbs, Mark Stickel, Douglas Appelt, and Paul Martin. Interpretation as abduction. *Artificial Intelligence*, 63(1-2):69–142, 1993.
- [JK79] Aravind Joshi and Steve Kuhn. Centered logic: The role of entity centered sentence representation in natural language inferencing. In *Proceedings of the 6th International Joint Conference in Artificial Intelligence*, pages 435–9, Tokyo, 1979.
- [JW81] Aravind Joshi and Scott Weinstein. Control of inference: Role of some aspects of discourse structure — centering. In *Proceedings of the 7th International Joint Conference in Artificial Intelligence*, pages 385–7, Vancouver, 1981. Association for Computational Linguistics.
- [Kam98] Megumi Kameyama. Intrasentential centering: A case study. In Marilyn Walker, Aravind Joshi, and Ellen Prince, editors, *Centering Theory in Discourse*, pages 89–112. Clarendon Press, Oxford, 1998.
- [Kam99] Megumi Kameyama. Stressed and unstressed pronouns: complementary preferences. In Peter Bosch and Rob van der Sandt, editors, *The Focus Book*. Cambridge University Press, 1999.
- [Kat80] Jerrold Katz. Chomsky on meaning. *Language*, 56:1–42, 1980.
- [Keh93] Andrew Kehler. Intrasentential constraints on intersentential anaphora in Centering Theory, 1993. Workshop on Centering Theory in Naturally Occurring Discourse, University of Pennsylvania.
- [Keh97] Andrew Kehler. Current theories of Centering for pronoun interpretation: A critical evaluation. *Computational Linguistics*, 23(3), 1997.
- [Kib00] Rodger Kibble. Cb or not Cb? Centering Theory applied to NLG, 2000. MS., University of Brighton.
- [Kip] Paul Kiparsky. Paradigm effects and opacity. ms.
- [KR93] Hans Kamp and Uwe Reyle. *From Discourse to Logic*. Kluwer, 1993.
- [Kun73] S. Kuno. *The Structure of Japanese Language*. MIT Press, Cambridge, Mass., 1973.
- [Kun87] S. Kuno. *Functional Syntax*. Chicago University Press, 1987.

- [LA] A. Lascarides and N. Asher. Discourse relations and common sense entailment.
- [Lam94] Knud Lambrecht. *Information Structure and Sentence Form*. Cambridge University Press, 1994.
- [Lan00] D. Terence Langendoen. An Optimality Theoretic account of the scope of operators, 2000. University of Arizona.
- [Lew79] David Lewis. Scorekeeping in a language game. *Journal of Philosophical Logic*, 8:339–359, 1979.
- [Nak97] Christine Nakatani. *The Computational Processing of Intonational Prominence: A Functional Prosody Perspective*. PhD thesis, Harvard University, 1997. Available as CRCT technical report TR-15-97.
- [PH90] Janet Pierrehumbert and Julia Hirschberg. The meaning of intonational contours in interpretation of discourse. In P. Cohen, J. Morgan, and M. Pollack, editors, *Intentions in Communication*. MIT Press, Cambridge, Massachusetts, 1990.
- [PS93] Alan Prince and Paul Smolensky. Optimality Theory: Constraint interaction in generative grammar. technical report 2. Technical report, Rutgers University Center for Cognitive Science, 1993.
- [PS94] Carl Pollard and Ivan Sag. *Head-Driven Phrase Structure Grammar*. University of Chicago Press and Stanford: CSLI Publications., 1994.
- [Rei82] Tanya Reinhart. Pragmatics and linguistics: An analysis of sentence topics. *Philosophica*, 27:53–94, 1982.
- [Rob98] Craige Roberts. The place of centering in a general theory of anaphora resolution. In Marilyn Walker, Aravind Joshi, and Ellen Prince, editors, *Centering Theory in Discourse*, pages 359–400. Clarendon Press, Oxford, 1998.
- [Sch99] Roger Schwarzschild. Givenness, AvoidF and other constraints on the placement of ac cent. *Natural Language Semantics*, 7(2):141–177, 1999.
- [Sel99] Peter Sells. Form and function in the typology of grammatical voice systems. In *Optimality-Theoretic Syntax*. MIT Press, Cambridge, 1999.
- [Sel00] Peter Sells. Alignment constraints in swedish clausal syntax, 2000. ms., University of Stanford.
- [SH99] Michael Strube and Udo Hahn. Functional Centering: Grounding referential coherence in information structure. *Computational Linguistics*, 25(3):309–344, 1999.
- [Shi00] Dingxu Shi. Topic and topic-comment in Mandarin Chinese. *Language*, 69:274–307, 2000.

- [Sid83] Candace L. Sidner. Focusing in the comprehension of definite anaphora. In M. Brady and R. C. Berwick, editors, *Computational Models of Discourse*, pages 267–330. MIT Press, Cambridge, MA, 1983.
- [Smo98] Paul Smolensky. Why syntax is different (but not really): Ineffability, violability and recoverability in syntax and phonology, 1998. Handout from talk at the Stanford University workshop: Is Syntax Different?
- [Str98] Michael Strube. Never look back: An alternative to Centering. In *Proceedings of the 17th International Conference on Computational Linguistics and the 36th Annual Meeting of the Association for Computational Linguistics*, pages 1251–1257, Montreal, 1998.
- [vdDdH98] Jaap van der Does and Helen de Hoop. Type-shifting and scrambled definites. *Journal of Semantics*, 15:393–416, 1998.
- [VV97] Enric Valduví and Maria Vilkuna. On rheme and kontrast. In Peter Culicover and Louise McNally, editors, *The limits of syntax*. Academic Press, New York, 1997.
- [WIC94] Marilyn Walker, Masayo Iida, and Sharon Cote. Japanese discourse and the process of centering. *Computational Linguistics*, 20/2:193–232, 1994.
- [Wil97] Edwin Williams. Blocking and anaphora. *Linguistic Inquiry*, 28:577–628, 1997.
- [WJP98] Marilyn Walker, Aravind Joshi, and Ellen Prince, editors. *Centering Theory in Discourse*. Clarendon Press, Oxford, 1998.
- [Zee99] Henk Zeevat. Explaining Presupposition Triggers, 1999.
- [Zee00] Henk Zeevat. The asymmetry of optimality theoretic syntax and semantics, 2000. Manuscript, University of Amsterdam.