Functional Centering— Grounding Referential Coherence in Information Structure

Michael Strube*
University of Pennsylvania

Udo Hahn†
Freiburg University

Considering empirical evidence from a free-word-order language (German) we propose a revision of the principles guiding the ordering of discourse entities in the forward-looking center list within the centering model. We claim that grammatical role criteria should be replaced by criteria that reflect the functional information structure of the utterances. These new criteria are based on the distinction between hearer-old and hearer-new discourse entities. We demonstrate that such a functional model of centering can be successfully applied to the analysis of several forms of referential text phenomena, viz. pronominal, nominal, and functional anaphora. Our methodological and empirical claims are substantiated by two evaluation studies. In the first one, we compare success rates for the resolution of pronominal anaphora that result from a grammatical-role-driven centering algorithm and from a functional centering algorithm. The second study deals with a new cost-based evaluation methodology for the assessment of centering data, one which can be directly derived from and justified by the cognitive load premises of the centering model.

1. Introduction

The problem of establishing referential coherence in discourse can be rephrased as the problem of determining the proper antecedent of a given anaphoric expression in the current or the preceding utterance(s) and the rendering of both as referentially identical (coreferential). This task can be approached in a very principled way by stating general constraints on the grammatical compatibility of the expressions involved (e.g., Haddock 1987; Alshawi 1992). Linguists have devoted a lot of effort to identifying conclusive syntactic and semantic criteria to reach this goal, e.g., for intrasentential anaphora within the binding theory part of the theory of Government and Binding (Chomsky 1981), or for intersentential anaphora within the context of the Discourse Representation Theory (Kamp and Reyle 1993).

Unfortunately, these frameworks fail to uniquely determine anaphoric antecedents in a variety of cases. As a consequence, referentially ambiguous interpretations have to be dealt with in those cases in which several alternatives fulfill all the required syntactic and semantic constraints. It seems that syntactic and semantic criteria constitute only necessary but by no means sufficient conditions for identifying the valid antecedent among several possible candidates. Hence, one is left with a preferential choice problem that falls outside of the scope of those strict grammaticality constraints relating to the level of syntax or semantics only. Its solution requires considering pat-

* Institute for Research in Cognitive Science, 3401 Walnut Street, Suite 400A, Philadelphia, PA 19104, USA
† Computational Linguistics Group, Text Understanding Lab, Werthmannplatz 1, 79085 Freiburg, Germany

terns of language use and, thus, introduces the level of discourse context and further pragmatic factors as a complementary description level.

Computational linguists have recognized the need to account for referential ambiguities in discourse and have developed various theories centered around the notion of discourse focus (Grosz 1977; Sidner 1983). In a seminal paper, Grosz and Sidner (1986) wrapped up the results of their research and formulated a model in which three levels of discourse coherence are distinguished—attention, intention, and discourse segment structure. While this paper gives a comprehensive picture of a complex, yet not explicitly spelled-out theory of discourse coherence, the centering model (Grosz, Joshi, and Weinstein, 1983, 1995) marked a major step in clarifying the relationship between attentional states and (local) discourse segment structure. More precisely, the centering model accounts for the interactions between local coherence and preferential choices of referring expressions. It relates differences in coherence (in part) to varying demands on inferences as required by different types of referring expressions, given a particular attentional state of the hearer in a discourse setting (Grosz, Joshi, and Weinstein 1995, 204–205). The claim is made then that the lower the inference load put on the hearer, the more coherent the underlying discourse appears.

The centering model as formulated by Grosz, Joshi, and Weinstein (1995) refines the structure of "centers" of discourse, which are conceived as the representational device for the attentional state at the local level of discourse. They distinguish two basic types of centers, which can be assigned to each utterance U_i—a single **backward-looking center**, $Cb(U_i)$, and a partially ordered set of discourse entities, the **forward-looking centers**, $Cf(U_i)$. The ordering on *Cf* is relevant for determining the *Cb*. It can be viewed as a salience ranking that reflects the assumption that the higher the ranking of a discourse entity in *Cf*, the more likely it will be mentioned again in the immediately following utterance. Thus, given an adequate ordering of the discourse entities in *Cf*, the costs of computations necessary to establish local coherence are minimized.

Given that the ordering on the *Cf* list is crucial for determining the *Cb*, it is no surprise that there has been much discussion among researchers about the ranking criteria appropriate for different languages. In fact, Walker, Iida, and Cote (1994) hypothesize that the *Cf* ranking criteria are the only language-dependent factors within the centering model. Though evidence for many additional criteria for the *Cf* ranking have been brought forward in the literature, to some extent consensus has emerged that grammatical roles play a major role in making ranking decisions (e.g., whether the referential expression appears as the grammatical subject, direct object, or indirect object of an utterance). Our own work on the centering model[1] (Strube and Hahn 1996; Hahn and Strube 1996) brings in evidence from German, a free-word-order language in which grammatical role information is far less predictive of the organization of centers than for fixed-word-order languages such as English. In establishing proper referential relations, we found the **functional information structure** of the utterances to be much more relevant. By this we mean indicators of whether or not a discourse entity in the current utterance refers to another discourse entity already introduced by previous utterances in the discourse. Borrowing terminology from Prince (1981, 1992), an entity that does refer to another discourse entity already introduced is called **discourse-old** or **hearer-old**, while an entity that does not refer to another discourse entity is called **discourse-new** or **hearer-new**.

1 This article is an extended and revised version of our contribution to the 1996 Annual Meeting of the Association for Computational Linguistics (Strube and Hahn 1996). It contains additional material from the doctoral thesis of the first author (Strube 1996a).

Based on evidence from empirical studies in which we considered German as well as English texts from different domains and genres, we make three contributions to the centering approach. The first, the introduction of functional notions of information structure into the centering model, is purely methodological in nature and concerns the centering approach as a theory of local coherence. The second deals with an empirical issue, in that we demonstrate how a functional model of centering can be successfully applied to the analysis of different forms of anaphoric text phenomena, namely pronominal, nominal, and functional anaphora. Finally, we propose a new evaluation methodology for centering data in terms of a cost-based evaluation approach that can be directly derived from and justified by the cognitive load premises of the centering model.

At the methodological level, we develop arguments that (at least for some free-word-order languages) grammatical role criteria should be replaced by **functional role** criteria, since they seem to more adequately account for the ordering of discourse entities in the *Cf* list. In Section 4, we elaborate on particular information structure criteria underlying such a functional center ordering. We also make a second, more general methodological claim for which we have gathered some preliminary, though still not conclusive evidence. Based on a reevaluation of centering analyses of some challenging language data that can be found in the literature on centering, we will argue that exchanging grammatical for functional criteria might also be a reasonable strategy for fixed-word-order languages. What makes this proposal so attractive is the obvious gain in the generality of the model—given a functional framework, fixed- and free-word-order languages might be accounted for by the *same* ordering principles.

The second major contribution of this paper is related to the unified treatment of different text coherence phenomena. It consists of an equally balanced treatment of intersentential (pro)nominal anaphora and inferables (also called functional, bridging, or partial anaphora). The latter phenomenon (cf. the examples in the next section and the in-depth treatment in Hahn, Markert, and Strube [1996]) is usually only sketchily dealt with in the centering literature, e.g., by asserting that the entity in question "is realized but not directly realized" (Grosz, Joshi, and Weinstein 1995, 217). Furthermore, the distinction between these two kinds of realization is not part of the centering mechanisms but delegated to the underlying semantic theory. We will develop arguments for how to discern inferable discourse entities and relate them properly to their antecedent at the center level. The ordering constraints we supply account for all of the types of anaphora mentioned above, including (pro)nominal anaphora (Strube and Hahn 1995; Hahn and Strube 1996). This claim will be validated by a substantial body of empirical data in Section 5.

Our third contribution relates to the way the results of centering-based anaphora resolution are usually evaluated. Basically, we argue that rather than counting resolution rates for anaphora or comparing isolated transition types holding among head positions in the center lists—preferred transition types stand for a high degree of local coherence, while less preferred ones signal that the underlying discourse might lack coherence—one should consider adjacent transition pairs and annotate such pairs with the **processing costs** they incur. This way, we define a dual theory-internal metric of inference load by distinguishing between "cheap" and "expensive" transition types. Based on this distinction, some transition types receiving bad marks in isolation are ranked "cheap" when they occur in the appropriate context, and vice versa.

The article is organized as follows: In Section 2, we introduce the different types of anaphora we consider subsequently, viz. pronominal, nominal, and functional anaphora. We then turn to the proposed modification of the centering model. After a brief introduction into what we call the "grammatical" centering model (actually, a recap of Grosz, Joshi, and Weinstein [1995]) in Section 3, we turn in Section 4 to our approach, the functional model of centering. In Section 5, we present the methodological framework and the empirical data from two evaluation studies we carried out. In Section 6, we relate our work to alternative approaches dealing with local text coherence. In Section 7, we discuss some remaining unsolved problems.

2. Types of Anaphoric Expressions

In this paper, we consider anaphora as a textual phenomenon only, and deal with anaphoric relations that hold between adjacent utterances (intersentential anaphora).[2] Text phenomena are a challenging issue for the design of a text parser for any text-understanding system, since recognition facilities that are imperfect or altogether lacking result in referentially incomplete, invalid, or incohesive text knowledge representation structures (Hahn, Romacker, and Schulz 1999). *Incomplete* knowledge structures emerge when references to already established discourse entities are simply not recognized, as in the case of conceptually neutral pronominal anaphora (e.g., *er*, 'it,' in example (1d) co-specifying with *316LT*, a particular notebook introduced in example (1a)). *Invalid* knowledge structures emerge when each entity that has a different denotation at the text surface is also treated as a formally distinct item at the level of text knowledge representation, although they all refer literally to the same entity. These false referential descriptions result from unresolved nominal anaphora (e.g., *Rechner*, 'computer' in example (1c) co-specifies with *316LT* in (1a)). Finally, *incohesive* or artificially fragmented knowledge structures emerge when entities that are linked by various conceptual relations at the knowledge level occur in a text such that an implicit reference to these relations can be made without the need for explicit signaling at the text surface level. Corresponding referential relations cannot be established at the text representation level, since these inferables remain unsolved (such as the relation between *Akkus*, 'rechargeable battery cell', and *316LT* in examples (1b) and (1a), respectively.[3] The linking conceptual relation between these two discourse elements has to be inferred in order to make it explicit at the level of text knowledge representation structures (for an early statement of this idea in terms of "bridging" inferences, see Clark [1975]).

Note an interesting asymmetric relationship between these three types of anaphora. Pronominal anaphora are constrained by morphosyntactic and grammatical agreement criteria between the pronoun and the antecedent,[4] and no conceptual constraints apply. Nominal anaphora are only constrained by number compatibility between the anaphoric expression and the antecedent, while at the conceptual level the anaphoric expression is related to its antecedent in terms of a conceptual generalization relation. Finally, no grammatical constraints apply to inferables, while conceptual constraints typically require a nongeneralization relation (e.g., part-whole) to hold between the inferable and its antecedent. Of course, contextual conceptual constraints are introduced for both nominal and pronominal anaphora by sortal requirements set up, e.g., by the case roles of the main verb.

2 We have also considered the role of anaphora within sentences. The d-binding criterion we have developed for resolving intrasentential anaphora is based on dependency grammar notions described in more detail in Strube and Hahn (1995).

3 Note that *Reserve-Batteriepack* in Example (1a) and *Akkus* in (1b) denote conceptually different discourse entities that cannot be coindexed.

4 See Jaeggli (1986) for special cases where this criterion is overruled.

Let us illustrate these different types of phenomena by considering the following text fragment:

Example 1

a. Ein Reserve-Batteriepack versorgt den *316LT* ca. 2 Minuten mit Strom.
[A reserve battery pack]$_{nom}$ – supplies – the [*316LT*]$_{acc}$ – for approximately 2 minutes – with power.
The *316LT* is supplied with power by a reserve battery pack for approximately 2 minutes.

b. Der Status des *Akkus* wird dem Anwender angezeigt.
[The status – [of the *rechargeable battery cell*]$_{gen}$]$_{nom}$ – is – [to the user]$_{dat}$ – signalled.
The status of the *rechargeable battery cell* is signalled to the user.

c. Ca. 30 Minuten vor der Entleerung beginnt der *Rechner* 5 Sekunden zu piepen.
Approximately 30 minutes – before discharge – starts – [the *computer*]$^{masc}_{nom}$ – for 5 seconds – to beep.
Approximately 30 minutes before discharge the *computer* beeps for 5 seconds.

d. 5 Minuten bevor *er* sich ausschaltet, fängt die Low-Battery-LED an zu blinken.
5 minutes – before – [*it*]$^{masc}_{nom}$ – itself – turns off – begins – [the low-battery-LED]$_{nom}$ – to flash.
5 minutes before *it* turns off, the low-battery-LED begins to flash.

Common to all the varieties of anaphora we discuss is the search for the proper antecedent in previous utterances, the correct determination of which is considered to be the task of the centering mechanism. The kinds of anaphora we treat can be distinguished, however, in terms of the criteria being evaluated for referentiality. In the case of inferables, the missing conceptual link must be inferred in order to establish local coherence between the utterances involved. In the surface form of utterance (1b) the information that *Akkus*, 'rechargeable battery cell', links up with *316LT* is missing, while, due to obvious conceptual constraints, it cannot link up with *Reserve-Batteriepack*, for example. The underlying relation can only be made explicit if conceptual knowledge about the domain, viz. the relation PART-OF between the concepts RECHARGEBATTERYCELL and 316LT, is available (see Hahn, Markert, and Strube [1996] for a detailed treatment of the resolution of inferables). In the case of nominal anaphors, a conceptual specialization relation has to be determined between the specific antecedent and the more general anaphoric expression, for example, between *316LT* and *Rechner*, 'computer', in (1a) and (1c), respectively. Finally, the resolution of pronominal anaphors need not take conceptual constraints into account at all, but is restricted to grammatical constraints, as illustrated by the masculine gender of *Rechner*, 'computer$_{masc}$', (co-specifying with *316LT*$_{masc}$) and *er* 'it'$_{masc}$, in (1c) and (1d), respectively.

Certainly, the types of phenomena we discuss cover only a limited range of anaphora. In particular, we leave out the whole range of quantificational studies on anaphora (in particular, the "hard" issues related to generalized quantifiers), deictic phenomena, etc., which significantly complicate matters. We return to these unresolved issues in Section 7.

Table 1
Cf ranking by grammatical roles.

subject > object(s) > other(s)

Table 2
Transition types.

<table>
<tr><th></th><th>$Cb(U_i) = Cb(U_{i-1})$</th><th>$Cb(U_i) \neq Cb(U_{i-1})$</th></tr>
<tr><td>$Cb(U_i) = Cp(U_i)$</td><td>CONTINUE</td><td rowspan="2">SHIFT</td></tr>
<tr><td>$Cb(U_i) \neq Cp(U_i)$</td><td>RETAIN</td></tr>
</table>

3. The Centering Model

The centering model (Grosz, Joshi, and Weinstein 1983, 1995) is intended to describe the relationship between local coherence and the use of referring expressions. The model requires two constructs, a single backward-looking center and a list of forward-looking centers, as well as a few rules and constraints that govern the interpretation of centers. It is assumed that discourses are composed of constituent segments (Grosz and Sidner 1986), each of which consists of a sequence of utterances. Each utterance U_i in a given discourse segment *DS* is assigned a list of forward-looking centers, $Cf(DS, U_i)$, and a unique backward-looking center, $Cb(DS, U_i)$. The forward-looking centers of U_i depend only on the discourse entities that constitute the *i*th utterance; previous utterances provide no constraints on $Cf(DS, U_i)$. A ranking imposed on the elements of the *Cf* reflects the assumption that the most highly ranked element of $Cf(DS, U_i)$, the **preferred center** $Cp(DS, U_i)$, will most likely be the $Cb(DS, U_{i+1})$. The most highly ranked element of $Cf(DS, U_i)$ that is finally realized in U_{i+1} (i.e., is associated with an expression that has a valid interpretation in the underlying semantic representation) is the actual $Cb(DS, U_{i+1})$. Since in this paper we will not discuss the topics of global coherence and discourse macro segmentation (for recent treatments of these issues, see Hahn and Strube [1997] and Walker [1998]), we assume a priori that any centering data structure is assigned an utterance in a given discourse segment and simplify the notation of centers to $Cb(U_i)$ and $Cf(U_i)$.

Grosz, Joshi, and Weinstein (1995) state that the items in the *Cf* list have to be ranked according to a number of factors including grammatical role, text position, and lexical semantics. As far as their discussion of concrete English discourse phenomena is concerned, they nevertheless restrict their ranking criteria to those solely based on grammatical roles, which we repeat in Table 1.

The centering model, in addition, defines transition relations across pairs of adjacent utterances (Table 2). These transitions differ from each other according to whether backward-looking centers of successive utterances are identical or not, and, if they are identical, whether they match the most highly ranked element of the current forward-looking center list, the $Cp(U_i)$, or not.

Grosz, Joshi, and Weinstein (1995) also define two rules on center movement and realization:

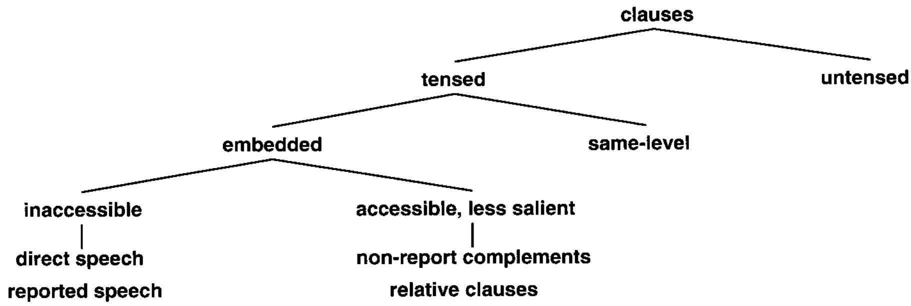

Figure 1
Kameyama's intrasentential centering categorization.

Rule 1
If any element of *Cf*(U_i) is realized by a pronoun in U_{i+1}, then the *Cb*(U_{i+1}) must be realized by a pronoun also.

Rule 2
Sequences of continuation are to be preferred over sequences of retaining; and sequences of retaining are to be preferred over sequences of shifting.

Rule 1 states that no element in an utterance can be realized by a pronoun unless the backward-looking center is realized by a pronoun, too. This rule is intended to capture one function of the use of pronominal anaphors—a pronoun in the *Cb* signals to the hearer that the speaker is continuing to refer to the same discourse. Rule 2 should reflect the intuition that a pair of utterances that have the same theme is more coherent than another pair of utterances with more than one theme. The theory claims, above all, that to the extent that a discourse adheres to these rules and constraints, its local coherence will increase and the inference load placed upon the hearer will decrease.

The basic unit for which the centering data structures are generated is the utterance *U*. Since Grosz, Joshi, and Weinstein (1995) and Brennan, Friedman, and Pollard (1987) do not give a reasonable definition of utterance, we follow Kameyama's (1998) method for dividing a sentence into several center-updating units (Figure 1). Her intrasentential centering mechanisms operate at the clause level. While tensed clauses are defined as utterances on their own, untensed clauses are processed with the main clause so that the *Cf* list of the main clause contains the elements of the untensed embedded clause. Kameyama further distinguishes, for tensed clauses, between sequential and hierarchical centering. Except for direct and reported speech (embedded and inaccessible to the superordinate level), nonreport complements, and relative clauses (both embedded but accessible to the superordinate level; less salient than the higher levels), all other types of tensed clauses build a chain of utterances at the same level.

3.1 A Centering Algorithm for Anaphora Resolution

Though the centering model was not originally intended to be used as a blueprint for anaphora resolution,[5] several applications tackling this problem have made use of

5 Aravind Joshi, personal communication.

Table 3
Basic centering algorithm.

1. If a pronoun in U_i is encountered, test the elements of the $Cf(U_{i-1})$ in the given order until an element under scrutiny satisfies all the required morphosyntactic, binding, and sortal criteria. This element is chosen as the antecedent of the pronoun.
2. When utterance U_i is completely read, compute $Cb(U_i)$ and generate $Cf(U_i)$; rank the elements according to agreed-upon preference criteria (such as the ones from Table 1).

the model, nevertheless. One interpretation is due to Brennan, Friedman, and Pollard (1987) who utilize Rule 2 for computing preferences for antecedents of pronouns (see Section 3.2). In this section, we will specify a simple algorithm that uses the *Cf* list directly for providing preferences for the antecedents of pronouns.

The algorithm (which we will refer to as the basic algorithm; Table 3) consists of two steps, which are triggered independently.

We may illustrate this algorithm by referring to the text fragment in example (2):[6]

Example 2

a. *The sentry* was not dead.

b. *He* was, in fact, showing signs of reviving ...

c. *He* was partially uniformed in a cavalry tunic.

d. *Mike* stripped *this* from *him* and donned *it*.

e. *He* tied and gagged *the man*, ...

Table 4 gives the centering analysis for this text fragment using the algorithm from Table 3.[7] Since (2a) is the first sentence in this fragment, it has no *Cb*. In (2b) and in (2c) the discourse entity SENTRY is referred to by the personal pronoun *he*. Since we assume a *Cf* ranking by grammatical roles in this example, SENTRY is ranked highest in these sentences (the pronoun always appears in subject position). In (2d), the discourse entity MIKE is introduced by a proper name in subject position. The pronoun *him* is resolved to the most highly ranked element of *Cf(2c)*, namely SENTRY. Since *Mike* occupies the subject position, it is ranked higher in the *Cf(2d)* than SENTRY. Therefore the pronoun *he* in (2e) can be resolved correctly to MIKE.

This example not only illustrates anaphora resolution using the basic algorithm from Table 3 but also incorporates the application of Rule 1 of the centering model. (2d) contains the pronoun *him*, which is the *Cb* of this utterance. In (2e), the *Cb* is also realized as a pronoun while SENTRY is realized by the definite noun phrase *the man*, which is allowed by Rule 1.

3.2 The BFP Algorithm

The centering algorithm described by Brennan, Friedman, and Pollard (1987, henceforth BFP algorithm) interprets the centering model in a certain way and applies it to the resolution of pronouns. The most obvious difference between Grosz, Joshi, and

6 With slight simplifications taken from the Brown Corpus cn03.

7 In the subsequent tables illustrating centering data, discourse entities, a notion at the representational level, are denoted by SMALLCAPS and appear on the left side of the colon, while the corresponding surface expressions, at the level of linguistic data, appear on the right side of the colon.

Table 4
Analysis for the text fragment in Example 2 according to the basic centering algorithm.

(2a)	*The sentry* was not dead. Cb: – Cf: [SENTRY: *sentry*]
(2b)	*He* was, in fact, showing signs of reviving ... Cb: SENTRY: *he* Cf: [SENTRY: *he*, SIGNS: *signs*]
(2c)	*He* was partially uniformed in a cavalry tunic. Cb: SENTRY: *he* Cf: [SENTRY: *he*, TUNIC: *tunic*]
(2d)	*Mike* stripped *this* from *him* and donned *it*. Cb: SENTRY: *him* Cf: [MIKE: *Mike*, TUNIC: *this, it*, SENTRY: *him*]
(2e)	*He* tied and gagged *the man*, ... Cb: MIKE: *he* Cf: [MIKE: *he*, SENTRY: *the man*]

Table 5
Transition types according to BFP.

	$Cb(U_i) = Cb(U_{i-1})$ OR $Cb(U_{i-1})$ undef.	$Cb(U_i) \neq Cb(U_{i-1})$
$Cb(U_i) = Cp(U_i)$	CONTINUE	SMOOTH-SHIFT
$Cb(U_i) \neq Cp(U_i)$	RETAIN	ROUGH-SHIFT

Weinstein (1983, 1995) and Brennan, Friedman, and Pollard (1987) is that the latter use two SHIFT transitions instead of only one: SMOOTH-SHIFT[8] requires the $Cb(U_i)$ to equal $Cp(U_i)$, while ROUGH-SHIFT requires inequality (Table 5). Brennan, Friedman, and Pollard (1987) also allow the $Cb(U_{i-1})$ to remain undefined.

Brennan, Friedman, and Pollard (1987) extend the ordering constraints in *Cf* in the following way: "We rank the items in *Cf* by obliqueness of grammatical relations of the subcategorized functions of the main verb: that is, first the subject, object, and object2, followed by other subcategorized functions, and finally, adjuncts." (p. 156). In order to apply the centering model to pronoun resolution, they use Rule 2 in making predictions for pronominal reference and redefine the rules as follows (quoting Walker, Iida, and Cote [1994]):

Rule 1′
If some element of $Cf(U_{i-1})$ is realized as a pronoun in U_i, then so is $Cb(U_i)$.

8 Brennan, Friedman, and Pollard (1987) call these transitions SHIFTING and SHIFTING-1. The more figurative names were introduced by Walker, Iida, and Cote (1994).

Table 6
BFP-algorithm.

1. **Generate possible *Cb-Cf* combinations.** In this step, all (plausible and implausible) assignments of pronouns to elements of the previous *Cf* are computed.
2. **Filter by constraints**, e.g., contra-indexing, sortal predicates, centering rules and constraints. This way, possible antecedents are filtered out because of morphosyntactic, binding, and semantic criteria. Also the realization of noun phrases in the current utterance (e.g., realization as a pronoun *vs.* realization as a definite noun phrase or proper name) comes into play.
3. **Rank by transition orderings.** This is the step, where the pragmatic constraints of centering apply. Basically, CONTINUE transitions are preferred, i.e., the antecedent of a pronoun is more likely to turn up as the *Cb* of the previous utterance than any other element of the *Cf*. In certain configurations, the algorithm includes a preference for parallelism in linguistic constructions.

Rule 2′
Transition states are ordered. CONTINUE is preferred to RETAIN is preferred to SMOOTH-SHIFT is preferred to ROUGH-SHIFT.

Their algorithm (Table 6) consists of three basic steps (as described by Walker, Iida, and Cote [1994]).[9]

In order to illustrate this algorithm, we use example (2) from above and supply the corresponding *Cb/Cf* data in Table 7. Let us focus on the interpretation of utterance (2e) where the centering data diverges when one compares the basic and the BFP algorithms. After step 2 (filtering), the algorithm has produced two readings, which are rated by the corresponding transitions in step 3. Since SMOOTH-SHIFT is preferred over ROUGH-SHIFT, the pronoun *he* is resolved to MIKE, the highest-ranked element of *Cf(2d)*. Also, Rule 1 would be violated in the rejected reading.

4. Principles of Functional Centering

The crucial point underlying functional centering is to relate the ranking of the forward-looking centers and the information structure of the corresponding utterances. Hence, a proper correspondence relation between the basic centering data structures and the relevant functional notions has to be established and formally rephrased in terms of the centering model. In this section, we first discuss two studies in which the information structure of utterances is already integrated into the centering model (Rambow 1993; Hoffman 1996, 1998). Using these proposals as a point of departure, we shall develop our own proposal—functional centering (Strube and Hahn 1996).

4.1 Integrating Information Structure and Centering

As far as the centering model is concerned, the first account involving information structure criteria was given by Kameyama (1986) and further refined by Walker, Iida, and Cote (1994) in their study on the use of zero pronouns and topic mark-

9 Walker, Iida, and Cote (1994) note that it is possible to improve the computational efficiency of the algorithm by interleaving generating, filtering, and ranking steps; cf. the version of the algorithm described by Walker (1998).

Table 7
Centering analysis for the text fragment in example (2) according to the BFP algorithm.

(2a)	*The sentry* was not dead. Cb: – Cf: [SENTRY: *sentry*]	–
(2b)	*He* was, in fact, showing signs of reviving ... Cb: SENTRY: *he* Cf: [SENTRY: *he*, SIGNS: *signs*]	CONTINUE
(2c)	*He* was partially uniformed in a cavalry tunic. Cb: SENTRY: *he* Cf: [SENTRY: *he*, TUNIC: *tunic*]	CONTINUE
(2d)	*Mike* stripped *this* from *him* and donned *it*. Cb: SENTRY: *him* Cf: [MIKE: *Mike*, TUNIC: *this, it*, SENTRY: *him*]	RETAIN
(2e)	*He* tied and gagged *the man*, ... Cb: MIKE: *he* Cf: [MIKE: *he*, SENTRY: *the man*]	SMOOTH-SHIFT
	Cb: ~~MIKE: *the man*~~ Cf: ~~[SENTRY: *he*, MIKE: *the man*]~~	~~ROUGH-SHIFT~~

ers in Japanese. This led them to augment the grammatical ranking conditions for the forward-looking centers by additional functional notions.

A deeper consideration of information structure principles and their relation to the centering model has been proposed in two studies concerned with the analysis of German and Turkish discourse. Rambow (1993) was the first to apply the centering methodology to German, aiming at the description of information structure aspects underlying scrambling and topicalization. As a side effect, he used centering to define the utterance's theme and rheme in the sense of the functional sentence perspective (FSP) (Firbas 1974). Viewed from this perspective, the theme/rheme-hierarchy of utterance U_i is determined by the $Cf(U_{i-1})$. Elements of U_i that are contained in $Cf(U_{i-1})$ are less rhematic than those not contained in $Cf(U_{i-1})$. He then concludes that the $Cb(U_i)$ must be the theme of the current utterance. Rambow does not exploit the information structure of utterances to determine the *Cf* ranking but formulates it on the basis of linear textual precedence among the relevant discourse entities.

In order to analyze Turkish texts, Hoffman (1996, 1998) distinguishes between the information structure of utterances and centering, since both constructs are assigned different functions for text understanding. A hearer exploits the information structure of an utterance to update his discourse model, and he applies the centering constraints in order to connect the current utterance to the previous discourse. Hoffman describes the information structure of an utterance in terms of topic (theme) and comment (rheme). The comment is split again into focus and (back)ground (see also Vallduví [1990] and Vallduví and Engdahl [1996]). Based on previous work about Turkish, Hoffman argues that, in this language, the sentence-initial position corresponds to the topic, the position that immediately precedes the verb yields the focus, and the remainder of the sentence is to be considered the (back)ground. Furthermore, Hoffman relates this notion of information structure of utterances to centering, claiming that the topic corresponds to the *Cb* in most cases—with the exception of segment-initial utterances, which do not have a *Cb*. Hoffman does not say anything about the relation between information structure and the ranking of the

Cf list. In her approach, this ranking is achieved by thematic roles (see also Turan [1998]).

Both Rambow (1983) as well as Hoffman (1996, 1998) argue for a correlation between the information structure of utterances and centering. Both of them find a correspondence between the *Cb* and the theme or the topic of an utterance. They refrain, however, from establishing a strong link between the information structure and centering as we suggest in our model, one that mirrors the influence of information structure in the way the forward-looking centers are actually ranked.

4.2 Functional Centering

Grosz, Joshi, and Weinstein (1995) admit that several factors may have an influence on the ranking of the *Cf* but limit their exposition to the exploitation of grammatical roles only. We diverge from this proposal and claim that, at least for languages with relatively free word order (such as German), the functional information structure of the utterance is crucial for the ranking of discourse entities in the *Cf* list. Originally, in Strube and Hahn (1996), we defined the *Cf* ranking criteria in terms of **context-boundedness**. In this paper, we redefine the functional *Cf* ranking criteria by making reference to Prince's work on the assumed familiarity of discourse entities (Prince 1981) and information status (Prince 1992). The term **context-bound** in Strube and Hahn (1996) corresponds to the term **evoked** used by Prince.[10]

We briefly list the major claims of our approach to centering. In the following sections, we elaborate on these claims, in particular the ranking of the forward-looking centers.

- The elements of the *Cf* list are ordered according to their information status. Hearer-old discourse entities are ranked higher than hearer-new discourse entities. The order of the elements of the *Cf* list for U_i provides the preference for the interpretation of anaphoric expressions in U_{i+1}.
- The first element of the $Cf(U_i)$, the preferred center, $Cp(U_i)$, is the discourse entity the utterance U_i is "about." In other words, the *Cp* is the center of attention.

In contrast to the BFP algorithm, the model of functional centering requires neither a backward-looking center, nor transitions, nor transition ranking criteria for anaphora resolution. For text interpretation, at least, functional centering also makes no commitments to further constraints and rules.

4.3 *Cf* Ranking Criteria in Functional Centering

In this section, we introduce the functional *Cf* ranking criteria. We first describe a basic version, which is valid for a wide range of text genres in which pronominal reference is the predominant text phenomenon. This is the type of discourse to which centering was mainly applied in previous approaches (see, for example, Walker's [1989] or Di Eugenio's [1998] test sets). We then describe the extended version of the functional *Cf* ranking constraints. The two versions differ with respect to the incorporation of (a subset of) inferables in the second version and, hence, with respect to the requirements

10 In Strube and Hahn (1996), we assumed that the information status of a discourse entity has the main impact on its salience. In particular, evoked discourse entities were ranked higher in the *Cf* list than brand-new discourse entities (using Prince's terminology). We also restricted the category of the most salient discourse entities to evoked (i.e., context-bound) discourse entities. In this article, we extend this category to hearer-old discourse entities, which includes, besides evoked discourse entities, unused ones (again, referring to Prince's terminology).

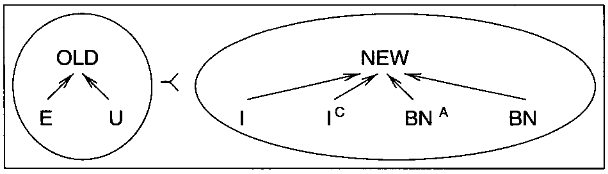

Figure 2
Information status and familiarity (basic version).

relating to the availability of world knowledge, which is needed to properly account for inferables. The extended version assumes a detailed treatment of a particular subset of inferables, so-called functional anaphora (in Hahn, Markert, and Strube [1996], functional anaphora are referred to as textual ellipses). We claim that the extended version of ranking constraints is necessary to analyze texts from certain genres, e.g., texts from technical or medical domains. In these areas, pronouns are used rather infrequently, while functional anaphors are the major text phenomena to achieve local coherence.

4.3.1 Basic *Cf* Ranking. Usually, the *Cf* ranking is represented by an ordering relation on a single set of elements, e.g., grammatical relations (as in Table 1). We use a layered representation for our criteria. For the basic *Cf* ranking criteria, we distinguish between two different sets of expressions, hearer-old discourse entities in U_i (OLD) and hearer-new discourse entities in U_i (NEW). These sets can be further split into the elements of Prince's (1981, 245) familiarity scale. The set of hearer-old discourse entities (OLD) consists of **evoked** (E) and **unused** (U) discourse entities, while the set of hearer-new discourse entities (NEW) consists of **brand-new** (BN) discourse entities. For the basic *Cf* ranking criteria, it is sufficient to assign **inferable** (I), **containing inferable** (I^C), and **anchored brand-new** (BN^A) discourse entities to the set of hearer-new discourse entities (NEW).[11] See Figure 2 for an illustration of Prince's familiarity scale and its relation to the two sets. Note that the elements of each set are indistinguishable with respect to their information status. Evoked and unused discourse entities, for example, have the same information status because they belong to the set of hearer-old discourse entities. So the basic *Cf* ranking in Figure 2 boils down to the preference of OLD discourse entities over NEW ones.

For an operationalization of Prince's terms, we state that evoked discourse entities are simply cospecifying (resolved anaphoric) expressions, i.e., pronominal and nominal anaphora, relative pronouns, previously mentioned proper names, etc. Unused discourse entities are proper names and titles. In texts, brand-new proper names are usually accompanied by a relative clause or an appositive that relates them to the hearer's knowledge. The corresponding discourse entity is evoked only after this elaboration. Whenever these linguistic devices are missing, we treat proper names as unused.[12] In the following, we give some examples of evoked, unused, and brand-new

11 Quoting Prince (1992, 305): "Inferrables are like Hearer-new (and, therefore, Discourse-new) entities in that the hearer is not expected to already have in his/her head the entity in question."
12 For examples of brand-new proper names and how they are introduced, see, for example, the beginning of articles in the "obituaries" section of the *New York Times*.

discourse entities, though in naturally occurring texts these phenomena rarely show up unadulterated.[13] The remaining categories will be explained subsequently.

Example 3

a. He lived his final nine years in one of [two rent-subsidized *buildings*]$_{BN}$ constructed especially for elderly survivors.

b. When the [*buildings*]$_E$ opened – one in 1964, one in 1970 – there were waiting lists.

c. Once, [*they*]$_E$ held 333 survivors.

In example (3a), *buildings* is introduced as a discourse-new discourse entity, which is brand-new (BN). In (3b), the definite NP *the buildings* cospecifies the discourse entity from (3a). Hence, *buildings* in (3b) is evoked (E), just as is *they* in (3c).

Certain proper names are assumed to be known by any hearer. Therefore, these proper names need no further explanation. *Winnie Madikizela Mandela* in example (4) is unused (U), i.e., it is discourse-new but hearer-old. Other proper names have to be introduced because they are discourse-new and hearer-new. In example (5), *Marianne Kador* is introduced by means of a lengthy appositive that relates the brand-new proper name to the knowledge of the hearer. In particular, the noun phrase *the apartment buildings* is discourse-old (see example (3)).

Example 4

[*A defiant Winnie Madikizela Mandela*]$_U$ testified for more than 10 hours today, dismissing all evidence that ...

Example 5

"He was an undervalued person all his life," said *Marianne Kador*, a social worker for Selfhelp Community Services, which operates the apartment buildings in Queens.

In Table 8, we define various sets, which are used for the specification of the *Cf* ranking criteria in Table 9. We distinguish between two different sets of discourse entities, hearer-old discourse entities (OLD) and hearer-new discourse entities (NEW).

For any two discourse entities (x, pos_x) and (y, pos_y), with x and y denoting the linguistic surface expression of those entities as they occur in the discourse, and pos_x and pos_y indicating their respective text position, $pos_x \neq pos_y$, in Table 9 we define the basic ordering constraints on elements in the forward-looking centers $Cf(U_i)$. For any utterance U_i, the ordering of discourse entities in the $Cf(U_i)$ that can be derived from the above definitions and the ordering constraints (1) to (3) are denoted by the relation "$\prec$".

Ordering constraint (1) characterizes the basic relation for the overall ranking of the elements in the *Cf*. Accordingly, any hearer-old expression in utterance U_i is given the highest preference as a potential antecedent for an anaphoric expression in U_{i+1}. Any

13 Examples (3) and (5)–(8) are from the *New York Times*, Dec. 11, 1997. ("Remembering one who remembered. Eugen Zuckermann, survivor, kept the ghosts of the holocaust alive," by Barry Bearak.) Example (4) is from the *New York Times*, Dec. 1, 1997. ("Winnie Mandela is defiant, calling accusations 'lunacy'," by Suzanne Daley.) We split complex sentences into the units specified by Kameyama (1998) following the categorization in Figure 1.

Table 8
Sets of discourse entities for the basic *Cf* ranking.

DE	:	the set of discourse entities in U_i
E	:	the set of evoked discourse entities in U_i
U	:	the set of unused discourse entities in U_i
OLD	:=	E ∪ U
NEW	:=	DE − OLD

Table 9
Basic functional ranking constraints on the *Cf* list.

1. If $x \in$ OLD and $y \in$ NEW, then $x \prec y$.
2. If $x, y \in$ OLD or $x, y \in$ NEW, then $x \prec y$, if $pos_x < pos_y$
3. If (1) or (2) do not apply, then x and y are unordered with respect to the *Cf*-ranking.

hearer-new expression is ranked below hearer-old expressions. Ordering constraint (2) captures the ordering for the sets OLD or NEW when they contain elements of the same type. In this case, the elements of each set are ranked according to their text position.

4.3.2 Extended *Cf* Ranking. While the basic *Cf* ranking criteria are sufficient for texts with a high proportion of pronouns and nominal anaphora (e.g., literary texts, newspaper articles about persons), it is necessary to refine the ranking criteria in order to deal with expository texts, e.g., test reports, discharge summaries. These texts usually contain few pronouns and are characterized by a large number of inferrables, which are often *the* major glue in achieving local coherence. In order to accommodate the centering model to texts from these genres, we distinguish a third set of expressions; **mediated discourse entities** in U_i (MED). On Prince's (1981) familiarity scale, the set of hearer-old discourse entities (OLD) remains the same as before, i.e., it consists of evoked (E) and unused (U) discourse entities, while the set of hearer-new discourse entities (NEW) now consists only of brand-new (BN) discourse entities. Inferable (I), containing inferable (I^C), and anchored brand-new (BN^A) discourse entities, which make up the set of mediated discourse entities, have a status between hearer-old and hearer-new discourse entities.[14] See Figure 3 for Prince's familiarity scale and its relation to the three sets. Again, the elements of this set are indistinguishable with respect to their information status—for instance, inferable and anchored brand-new discourse entities have the same information status because they belong to the set of mediated discourse entities. Hence, the extended *Cf* ranking, depicted in Figure 3, will prefer OLD discourse entities over MEDiated ones, and MEDiated ones will be preferred over NEW ones.

We assume that the difference between containing inferables and anchored brand-new discourse entities is negligible. (It was not well defined in Prince [1981] and in

14 Again, quoting Prince (1992, 305–306): "Inferrables are thus like Hearer-old entities in that they rely on certain assumptions about what the hearer does know, e.g. that buildings typically have doors [. . .], and they are like Discourse-old entities in that they rely on there being already in the discourse-model some entity to trigger the inference [. . .]."

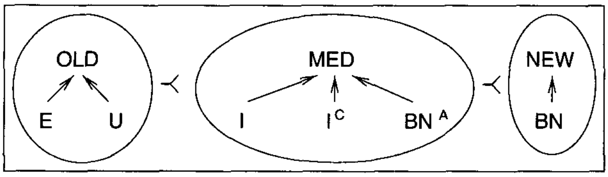

Figure 3
Information status and familiarity (refined version).

Prince [1992] she abandoned the second term.) Therefore, we conflate them into the category of anchored brand-new discourse entities. These discourse entities require that the anchor modifies a brand-new head and that the anchor is either an evoked or an unused discourse entity. In the following, we give examples of inferrables and anchored brand-new discourse entities.

Example 6

a. By *his* teen-age years, the distorted mentality of anti-Semitism was in full warp.

b. [*The family*]$_I$ was expelled to Hungary in 1939 ...

In example 6 the relation between the definite NP *the family* and the context has to be inferred, therefore *the family* belongs to the category inferable (I). It is marked by definiteness but it is not anaphoric since there is no anaphoric antecedent. Though inferables are often marked by definiteness, it is possible that they are indefinite, like *an uncle* in example (7b).

Example 7

a. *He* shared this bounty with his father

b. but [*a sickly uncle*]$_I$ was left to remain hungry.

Anchored brand-new (BN^A) discourse entities as in example (8) are heads of phrases whose modifiers relate (anchor) them to the context.

Example 8

a. *He* had already lost too many companions.

b. [[His]$_E$ *fiancée*]$_{BN^A}$ had died in a car wreck.

With respect to inferables, there exist only a few computational treatments, all of which are limited in scope. We here restrict inferables to the particular subset defined by Hahn, Markert, and Strube (1996), which we call functional anaphora (FA). In the following, we will limit our discussion of inferables to those which figure as functional anaphors. In Table 10, we define the sets needed for the specification of the extended *Cf* ranking criteria in Table 11. We distinguish between three different sets of discourse entities; hearer-old discourse entities (OLD), mediated discourse entities (MED), and hearer-new discourse entities (NEW). Note that the antecedent of a functional anaphor (the inferred discourse entity) is included in the set of hearer-old discourse entities.

Table 10
Sets of discourse entities for the extended *Cf* ranking.

DE	:	the set of discourse entities in U_i
E	:	the set of evoked discourse entities in U_i
U	:	the set of unused discourse entities in U_i
FA^{ante}	:	the set of antecedents of functional anaphors in U_i
FA	:	the set of functional anaphors in U_i
BN^A	:	the set of anchored brand-new discourse entities in U_i
OLD	:=	$E \cup U \cup FA^{ante}$
MED	:=	$FA \cup BN^A$
NEW	:=	DE − (MED ∪ OLD)

Table 11
Extended functional ranking constraints on the *Cf* list.

1. If $x \in$ OLD and $y \in$ MED, then $x \prec y$.
   If $x \in$ OLD and $y \in$ NEW, then $x \prec y$.
   If $x \in$ MED and $y \in$ NEW, then $x \prec y$.
2. If $x, y \in$ OLD, or $x, y \in$ MED, or $x, y \in$ NEW, then $x \prec y$, if $pos_x < pos_y$
3. If (1) or (2) do not apply, then x and y are unordered with respect to the *Cf*-ranking.

For any two discourse entities (x, pos_x) and (y, pos_y), with x and y denoting the linguistic surface expression of those entities as they occur in the discourse, and pos_x and pos_y indicating their respective text position, $pos_x \neq pos_y$, in Table 11 we define the extended functional ordering constraints on elements in the forward-looking centers $Cf(U_i)$. In the following, for any utterance U_i, the ordering of discourse entities in the $Cf(U_i)$ that can be derived from the above definitions and the ordering constraints (1) to (3) are denoted by the relation "$\prec$".

Ordering constraint (1) characterizes the basic relation for the overall ranking of the elements in the *Cf*. Accordingly, any hearer-old expression in utterance U_i is given the highest preference as a potential antecedent for an anaphoric or functional anaphoric expression in U_{i+1}. Any mediated expression is ranked just below hearer-old expressions. Any hearer-new expression is ranked lowest. Ordering constraint (2) fixes the ordering when the sets OLD, MED, or NEW contain elements of the same type. In these cases, the elements of each set are ranked according to their text position.

In Table 12 we show the analysis of text fragment (2) using the basic algorithm see Table 3) with the basic functional *Cf* ranking constraints (see Table 9). The fragment starts with the evoked discourse entity SENTRY in (2a) (the definiteness of the NP indicates that it was already mentioned earlier in the text). The pronouns *he* in (2b) and (2c) are evoked, while *signs* and *tunic* are brand-new. We assume *Mike* in (2d) to be evoked, too (MIKE is the main character of that story). MIKE is the leftmost evoked discourse entity in (2d), hence ranked highest in the *Cf(2d)* and the most preferred antecedent for the pronoun *he* in (2e).

5. Evaluation

In this section, we discuss two evaluation experiments on naturally occurring data. We first compare the success rate of the functional centering algorithm with that of the BFP algorithm. This evaluation uses the basic *Cf* ranking constraints from Table 9.

Table 12
Analysis for text fragment in example (2) according to the model of functional centering.

(2a)	*The sentry* was not dead. Cb: – Cf: [SENTRY$_E$: *sentry*]
(2b)	*He* was, in fact, showing signs of reviving ... Cb: SENTRY$_E$: *he* Cf: [SENTRY$_E$: *he*, SIGNS:$_{BN}$ *signs*]
(2c)	*He* was partially uniformed in a cavalry tunic. Cb: SENTRY$_E$: *he* Cf: [SENTRY$_E$: *he*, TUNIC$_{BN}$: *tunic*]
(2d)	*Mike* stripped *this* from *him* and donned *it*. Cb: SENTRY$_E$: *him* Cf: [MIKE$_E$: *Mike*, TUNIC$_E$: *this, it*, SENTRY$_E$: *him*]
(2e)	*He* tied and gagged *the man*, ... Cb: MIKE$_E$: *he* Cf: [MIKE$_E$: *he*, SENTRY:$_E$ *the man*]

We then introduce a new cost-based evaluation method, which we use for comparing the extended *Cf* ranking constraints from Table 11 with several other approaches.

5.1 Success Rate Evaluation

5.1.1 Data. In order to compare the functional centering algorithm (i.e., the basic algorithm from Table 3 operating with the basic functional *Cf* ranking constraints from Table 9) with the BFP algorithm, we analyzed a sample of English and German texts. The test set (Table 13) consisted of the beginnings of three short stories by Ernest Hemingway,[15] three articles from the *New York Times (NYT)*,[16] the first three chapters of a novel by Uwe Johnson,[17] the first two chapters of a short story by Heiner Müller,[18] and seven articles from the *Frankfurter Allgemeine Zeitung (FAZ)*.[19]

15 Hemingway, Ernest. 1987. *The Complete Short Stories of Ernest Hemingway*. Scribner, New York. ("An African story," pages 545–554; "Soldier's home," pages 111–116; "Up in Michigan," pages 59–62.)

16 (i) *New York Times*, Dec. 7, 1997. ("Shot in head, suspect goes free, then to college," by Jane Fritsch, pages A45–48.) (ii) *New York Times*, Dec. 1, 1997. ("Winnie Mandela is defiant, calling accusations 'lunacy'," by Suzanne Daley, pages A1–12.) (iii) *New York Times*, Dec. 11, 1997. ("Remembering one who remembered. Eugen Zuckermann, survivor, kept the ghosts of the holocaust alive," by Barry Bearak, pages B1–8.)

17 Johnson, Uwe. 1965. *Zwei Ansichten*. Suhrkamp Verlag, Frankfurt am Main.

18 Müller, Heiner. 1974. *Geschichten aus der Produktion 2*. Rotbuch Verlag, Berlin. ("Liebesgeschichte," pages 57–62.)

19 *FAZ*, Aug. 28, 1997. ("Die gute Nachricht ist: Wir können gewinnen. New Yorks früherer Polizeipräsident in Berlin," by Konrad Schuller.) (ii) *FAZ*, Nov. 3, 1997. ("Bürgermeister Giuliani steht vor einer fast sicheren Wiederwahl," by Verena Leucken.) (iii) *FAZ*, Sept. 9, 1997. ("Wir haben viel voneinander lernen können," by Claus Peter Müller.) (iv) *FAZ*, Sept. 10, 1997. ("Die Mutter der Meinungsforschung im Streit. Ist Elisabeth Noelle-Neumann eine unverbesserliche Deutsche?" by Kurt Reumann.) (v) *FAZ*, Aug. 4, 1997. ("Der zarte Riese, Geisterhaftes Klanglicht und ein Zug ins Weite: Zum Tode von Swjatoslaw Richter," by Gerhard R. Koch.) (vi) *FAZ*, Sept. 2, 1997. ("Glaubwürdiger als der Königssohn. Der Oppositionspolitiker Sam Rainsy kämpft für das bessere Kambodscha," by Erhard Haubold.) (vii) *FAZ*, Sept. 3, 1997. ("Bald das Ende des Vorsitzenden Wagner? Wechsel an der Spitze der CDU-Fraktion," by Peter Jochen Winters.)

Table 13
Test set for success rate evaluation.

	Hemingway	NYT	English	Writers	FAZ	German
3rd pers. & poss. pron.	274	302	576	299	320	619
sentences	153	233	386	186	394	580
words	2785	4546	7331	3195	8005	11200

5.1.2 Method. The evaluation was carried out manually by the authors, supported by a small-scale discourse annotation tool. We used the following guidelines for our evaluation: We did not assume any world knowledge as part of the anaphora resolution process. Only agreement criteria and sortal constraints were applied. We did not account for false positives and error chains, but marked the latter (see Walker 1989).

We use Kameyama's (1998) specifications for dealing with complex sentences (for a description, see Section 3). Following Walker (1989), a discourse segment is defined as a paragraph unless its first sentence has a pronoun in subject position or a pronoun whose syntactic features do not match the syntactic features of any of the preceding sentence-internal noun phrases. Also, at the beginning of a segment, anaphora resolution is preferentially performed within the same utterance. According to the preference for intersentential candidates in the original centering model, we defined the following anaphora resolution strategy (which is not the best solution for the anaphora resolution problem either, but sufficient for the purposes of the evaluation):

1. Test elements of $Cf(U_{i-1})$—according to the BFP algorithm, or the functional centering (henceforth abbreviated as FunC) algorithm.
2. Test elements of U_i, which precede the pronoun, left-to-right.
3. Test elements of $Cf(U_{i-2})$, $Cf(U_{i-3})$, ... in the given order.

Since clauses are short in general, step 2 of the algorithm only rarely applies.

5.1.3 Results. The results of our evaluation are given in Table 14. The first row gives the number of third person pronouns and possessive pronouns in the data. The upper part of the table shows the results for the BFP algorithm, the lower part those for the FunC algorithm. Overall, the data are consistently in favor of the FuncC algorithm, though no significance judgments can be made (the data were not drawn as a random sample). The overall error rate of each approach is given in the rows labeled as "wrong". We also tried to determine the major sources of errors (see the nonbold sections in Table 14), and were able to distinguish three different types. One class of errors relates to the algorithm's strategy. In the case of the BFP algorithm, the corresponding row also contains the number of ambiguous cases generated by this algorithm (we counted ambiguities as errors, since FunC produced only one reading in these cases). A second class of errors results from error chains, mainly caused by the strategy of each approach or by ambiguities in the BFP algorithm. A third error class is caused by the intersentential specifications, e.g., the correct antecedent is not accessible because it is realized in an embedded clause (reported speech). Finally, other errors were mainly caused by split antecedents (plural pronouns referring to a couple of antecedents in singular), reference to events (or propositions), and cataphora.

Table 14
Evaluation results for success rates.

	Hemingway	NYT	English		Writers	FAZ	German	
3rd pers. & poss. pron.	**274**	**302**	**576**		**299**	**320**	**619**	
correct	**193**	**245**	**438**	**(76%)**	**236**	**227**	**463**	**74,8%**
wrong	**81**	**57**	**138**	**(24%)**	**63**	**93**	**156**	**(25,2%)**
wrong (strategy)	20	8	28		10	27	37	
wrong (error chains)	29	15	44		22	28	50	
wrong (intersentential)	17	27	44		18	24	42	
wrong (others)	15	7	22		13	14	27	
correct	**214**	**252**	**466**	**(80,9%)**	**248**	**270**	**518**	**(83,7%)**
wrong	**60**	**50**	**110**	**(19,1%)**	**51**	**50**	**101**	**(16,3%)**
wrong (strategy)	8	3	11		3	3	6	
wrong (error chains)	18	13	31		18	6	24	
wrong (intersentential)	18	27	45		17	27	44	
wrong (others)	16	7	23		13	14	27	

5.1.4 Interpretation. While the rate of errors caused by the specifications for complex sentences and by other reasons is almost identical (the small difference can be explained by false positives), there is a remarkable difference between the algorithms with respect to strategic errors and error chains. Strategic errors occur whenever the preference given by the algorithm under consideration leads to an error. Most of the strategic errors implied by the FunC algorithm also show up as errors for the BFP algorithm. We interpret this finding as an indication that these errors are caused by a lack of semantic or world knowledge. The remaining errors of the BFP algorithm are caused by the strictly local definition of its criteria and because the BFP algorithm cannot deal with some particular configurations leading to ambiguities. The FunC algorithm has fewer error chains not only because it yields fewer strategic errors, but also because it is more robust with respect to real texts. An utterance U_i, for instance, which intervenes between U_{i-1} and U_{i+1} without any relation to U_{i-1} does not affect the preference decisions in U_{i+2} for FunC, although it does affect them for the BFP algorithm, since the latter cannot assign the $Cb(U_{i+1})$. Also, error chains are sometimes shorter in the FunC analyses.

Example (9) illustrates how the local restrictions as defined by the original centering model and the BFP algorithm result in errors and lead to rather lengthy error chains (see Table 15 for the corresponding centering analysis). The discourse entity SENTENCE, which is cospecified by the pronoun *er*, 'it'$_{masc}$, in (9b), is the *Cb(9b)*. Therefore, it is the most preferred antecedent for the pronoun *ihn* in (9c), which causes a strategic error. This error, in turn, is the reason for a consequent error in (9d), because there are no semantic cues that enforce the correct interpretation, i.e., the coreferentiality between *ihn* and *Giuliani*. The possible interruption of the error chain, indicated by the alternative interpretation in (9c), is ruled out, however, by the preference for RETAIN over ROUGH-SHIFT transitions (cf. Rule 2′).

Example 9

a. Der Satz, mit dem Ruth Messinger eine der Fernsehdebatten im Bürgermeisterwahlkampf in New York eröffnete, wird der einzige sein, der von ihr in Erinnerung bleibt.

Table 15
BFP results for example (9).

(9a)	Cb: – Cf: [SENTENCE: *Satz, dem, der, der,* RUTH: *Ruth Messinger, ihr,* DEBATES: *Fernsehdebatten,* RACE: *Bürgermeisterwahlkampf,* NEW YORK: *New York,* RECOLLECTION: *Erinnerung*]	–
(9b)	Cb: SENTENCE: *er* Cf: [SENTENCE: *er,* VICTORY: *Wahlsieg,* GIULIANI: *Rudolph Giuliani*]	CONTINUE
(9c)	Cb: SENTENCE: *ihn* Cf: [NEWSPAPERS: *Zeitungen,* SENTENCE: *ihn,* NEW YORK: *Stadt*]	RETAIN
	Cb: ~~GUILIANI: *ihn*~~ Cf: ~~[NEWSPAPERS: *Zeitungen,* GIULIANI: *ihn,* NEW YORK: *Stadt*]~~	~~ROUGH-SHIFT~~
(9d)	Cb: SENTENCE: *ihm* Cf: [UNIONS: *Gewerkschaften,* SENTENCE: *ihm*]	RETAIN

[The sentence]$^{masc}_{subject}$, with which Ruth Messinger – one of the TV debates – opened, – will – the only one – be, – which – of her – in memory – remains.
The sentence, with which Ruth Messinger opened one of the TV debates, will be the only one, which will be recollected of her.

b. Am nahezu sicheren Wahlsieg des Amtsinhabers Rudolph Giuliani am Dienstag wird *er* nichts ändern.
[Of the almost certain – victory in the election – of [the officeholder Rudolph Giuliani]^{masc}]$^{masc}_{adjunct}$ – on Tuesday – will – [*it*]$^{masc}_{subject}$ – nothing – alter.
Of the officeholder Rudolph Giuliani's almost certain victory in the election on Tuesday, *it* will alter nothing.

c. Alle Zeitungen der Stadt unterstützen *ihn.*
[All – newspapers of the city]$_{subject}$ – support – [*him*]$^{masc}_{direct-object}$.
He is supported by all newspapers of the city.

d. Die Gewerkschaften stehen hinter *ihm.*
[The unions]$_{subject}$ – stand behind – [*him*]$^{masc}_{indirect-object}$.
He is backed up by the unions.

The nonlocal definition of hearer-old discourse entities enables the FunC algorithm to compute the correct antecedent for the pronoun *ihn* in (9c) preventing it from running into an error chain (see Table 16 for the functional centering data). GIULIANI, who was mentioned earlier in the text, is the leftmost evoked discourse entity in (9b) and therefore the most preferred antecedent for the pronoun in (9c), though there is a pronoun of the same gender in (9b).

We encountered problems with Kameyama's (1998) specifications for complex sentences. The differences between clauses that are accessible from a higher syntactic level and clauses that are not could not be verified by our analyses. Also, her approach is sometimes too coarse-grained (i.e., there are still antecedents within one utterance), and sometimes too fine-grained.[20]

20 An alternative to Kameyama's intrasentential centering, which overcomes these problems and leads to

Table 16
FunC results for example (9).

(9a)	Cf:	[SENTENCE$_E$: *Satz, dem, der, der,* RUTH$_E$: *Ruth Messinger,* RACE$_E$: *Bürgermeisterwahlkampf,* NEW YORK$_E$: *New York,* DEBATES$_{BNA}$: *Fernsehdebatten,* RECOLLECTION$_{BN}$: *Erinnerung*]
(9b)	Cf:	[GIULIANI$_E$: *Rudolph Giuliani,* SENTENCE$_E$: *er,* VICTORY$_{BNA}$: *Wahlsieg*]
(9c)	Cf:	[NEW YORK$_E$: *Stadt,* GIULIANI$_E$: *ihn,* NEWSPAPERS$_{BNA}$: *Zeitungen*]
(9d)	Cf:	[GIULIANI$_E$: *ihm,* UNIONS$_{BN}$: *Gewerkschaften*]

Table 17
Test set for cost-based evaluation.

	IT	Spiegel	Müller	Σ
(pro)nominal anaphors	308	102	153	563
functional anaphors	294	25	20	339
sentences	451	82	87	620
words	5542	1468	867	7877

5.2 Cost-based Evaluation

5.2.1 Data. The test set for our second evaluation experiment consisted of three different text genres: 15 product reviews from the information technology (IT) domain, one article from the German news magazine *Der Spiegel*, and the first two chapters of a short story by the German writer Heiner Müller.[21] Table 17 summarizes the total number of (pro)nominal anaphors, functional anaphors, utterances and words in the test set.

5.2.2 Method (Distribution of Transition Types). Given these sample texts, we compared three approaches to the ranking of the *Cf*: a model whose ordering principles are based on grammatical role indicators only (see Table 1); an "intermediate" model, which can be considered a "naive" approach to free-word-order languages; and the functional model based on the information structure constraints stated in Table 11. For reasons discussed below, slightly modified versions of the naive and the grammatical approaches will also be considered. They are characterized by the additional constraint that antecedents of functional anaphors are ranked higher than the functional anaphors themselves. As in Section 5.1, the evaluation was carried out manually by the authors.

Since most of the anaphors in these texts are nominal anaphors, the resolution of which is much more restricted than that of pronominal anaphors, the success rate for the whole anaphora resolution process is not distinctive enough for a proper evaluation of the functional constraints. The reason for this lies in the fact that nominal anaphors are far more constrained by conceptual criteria than pronominal ones. Thus, the chance of properly resolving a nominal anaphor, even when ranked at a lower position in the center lists, is greater than for pronominal anaphors. By shifting our evaluation criteria away from resolution success data to structural conditions reflecting the proper ordering of center lists (in particular, we focus on the most highly ranked item of the forward-looking centers), these criteria are intended to compensate for the

a significant improvement in the results, is proposed in Strube (1998).
21 Müller, Heiner. 1974. *Geschichten aus der Produktion* 2. Rotbuch Verlag, Berlin. ("Liebesgeschichte," pages 57–62.)

Table 18
Quantitative distribution of centering transitions.

	Transition Types	Naive	Naive & $FA^{ante} > FA$	Grammatical	Grammatical & $FA^{ante} > FA$	FunC
IT	CONTINUE	49	167	102	197	309
	RETAIN	269	158	226	131	25
	SMOOTH-SHIFT	32	41	24	35	51
	ROUGH-SHIFT	39	23	37	26	4
Spiegel	CONTINUE	17	28	37	43	50
	RETAIN	42	32	28	23	12
	SMOOTH-SHIFT	9	9	7	8	13
	ROUGH-SHIFT	7	6	3	1	0
Müller	CONTINUE	31	31	32	32	36
	RETAIN	19	19	18	18	15
	SMOOTH-SHIFT	15	17	15	16	18
	ROUGH-SHIFT	14	12	14	13	10
Σ	CONTINUE	97	226	171	272	395
	RETAIN	330	209	272	172	52
	SMOOTH-SHIFT	56	67	46	59	82
	ROUGH-SHIFT	60	41	54	40	14

high proportion of nominal anaphora in our sample. Table 5 enumerates the types of centering transitions we consider.

5.2.3 Results (Distribution of Transition Types). In Table 18, we give the numbers of centering transitions between the utterances in the three test sets. The first column contains those generated by the naive approach (such a proposal was made by Gordon, Grosz, and Gilliom [1993] as well as by Rambow [1993], who, nevertheless, restricts it to the German middlefield). We simply ranked the elements of C_f according to their text position. While it is usually assumed that the functional anaphor (FA) is ranked above its antecedent (FA^{ante}) (Grosz, Joshi, and Weinstein 1995, 217), we assume the opposite. The second column contains the results of this modification with respect to the naive approach. In the third column of Table 18, we give the numbers of transitions generated by the grammatical constraints (Table 1) stated by Grosz, Joshi, and Weinstein (1995, 214, 217). The fourth column supplies the results of the same modification as was used for the naive approach, namely, antecedents of functional anaphors are ranked higher than the corresponding anaphoric expressions. The fifth column shows the results generated by the functional constraints from Table 11.

5.2.4 Interpretation (Distribution of Transition Types). The centering model assumes a preference order among transition types—CONTINUE ranks above RETAIN and RETAIN ranks above SHIFT. This preference order reflects the presumed inference load put on the hearer to *coherently* decode a discourse. Since the functional approach generates more CONTINUE transitions (see Table 18), we interpret this as preliminary evidence that this approach provides for a more efficient processing than its competitors. In particular, the observation of a predominance of CONTINUEs holds irrespective of the various text genres we considered for functional centering and, to a lesser degree, for the modified grammatical ranking constraints.

5.2.5 Method (Costs of Transition Types). The arguments we have given so far do not seem to be entirely convincing. Counting single occurrences of transition types, in general, does not reveal the entire validity of the center lists. Considering adjacent transition pairs as an indicator of validity should give a more reliable picture, since depending on the text genre considered (e.g., technical vs. news magazine vs. literary texts), certain sequences of transition types may be entirely plausible though they include transitions which, when viewed in isolation, seem to imply considerable inferencing load (Table 18). For instance, a CONTINUE transition that follows a CONTINUE transition is a sequence that requires the lowest processing costs. But a CONTINUE transition that follows a RETAIN transition implies higher processing costs than a SMOOTH-SHIFT transition following a RETAIN transition. This is due to the fact that a RETAIN transition ideally predicts a SMOOTH-SHIFT in the following utterance. Hence, we claim that no one particular centering transition should be preferred over another. Instead, we advocate the idea that certain centering transition pairs are to be preferred over others. Following this line of argumentation, we propose here to classify all occurrences of centering transition pairs with respect to the "costs" they imply. The cost-based evaluation of different C_f orderings refers to evaluation criteria that form an intrinsic part of the centering model.

Transition pairs hold for three immediately successive utterances. We distinguish between two types of transition pairs, cheap ones and expensive ones.

- A transition pair is cheap if the backward-looking center of the current utterance is correctly predicted by the preferred center of the immediately preceding utterance, i.e., $C_b(U_i) = C_p(U_{i-1})$.
- A transition pair is expensive if the backward-looking center of the current utterance is not correctly predicted by the preferred center of the immediately preceding utterance, i.e., $C_b(U_i) \neq C_p(U_{i-1})$.

In particular, chains of the RETAIN transition in passages where the C_b does not change (passages with constant theme) show that the grammatical ordering constraints for the forward-looking centers are not appropriate.

5.2.6 Results (Costs of Transition Types). The numbers of centering transition pairs generated by the different approaches are shown in Table 19. In general, the functional approach reveals the best results, while the naive and the grammatical approaches work reasonably well for the literary text, but exhibit a remarkably poorer performance for the texts from the IT domain and, to a lesser degree, from the news magazine. The results for the latter approaches improve only slightly with the modification of ranking the antecedent of an functional anaphor (FA^{ante}) above the functional anaphor itself (FA). In any case, they do not compare to the results of the functional approach.

5.3 Extension of the Centering Transitions

Our use of the centering transitions led us to the conclusion that CONTINUE and SMOOTH-SHIFT are not completely specified by Grosz, Joshi, and Weinstein (1995) and Brennan, Friedman, and Pollard (1987). According to Brennan, Friedman, and Pollard's definition, it is possible that a transition is labeled SMOOTH-SHIFT even if $Cp(U_i) \neq Cp(U_{i-1})$. Such a SHIFT is less smooth, because it contradicts the intuition that a SMOOTH-SHIFT fulfills what a RETAIN predicted. The same applies to a CONTINUE with this characteristic. Hence, we propose to extend the set of transitions as shown in Ta-

Table 19
Cost values for centering transition pair types.

	Cost Type	Naive	Naive & $FA^{ante} > FA$	Grammatical	Grammatical & $FA^{ante} > FA$	FunC
IT	cheap	72	180	129	236	321
	expensive	317	209	260	153	68
Spiegel	cheap	25	36	45	51	62
	expensive	50	39	30	24	13
Müller	cheap	45	48	46	48	55
	expensive	34	31	33	31	24
Σ	cheap	142	264	220	335	438
	expensive	401	279	323	208	105

Table 20
Revised transition types.

	$Cb(U_i) = Cb(U_{i-1})$ OR $Cb(U_{i-1})$ undef.	$Cb(U_i) \neq Cb(U_{i-1})$
$Cb(U_i) = Cp(U_i)$ AND $Cp(U_i) = Cp(U_{i-1})$	CONTINUE	SMOOTH-SHIFT
$Cb(U_i) = Cp(U_i)$ AND $Cp(U_i) \neq Cp(U_{i-1})$	EXP-CONTINUE	EXP-SMOOTH-SHIFT
$Cb(U_i) \neq Cp(U_i)$	RETAIN	ROUGH-SHIFT

Table 21
Costs for transition pairs.

	CONT.	EXP-CONT.	RET.	SMOOTH-S.	EXP-SMOOTH-S.	ROUGH-S.
–	**cheap**	–	exp.	–	–	–
CONT.	**cheap**	–	**cheap**	exp.	–	exp.
EXP-CONT.	exp.	–	exp.	exp.	–	exp.
RET.	exp.	exp.	exp.	**cheap**	exp.	exp.
SMOOTH-S.	**cheap**	exp.	exp.	exp.	exp.	exp.
EXP-SMOOTH-S.	exp.	exp.	exp.	exp.	exp.	exp.
ROUGH-S.	exp.	exp.	exp.	**cheap**	exp.	exp.

ble 20. The definitions of CONTINUE and SMOOTH-SHIFT are extended by the condition that $Cp(U_i) = Cp(U_{i-1})$, while EXP-CONTINUE and EXP-SMOOTH-SHIFT (expensive CONTINUE and expensive SMOOTH-SHIFT) require the opposite. RETAIN and ROUGH-SHIFT fulfill $Cp(U_i) \neq Cp(U_{i-1})$ without further extensions.

Table 21 contains a complete overview of the transition pairs. Only those whose second transition fulfills the criterion $Cp(U_i) = Cp(U_{i-1})$ are labeled as "cheap."

5.4 Redefinition of Rule 2

Grosz, Joshi, and Weinstein (1995) define Rule 2 of the centering model on the basis of sequences of transitions. Sequences of CONTINUE transitions are preferred over sequences of RETAIN transitions, which are preferred over sequences of SHIFT transitions. Brennan, Friedman, and Pollard (1987) utilize this rule for anaphora resolution but restrict it to single transitions. Based on the preceding discussion of cheap and expensive transition pairs, we propose to redefine Rule 2 in terms of the costs of transition types.[22] Rule 2 then reads as follows:

Rule 2″ Cheap transition pairs are preferred over expensive ones.

We believe that this definition of Rule 2 allows for a far better assessment of referential coherence in discourse than a definition in terms of sequences of transitions.

For anaphora resolution, we interpret Rule 2″ such that the preference for antecedents of anaphors in U_i can be derived directly from the $Cf(U_{i-1})$. The higher a discourse entity is ranked in the *Cf*, the more likely it is the antecedent of a pronoun. We see the redefinition of Rule 2 as the theoretical basis for a centering algorithm for pronoun resolution that simply uses the *Cf* as a preference ranking device like the basic centering algorithm shown in Table 3. In this algorithm, the metaphor of costs translates into the number of elements of the *Cf* that have to be tested until the correct antecedent is found. If the *Cp* of the previous utterance is the correct one, then the costs are indeed very low.

5.5 Does Functional Centering Provide a More Satisfactory Explanation of the Data?

We were also interested in finding out whether the functional criteria we propose might explain the linguistic data in a more satisfactory way than the grammatical-role-based criteria discussed so far. So, we screened sample data from the literature, which were already annotated by centering analyses (for English, we considered all examples discussed in Grosz, Joshi, and Weinstein [1995] and Brennan, Friedman, and Pollard [1987]). We achieved consistent results for the grammatical and the functional approach for all the examples contained in Grosz, Joshi, and Weinstein (1995) but found diverging analyses for some examples discussed by Brennan, Friedman, and Pollard (1987). While the RETAIN-SHIFT combination in examples (10c) and (10d′) (slightly modified from Brennan, Friedman, and Pollard [1987, 157]) did not indicate a difference between the approaches, for the RETAIN-CONTINUE combination in examples (10c) and (10d), the two approaches led to different results (see Table 22 for the BFP algorithm and Table 23 for the FunC algorithm).

Example 10

a. *Brennan* drives an Alfa Romeo.

b. *She* drives too fast.

c. *Friedman* races *her* on weekends.

d. *She* often wins.

d′. *She* often beats *her*.

Within the functional approach, the proper name *Friedman* is unused and, therefore, the leftmost hearer-old discourse entity of (10c). Hence, FRIEDMAN is the most preferred antecedent for the pronoun *she* in (10d) and (10d′).

22 See Di Eugenio (1998) for a discussion regarding certain pairs of transitions and their relation to zero vs. strong pronouns.

Table 22
BFP interpretation for example (10)—The "Friedman" scenario.

(10a)	Cb: Cf:	– [BRENNAN: *Brennan*, ALFA ROMEO: *Alfa Romeo*]	–
(10b)	Cb: Cf:	[BRENNAN: *she*] [BRENNAN: *she*]	CONTINUE
(10c)	Cb: Cf:	[BRENNAN: *her*] [FRIEDMAN: *Friedman*, BRENNAN: *her*]	RETAIN
(10d)	Cb: Cf:	[BRENNAN: *she*] [BRENNAN: *she*]	CONTINUE
	Cb: Cf:	[~~FRIEDMAN: *she*~~] [~~FRIEDMAN: *she*~~]	~~SMOOTH-SHIFT~~
(10d′)	Cb: Cf:	[FRIEDMAN: *she*] [FRIEDMAN: *she*, BRENNAN: *her*]	SMOOTH-SHIFT
	Cb: Cf:	[~~FRIEDMAN: *her*~~] [~~BRENNAN: *she*, FRIEDMAN: *her*~~]	~~ROUGH-SHIFT~~

Table 23
FunC interpretation for example (10)—The "Friedman" scenario.

(10a)	Cf:	[BRENNAN$_U$: *Brennan*, ALFA ROMEO$_{BN}$: *Alfa Romeo*]
(10b)	Cf:	[BRENNAN$_E$: *she*]
(10c)	Cf:	[FRIEDMAN$_U$: *Friedman*, BRENNAN$_E$: *her*]
(10d)	Cf:	[FRIEDMAN$_E$: *she*]
(10d′)	Cf:	[FRIEDMAN$_E$: *she*, BRENNAN$_E$: *her*]

But is subjecthood really the decisive factor? When we replace *Friedman* with a hearer-new discourse entity, e.g., *a professional driver*, as in (10c′),[23] then the procedures generate inconsistent results, again. In the BFP algorithm, the ranking of the *Cf* list depends only on grammatical roles. Hence, DRIVER is ranked higher than BRENNAN in the *Cf(10c′)*. In (10d), the pronoun *she* is resolved to BRENNAN because of the preference for CONTINUE over SMOOTH-SHIFT. In (10d′), *she* is resolved to DRIVER because SMOOTH-SHIFT is preferred over ROUGH-SHIFT (see Table 24).

10c′. A professional driver races *her* on weekends.

Within the functional approach, the evoked phrase *her* in (10c′) is ranked higher than the brand-new phrase *a professional driver*. Therefore, the preference changes between example (10c) and (10c′). In (10d) and (10d′) the pronoun *she* is resolved to BRENNAN, the discourse entity denoted by *her* (see Table 25).

We find the analyses of functional centering to match our intuitions about the underlying referential relations more closely than those that are computed by grammatically based centering approaches. Hence, in the light of this still preliminary evidence, we answer the question we posed at the beginning of this subsection in the affirmative—functional centering indeed explains the data in a more satisfying manner than other well-known centering principles.

23 We owe this variant to Andrew Kehler. This example may misdirect readers because the phrase *a professional driver* might be assigned the "default" gender masculine. Anyway, this example—like the original example—seems not to be felicitous English and has only illustrative character.

Table 24
BFP interpretation for Example (10)—The "driver" scenario.

(10a)	Cb: – Cf: [BRENNAN: *Brennan*, ALFA ROMEO: *Alfa Romeo*]	–
(10b)	Cb: [BRENNAN: *she*] Cf: [BRENNAN: *she*]	CONTINUE
(10c′)	Cb: [BRENNAN: *her*] Cf: [DRIVER: *driver*, BRENNAN: *her*]	RETAIN
(10d)	Cb: [BRENNAN: *she*] Cf: [BRENNAN: *she*]	CONTINUE
	Cb: [~~DRIVER: *she*~~] Cf: [~~DRIVER: *she*~~]	~~SMOOTH-SHIFT~~
(10d′)	Cb: [DRIVER: *she*] Cf: [DRIVER: *she*, BRENNAN: *her*]	SMOOTH-SHIFT
	Cb: [~~DRIVER: *her*~~] Cf: [~~BRENNAN: *she*, DRIVER: *her*~~]	~~ROUGH-SHIFT~~

Table 25
FunC interpretation for Example (10)—The "driver" scenario.

(10a)	Cf:	[BRENNAN$_U$: *Brennan*, ALFA ROMEO$_{BN}$: *Alfa Romeo*]
(10b)	Cf:	[BRENNAN$_E$: *she*]
(10c′)	Cf:	[BRENNAN$_E$: *her*, DRIVER$_{BN}$: *driver*]
(10d)	Cf:	[BRENNAN$_E$: *she*]
(10d′)	Cf:	[BRENNAN$_E$: *she*, DRIVER$_E$: *her*]

5.6 Summary of Evaluation

To summarize the results of our empirical evaluation, we claim, first, that our proposal based on functional criteria leads to substantially improved and—with respect to the inference load placed on the text understander, whether human or machine—more plausible results for languages with free word order than the structural constraints given by Grosz, Joshi, and Weinstein (1995) and those underlying the naive approach. We base these observations on an evaluation study that considers transition pairs in terms of the inference load specific pairs imply. Second, we have gathered preliminary evidence, still far from conclusive, that the functional constraints on centering seem to explain linguistic data more satisfactorily than the common grammar-oriented constraints. Hence, we hypothesize that these functional constraints might constitute a general framework for treating free- *and* fixed-word-order languages by the same methodology. This claim, without doubt, has to be further substantiated by additional cross-linguistic empirical studies.

The cost-based evaluation we focused on in this section refers to evaluation criteria that form an intrinsic part of the centering model. As a consequence, we have redefined Rule 2 of the Centering Constraints (Grosz, Joshi, and Weinstein 1995, 215) appropriately. We replaced the characterization of a preference for sequences of CONTINUE over sequences of RETAIN and, similarly, sequences of RETAIN over sequences of SHIFT by one in which cheap transitions are to be preferred over expensive ones.

6. Comparison with Related Approaches

6.1 Focus-based Approaches

Approaches to anaphora resolution based on focus devices partly use the information status of discourse entities to determine the current discourse focus. However, a common area of criticism of these approaches is the diversity of data structures they require. These data structures are likely to hide the underlying linguistic regularities, because they promote the mix of preference and data structure considerations in the focusing algorithms. As an example, Sidner (1983, 292ff.) distinguishes between an *Actor Focus* and a *Discourse Focus*, as well as corresponding lists, viz. *Potential Actor Focus List* and *Potential Discourse Focus List*. Suri and McCoy (1994) in their RAFT/RAPR approach use grammatical roles for ordering the focus lists and make a distinction between *Subject Focus*, *Current Focus*, and corresponding lists. Both focusing algorithms prefer an element that represents the *Focus* to the elements in the list when the anaphoric expression under consideration is not the agent (for Sidner) or the subject (for Suri and McCoy). Relating these approaches to our proposal, they already exhibit a weak preference for a single hearer-old (more precisely, evoked) discourse element. Dahl and Ball (1990), describing the anaphora resolution module of the PUNDIT system, improve the focusing mechanism by simplifying its underlying data structures. Thus, their proposal is more closely related to the centering model than any other focusing mechanism. Furthermore, if there is a pronoun in the sentence for which the *Focus List* is built, the corresponding evoked discourse entity is shifted to the front of the list. The following elements of the *Focus List* are ordered by grammatical roles again. Hence, their approach still relies upon grammatical information for the ordering of the centering list, while we use only the functional information structure as the guiding principle.

6.2 Heuristics

Given its embedding in a cognitive theory of inference loads imposed on the hearer and, even more importantly, its fundamental role in a more comprehensive theory of discourse understanding based on linguistic, attentional, and intentional layers, the centering model can be considered the first principled attempt to deal with preference orders for plausible antecedent selection for anaphors. Its predecessors were entirely heuristic approaches to anaphora resolution. These were concerned with various criteria—beyond strictly grammatical constraints such as agreement—for the optimization of the referent selection process based on preferential choices. An elaborate description of several of these preference criteria is supplied by Carbonell and Brown (1988) who discuss, among others, heuristics involving case role filling, semantic and pragmatic alignment, syntactic parallelism, syntactic topicalization, and intersentential recency. Given such a wealth of criteria one may either try to order them *a priori* in terms of importance or—as was proposed by the majority of researchers in this field—define several scoring functions that compute flexible orderings on the fly. These combine the variety of available evidence, each one usually annotated by a specific weight factor, and, finally, map the weights to a single salience score (Rich and LuperFoy 1988; Hajičová, Kuboň, and Kuboň 1992; Lappin and Leass 1994)

These heuristics helped to improve the performance of discourse-understanding systems through significant reductions of the available search-space for antecedents. Their major drawback is that they require a great deal of skilled hand-crafting that, unfortunately, usually does not scale in broader application domains. Hence, proposals were made to replace these high-level "symbolic" categories by statistically interpreted occurrence patterns derived from large text corpora (Dagan and Itai 1990). Preferences then reflect patterns of statistically significant lexical usage rather than introspective abstractions of linguistic patterns such as syntactic parallelism or pragmatic alignment.

Among the heuristic approaches to anaphora resolution, those which consider the identification of heuristics a machine learning (ML) problem are particularly interesting, since their heuristics dynamically adapt to the textual data. Furthermore, ML procedures operate on incomplete parses (hence, they accept noisy data), which distinguishes them from the requirements of perfect information and high data fidelity imposed by almost any other anaphora resolution scheme. Connolly, Burger, and Day (1994) treat anaphora resolution as an ML classification problem and compare seven classifier approaches with the solution quality of a naive hand-crafted algorithm whose heuristics incorporate the well-known agreement and recency indicators. Aone and Bennett (1996) outline an approach where they consider more than 60 features automatically obtained from the machinery of the host natural language processing system the learner is embedded in. The features under consideration include lexical ones like categories, syntactic ones like grammatical roles, semantic ones like semantic classes, and text positional ones, e.g., the distance between anaphor and antecedent. These features are packed in feature vectors—for each pair of an anaphor and its possible antecedent—and used to train a decision tree, employing Quinlan's C4.5 algorithm (Aone and Bennett 1996), or a whole battery of alternative classifiers in which hybrid variants yield the highest scores (Connolly, Burger, and Day 1994). Though still not fully worked out, it is interesting to note that in both studies ML-derived heuristics tend to outperform those that were carefully developed by human experts (similar results are reported by Cardie [1992] with respect to learning resolution heuristics for relative pronouns pertaining to a case-based learning procedure). This indicates, at least, that heuristically based methods using simple combinations of features benefit from being exposed to and having to adapt to training data. ML-based mechanisms might constitute an interesting perspective for the further tuning of ordering criteria for the forward-looking centers.

These mixed heuristic approaches, using multidimensional metrics for ranking antecedent candidates, diverge from the assumption that underlies the centering model that a single type of criterion—the attentional state and its representation in terms of the backward- and forward-looking centers—is crucial for referent selection. By incorporating functional considerations in terms of the information structure of utterances into the centering model we actually enrich the types of knowledge that go into centered anaphora resolution decisions, i.e., we extend the "dimensionality" of the centering model, too. But unlike the numerical scoring approaches, our combination remains at the symbolic computation level, preserves the modularity of criteria, and, in particular, is linguistically justified. Although functional centering is not a complete theory of preferential anaphora resolution, one should clearly stress the different goals behind heuristics-based systems, such as the ones just discussed, and the model of centering. Heuristic approaches combine introspectively acquired descriptive evidence and attempt to optimize reference resolution performance by proper evidence "engineering". This is often done in an admittedly ad hoc way, requiring tricky retuning when new evidence is added (Rich and LuperFoy 1988). On the other hand, many of these systems work in a real-world environment (Rich and LuperFoy 1988; Lappin and Leass 1994; Kennedy and Boguraev 1996) in which noisy data and incomplete, sometimes even faulty, analysis results have to be accounted for. The centering model differs from these considerations in that it aims at unfolding a unified theory of discourse coherence at the linguistic, attentional, and intentional level (Grosz and Sidner 1986); hence, the search for a more principled, theory-based solution, but also the need for (almost) perfect linguistic analyses in terms of parsing and semantic interpretation.

7. Conclusion

In this paper, we provided a novel account for ordering the forward-looking center list, a major construct of the centering model. The new formulation is entirely based on functional notions, grounded in the information structure of utterances in a discourse.

We motivated our proposal by the constraints that hold for a free-word-order language such as German and derived our results from empirical studies of real-world texts. We also augmented the ordering criteria of the forward-looking center list such that it accounts not only for (pro)nominal anaphora but also for inferables (restricted to the subset of functional anaphora), an issue that, up to now, has only been sketchily dealt with in the centering framework. The extensions we proposed were validated by the empirical analysis of various texts of considerable length selected from different domains and genres. The "evaluation metric" we used refers to a new cost-based model of interpreting the validity of centering data. The distinction between cognitively cheap and expensive transition pairs led us to replace Rule 2 from the original model by a formulation that explicitly incorporates this cost-oriented distinction.

A resolution module for (pro)nominal anaphora (Strube and Hahn 1995) and one for functional anaphora (Hahn, Markert, and Strube 1996) based on this functional centering model has been implemented as part of PARSETALK, a comprehensive text parser for German (Hahn, Schacht, and Bröker 1994; Hahn, Neuhaus, and Bröker 1997) in our group. All these modules are fully operational and integrated within the text-understanding backbone of SYNDIKATE, a large-scale text knowledge acquisition system for the two real-world domains of information technology (Hahn and Schnattinger 1998) and medicine (Hahn, Romacker, and Schulz 1999).

Despite the progress made so far, many research problems remain open for further consideration in the centering framework. The following list mentions only the most pertinent issues that have come to our attention and complements the list given by Grosz, Joshi, and Weinstein (1995):

1. The centering model is rather agnostic about the intricacies of complex sentences such as relative clauses, subordinate clauses, coordinations, and complex noun phrases. The problem caused by these structures for the centering model is how to decompose a complex sentence into center-updating units and how to process complex utterances consisting of multiple clauses. A first proposal is due to Kameyama (1998) who breaks a complex sentence into a hierarchy of center-updating units. Furthermore, she distinguishes several types of constructions in order to decide which part of the sentence is relevant for the resolution of an intersentential anaphor in the following sentence. Strube (1996b) (with respect to centering) and Suri and McCoy (1994) (with respect to the focus model) describe similar approaches and provide algorithms for the interaction of the resolution of inter- and intrasentential anaphora, but the topic has certainly not been dealt with exhaustively. The problem of complex NPs was pointed out by Walker and Prince (1996). Since the grammatical functions in a sentence may be realized by a complex NP, it is not clear how to rank these phrases in the C_f list. Walker and Prince (1996) propose a "working hypothesis" based on the surface order. Strube (1998) provides a complete specification for dealing with complex sentences, but this approach departs significantly from the centering model.

2. It seems that there exist only a few fully operational implementations of centering-based algorithms, since the interaction of the algorithm with global and local ambiguities generated by a sentence parser has not received much attention until now. A first proposal for how to deal with center ambiguity in an incremental text parser has been made by Hahn and Strube (1996).

3. The centering model covers the standard cases of anaphora, i.e., pronominal and nominal anaphora and even functional anaphora based on the proposal we have developed in this article. It does not, however, take into account several "hard" issues such as plural anaphora, generic definite noun phrases, propositional anaphora, and deictic forms (but see Eckert and Strube [1999] for a treatment of discourse-deictic anaphora in dialogues within a centering-type framework). These shortcomings might be traced back to the fact that the centering model, up to now, did not consider the role of the (main) verb of the utterance under scrutiny. Other cases, such as VP anaphora (Hardt 1992), temporal anaphora (Kameyama, Passonneau, and Poesio 1993; Hitzeman, Moens, and Grover 1995) have already been examined within the centering model. The particular phenomenon of paycheck anaphora is described by Hardt (1996), though he uses only a rather simplified centering model for this work. Other cases are only dealt with in the focusing framework such as propositional anaphora (Dahl and Ball 1990).

4. Evaluations of the centering model have so far only been carried out manually. This is clearly no longer rewarding, so appropriate computational support environments have to be provided. What we have in mind is a kind of discourse structure bank and associated workbenches comparable to grammar workbenches and parse treebanks. Aone and Bennett (1994), for example, report on a GUI-based Discourse Tagging Tool (DTT) that allows a user to link an anaphor with its antecedent and specify the type of the anaphor (e.g., pronoun, definite NP, etc.). The tagged result can be written out to an SGML-marked file. Arguing for the need for discourse taggers, this also implies the development of a discourse structure interlingua (some sort of Discourse Structure Mark-up Language) for describing discourse structures in a common format in order to ease nonproblematic exchange and world-wide distribution of discourse structure data sets. Such an environment would provide excellent conditions for further testing, for example, of our assumption that the information structure constraints we suggest might apply in a universal manner.

5. Centering theory, so far, is a model of local coherence in the minimal sense, i.e., it allows only the consideration of immediately adjacent centering structures for establishing proper referential links. In order to extend that theory to the level of global coherence, various steps have to be taken.

   - At the referential level, mechanisms have to be introduced to account for reference relationships that extend beyond the immediately preceding utterance. Empirical evidence for such phenomena exists in the literature and we also found the need to have such a mechanism available for longer texts. The extension of functional centering to these phenomena is presented in Hahn and Strube (1997), while Walker (1998) builds upon the centering algorithm described in Brennan, Friedman, and Pollard (1987).
   - At the level of discourse pragmatics, a richer notion than mere reference between terms is needed to account for coherence relations such as those aimed at by Rhetorical Structure Theory

(Mann and Thompson 1988). In addition, an explicit relation to basic notions from speech act theory is also missing, though it should be considered vital for the global coherence of discourse (Grosz and Sidner 1986). In general, it might become increasingly necessary to integrate very deep forms of reasoning, perhaps even nonmonotonic (Dunin-Keplicz and Lukaszewicz 1986) or abductive inference mechanisms (Nagao 1989), into the anaphora resolution process. This might become a sheer necessity when incrementality of processing receives a higher level of attention in the centering community.

Acknowledgments
We would like to thank our colleagues from the Computational Linguistics Group in Freiburg and at the University of Pennsylvania for fruitful discussions, in particular Norbert Bröker, Miriam Eckert, Aravind Joshi, Manfred Klenner, Nobo Komagata, Katja Markert, Peter Neuhaus, Ellen Prince, Rashmi Prasad, Owen Rambow, Susanne Schacht, and Bonnie Webber. We also owe special thanks to the four reviewers whose challenges and suggestions have considerably improved the presentation of our ideas about functional centering in this article. The first author was partially funded by *LGFG Baden-Württemberg*, a post-doctoral grant from *DFG* (Str 545/1-1) and a post-doctoral fellowship award from the Institute for Research in Cognitive Science at the University of Pennsylvania (NSF SBR 8920230).

References

Alshawi, Hiyan. 1992. Resolving quasi logical forms. In H. Alshawi, editor, *The Core Language Engine*. MIT Press, Cambridge, MA, pages 187–216.

Aone, Chinatsu and Scott W. Bennett. 1994. Discourse tagging tool and discourse-tagged multilingual corpora. In *Proceedings of the International Workshop on Sharable Natural Language Resources*, pages 71–77, Ikoma, Nara, Japan, August.

Aone, Chinatsu and Scott W. Bennett. 1996. Applying machine learning to anaphora resolution. In S. Wermter, E. Riloff, and G. Scheler, editors, *Connectionist, Statistical and Symbolic Approaches to Learning for Natural Language Processing*. Springer, Berlin, pages 302–314.

Brennan, Susan E., Marilyn W. Friedman, and Carl J. Pollard. 1987. A centering approach to pronouns. In *Proceedings of the 25th Annual Meeting*, pages 155–162, Association for Computational Linguistics, Stanford, CA, July.

Carbonell, Jaime G. and Ralph D. Brown. 1988. Anaphora resolution: A multi-strategy approach. In *Proceedings of the 12th International Conference on Computational Linguistics*, volume 1, pages 96–101, Budapest, Hungary, August.

Cardie, Claire. 1992. Learning to disambiguate relative pronouns. In *Proceedings of the 10th National Conference on Artificial Intelligence*, pages 38–43, San José, CA, July.

Chomsky, Noam. 1981. *Lectures on Government and Binding*. Foris, Dordrecht.

Clark, Herbert H. 1975. Bridging. In *Proceedings of the Conference on Theoretical Issues in Natural Language Processing*, pages 169–174, Cambridge, MA, June.

Connolly, Dennis, John D. Burger, and David S. Day. 1994. A machine learning approach to anaphoric reference. In *Proceedings of the International Conference on New Methods in Language Processing*, pages 255–261, Manchester, England, September.

Dagan, Ido and Alon Itai. 1990. Automatic processing of large corpora for the resolution of anaphora references. In *Proceedings of the 13th International Conference on Computational Linguistics*, volume 3, pages 330–332, Helsinki, Finland, August.

Dahl, Deborah A. and Catherine N. Ball. 1990. Reference resolution in PUNDIT. In P. Saint-Dizier and S. Szpakowicz, editors, *Logic and Logic Grammars for Language Processing*. Ellis Horwood, Chichester, England, pages 168–184.

Di Eugenio, Barbara. 1998. Centering in Italian. In M. A. Walker, A. K. Joshi, and E. F. Prince, editors, *Centering Theory in Discourse*, Oxford University Press, Oxford, England, pages 115–137.

Dunin-Keplicz, Barbara and Witold

Lukaszewicz. 1986. Towards discourse-oriented nonmonotonic system. In *Proceedings of the 11th International Conference on Computational Linguistics*, pages 504–506, Bonn, Germany, August.

Eckert, Miriam and Michael Strube. 1999. Resolving discourse deictic anaphora in dialogues. In *Proceedings of the 9th Conference of the European Chapter of the Association for Computational Linguistics*, pages 37–44, Bergen, Norway, June.

Firbas, Jan. 1974. Some aspects of the Czechoslovak approach to problems of functional sentence prespective. In F. Daneš, editor, *Papers on Functional Sentence Perspective*. Academia, Prague, pages 11–37.

Gordon, Peter C., Barbara J. Grosz, and Laura A. Gilliom. 1993. Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, 17:311–347.

Grosz, Barbara J. 1977. The representation and use of focus in a system for understanding dialogs. In *Proceedings of the 5th International Joint Conference on Artificial Intelligence*, volume 1, pages 67–76, Cambridge, MA, August.

Grosz, Barbara J., Aravind K. Joshi, and Scott Weinstein. 1983. Providing a unified account of definite noun phrases in discourse. In *Proceedings of the 21st Annual Meeting*, pages 44–50, Cambridge, MA, June. Association for Computational Linguistics.

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

Grosz, Barbara J. and Candace L. Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204.

Haddock, Nicholas J. 1987. Incremental interpretation and combinatory categorial grammar. In *Proceedings of the 10th International Joint Conference on Artificial Intelligence*, volume 2, pages 661–663, Milan, Italy, August.

Hahn, Udo, Katja Markert, and Michael Strube. 1996. A conceptual reasoning approach to textual ellipsis. In *Proceedings of the 12th European Conference on Artificial Intelligence*, pages 572–576, Budapest, Hungary, August.

Hahn, Udo, Peter Neuhaus, and Norbert Bröker. 1997. Message-passing protocols for real-world parsing: An object-oriented model and its preliminary evaluation. In *Proceedings of the 5th International Workshop on Parsing Technologies*, pages 101–112, Massachusetts Institute of Technology, Cambridge, MA, September.

Hahn, Udo, Martin Romacker, and Stefan Schulz. 1999. Discourse structures in medical reports—watch out! The generation of referentially coherent and valid text knowledge bases in MEDSYNDIKATE system. *International Journal of Medical Informatics*, 53(1):1–28.

Hahn, Udo, Susanne Schacht, and Norbert Bröker. 1994. Concurrent, object-oriented natural language parsing: The PARSETALK model. *International Journal of Human-Computer Studies*, 41(1/2):179–222.

Hahn, Udo and Klemens Schnattinger. 1998. Towards text knowledge engineering. In *Proceedings of the 15th National Conference on Artificial Intelligence & the 10th Conference on Innovative Applications of Artificial Intelligence*, pages 524–531, Madison, WI, July.

Hahn, Udo and Michael Strube. 1996. Incremental centering and center ambiguity. In *Proceedings of the 18th Annual Conference of the Cognitive Science Society*, pages 568–573, LaJolla, CA, July.

Hahn, Udo and Michael Strube. 1997. Centering in-the-large: Computing referential discourse segments. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and the 8th Conference of the European Chapter of the Association for Computational Linguistics*, pages 104–111, Madrid, Spain, July.

Hajičová, Eva, Vladislav Kuboň, and Petr Kuboň. 1992. Stock of shared knowledge: A tool for solving pronominal anaphora. In *Proceedings of the 15th International Conference on Computational Linguistics*, volume 1, pages 127–133, Nantes, France, August.

Hardt, Daniel. 1992. An algorithm for VP ellipsis. In *Proceedings of the 30th Annual Meeting*, pages 9–14, Newark, DE, June–July. Association for Computational Linguistics.

Hardt, Daniel. 1996. Centering in dynamic semantics. In *Proceedings of the 16th International Conference on Computational Linguistics*, volume 1, pages 519–524, Copenhagen, Denmark, August.

Hitzeman, Janet. Marc Moens, and Claire Grover. 1995. Algorithms for analysing the temporal structure of discourse. In *Proceedings of the 7th Conference of the European Chapter of the Association for Computational Linguistics*, pages 253–260, Dublin, Ireland, March.

Hoffman, Beryl. 1996. Translating into free word order languages. In *Proceedings of the 16th International Conference on*

Computational Linguistics, volume 1, pages 556–561, Copenhagen, Denmark, August.

Hoffman, Beryl. 1998. Word order, information structure, and centering in Turkish. In M. A. Walker, A. K. Joshi, and E. F. Prince, editors, *Centering Theory in Discourse*, pages 251–271, Oxford University Press, Oxford, England.

Jaeggli, Osvaldo. 1986. Arbitrary plural pronominals. *Natural Language and Linguistic Theory*, 4:43–76.

Kameyama, Megumi. 1986. A property-sharing constraint in centering. In *Proceedings of the 24th Annual Meeting*, pages 200–206, New York, NY, June. Association for Computational Linguistics.

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

Kameyama, Megumi, Rebecca Passonneau, and Massimo Poesio. 1993. Temporal centering. In *Proceedings of the 31st Annual Meeting*, pages 70–77, Columbus, OH, June. Association for Computational Linguistics.

Kamp, Hans and Uwe Reyle. 1993. *From Discourse to Logic. Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Kluwer, Dordrecht.

Kennedy, Christopher and Branimir Boguraev. 1996. Anaphora for everyone: Pronominal anaphora resolution without a parser. In *Proceedings of the 16th International Conference on Computational Linguistics*, volume 1, pages 113–118, Copenhagen, Denmark, August.

Lappin, Shalom and Herbert J. Leass. 1994. An algorithm for pronominal anaphora resolution. *Computational Linguistics*, 20(4):535–561.

Mann, William C. and Sandra A. Thompson. 1988. Rhetorical Structure Theory: Toward a functional theory of text organization. *Text*, 8(3):243–281.

Nagao, Katashi. 1989. Semantic interpretation based on the multi-world model. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, pages 1,467–1,473, Detroit, MI, August.

Prince, Ellen F. 1981. Towards a taxonomy of given-new information. In P. Cole, editor, *Radical Pragmatics*. Academic Press, New York, NY, pages 223–255.

Prince, Ellen F. 1992. The ZPG letter: Subjects, definiteness, and information-status. In W. C. Mann and S. A. Thompson, editors, *Discourse Description: Diverse Linguistic Analyses of a Fund-Raising Text*. John Benjamins, Amsterdam, pages 295–325.

Rambow, Owen. 1993. Pragmatic aspects of scrambling and topicalization in German. In *Workshop on Centering Theory in Naturally-Occurring Discourse*. Institute for Research in Cognitive Science (IRCS), University of Pennsylvania, Philadelphia, PA, May.

Rich, Elaine and Susann LuperFoy. 1988. An architecture for anaphora resolution. In *Proceedings of the 2nd Conference on Applied Natural Language Processing*, pages 18–24, Austin, TX, February.

Sidner, Candace L. 1983. Focusing in the comprehension of definite anaphora. In M. Brady and R. C. Berwick, editors, *Computational Models of Discourse*. MIT Press, Cambridge, MA, pages 267–330.

Strube, Michael. 1996a. *Funktionales Centering*. Ph.D. thesis, Albert-Ludwigs-Universität Freiburg, Freiburg.

Strube, Michael. 1996b. Processing complex sentences in the centering framework. In *Proceedings of the 34th Annual Meeting*, pages 378–380, Santa Cruz, CA, June. Association for Computational Linguistics.

Strube, Michael. 1998. Never look back: An alternative to centering. In *COLING-ACL '98: 36th Annual Meeting of the Association for Computational Linguistics and the 17th International Conference on Computational Linguistics*, Montreal, Quebec, Canada, volume 2, pages 1,251–1,257.

Strube, Michael and Udo Hahn. 1995. PARSETALK about sentence- and text-level anaphora. In *Proceedings of the 7th Conference of the European Chapter of the Association for Computational Linguistics*, pages 237–244, Dublin, Ireland, March.

Strube, Michael and Udo Hahn. 1996. Functional centering. In *Proceedings of the 34th Annual Meeting*, pages 270–277, Santa Cruz, CA, June. Association for Computational Linguistics.

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

Turan, Ümit Deniz. 1998. Ranking forward-looking centers in Turkish: Universal and language specific properties. In M. A. Walker, A. K. Joshi, and E. F. Prince, editors, *Centering in*

Discourse. Oxford University Press, Oxford, England, pages 138–160.
Vallduví, Enric. 1990. *The Informational Component*. Ph.D. thesis, Department of Linguistics, University of Pennsylvania, Philadelphia, PA.
Vallduví, Enric and Elisabet Engdahl. 1996. The linguistic realization of information packaging. *Linguistics*, 34:459–519.
Walker, Marilyn A. 1989. Evaluating discourse processing algorithms. In *Proceedings of the 27th Annual Meeting*, pages 251–261, Vancouver, B.C., Canada, June. Association for Computational Linguistics.
Walker, Marilyn A. 1998. Centering anaphora resolution, and discourse structure. In M. A. Walker, A. K. Joshi, and E. F. Prince, editors, *Centering Theory in Discourse*. Oxford University Press, Oxford, England, pages 401–435.
Walker, Marilyn A., Masayo Iida, and Sharon Cote. 1994. Japanese discourse and the process of centering. *Computational Linguistics*, 20(2):193–233.
Walker, Marilyn A. and Ellen F. Prince. A bilateral approach to givenness: A hearer-status algorithm and a centering algorithm. In T. Fretheim and J. K. Gundel, editors, *Reference and Referent Accessibility*. John Benjamins, Amsterdam, pages 291–306.