Combining Centering-Based Models of Salience and Information Structure for Resolving Intersentential Pronominal Anaphora

Costanza Navarretta
Centre for Language Technology, Copenhagen University

In many computational approaches for resolving intersentential pronominal anaphora, the degree of salience of entities is identified by their degree of givenness in the addressee's discourse model, so that *given* (i.e. known, familiar) entities are assigned the highest degree of salience. The most salient entities are chosen as antecedents of pronouns. Centering-based resolution models also adopt this approach (Grosz *et al.*, 1995). A different point of view is taken by Hajičová *et al.* (1990) who assume that discourse elements in the focal part of an utterance in Information Structure terms have the highest degree of salience. These elements often correspond to new information. Analysing Danish discourse we found that these apparently contrasting interpretations of salience are both valid, but in different contexts. We propose a unified approach combining Centering-based models of salience with Hajičová *et al.*'s proposal.

1 Introduction

Cognitive-based theories on the use of referring expressions presuppose that the speaker makes some assumptions about the status of entities in the addressee's mental state and that these assumptions influence her/his choice of referring expressions, i.a. (Prince, 1981; Ariel, 1994; Gundel *et al.*, 1993; Givón, 1976, 1979, 1983). Although the theories focus on different aspects, they all conclude that pronominal anaphors refer to those entities in discourse which are most easily accessible because they are more *given* (known) in the addressee's discourse model. The different theories classify these entities as being *familiar* (Prince, 1981, 1992), *in focus* (Gundel *et al.*, 1993, 2001), *topic prominent* (Givón, 1979) or *accessible* (Ariel, 1994).

The fact that pronouns usually refer to the most accessible elements in discourse is also presupposed by computational approaches for resolving intersentential pronominal anaphora.[1] In these approaches, the degree of accessibility of elements in the addressee's discourse model is connected to their degree of salience, so that the most salient discourse elements are also the most easily accessible. Inspired by various cognitive-based theories, computational approaches use different models of salience but, as we will discuss in the paper, the majority of models identify high degree of salience

[1] In this paper we only look at entities introduced in discourse by nominals and by pronominal antecedents, we simply mean the nominals in the preceding utterance which corefer with the intersentential anaphors.

with high degree of *givenness*, where *given* means *known*. This is also the case for the various algorithms which have been proposed inside the popular Centering framework (Grosz *et al.*, 1995), which we choose as examples of givenness-based models of salience.

Analysing the use of intersentential pronominal anaphora in Danish written and spoken discourse, however, we found a number of cases, not involving grammatical parallelism, in which nominals providing new information were the preferred antecedents of pronominal anaphors, although they competed with known candidate antecedents. Many of these antecedents providing new information occurred in specific types of syntactic construction and in these constructions the preferred pronominal antecedents were more frequently nominals providing new information than "known" candidate antecedents. The percentage of this phenomenon in the analysed Danish texts and dialogues is given in Section 4. An example of a "not given" nominal being the pronominal antecedent instead of the more known antecedent candidate is given in (1).

(1) [Dommeren i Hørsholm]$_i$ fængslede i onsdags [en 28-årig mand fra Århus]$_k$. [Han]$_k$ sigtes for sammen med [en 44-årig, der blev fængslet før påske,]$_j$ at være gået ind på samlingen og uantastet at have taget de to billeder ned fra væggen.
"[The judge on duty in Hørsholm]$_i$ arrested last Wednesday [a 28-year old man from Århus]$_k$. [He]$_k$ is charged for, together with [a 44-year old man, who was arrested before Easter,]$_j$ having entered the gallery and unchallenged having taken the two pictures from the wall." [BERLINGSKE]

In the example the antecedent of the pronoun *han* (he) is the indefinite object *en 28-årig mand fra Århus* (a 28-year old man from Århus) and not the subject definite candidate *dommervagten i Hørsholm* (the judge on duty in Hørsholm). The latter nominal is considered to be the most given and then also the most accessible element in most models of salience. The only model which assigns the highest degree of salience to elements which are not the most given ones is the one proposed by Hajičová & Vrbová (1982) and Hajičová *et al.* (1990). In this model, Hajičová *et al.* presuppose that elements in the focal part of utterances in Information Structure terms have the highest degree of salience. These focal elements represent, in most cases, new information. Hajičová *et al.*'s proposal is very interesting because it can account for data like the one presented in example (1). However, it is also problematic especially because it is not always true that nominals in the focal part of an utterance are the most salient elements.

We believe that the two apparently contrasting models of salience proposed by Grosz *et al.* (1995) and Hajičová *et al.* (1990) are both valid, but in different contexts. In this paper we present a novel approach combining the two models.

Although in this paper we focus on the degree of givenness of entities in discourse as a measure of their degree of salience, we do not state that givenness is the only factor influencing salience, neither that salience is the only aspect to be considered when resolving pronominal anaphora. This is discussed further in Sections 5 and 7.

The paper is organised as follows. In Section 2, we describe how the degree of salience of discourse elements is modelled in various Centering-based approaches and how all these models assign the highest degree of salience to given elements. In Section 3, we present Hajičová *et al.*'s (1990) model of salience and discuss why it is problematic from an applied point of view.

In Section 4, we look at some examples of Danish discourse containing anaphors whose antecedents are not the most given candidate elements, and discuss how different Centering-based models resolve these anaphors. Finally, we present the results of an analysis of pronominal antecedents in Danish written and spoken corpora. In Section 5, we propose our approach combining a Centering-based model of salience with Hajičová's proposal and describe the results of a survey of the uses of pronominal anaphors which confirm our proposal. In Section 6, we shortly outline other pronominal anaphora resolution approaches which assign high prominence to some types of focal information. Finally, in Section 7, we give a summary of the paper and make some concluding remarks.

2 Modelling Salience in Centering

The Centering theory (Grosz *et al.*, 1995) has been quite influential because of its simplicity and because some of its basic assumptions are quite intuitive, are confirmed by cognitive studies of pronominal anaphors and can account for many anaphoric occurrences as shown by a number of tests on more languages, i.a. (Strube, 1998; Tetreault, 2001). The Centering theory assumes that discourse tends to be "about" few salient entities at a time, the so-called *centres*, and that intersentential pronominal anaphors often refer to the most salient of these centres. The theory presupposes Grosz and Sidner's (1986) three-level discourse model according to which the intentions behind discourse allow to divide discourse in discourse segments which exhibit global coherence. Centering models local reference, i.e. entities inside a discourse segment.

In the original formulation of Centering, Grosz *et al.* (1995) assign to an utterance U_n a set of forward-looking centres, $C_f(U_n)$, corresponding to the entities which can be referred to in the following utterance. The elements in $C_f(U_n)$ are partially ordered according to their prominence (salience). The highest ranked element in $C_f(U_n)$ is called the *preferred centre* $C_p(U_n)$,

following Brennan *et al.* (1987). The highest ranked element in $C_f(U_n)$ which was also realised in the preceding utterance U_{n-1} is called the backward-looking centre $(C_b(U_n))$.

If one of the elements in $C_f(U_{n-1})$ is realised by a pronoun in U_n, then the $C_b(U_n)$ must also be realised by a pronoun.

Grosz *et al.* also assume that, inside a discourse segment, the addressees perceive utterances in which the speakers continue speaking about the same entities as more coherent than utterances in which speakers change the focus of attention. This assumption is implemented by the ranking of center transition states between pairs of utterances. In (Grosz *et al.*, 1995), the highest ranked transition is *centre continuation*. In centre continuation the backward-looking centre is the preferred centre in U_n and co-refers with the backward-looking centre in U_{n-1}, i.e. $C_b(U_n)=C_b(U_{n-1})=C_p(U_n)$. *Centre retainment*, $C_b(U_n)=C_b(U_{n-1})\neq C_p(U_n)$ is ranked less than *centre continuation* in the transition ranking hierarchy, but it precedes *centre shifting*, where the C_b in two adjacent utterances are not the same. The transition state hierarchy is illustrated below:

continue > retain > centre

Figure 1: Hierarchy of transition states

Brennan *et al.* (1987) extend the hierarchy of transition states, but they still assume that centre continuation and centre retainment are more coherent than centre shifting states. In our opinion, the assumption that continuing speaking about the same elements is perceived as more coherent than shifting centre of attention is not very intuitive. It is true that discourse tends to be about some few entities at a time, but it is also natural that speakers change the focus of attention. We will show later in this paper that this shift is often as coherent as centre continuation, because it is announced to the addressee.

Grosz *et al.* (1995) recognise that many factors contribute to the ordering of the forward-looking centres, but for practical reasons all Centering-based algorithms use simple models of salience. Grosz *et al.* rank elements according to their order of occurrence in the utterance. In English, this order often corresponds to the hierarchy of grammatical roles. This hierarchy has been proposed as the preferred syntactic structure for describing the topics in discourse, i.e. the elements discourse "is mainly about", see i.a. Givón's (1979) Topicality hierarchy.

Brennan *et al.* (1987) and Kameyama (1998) adopt the hierarchy of grammatical roles to rank forward-looking centres in their Centering-based

algorithms.[2] The hierarchy of grammatical roles used by Brennan *et al.* is illustrated in Figure 2.

subject > first-object > second-object > other complements > adjuncts

Figure 2: Brennan et al.'s hierarchy of grammatical roles

Strube and Hahn (1996) present a so-called functional Centering model for ordering discourse elements according to their degree of salience. They use the information structure terms proposed by Daneš (1968) who distinguishes between given information, *theme*, i.e. the already known information that the discourse is mainly about, and *rheme* or new information, i.e. information that has just been introduced in discourse.

In Strube and Hahn's interpretation of Centering, the $C_b(U_n)$ corresponds to given information, while the highest ranked element in U_n, the $C_p(U_n)$, is the theme of the utterance. The *theme/rheme* hierarchy is determined by the elements in U_n and U_{n-1}. Elements which are contained in both $C_f(U_{n-1})$ and in $C_f(U_n)$ are thematic and Strube and Hahn call them *bound elements*. Bound elements are ranked higher than rhematic, or *unbound elements*, i.e. elements that are in $C_f(U_n)$ but not in $C_f(U_{n-1})$. Strube and Hahn also propose a ranking order for the various types of bound element, while they rank elements belonging to the same type according to their order of occurrence in the utterance, so the leftmost elements have the highest prominence. The three-levels of ranking in Strube and Hahn's model are given in Figure 3.

bound elements > unbound elements
anaphora > (possessive pronoun *xor* elliptical antecedent)> (elliptical expression *xor* head of anaphoric expression)
nom head 1 > nom head 2 > . . .> nom head n

Figure 3: Ranking of information structure patterns

In all three ranking levels, the elements introduced in discourse earlier, and thus with a higher degree of givenness, are ranked higher than those elements which have just been introduced in it. Strube and Hahn's approach has the advantage of extending the Centering framework to free word-order languages, such as German, where the order of discourse elements does not correspond to their grammatical role in utterances. Although Strube and Hahn assume an information structure based model for measuring salience, they do not consider the possibility of unbound elements having higher degree of salience than bound elements.

[2] Kameyama's (1996) model is much more complex than that proposed by Grosz *et al.* (1995) and Brennan *et al.* (1987). Kameyama distinguishes, among other things, an input and an output attentional state. In her model the hierarchy of grammatical roles is used to rank discourse elements in the output attentional state.

An approach similar to that proposed by Strube and Hahn is followed by Hoffman (1998). She investigates the pronominal anaphors in Turkish taking into account the information structure of sentences. Hoffmann concludes that in Turkish, the backward-looking centre preferentially co-refers with discourse elements in the topic part of utterances.

Another functional model of salience degree is used by Strube (1998). Strube's model is an operationalised version of the *Familiarity scale* proposed by Prince's (1981). In the *Familiarity scale*, Prince models to which degree information is assumed by the speaker to be known to the addressee, where known means familiar in the hearer's model:

Evoked **>** ***Unused*** **>** ***Noncontaining Inferable*** **>** ***Containing Inferable*** **>** ***Brand-New Anchored*** **>** ***Brand-New Unanchored***

Figure 4: Prince's Familiarity scale

Strube's operationalised version of the model is shown below:

OLD (pronominal and nominal anaphors, previously mentioned proper names, relative pronouns, appositives, proper names, titles) **>** **MEDIATED** (inferables) **>** **NEW** (indefinites)

Figure 5: Strube's model

In Strube's model, discourse elements classified as OLD are ranked higher than those classified as MEDIATED and NEW. Two elements of the same type are reciprocally ranked according to their order of occurrence in the utterance, the element mentioned earliest being assigned the highest prominence. Therefore, also in Strube's model given (OLD) elements are always ranked higher than non-given (NEW) elements.

In conclusion, in all the centering-based models of salience we have discussed, the criterion for ranking elements in the utterance connects high degree of salience of discourse elements with high degree of givenness in the addressee's model. Because the different Centering-based models identify degree of givenness with different phenomena, they sometimes rank elements differently. However, none of these models takes into account the fact that speakers can mark as salient elements that do not have a high degree of givenness in the addressee's discourse model.[3] In the following section, we present Hajičová *et al.*'s model of salience where the highest degree of salience of entities is not necessarily connected with the highest degree of givenness.

[3] In our account we have only taken into consideration how Centering-based approaches model the degree of salience of entities. The various approaches also follow different resolution strategies and they do not cover exactly the same types of phenomenon. However, discussing these aspects is out of the scope of the present paper.

3 Hajičová's model of salience

Hajičová and Vrbová (1982) propose a model that assigns the highest degree of salience to discourse elements which represent new information. The model is operationalised in (Hajičová *et al.*, 1990).

Hajičová *et al.* presuppose the information structure of utterances proposed by Sgall *et al.* (1986). Sgall *et al.* recognise a topic/focus dichotomy in sentences, which they call the topic/focus articulation, **TFA**, of the sentences. The terms *topic* and *focus* are used with various meanings not only in different fields, but also in the information structure literature. In the following, we use them as in (Sgall *et al.*, 1986), where *topic* is assumed to correspond to given, known, bound information, the *theme* in (Daneš, 1968), while *focus* corresponds to new, unbound information the *rheme*, or as suggested by Vallduví and Engdahl (1995: 462) to "the informative, newsy, dominant, or contrary-to-expectation part" of an utterance. Thus, focus is used differently than in (Gundel *et al.*, 1993) where elements which are "in focus" are those elements in an utterance whose referents are at the current centre of attention (corresponding to the *focus* of attention in (Grosz & Sidner, 1986)). According to Sgall *et al.*'s (1986) context-unbound nominals are always part of the focus, while context-bound nominals are, in most cases, part of the topic. Elements in the topic or in the focal part of an utterance can be ordered according to their degree of dynamism in the sentence.

In the algorithm proposed by Hajičová *et al.* (1990), the degree of accessibility of elements in the addressee's discourse model is identified with their degree of salience and is implemented by weights. Differing from other salience models, Hajičová *et al.* assign the highest degree of salience to the nominal phrases referred to in the focal part of an utterance *U*, with these phrases being given the highest accessibility weight, *w=max*. The nominal phrases in the topic part of *U* are activated one degree less than those referred to in its focal part, i.e. *w=max-1*. A pronominal reference to an element in the topic part of *U* retains the degree of activation of the element in the discourse model.

The activation of elements not mentioned in *U* fades away and it fades away most quickly for those elements which had the highest activation in the model. Elements whose activation weight differs only by one compete as preferred antecedents of pronominal anaphors. This is exactly the case for antecedents in the focal and in the topic part of *U*.

The suggestion that elements in the focal part of an utterance, which often correspond to new information, have the highest degree of salience

distinguishes Hajičová *et al.*'s approach from other salience models. Hajičová *et al.*'s approach is original and can account for some types of anaphor, such as the one discussed in Section 1, which are often incorrectly resolved by givenness-based salience models. However, Hajičová *et al.*'s proposal is problematic from an applied point of view. In the first place, it is difficult to determine the TFA of all utterances. Secondly, focal candidate antecedents are ranked highest in Hajičová *et al.*'s model, but they compete with topic candidate antecedents in their resolution system, because their accessibility weights differ only by one. Finally, the data does not confirm that all entities referred to in the focal part of an utterance have the highest degree of accessibility. On the contrary, the analysis of the Danish data indicates that there are only a restricted number of constructions where the entities referred to in the focal part of an utterance have so high degree of salience that they should be proposed as the preferred antecedents in applied systems. We discuss these empirical data in the following section.

4 Danish Data

We have analysed the occurrences of intersentential pronominal anaphora in Danish texts and transcriptions of naturally occurring conversations. The texts are taken from newspaper collections, Berlingske Tidende 1992 and 1999, henceforth BERLINGSKE, a collection of computer manuals and novels. The analysed dialogues belong to the BYSOC corpus, collected under "Projekt Bysociolingvistik" (Project Urban Sociolinguistics) (Gregersen & Pedersen, 1991; Henrichsen, 1998), to the PID corpus collected under "Projekt Indvandrerdansk" (Project Immigrant Danish) (Jensen, 1989) and to the corpus "Samtale hos Lægen" ("Talking with the doctor") (Duncker & Hermann, 1996), henceforth SL. The BYSOC and the PID collections consist of the recordings of everyday conversations. The SL collection consists of the recording of 43 dialogues between Danish adult patients and their GPs. In the texts and dialogues, we found a number of intersentential pronominal anaphors with more candidate antecedents where the antecedent chosen by two humans, also on the basis of the context, is the least given nominal according to givenness-based salience ranking. We only considered examples occurring inside discourse segments. In the texts, discourse segments were identified by paragraphs as in i.a. (Strube, 1998; Tetreault, 2001) while in the dialogues they were manually marked mainly on the basis of the context and cue words. As indicated in Section 1, we did not consider cases involving grammatical parallelism, according to which in adjacent utterances with parallel grammatical complements, the preferred antecedent of an anaphor in the second utterance is

the linguistic expression in the first utterance with the same grammatical function. Parallelism has been discussed in i.a. (Asher, 1993; Kameyama, 1996; Kehler, 2000). We discuss examples of parallelism in Section 5.

In (2), examples from the Danish corpora of pronominal antecedents presenting new information are given.

(2) a. *[Chefen]$_i$ fik kun [en søn]$_k$ og [han]$_k$ gad i hvert fald ikke*
[Boss-defin]$_i$ got only [one son]$_k$ and [he]$_k$ wanted surely not
videreføre familieforetagendet.
carry on familybusiness-defin.
"[The boss]$_i$ got only [one son]$_k$ and [he]$_k$ surely did not want to carry on the family business". [SL]

b. *Med [Peter]$_I$ sad der altid [en enkelt mand fra "den lokale" I*
With [Peter]$_I$ sat there always [one man from "the local pub" in
Flensburg]$_k$ og [han]$_k$ var aldrig med til udekampene.
Flensburg]$_k$, and [he]$_k$ came never to away matches-defin.
"There was always sitting [one man from "the local pub" in Flensburg]$_k$ with [Peter]$_i$, and [he]$_k$ never came to the away matches." [BERLINGSKE]

c. *[Igor]$_I$ talte med [en mand]$_k$ udenfor Irma. [Han]$_k$ var stor og havde uredt hår.*
"[Igor]$_I$ spoke with [a man]$_k$ outside Irma. [He]$_k$ was big and had ruffled hair."
[BERLINGSKE]

d. *Og så var der [patient-chaufføren Duddi]$_i$, [der]$_i$ kørte*
And then was there [patient-chauffeur-defin Duddi]$_i$, [who]$_i$ drove
[en mand]$_k$ hjem fra sygehuset. [Han]$_k$ havde været indlagt,
[a man]$_k$ home from hospital-defin. [He]k had been hospitalised,
for [han]$_k$ fik [sin]$_k$ fod i plæneklipperen.
because [he]$_k$ got [his]$_k$ foot in lawn mower-defin.
"And then there was [the patient-chauffeur Duddi]$_i$, [who]$_i$ drove [a man]$_k$ home from the hospital. [He]$_k$ had been hospitalised, because [he]$_k$ had got a foot in the lawn mower." [BERLINGSKE]

e. speaker 1: *hvem hvem arbejdede [din mor]$_i$ med*
whom... whom worked [your mother]$_i$ with
"with whom... whom did [your mother]$_i$ work"
speaker 2: *[Hun]$_i$ arbejdede med [vores nabo]$_k$*
"[She]$_i$ worked with [our neighbour]$_k$"
[Hun]$_k$ var enke ... havde tre sønner
"[She]$_k$ was a widow... had three sons" [BYSOC]

In example (2a), the antecedent of the pronominal anaphor *han* (he) is the indefinite nominal *en søn* (one son), the object of the preceding utterance. The second candidate antecedent is the definite nominal *chefen* (the boss), the subject of the utterance. All centering-based approaches, discussed in Section 2, prefer the definite subject *chefen* (the boss) as antecedent. Grosz *et al.*, (1995) rank it highest because it is the first occurring candidate antecedent in U_{n-1}. Brennan *et al.*, (1987) and Kameyama (1996) rank *chefen* highest because it is

the subject. In Strube and Hahn's (1996) model, context-bound elements are preferred to context-unbound ones as anaphoric antecedents. The definite *chefen* (the boss) is bound while the indefinite *en søn* (a son) is not. Similarly, in Strube's model definite nominals are classified as OLD information, which are ranked higher than indefinite nominals, classified as NEW.

In example (2b), the antecedent of the pronoun *han* (he) is the indefinite nominal *en enkelt mand fra "den lokale" i Flensburg* (a man from "the local pub" in Flensburg) and not the more given proper noun *Peter*. Both Kameyama's and Brennan *et al.*'s models indicate the indefinite nominal as the antecedent because it is the subject of the utterance. In Grosz *et al.*'s model, the topic-fronted nominal[4] *Peter* is chosen as antecedent because it occurs in the utterance before the competing subject nominal. In the models proposed by Strube and Hahn and by Strube, proper names are ranked higher than indefinite nominals and therefore *Peter* is chosen as the pronominal antecedent. Concluding the discussed Centering-based models resolve the anaphor in example (2b) in different ways.

In example (2c), all Centering-based models rank the subject proper name *Igor* higher than the indefinite object *en mand* (a man).

In example (2d), the antecedents of the pronoun *han* (he) is the indefinite object *en mand* (a man) and not the subject relative pronoun *der* (that) co-referring with the nominal *patientchafføren Duddi* (the patient chauffeur Duddi).[5] All Centering-based algorithms rank the subject relative pronoun highest.

In the last example, (2e), the antecedent of the second occurrence of the pronoun *hun* (she) is *vores nabo* (our neighbour), the object in the preceding utterance. Instead, all Centering based models choose as antecedent the entity referred to by the first occurrence of the pronoun *hun*, which co-refers with the nominal *din mor* (your mother). In conclusion, in all the examples in (2), the antecedents of pronominal anaphors are less "given" than the competing candidate antecedents in the utterances. All these less given antecedents occur in the focal part of the utterances. In the first four examples, (2a)-(2d), the focal antecedents are context-unbound elements. In example (2a), the indefinite nominal *en søn* (a son) follows the rhematiser or focusing adverbial *kun* (only) (Quirk *et al.*, 1985). In example (2b), the indefinite nominal *en enkelt mand fra*

4 We propose that topic-fronted nominals – usually not prosodically marked – should be distinguished from topicalised entities, which are prosodically marked. Only the latter are focal elements. This distinction is noticed for Swedish by Vallduví and Engdahl (1995) and, in our opinion, is also valid in Danish.

5 In this example we presuppose that the Centering algorithms are applied to intrasentential clauses as proposed by (Kameyama, 1998). This is not necessary for Strube's (1998) algorithm which applies to both intrasentential and intersentential anaphors.

"den lokale" i Flensburg (a man from "the local pub" in Flensburg) occurs in an existential construction. In examples (2c) and (2d), the focal entity is the indefinite nominal object, *en mand* (a man). In example (2e), the antecedent *vores nabo* (our neighbour) is context-bound, but is less "given" than the competing personal pronoun *hun* (she). Furthermore, the nominal *vores nabo* also presents new information, i.e. provides the information which was asked for in the preceding question and is thus the focus of the utterance. Usually focal information is also prosodically marked in spoken language, see i.a. (Sgall *et al.*, 1986; Vallduví & Engdahl, 1995).

The anaphors in (2) can be accounted for by Hajičová and Vrbová's (1982) and Hajičová *et al.*'s (1990) model which assigns the highest degree of salience to elements in the focal part of an utterance. However, not all elements in the focal part of an utterance are the antecedents of pronominal anaphora. The data indicates that only in specific contexts, nominals in the focal part of an utterance have so high degree of salience that they should be chosen as the preferred anaphoric antecedents. More precisely, the majority of the elements in the focal part of an utterance which are the antecedents of pronominal anaphors in ambiguous contexts, i.e. in contexts with more given candidate antecedents, occurred in a restricted number of construction types in our data. Most of these constructions have also been recognised as focus-marking in the English and/or Danish information structure literature i.a. (Vallduví & Engdahl, 1995; Togeby, 1993; Paggio, 1997; Kruijff, 2001). They comprise *there*-constructions,[6] topicalised constructions, clefts, nominals providing information asked for in the preceding question, nominals preceded by a focusing adverbial, nominal indefinite objects occurring in particular positions in the utterances. In Table 1, the percentage of focal antecedents preferred to more given antecedents in each type of the above constructions in the analysed texts and dialogues is indicated.[7]

Focal antecedents preference

Construction type	texts	Dialogues
There-constructions	98%	97%
Clefts	100%	100%
topicalised constructions	86%	91%
constructions with focusing adverbs	100%	97%
context-marked focus in question/answer	--	100%
constructions with indefinite nominal objects in object-nominal position	59%	65%

Table 1: Focal preference in focal-marked constructions

[6] *Der*-constructions in Danish.

[7] There were two examples in the data where a topicalised nominal preceded a *there*-construction. In one case, the pronominal antecedents were the topicalised nominals; in the other, the indefinite nominals in the *there*-construction.

As the values in the table indicate, focality preference is not equally stronger in all types of construction and is weakest in the case of indefinite nominal objects in object-nominal position.

In the following section, we propose an account of the anaphors in these constructions combining centering with focal preferences.

5 Our Proposal

On the basis of the analysed data, we propose that nominal elements identified as focal in the particular constructions presented in Section 4 not only belong to the focal part of an utterance, but are the main focus of it. The main focus of an utterance may correspond to what Sgall *et al.* (1986) call *focus proper*, i.e. the most dynamic element in the focal part of an utterance, "carrying the intonation centre" (Sgall *et al.*, 1986:178). The focus proper can be considered the opposite of what Sgall *et al.* call the *topic proper*, i.e. the less dynamic element in the topic part of the utterance.

In our opinion, only nominals which are the *focus proper* in an utterance have the highest degree of salience. On the basis of the analysed data we propose the following tentative list of entities which can usefully be ranked as the most prominent candidate antecedents of pronominal anaphors.

1. Entities referred to by nominals which are focally marked structurally. In Danish, structural marking of focus occurs in clefts, existential and topicalised utterances. The preferred antecedent in example (2b) is the indefinite nominal in an existential construction.
2. Entities referred to by nominals that follow focusing adverbs. These adverbs, include additives such as *også* (also) and restrictives such as *kun* (only). The antecedent in example (2a) belongs to this group.
3. Entities focally marked by prosodic marking and/or by the context.[8] This is the case in question/answer pairs as in example (2e). In this example the focus proper, *vores nabo* (our neighbour), provides the information asked for in the preceding question. These types of anaphoric antecedents are quite frequent in our dialogues.
4. Objects which have just been introduced in the discourse by indefinite nominals and which occur in the object-nominal position (Togeby, 1993).[9] Examples (2c) and (2d) belong to this group.

It is relatively easy to recognise the majority of these constructions in Danish. We propose that the accessibility of discourse elements is by default connected with the concept of givenness as assumed by the Centering theory.

[8] It should be noticed that in the dialogue transcriptions used above prosodic information was only available in some cases.

[9] The Danish word order has been described using the so-called *Feltskema* (Field schema) proposed by Diderichsen (1957; (1946)).

This is also the case for entities referred to in the focal part of utterances, which are not the focus proper. The accessibility of given elements, however, is overridden by the accessibility of elements which are the focus proper in the utterances they occur in. In a few cases, the focus proper is indicated by the context, while in the majority of cases, the speakers explicitly change the degree of accessibility of elements in utterances by marking them as salient with information structure related devices. These devices, in Danish as in many other languages, comprise word order, prosodic marking and syntax.

Our tentative list of focus-marking constructions is mainly based on empirical data and it is confirmed by the information structure literature. However, it is not always clear in texts and dialogues whether the focal-marked antecedents are chosen because of their salience or because of the context. To verify our hypothesis that main foci have the highest degree of accessibility we conducted a survey of the use of intersentential pronominal anaphora. In the survey, we isolated groups of preference types in constructed examples. Some of these examples are discussed in Section 5.1. Our work is inspired by Kameyama (1996) who in a survey of English pronouns studies how Centering-based preferences interact with parallelism and common sense knowledge.

5.1 Verifying Our Hypothesis

In our survey of the use of Danish pronouns, we asked 32 native speakers of Danish, the *informants* henceforth, to choose the preferred antecedents in a number of constructed examples. Less than half of the informants were linguists. If the informants could not choose a preferred antecedent, they had to signal this impossibility. Most of the constructed examples are variations of utterances found in our texts or dialogues.

In the survey, among other things, we investigated the relation between givenness and focality preferences in examples where a pronominal anaphor has two competing candidate antecedents, one being the focus proper, the other being a nominal which is more given according to givenness-based definitions of salience. In particular, we considered cases where the focus proper is a NEW entity (an indefinite nominal) and the competing antecedent is an OLD entity (a proper name or a definite nominal), according to the Familiarity scale proposed by Prince (1981) and implemented in Strube's centering algorithm (Strube, 1998). The reason to focus on these cases was that, as indicated by Table 1, focality preference is less strong in cases where a given subject nominal competes with an indefinite object nominal in object-nominal position. We also investigated (i) cases where the NEW focus proper and the OLD candidate antecedents have different syntactic roles and/or occur in different positions and

(ii) the relation between givenness, focality and parallelism. In the survey, we also examined other factors influencing anaphora resolution, such as animacy, recency and lexical knowledge. We do not discuss these aspects in this paper. World knowledge and conventional presuppositions are of course the strongest preference of all, but we attempted to minimize their influence in our examples by constructing as "neutral" and context-isolated utterances as possible.[10]

The survey examples relevant to this paper are listed in the following.

A. *Der sad en mand ved siden af Peter i toget Han så træt ud.*
There sat a man next to Peter on train-defin. He looked tired.
"A man sat next to Peter on the train. He looked tired."

B. *Peter snakkede med en gammel mand i toget. Han så meget sur ud.*
"Peter talked with an old man on the train. He looked very angry."

C. 1. speaker:
Hvem mødte Peter på gaden i går?
"Whom did Peter meet in the street yesterday?"
2. speaker:
Peter mødte Søren. Han havde travlt.
"Peter met Søren. He was busy."

D. *En journalist genkendte forsvarsministeren. Han begyndte at løbe.*
"A journalist recognised the minister of defence. He began to run."

E. *Forsvarsministeren blev genkendt af en journalist. Han var meget overrasket.*
"The minister of defence was recognised by a journalist. He was very surprised."

F. *Peter mødte Søren på gaden. Han hilste på ham.*
"Peter met Søren in the street. He greeted him."

G. *Peter så en mand på gaden. Han råbte til ham.*
"Peter saw a man in the street. He shouted at him."

H. *Peter mødte en mand på gaden. Han hilste på ham.*
"Peter met a man in the street. He greeted him."

I. *Peter mødte en mand i toget. Maria hilste på ham.*
"Peter met a man on the train. Maria greeted him."

In example **A**, the focus proper of the first utterance is the indefinite subject nominal *en mand* (a man) in the *there*-construction. Strube and Hahn's and Strube's algorithm chooses the proper name *Peter* as antecedent of the pronoun *han* (he). Brennan *et al.*'s, Kameyama's and Grosz *et al.*'s algorithms choose *en mand* (a man) as antecedent.

In **B**, the focus proper is the prepositional object *en gammel mand* (an old man), while the given candidate antecedent is the subject *Peter*, which is proposed as antecedent by all Centering algorithms.

[10] The whole survey is described in Navarretta (2002).

In **C**, both candidate antecedents of the pronoun *han* (he) are proper names, thus OLD elements, but the focus proper, *Søren*, occurs after the subject antecedent *Peter*, which is chosen as antecedent in all Centering algorithms.

In example **D**, as in **A**, the subject is an indefinite nominal, but in **D** this subject does not occur in a typical focal position. Brennan *et al.*'s, Kameyama's and Grosz *et al.*'s algorithms choose the indefinite subject nominal *en journalist* (a journalist) as antecedent of the pronoun *han* (he), while in Strube and Hahn's and in Strube's algorithm the preferred antecedent is the definite object *forsvarsministeren* (the ministry of defence).

The first utterance of example **E** contains a passive construction, thus the syntactic subject is not the agent. The agent is expressed by an indefinite nominal. The subject is a definite nominal. Passivisation alters the information structure of the active clause. In all the considered algorithms, the syntactic definite subject *forsvarsministeren* (the minister of defence) is chosen as the antecedent of the pronoun *han* (he).

In examples **F**-**I**, the relation between focality preference and parallelism is investigated.

In example **F**, the two candidate antecedents are both proper names. The first utterance in **F** has the same semantic content as the second utterance in **C**, but the topic/focus articulation of the two utterances is quite different. In examples **G**-**H**, there are two pronominal anaphors in parallel position to the two candidate antecedents in the preceding utterance. One candidate is the focus, the other candidate is a more given entities as in examples **A** and **B**. In example **I**, we investigate a case where parallelism competes with subject-antecedent preference which, according to the results presented in (Kameyama, 1996), can overrule parallelism.

All the Centering-based algorithms choose the subject proper nominal *Peter* as antecedent of the first pronoun *han* (he) in examples **G**-**H**, and of the pronoun *ham* (him) in example **I**, because it is the subject, precedes the other nominal or is the most given candidate antecedent.

5.2 Results of the Survey

The results of the survey are shown in Table 2. In the table the number of informants that chose each candidate antecedent in the examples is shown. The sign '*?*' indicates that the informants could not choose a preferred interpretation. In the last two columns of the table, we give the $\chi^2_{df=1}$ significance and the level of preference *p* for each example. The $\chi^2_{df=1}$ significance is computed by adding an evenly divided number of the answer "unclear" (in Table 2 indicated by "?") to each explicitly selected answer.

Significance is calculated with Pearson's correlation coefficient, see (Woods *et al.*, 1986). Being a two-sided test we have doubled the *p*-value, before calculating significance. Preference is considered significant if $p < .05$, weakly significant if $.05 < p < .10$ and insignificant if $.10 < p$.

The interpretations of examples **A**-**E** confirm that foci have the highest degree of salience. Kameyama (1988) consider empathy as a salience factor explaining the pronominal antecedents in Japanese utterances similar to examples **A** and **B**. Empathy is defined in Kuno (1987: 206) as "the speaker's identification, which may vary in degree, with a person/thing that participates in the event or state that he describes in a sentence". It is possible that the focal nominals in examples **A** and **B** can also be accounted for as cases of empathy, but they are still focal entities. Furthermore, it is not possible to determine cases of empathy in Danish without a deep analysis of discourse. In the paper we have exclusively focused on phenomena which can be identified on the basis of syntactic phenomena.

In example **D**, there is no significant preference for any antecedent. The antecedent subject *en* journalist (a journalist) is an indefinite nominal, thus a newly introduced entity, but it is not the focus proper, as it competes in salience degree with the *given* object *forsvarsministeren* (the minister of defence).

	Answers						Preference	
	1. antecedent		2. antecedent		unclear		$\chi^2_{df=1}$	*p*
A	mand	32	Peter	0	?	0	32	$p < .001$
B	Peter	2	Gammel mand	30	?	0	24.05	$p < .001$
C	Peter	0	Søren	32	?	0	32	$p < .001$
D	journalist	14	forsvarminister	13	?	5	0.03	$.40 < p < .50$
E	forsvarminister	26	journalist	5	?	1	16.53	$p < .001$
F	Peter hilste S.	31	Søren hilste P.	1	?	0	28.12	$p < .001$
G	Peter råbte til m.	27	mand råbte til P.	5	?	0	15.12	$p < .001$
H	P. hilste mand	29	mand hilste P.	2	?	1	22.78	$p < .001$
I	Peter	4	Gammel mand	16	?	12	4.50	$.10 < p < ..20$

Table 2: Survey results

In example **E**, the known subject nominal *forsvarsministeren* (the minister of defence) is preferred to the agent *en journalist* (a journalist) which is unknown, but which has a high status because of its thematic role. This example confirms many cases in the data which indicate that in Danish, the hierarchy of grammatical complements is more relevant to anaphora resolution than the hierarchy of thematic roles. This is also the case in English according to Kameyama (1996). In both **D** and **E**, common sense knowledge may also have influenced the results.

In the examples **F**-**H**, the two pronouns *han* (he) and *ham* (him) are interpreted according to parallelism preference. The answers to **G** and **H**, compared to those in examples **A** and **B**, show furthermore that parallelism preference overrules focality preference. In the interpretation of **I**, parallelism only competes with subject antecedent preference (the preference is not significant).

In conclusion, the results of the survey confirm the hypothesis that focality preference is a stronger preference than givenness and that both preferences are overridden by parallelism. Other cases of parallelism in Danish are discussed in (Navarretta, 2002).

5.3 Combining Givenness and Focality

We have proposed that givenness preference is valid by default when resolving pronominal anaphora. However, this preference is overridden by focality preference, i.e. the salience ranking proposed by a givenness-based model is overridden if there is a focus proper candidate antecedent in the explicitly focally-marked constructions we have tentatively listed. In the following we give an example of how givenness preference, implemented by a Centering-based model of salience, can be combined with focality preference. We use the hierarchy of grammatical complements as a givenness-based model of salience (Brennan *et al.*, 1987). In the model, focality preference is simply expressed by putting the focus proper in front of the list of forward-looking centers $C_f\ (U_{n-1})$ as illustrated in Figure 5.

FOCUS PROPER > SUBJECT > OBJECT/PrepOBJECT > OBJECT 2 > OTHER COMPLEMENTS > ADJUNCTS

Figure 5: Hierarchy of verbal complements with focality preference

The fact that focality preference overrides givenness in determining the degree of salience of elements in an utterance has also consequences for the assumption in Centering that continuing speaking about the same elements in discourse is perceived as more coherent than changing the centre of attention. As described in Section 2, interrupting the centering chain of reference in Centering results in a *shifting* transition state which being assumed to be less coherent than *continuing* and *retaining* transition states is assigned a lower rank than them. In our opinion, this is not correct. In the majority of cases, discourse is coherent and because speakers explicitly change the "focus of attention" using information structure related devices this shift is as coherent as centre continuation. Therefore, coherence in discourse is not only expressed by the fact that speakers continue speaking about the same centres for a while, but also by other phenomena such as information structure or relations holding between

discourse units, which can be used to discover parallelism and other phenomena, as proposed in i.a. (Hobbs, 1979; Asher, 1993; Kehler, 2000). Explicitly focally marked elements in the focal-marking constructions which we have described, can easily be recognised in Danish. However, there are utterances where the main focus can only be identified by analysing the context of discourse. This analysis requires much more sophisticated processing techniques than those used in simple resolution approaches such as Centering-based ones.

6 Related Approaches

Hajičová *et al.*'s proposal that entities referred to in the focal part of utterances have the highest degree of salience is unique. However, the fact that elements in the focal part of an utterance can be very accessible in the addressee's discourse model is confirmed by psycholinguistic experiments conducted by Arnold (1998). Arnold tests the accessibility of focal nominals in clefts and compares their accessibility with the accessibility of subjects which in all linguistic theories are considered to be "very" *given*. Her experiments indicate that both the focus of clefts and the grammatical subject increase the accessibility of their referents. Arnold also investigates reference to foci of clefts and subjects in a corpus. The results of her analysis indicate that the referents of both are highly likely to be referred to again.

Some focal constructions have also been recognised as special in various resolution approaches, although these approaches do not explicitly refer to a general theory of focality preference.

Sidner (1983) suggests that the discourse focus, which in her resolution framework is the most prominent entity at that point of discourse, is explicitly indicated in cleft-, pseudocleft- and *there*-constructions.

Mitkov (1998) recognises topicalisation as one of the many factors influencing anaphora resolution. Furthermore, in his proposal, the objects in a number of verbal constructions in certain types of text also receive a high score as candidate pronominal antecedents.

Fraurud (1992) proposes a simple algorithm for resolving intersentential pronominal anaphora in Swedish. In her proposal, the highest ranked antecedent in a sentence is the subject, but she uses recency as a second ranking criterion. This means that rightmost nominals are preferred to leftmost ones. In Swedish, (Vallduví & Engdahl, 1995), as in Danish (Togeby, 1993; Paggio, 1997), focal elements tend to occur in the final part of the utterance while topic elements preferentially occur in the beginning of an utterance. Hence, Fraurud's

algorithm, in some cases, ranks nominals in the focal part of the utterance higher than nominals in the topic part.

7 *Summary and Concluding Remarks*

The analysis of Danish discourse indicates that entities referred to in the focal part of an utterance in information structure terms are, in some specific cases, more salient than entities referred to in the topic part of the utterance. Nominals in the topic part of an utterance often correspond to the most given entities in the addressee's discourse model. They are preferred as antecedents in most models of salience because the models identify degree of salience with degree of givenness in the addressee's discourse model. In these models, antecedents representing newly introduced information, which correspond to focal information, are assigned the lowest degree of salience. Only few types of focal nominals are sometimes chosen as antecedents in some of the proposed models, not because they are focal, but because they are the leftmost candidate antecedents in the preceding utterance or because they are the subjects of the utterance.

The only model of salience explicitly assigning the highest degree of salience, and then of accessibility, to entities referred to in the focal part of the utterance is that proposed by Hajičová *et al.* However, in Hajičová *et al.*'s model all nominals in the focal part of an utterance are assigned the highest degree of salience. Furthermore, in their resolution system focal elements only compete with given (topic) elements as antecedents of intersentential pronominal anaphora.

Centering-based models identifying degree of salience with degree of givenness do not account for the high prominence of focal elements, but, on the other side, Hajičová's assumption that all focal elements have the highest degree of salience is not always true.

In this paper we proposed that accessibility by default is connected with *givenness* as assumed in Centering, but speakers can explicitly change the degree of accessibility of elements in discourse by marking them as salient with information structure related devices. Only when speakers explicitly mark nominals as the main focus of an utterance, these nominals have the highest degree of salience and can be chosen as antecedents of anaphors. In these cases, the shift of focus of attention is, in our opinion, as coherent as continuing speaking about the same elements, because it is pre-announced to the addressee. Also in this aspect, our proposal departs from the original formulation of Centering. A tentative list of constructions in which this explicit focus-marking

occurs was also given in the paper together with a model of salience combining the Centering-based approach with focality preference.

Although in this paper we have focused on givenness and focality as indicators of the salience of entities in discourse, other aspects such as animacy (see i.a. (Fraurud, 1992)) influence salience. Furthermore, factors such as parallelism and world knowledge are stronger than salience-based preferences and should be applied after salience-based resolution in ambiguous cases, i.a. (Sidner, 1983; Kameyama, 1998). However, because world knowledge and conventional presuppositions require a much more deep analysis of discourse than simple syntactic structure, they are less attractive than salience-based models in applied approaches.

In the last part of the paper, we shortly presented the results of psycholinguistic studies which confirm the hypothesis that some types of foci have high degree of accessibility and we listed resolution algorithms that give high accessibility ranking to some focal phenomena, i.a. (Sidner, 1983; Fraurud, 1992; Mitkov, 1998). Our proposal is new because it generalises the relation between givenness and focality preferences and relate them to parallelism.

We have only analysed Danish data. Our proposal should be verified on other languages.

References

Ariel, M. 1994. "Interpreting Anaphoric Expressions: A Cognitive Versus a Pragmatic Approach." *Journal of Linguistics*, 30(1):3–40.

Arnold, J. F. 1998. *Reference Form and Discourse Patterns*. Ph.D. thesis, Stanford University.

Asher, N. 1993. "Reference to Abstract Objects in Discourse." *Studies in Linguistics and Philosophy* 50. Dordrecht: Kluwer Academic Publishers.

Brennan, S. F., M. W. Friedman and C. J. Pollard 1987. "A Centering Approach to Pronouns." *Proc. of the 25th Annual Meeting of the ACL (ACL'87)*. 155–162. California, Stanford University.

Daneš, F. 1968. "Some thoughts on the semantic structure of the sentence." *Lingua*, 21:55–69.

Diderichsen, P. 1957[1946]. *Elementær Dansk Grammatik*. Copenhagen: Gyldendal.

Duncker, D. and J. Hermann 1996. "Patientord og lægeord - særord eller fællesord?" *Månedsskrift for Praktisk Lægegerning - Tidsskrift for Praktiserende Lægers Efteruddannelse*. 1019–1030.

Fraurud, K. 1992. *Processing Noun Phrases in Natural Discourse*. Department of Linguistics - Stockholm University.

Givón, T. 1976. "Topic, Pronoun and Grammatical Agreement." Charles N. Li (ed.), *Subject and Topic*. New York: Academic Press. 149–188.

---------- 1979. *On Understanding Grammar*. New York: Academic Press.

---------- (ed.) 1983. *Topic Continuity in Discourse: A Quantitative Cross-Language Study*. John Benjamin, Amsterdam.

Gregersen, F. and I. L. Pedersen (eds.) 1991. *The Copenhagen study in urban sociolinguistics*. Reitzel.

Grosz, B. and C. L. Sidner 1986. "Attention, Intentions, and the Structure of Discourse." *Computational Linguistics*, 12(3):175–284.

----------, A. K. Joshi and S. Weinstein 1995. "Centering: A Framework for Modeling the Local Coherence of Discourse." *Computational Linguistics*, 21(2):203–225.

Gundel, J. K., N. Hedberg and R. Zacharski 1993. "Cognitive status and the form of referring expressions in discourse." *Language*, 69(2):274–307.

----------, M. Hegarty and K. Borthen 2001. "Information structure and pronominal reference to clausal introduced entities." *Proc. of the ESSLLI 2001 Workshop on Information Structure, Discourse Structure and Discourse Semantics*, Helsinki. http://www.coli.uni-sb.de/ korbay/esslli01-wsh/index.html.

Hajičová, E. and J. Vrbová 1982. "On the Role of the Hierarchy of Activation in the Process of Natural Language Understanding." J. Horecký (ed.) *Proc. of the 9th Int. Conference on Computational Linguistics*, 107–113, Prague-Amsterdam.

----------, P. Kuboň and V. Kuboň 1990. "Hierarchy of Salience and Discourse Analysis and Production." H. Karlgren (ed.), *Proc. of the 13th Int. Conference on Computational Linguistics (COLING'90)*, volume III, 144–148, Helsinki.

Henrichsen, P. J. 1998. "Peeking Into the Danish Living Room." CST (ed.), *Nodalida '98 Proc. - The 11th Nordic Conference on Computational Linguistics*. Center for Sprogteknologi and Department of General and Applied Linguistics-University of Copenhagen.

Hobbs, J. R. 1979. "Coherence and Coreference." *Cognitive Science*, 3(1):67–90.

Hoffmann, B. 1998. "Word order, information structure and centering in Turkish." M. Walker, A. K. Joshi and E.Prince (eds.), *Centering in Discourse*. Oxford: Oxford University Press.

Jensen, K. A. 1989. *Projekt invandrerdansk*. Technical report, Copenhagen University.

Kameyama, M. 1988. "Japanese zero pronominal binding, where syntax and discourse meet." William Poser (ed.), Papers from the 2nd Int. Workshop on Japanese Syntax. Stanford: CSLI.

---------- 1996. "Indefeasible Semantics and Defeasible Pragmatics." M. Kanazawa, C. Piñon and H. de Stwart (eds.), *Quantifiers, Deduction and Context*, 111–138. Stanford: CSLI.

---------- 1998. "Intrasentential centering: A case study." M. Walker, A. Joshi and E. Prince (eds.), *Centering Theory in Discourse*, pages 89–112. Oxford Univ. Press, Oxford, U.K.

Kehler, A. 2000. "Coherence and the Resolution of Ellipsis." *Linguistics and Philosophy*, 23:533–575.

Kruijff, G.-J. M. 2001. *A Categorial-Modal Architecture of Informativity: Dependency Grammar Logic and Information Structure*. Ph.D. thesis, Charles University, Prague.

Kuno, S. 1987. *Functional syntax: anaphora, discourse, and empathy*. Chicago: University of Chicago Press.

Mitkov, R. 1998. "Robust pronoun resolution with limited knowledge." *Proc. of the 36th Annual Meeting of the ACL and the 17th Int. Conference on Computational Linguistics (COLING-ACL'98)*, 869–875, Montreal.

Navarretta, C. 2002. *The Use and Resolution of Intersentential Pronominal Anaphora in Danish Discourse*. Ph.D. thesis, Center for Language Technology and Department of General and Applied Linguistics, Copenhagen University.

Paggio, P. 1997. *The Treatment of Information Structure in Machine Translation*. Ph.D. thesis, University of Copenhagen.

Prince, E. F. 1981. "Toward a taxonomy of given-new information." P. Cole (ed.), *Radical Pragmatics*, 223–255. New York: Academic Press.

---------- 1992. "The ZPG Letter: Subjects, Definiteness, and Information-status." W. Mann and S. A. Thompson (eds.), *Discourse Description. Diverse Linguistic Analyses of a Fund-Raising Text.*, 295–325. Amesterdam: John Benjamins.

Quirk, R., S. Greenbaum, G. Leech and J. Svartvik 1985. *A Comprehensive Grammar of the English Language*. Longman.

Sgall, P., E. Hajičová and J. Panevová 1986. *The Meaning of the Sentence in its Semantic and Pragmatic Aspects*. Dordrecht: Reidel.

Sidner, C. 1983. "Focusing in the Comprehension of Definite Anaphora." M. Brady and R. Berwick (eds.), *Computational Models of Discourse*, 267–330. MIT Press.

Strube, M. and U. Hahn 1996. "Functional Centering." *Proc. of the 34th Int. Conference on Computational Linguistics (ACL'96)*, 270–277, Santa Cruz, CA.

---------- 1998. "Never Look Back: An Alternative to Centering." *Proc. of the 36th Annual Meeting of the ACL and the 17th Int. Conference on Computational Linguistics (COLING-ACL'98)*, vol. II, 1251–1257, Montreal, Quebec, Université de Montréal.

Tetreault, J. R. 2001. "A Corpus-Based Evaluation of Centering and Pronoun Resolution." *Computational Linguistics*, 27(4):507–520.

Togeby, O. 1993. *PRAXT - Pragmatisk tekstteori 1-2*. Aarhus Universitetsforlag 1993.

Vallduví, E. and E. Engdahl 1995. "The linguistic realization of information packaging." *Linguistics*, 34(3):459–519.

Woods A., P. Fletcher and A. Hughes 1986. "Statistics in Language Studies." *Cambridge Textbooks in Linguistics*. Cambridge: Cambridge University Press.