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

Costanza Navarretta

Center for Sprogteknologi
Njalsgade 80, 2300 Copenhagen S, Denmark
costanza@cst.dk

Abstract

In many approaches for resolving intersentential pronominal anaphora the degree of salience of entities determines their accessibility in the addressee's discourse model. Degree of salience is identified by the degree of givenness of entities in the addressee's cognitive model, so that given/known entities have highest degree of salience. The most given entities are chosen as antecedents of pronouns. Centering-based resolution models also adopt this approach. A different point of view is taken by Hajičovà and Vrbovà (1982), who suggest that entities in the focal part of an utterance in Information Structure (IS) terms have highest degree of salience. These entities often correspond to new information. Analysing Danish discourse we have found that these apparently contrasting interpretations of salience are both valid, but in different contexts. We propose a model of accessibility where salience is by default connected with givenness, but where this default can be explicitly overruled in contexts where the speaker explicitly marks entities as salient by IS-related devices. These explicitly marked focal entities have highest degree of salience. A tentative list of the IS-related devices and contexts in which the shift of attentional focus is explicitly marked in Danish is proposed, based on naturally occurring discourse and IS-literature.

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). Although the theories focus on different aspects, all recognise that pronominal anaphors refer to those entities in discourse, which are easiest accessible in the addresse's discourse model.

The fact that pronouns usually refer to the most accessible entities in discourse is also presupposed by computational approaches for resolving intersentential pronominal anaphora resolution. In these approaches the degree of accessibility of entities in the addresse's discourse model is often connected to their salience, so that the most salient entities are also the easiest accessible. Inspired by various cognitive-based theories, computational approaches use different models of salience.

That discourse "is about" few salient (*central*) entities at a time is also one of the main assumptions behind the popular Centering theory (Grosz et al., 1986; 1995). The theory assumes the discourse structure described in (Grosz and Sidner, 1986) and accounts for local reference, i.e. reference among utterances inside a discourse segment. Centering models the fact that intersentential pronominal anaphors tend to refer to the most salient entities at each point of discourse. Thus determining the degree of salience of entities is fundamental in the theory. Inside the Centering framework various models of salience have been proposed. Common to all these models is the fact that entities that have been dealt with for a while in discourse, and thus are *known* or *given* in the addressee's mental model, are considered to have highest degree of salience.

Analysing the use of intersentential pronominal anaphora in Danish written and spoken discourse, we found a number of cases in which newly introduced entities were the antecedents of pronominal anaphors, although there were *known* candidate antecedents at that point of discourse. A model of salience that considers *new* information as being most salient is that proposed by Hajičová and Vrbová (1982) and operationalised by Hajičová et al. (1992).

Also Hajičová et al. (1992) identify the degree of accessibility of entities in the adressee's discourse model with their degree of salience. However, differing from other approaches, Hajičová et al. state that entities referred to in the focal part of an utterance in information structure (IS) terms[1] have highest degree of salience and thus are assigned the highest accessibility value. The theory of information structure which Hajičová et al. presuppose is that described in (Sgall et al., 1986). Because entities in the focal part of an utterance in IS terms often, although not necessarily, correspond to *new* information, Hajičová et al.'s (1982) proposal departs from the view followed in the Centering framework, where high degree of salience equals *givenness* in the addressee's mental model. Although the proposal can account for the Danish data, in its original formulation it is problematic from a computational point of view. Firstly, it is not always easy to determine which entities are in the focal part of an utterance. Secondly, although Hajičová et al. assign highest degree of salience to entities in the focal part of an utterance, they do not assume that these entities are the preferred antecedents of successive pronouns. In cases where there are both *known* candidate antecedents and less known, but focal candidate antecedents Hajičová et al. only recognise the ambiguity, although they rank focal antecedents highest.

The analysis of the Danish data indicates that there are only a number of cases where the entities in the focal part of an utterance have so high degree of salience, that they

---

[1] The word *focus* is used with different meanings in different fields ando also in the IS literature. In this paper we use it as proposed by Sgall et al. (1986) or by Vallduví and Engdahl (1995). The latter define *focus* as "the informative, new, dominant, or contrary-to-expectation part" of an utterance.

should be proposed as the preferred antecedents in applied systems. These constructions have been identified in the IS literature as explicitly indicating focal information in English and/or in Danish. In the remaining cases *known* antecedents should be preferred to antecedents in the focal part of the utterance, in line with the strategy proposed in Centering-based approaches.[2]

The paper is organised as follows. In section 2 we describe how the degree of salience of entities is modelled in different Centering-based algorithms and how the models assign highest degree of salience to known entities. In section 4 we present examples of Danish discourse where focus-marked entities are the preferred antecedents to pronouns while in section 4 we shortly present Hajičová et al.'s (1992) model of salience. In section 5 we propose a more restrictive definition of focal entities having the highest degree of salience than that proposed by Hajičová et al. In the section we also discuss how this model fits into the Centering framework. In section 6 we shortly outline other studies of pronominal anaphora resolution, which, although indirectly, use IS-related aspects in their resolution strategy. Finally, in section 7, we 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, such as that discourse tends to be about few salient entities at a time, the *centers*, and that intersentential pronominal anaphors often refer to the most salient of these centers, are quite intuitive.

Grosz et al. assign to an utterance U_{n-1} a set of *forward-looking centers* ($C_f(U_{n-1})$), corresponding to the entities which can be referred to in the following utterance. The elements in $C_f(U_{n-1})$ are partially ordered according to their prominence. The highest ranked element in $C_f(U_{n-1})$ is called the preferred center $C_p(U_{n-1})$, follwing Brennan et al. (1987). The highest ranked element in the set which is realised in the following utterance U_n is called the *backward-looking center* ($C_b(U_n)$). 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.

The assumption that people continue to speak about the same entities for a while is also reflected in the ranking of center transition states between pairs of utterances. The highest ranked transition is center continuation where $C_b(U_n)$ is both the preferred center in U_n and the same center as in U_{n-1}, i.e. $C_b(U_n)= C_b(U_{n-1})= C_p(U_n)$. Center retainment, $C_b(U_n)= C_b(U_{n-1}) \neq C_p(U_n)$, follows in the transition ranking hierarchy and precedes center shifting, where the C_b in two adjacent utterances is not the same.

Grosz et al. (1995) recognise that many factors contribute to the ordering of the forward-looking centers, In practice all Centering-based algorithms use simple models of salience. Grosz et al. rank entities according to their order of occurrence in the utterance. In English this order often corresponds to the hierarchy of grammatical roles which has been proposed as the preferred syntactic

structure for describing the topics in discourse, i.e. what the discourse "is mainly about" (see i.a. Givón (1979) with his Topicality hierarchy).

Brentan et al. (1987) and Kameyama (1998) use such a hierarchy of grammatical roles to rank entities according to their salience in their versions of the Centering framework. The hierarchy of grammatical roles used by Brennan et al. is illustrated in figure 1.

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

Figure 1. The hierarchy of grammatical roles

Strube and Hahn (1996) present a so-called functional Centering model for ordering entities according to their salience degree. They use the IS terms proposed by Daneš (1974) who distinguishes between *given* information, *theme*, i.e. the given information that the discourse is mainly about[3], and *rheme/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 Cf (Un-1) and in Cf (Un), are thematic and are called bounded elements. They are ranked higher than rhematic elements, called unbounded 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 different types of bound element. Finally elements of the same type are ranked according to their order of occurence in the utterance, so the leftmost elements are ranked highest. Thus, also in Strube and Hahn's model entities introduced in discourse earlier are ranked higher than entities, which have just been introduced in it. Strube and Hahn's approach has the advantage of extending the Centering framework to a free-order language, such as German, where the order of entities does not correspond to their grammatical role. But, although Strube and Hahn assume an IS-based model for measuring salience, they do not consider the possibility of unbound entities having higher degree of salience than bound entities. A similar approach for Turkish is taken by Hoffman (1998), who looks at the IS of utterances, but also concludes that the *backward-looking center* preferentially corefers with entities in the topic part of utterances.

Strube (1998) proposes a similar model of salience degree inside the Centering framework. The model is an operationalised version of Prince's (1981) **Familiarity scale**. In her scale Prince models to which degree information is assumed by the speaker to be known to the addressee. Prince's Familiarity scale is shown in figure 2.

Evoked > Unused > Noncontaining Inferable > Containing Inferable > Brand-new Anchored > Brand-new Unanchored

Figure 2. Prince's familiarity scale

---

[2] *Antecedents* of pronominal anaphors in these paper are simply those nominals which immediately corefer with the anaphors in the preceding discourse.

[3] The *theme* is called the *topic proper*, by Sgall et al. (1986).

Strube's operationalised version of the model is shown in figure 3.

OLD (pronominal and nominal anaphors, previously mentioned proper names, relative pronouns, appositives, proper names and titles) > MEDIATED (inferables, Brand-New Anchored) > NEW (indefinite nominals)

Figure 3. Strube's version of the familiarity scale

Entities classified as OLD are ranked higher than those classified as MEDIATED and NEW. Two entities of the same type are reciprocally ranked according to their order of occurence in the utterance, the entity mentioned earliest being assigned highest prominence.

Concluding, in all the presented centering-based models the major criterion for ranking entities according to their degree of salience in the utterance is to which extent they are known in the adressee's model. Because the models identify degree of givenness with different phenomena, however, they do in some cases, rank entities differently in the same contexts. However, no model takes into account the fact that the speaker can explicitly mark as salient entities that have a low degree of givenness by IS-related devices.[4]

In the following section we present and discuss Danish examples illustrating that entities with low degree of givenness in particular cases are the preferred pronominal antecedents.

3. Danish Data

Analysing the use of intersentential pronominal anaphora in a number of Danish written texts and transcriptions of naturally occurring conversations, [5] we found a number of intersentential pronominal anaphors with more candidate antecedents. According to the Centering-based models, one of these antecedents was highest ranked being more *given* in the adresse's mental model. However in many cases, the antecedent preferred by two humans, was the latess "known" entity, which also corresponded to the focal part of the utterance in IS-terms. [6]

In the following we discuss some examples from the data, illustrating this phenomenon. The correct antecedent

of each pronominal anaphors is considered the one identified by two humans, also on the basis of the context.

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

2. *Med Peter$_i$ sad der altid [en enkelt mand fra "den lokale" i Flensburg]$_k$, og han$_k$ var aldrig med til udekampene.*
   (lit. With Peter$_i$ sat there always [a single man from "the local" in Flensburg]$_k$, and he$_k$ never came to away matches-defin)
   (With Peter$_i$ [a single man from "the local" in Flensburg]$_k$ was always sitting, and he$_k$ never came to the away matches.)
   Berlingske

3. *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

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

5. A: *hvem...hvem arbejdede [din mor]$_i$ med?*
      (lit. with whom... whom worked [your mother]$_i$)
      (with whom... whom did [your mother]$_i$ work)
   B: *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 1 the antecedent of the pronominal anaphor *han* (he) is the indefinite nominal *en søn* (a 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, which we discussed in section 2, prefer *chefen* (the boss) as antecedent. Grosz et al. (1995) rank it highest because it occurs before the other nominal in the utterance. Brennan et al. (1987) and Kameyama (1996) rank it highest because it is the subject of the utterance. Strube and Hahn, instead, prefer context-bound entities, in this case the definite nominal *chefen*, to context-unbound ones, the indefinite nominal *en søn*. Similarly in Strube's model a definite nominal is an OLD entity, which is preferred to the indefinite nominal, classified as NEW.

---

[4] It must be noticed that in this section we have only taken into consideration how salience of entities is modelled in the various Centering-based approaches. The various approaches have also different resolution strategies and the number of phenomena they cover is not the same, but it is out of the present scope to discuss these differences.

[5] The texts we analysed were taken from a newspaper collection, Berlingske Tidende 1999, henceforth Berlingske. The dialogues belong to the Bysoc corpus, collected in 1987 under "Projekt Bysociolingvistik" (Project Urban Sociolinguistics) (Gregersen et al., 1991; Henrichsen,1998) and to the collection "Projekt Indvandrerdansk" (Project Immigrant Danish) collected in 1987, henceforth Pid (Jensen, 1989).

[6] We did not consider utterances involving grammatical parallelism, because it seems that in Danish, as in English, parallelism is a stronger preference than preferences related to the attentional state as it is the case for Centering-based salience models, see especially (Kameyama, 1996).

In example 2 the antecedent of the pronoun *han* (he) is the indefinite nominal *en enkelt mand fra "den lokale" i Flensburg* (a single man from "the local" in Flensburg). In both Kameyama's and Brennan et al.'s models this nominal is the preferred antecedent because it is the subject of the utterance, while in Grosz et al.'s model the topic-fronted nominal, *Peter*, would be chosen. Also in the models proposed by Strube and Hahn and by Strube, the entity referred to by a proper name is ranked higher than an indefinite nominal and is then chosen as antecedent of the pronoun. In Strube and Hahn's model in fact

In example 3 all the Centering-based models rank the subject proper name, *Igor*, higher than the indefinite object *en mand* (a man) which was chosen as antecedent by the two humans.

In example 4 there are two candidate antecedents of the pronoun *han* (he). [7] The first antecedent is the relative pronoun *der* (that) coreferring with the nominal *patient-chafføren "Duddi"* (the patient chauffeur "Duddi"). The pronouns is the subject of the relative clause. The second antecedent is the indefinite nominal *en mand* (a man), the object of the relative clause. The Centering-based strategies rank the subject relative pronoun highest, while the humans chose as antecedent the indefinite nominal *en mand*.

In the last example, 5, the preferred antecedent of the second occurrence of the pronoun *hun* (she) is the object in the preceding utterance, *vores nabo* (our neighbour). Instead, all Centering-based models choose as antecedent the entity referred to by the first occurrence of the pronoun *hun*, which corefers with the nominal *din mor* (your mother).

In all the discussed examples the pronominal antecedents proposed by the humans are the focus of their utterance. In all cases, except example 5, these entities are also context-unbound.

In examples 1 the indefinite nominal *en søn* (a son) is preceded by the rhematiser or focusing (Quirk et al., 1985) adverbial *kun* (only). In example 2 the indefinite nominal *en enkelt mand fra "den lokale" i Flensburg* (a single man from "the local" in Flensburg) occurs in an existential construction. In examples 3 and 4 the focal entity is the indefinite nominal object, while in example 5 the nominal *vores nabo* (our neighbour) provides the information which is asked for in the preceding question. Usually this information is also prosodic marked in spoken language, see i.a. (Sgall et al. 1986; Vallduví and Engdahl, 1995).

These examples seem to confirm Hajičová's (Hajičová and Vrbová, 1982; Hajičová et al. 1990) proposal that entities in the focal part of an entity have highest degree of salience.

4. Hajičová's model of salience

Hajičová proposes a model of degree of salience, which takes into account the IS of utterances, as defined by Sgall et al. (1986). Sgall et al. recognise a topic/focus dychotomy in sentences, which they call the topic/focus articulation, TFA, of the sentence. According to their proposal 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 particular Sgall et al. (1986) distinguish a *focus proper*, i.e. the most dynamic entity in the focal part, "carrying the intonation center" [p.178] and a *topic proper*, the less dynamic element in the topic part of the sentence corresponding to the *theme* in Daneš terminology.

Hajičová et al. define salience as activation or foregrounding of the elements in the hearer's and speaker's shared discourse model (SM), so that the items with highest degree of salience are the most easily accessible in memory. The nominal phrases referred to in the focal part of an utterance *U* are assigned the highest degree of salience. The nominal phrases in the topic part of *U* are activated one degree less than those referred to in the focal part. A pronominal reference to an element in the topic part of *U* retains the degree of activation of the element in SM. 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 SM. Elements whose degree of activation differs only by 1 compete as preferred antecedents of pronominal anaphors and these is the case for antecedents in the focal and in the topic part of *U*.

The fact that entities in the focal part of an utterance, which often correspond to unbound entities, have highest degree of salience distinguishes the approach presented by Hajičová et al. from the majority of the salience models proposed in the literature. However, as discussed in section 1, the model of salience proposed by Hajičová is problematic from an applied point of view, because it is difficult to determine the TFA structure of utterances and because they only recognise that focal candidate antecedents compete with known antecedents in case of ambiguous contexts. Another problem with Hajičová's proposal is that it states that all entities in the focal part of discourse have highest degree of salience. The analysis of Danish discourse indicates that this is not always the case. We discuss this point further in the following section.

5. Salient focal entities

The analysis of the Danish data indicates that only entities which are explicitly focally marked by the speaker have highest degree of salience and thus are the most probable antecedents of intersentential pronominal anaphors. These entities correspond to what Sgall et al. (1986) call the focus proper. Although prosodic information is not relevant in written language and this information was not available in the dialogue transcriptions we used, all the focal entities which were chosen as antecedents in the data occurred in constructions which in the English and/or Danish IS-literature have been recognised as explicitly indicating the focus of utterances. On the basis of the analysed data and of especially Danish IS studies (Togeby 1992; Paggio, 1997) we propose the following tentative list of explicitly focally marked entities which can usefully be ranked highest among candidate antecedents of pronominal anaphors.

- Entities referred to by nominals which are focally marked structurally. In Danish structural marking of focus occurs in clefts, existential and

---

topicalised utterances.[8] The preferred antecedent in example 2 is the indefinite nominal in an existential construction.

- 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 1 falls under this category.
- Entities marked as focus proper by prosodic marking and/or by the context. This is clearly the case in question/answer pairs as illustrated in example 5. Here the focus proper *vores nabo* provides the information asked for in the preceding question. These cases were quite frequent in the transcriptions of dialogues.
- Objects which have just introduced in the discourse by indefinite nominals. More precisely Togeby (1993) proposes that indefinite nominals occurring in the object-nominal position[9] in the Danish utterance are usually the focus of the utterance. In this position occur the first object and obliques. Examples 3 and 4 belong to this group.

These explicitly marked entities correspond to what Sgall et al. (1986) call *focus proper* (section 3), i.e. the most dynamic element in the utterance.

It is possible to deal with this phenomenon in the Centering framework. In particular we propose that accessibility by default is connected with the concept of givenness as assumed by the Centering theory. This is also the case for entities in the focal part of utterances, which are not the focus proper. However, speakers can explicitly change the salience of entities in discourse by marking entities as salient with IS-related devises. These devises in Danish have been mainly recognised in word order, prosodic marking and syntax.

In the original formulation of the Centering theory interrupting the centering chain of reference is possible, but it results in a shifting transition state which is less coherent and then has a lower rank than continuing and retaining transition states (see section 2). Thus a more "given" antecedent is always preferred to a less "given" one in ambiguous contexts.

In our opinion when speakers change the "focus of attention" (using Grosz and Sidner's (1986) terminology) by IS-related devices this shift is as coherent as the states in which speakers continue speaking dealing about the same entities, because the shift is explicitly announced to the adressee. Thus coherence in discourse is not only given by the fact that speakers continue speaking about the same centers for a while, but can also be given by IS-related devises which inform the adressee that the center is changing. Focus marked by particular syntactic constructions can be easily recognised in practice. However, in some cases, the focus proper can only be identified by analysing the context of discourse, which

requires much more sophisticated processing techniques than those used in simple resolution approaches. Finally we must notice that although we propose that information on the focus proper of utterances can improve the resolution strategy of centering-based algorithms, we do not suggest that centering and focus marking are the only coherence signs in language. Factors such as the relations holding between discourse units, are also important coherence markers, see i.a. (Hobbs, 1979, Asher 1993) indispensable to discover phenomena such as parallelism, which also influences anaphora resolution.

6. Related approaches

Hajičová et al.'s proposal that entities in the focal part of discourse have highest degree of salience is unique. However, some construction types identified as focal markers in the IS literature have been recognised as special in quite different approaches to pronominal anaphora resolution, although this approaches do not refer to the IS of utterance explicitly.

Sidner (1983) states 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[10] as one of the many factors influencing anaphora resolution. Furthermore in his proposal the objects in particular verbal constructions in certain types of texts are also the preferred pronominal antecedents.

Fraurud (1992) proposes a simple algorithm for resolving intersentential pronominal anaphors in Swedish. In her proposal the highest ranked antecedent in a sentence is the subject, as in many other approaches. However, recency is the second criterion for ranking antecedents, so that the most recent antecedent with respect to the anaphor is chosen, unless it is competing with a subject candidate. Thus rightmost nominals are preferred to leftmost ones. In Swedish (Vallduví and Engdahl, 1995), as in Danish, (Togeby, 1993; Paggio, 1997) focal entities tend to occur in the final part of the utterance. Topic entities, instead, preferentially occur in the beginning of an utterance. Thus, in practice, Fraurud's algorithm often ranks entities in the focal part of the utterance higher than entities in the topic part.

7. Concluding Remarks

The analysis of Danish discourse indicates that entities in the focal part of an utterance in IS-terms are in some specific cases more salient than entities in the topic part of utterance. Nominal in the topic part of an utterance often correspond to the most *known* entities at that point in the addressee's discourse model. They are preferred as antecedents in most models of salience of entities, here included models in the Centering framework, which identify degree of salience with degree of givenness in the adressee's discourse model. In these models antecedents representing newly introduced information, which correspond to focus information in IS, are instead assigned lowest degree of salience. Only few types of focal entities are chosen as antecedents in some of the proposed models,

[8] In Danish, as in Swedish (Vallduví and Engdahl, 1995) one should distinguish between topic-fronted entities, which are not the focus of the utterance and topicalised entities, which are. The latter are usually prosodic marked, whereas the former are not.
[9] The Danish word order has been described using a so-called Field Schema *Feltskema* (Diderichsen, 1957). Various clausal constituents occupy precise positions in the schema.

[10] Also *topicalisation* is used in literature with different meanings. Mitkov uses it as the phenomenon of emphasising or contrasting some information.

because the focal entities occur in the front of the utterance or are the subjects (existential constructions and front-positioned topicalisation).

The only model of salience stating that the entities in the focal part of the utterance have highest degree of attention is that proposed by Hajičová, but in her model all focal entities compete with more given entities as antecedents of intersentential pronominal anaphora.

In this paper we proposed that Centering-based models identifying degree of salience with degree of givenness are also valid in Danish with the exception of constructions where the speaker explicitly marks entities as salient by IS-related devises. These entities correspond to what Sgall et al. call focus proper, i.e. the most dynamic focal entities in the utterance. These shifts in attention are as coherent as centering continuation because the speaker marks them explicitly in the utterance preceding the intersentential anaphor. Finally in the paper we also proposed a tentative list of constructions which in Danish indicate the focus proper nominal.

8. References

Ariel, M., 1994. Interpreting anaphoric expressions: a cognitive versus a pragmatic approach, *Journal of Linguistics* 30(1),3-40.

Asher, N., 1993. *Reference to Abstract Objects in Discourse*. Vol. 50 of Studies in Linguistics and Philosophy, Kluwer Academic Publishers, Dordrecht, the Netherlands.

Brennan, S.F., M.W. Friedman, C.J. Pollard, 1987. A Centering Approach to Pronouns. In *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics (ACL'87)*, Stanford University, California, USA, pp.155-162.

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

Diderichsen, P., 1957. *Elementær Dansk Grammatik*, Gyldendal, Copenhagen.

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

Givón, T., 1979. *On Understanding Grammar*, Academic Press, New York, N.Y.

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

Grosz, B., A.K. Joshi, S. Weinstein, 1986. Towards a computational theory of discourse interpretation. Unpublished ms.

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

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

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

Hajičová, E. and J. Vrbová, 1982. On the Role of the Hierarchy of Activation in the Process of Natural Language Understanding. In J. Horecký (ed.), *Proceedings of the Ninth International Conference on Computational Linguistics* (pp.107-113), Prague-Amsterdam.

Hajičová, E., P. Kubon, V. Kubon, 1990. Hierarchy of Salience and Discourse Analysis and Production. In H. Karlgren (ed.), *Proceedings of the 13th International Conference on Computational Linguistics* - COLING' 90, III-144-148, Helsinki.

Hobbs, J.R., 1978. *Why Is Discourse Coherent?* Technical Note 176, AI Center, SRI International, 333 Ravenswood Ave., Menlo Park, CA 94025.

Hoffmann, B., 1998. Word Order, Information Structure, and Centering in Turkish. In M.Walker, A. Joshi E. Prince (eds.), *Centering Theory in Discourse*, Oxford University Press, Oxford, UK.

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

Kameyama, M., 1996. Indefeasible Semantics and Defeasible Pragmatics. In M. Kanazawa, C. Pinon, H. deStwart (eds.), *Quantifiers, Deduction and Context*, CSLI, Stanford, CA (pp. 111-138).

Kameyama, M., 1998. Intrasentential centering: A case study. In M.Walker, A. Joshi E. Prince (eds.), *Centering Theory in Discourse*, Oxford University Press, Oxford, UK, pp.89-112.

Mitkov, R., 1998. Robust pronoun resolution with limited knowledge. In Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and the 17th International Conference on Computational Linguistics (COLING-ACL'98), Montreal, Canada (pp. 869-875).

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

Prince, E.F., 1981. Toward a taxonomy of given-new information. In P. Cole (ed.), *Radical Pragmatics* (pp. 223-255), Academic Press.

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

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

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

Strube, M., 1998. Never Look Back: An Alternative to Centering. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and the 17th International Conference on Computational Linguistics (COLING-ACL'98)*, Vol.II, Université de Montréal, Montreal, Quebec, Canada, pp.1251-1257.

Strube, M. and U. Hahn, 1996. Functional Centering. In Proceedings of and the 34th International Conference on Computational Linguistics (ACL'96), Santa Cruz, Ca (pp. 270-277).

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

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