An Algorithm for Resolving Individual and Abstract Anaphora in Danish Texts and Dialogues

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

Abstract

This paper describes the DAR-algorithm for resolving intersentential pronominal anaphors referring to individual and abstract entities in Danish texts and dialogues. Individual entities are resolved combining models which identify high degree of salience with high degree of givenness (topicality) of entities in the hearer's cognitive model, e.g. (Grosz et al., 1995), with Hajičová et al.'s (1990) salience account which assigns the highest degree of salience to entities in the focal part of an utterance in Information Structure terms. These focal entities often introduce new information in discourse. Anaphors referring to abstract entities are resolved with an extension of the algorithm presented by Eckert and Strube (2000). Manual tests of the DAR-algorithm and other well-known resolution algorithms on the same data show that DAR performs significantly better on most types of anaphor.

1 Introduction

Most intersentential anaphor resolution algorithms exclusively account for pronominal anaphors with individual nominal antecedents (henceforth IPAs) in texts. Less attention has been given to pronominal anaphors which refer to abstract entities evoked by verbal phrases, clauses or discourse segments (henceforth APAs). However APAs are quite common in English dialogues, see i.a. (Byron and Allen, 1998). Recently two algorithms for resolving APAs and IPAs in specific English dialogues have been proposed: Eckert and Strube's (2000) ES00, Byron's (2002) PHORA. APAs are also frequent in Danish. We found that 15% of all pronominal anaphors in our texts were APAs, while they constituted 48% of the anaphors in the analysed dialogues. Furthermore third-person singular pronouns in neuter gender which can be IPAs or APAs were APAs

in two-third of the cases in both texts and dialogues.

In this paper we describe an algorithm, called DAR, for resolving intersentential IPAs and APAs in Danish.[1] Unlike ES00 and PHORA, DAR applies to both texts and dialogues.

Differing from most resolution algorithms, DAR correctly accounts for the resolution of pronouns referring to newly introduced information, as it is the case in examples **(1)** and **(2)**.

(1) [*Chefen*]$_i$ *fik kun* [*en søn*]$_k$ *og* [*han*]$_k$ *gad i hvert fald ikke videreføre familieforetagendet.* [PID]
([The boss]$_i$ had only [one son]$_k$ and [he]$_k$ surely did not want to carry on the family business.)

(2) A: *hvem...hvem arbejdede* [*din mor*]$_i$ *med?*
(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* [BYSOC]
([She]$_k$ was a widow... had three sons)

In **(1)** the antecedent of the pronoun *han* (he) is the indefinite object and not the more "given" definite subject. In **(2)** the antecedent of the second occurrence of the pronoun *hun* (she) is the object *vores nabo* (our neighbour) which provides the information requested in the preceding question. This nominal is assigned lower prominence than the subject pronoun *hun* (she) in most salience models. To account for this type of data the DAR-algorithm proposes a novel strategy combining two apparently contrasting accounts of salience of entities (Navarretta, 2002a). The first account, e.g. (Grosz et al., 1995), assigns the highest degree of salience to the most known (topical) entities in the discourse model, the second assigns the highest degree of salience to entities in the focal part of utterances in Information Structure terms which, often, represent new information (Hajičová et

---

[1] DAR presupposes that intrasentential anaphors are correctly resolved. At present no resolution algorithm accounts for all uses of Danish intrasentential pronouns.

al., 1990).

DAR was developed on the basis of the uses of pronouns in three text collections and three corpora of naturally-occurring dialogues. The texts comprise computer manuals, henceforth EDB, novels and newspaper articles. The dialogue collections are SL (Duncker and Hermann, 1996), consisting of recorded conversations between GPs and their patients, the BYSOC corpus (Gregersen and Pedersen, 1991) and the PID corpus (Jensen, 1989) both containing recorded conversations about everyday subjects.

In the paper we first present related work (section 2) then we discuss the background for our proposal (section 3). In section 4 the DAR-algorithm is described. In section 5 we present some tests of the algorithm, evaluate it and compare its performance with the performance of other known algorithms. Finally, in section 6, we make some concluding remarks.

2 Related Work

The two algorithms for resolving IPAs and APAs in English dialogues, ES00 and PHORA, recognise IPAs and APAs on the basis of semantic constraints on the argument position occupied by the anaphors. Both algorithms account for differences in reference between personal and demonstrative pronouns. In ES00 demonstrative pronouns preferentially refer to abstract entities, while personal pronouns preferentially refer to individual ones. ES00 resolves IPAs applying Strube's (1998) algorithm. In PHORA the antecedents of personal pronouns are searched for looking at their degree of salience which is implemented by word order as in (Grosz et al., 1995). Demonstratives, instead, are searched for in the list of *activated* entities (Gundel et al., 1993) containing non NP antecedents, which are assumed to be less salient. In PHORA demonstratives can also refer to Kinds.

ES00 requires that the structure of dialogues has been marked. Byron's PHORA-algorithm does not rely on predefined dialogue structure, but only searches for abstract antecedents of APAs in the sentence preceding the anaphor. Thus it does not account for APAs referring to larger discourse segments. PHORA relies on both semantic knowledge and a model of speech acts and accounts for more phenomena than ES00. Differing from ES00, PHORA has been implemented. A very different strategy for resolving IPAs and APAs in spoken dialogues is proposed in (Strube and Müller, 2003). We will not further discuss this proposal, but Strube and Müller's machine learning approach is an interesting attempt to automatically resolve anaphors without relying on any domain specific resource or preannotated data.

3 Background for DAR

In most applied approaches pronominal anaphora resolution is equivalent to determining the antecedent domain and choosing the most prominent or *salient* antecedent among possible candidates. Although there is not always an identity relation between linguistic antecedents and referents, we also follow this strategy, well aware that it is particularly problematic for APAs. In fact, the same linguistic expression can evoke different abstract objects depending on the context in which the APA occurs, see (Webber, 1991).

Determining the degree of salience of discourse elements, henceforth DEs, is essential to anaphor resolution because personal pronouns refer to the most salient candidate antecedent that matches the given predication (Sidner, 1983). Nearly all salience-based models identify high degree of salience with high degree of givenness of DEs. In fact, although the various algorithms use different criteria for ranking DEs such as linear order, hierarchy of grammatical roles, information structure, Prince's Familiarity Scale (Prince, 1981), they all assign the highest prominence to the DEs which are most topical, known, bound, familiar and thus *given*, i.a. (Grosz et al., 1995; Brennan et al., 1987; Strube and Hahn, 1996; Strube, 1998). Grosz et al. (1995) also suggest that continuing speaking about the same elements in a discourse segment is perceived as more coherent than shifting the focus of attention. They implement this by the following ranking of transition states:
continue > retain > shift.

One salience model departs from the *givenness*[2] assumption. It has been proposed by Hajičová et al. (1990) and assigns the highest degree of salience to DEs in the focal part of an utterance in information structure terms (Sgall et al., 1986). These entities often represent new information. Hajičová et al.'s approach is original and can account for the data in **(1)** and **(2)**. However, it is problematic from an applied point of view. In the first place it is difficult to

---

[2]Here givenness subsumes concepts such as topicality and familiarity.

determine the information structure of all utterances. Secondly, focal candidate antecedents are ranked highest in Hajičová et al.'s model, but they still compete with *given* candidate antecedents in their system. Finally the data does not confirm that all entities in the focal part of an utterance have the highest degree of accessibility.

We agree with Hajičová's insight, but in order to operationalise the role of focality in resolution in a reliable way we propose the following. Accessibility by default is connected with *givenness* as assumed in most resolution algorithms. However, speakers can explicitly change the degree of accessibility of entities in discourse by marking them as salient with information structure related devices. These entities represent the main focus of an utterance, have the highest degree of salience and are, in the majority of cases, the preferred antecedents of anaphors. In these cases the shift of focus of attention is, in our opinion, as coherent as continuing speaking about the same entities, because it is preannounced to the addressee. On the basis of the data we propose a list of identifiable constructions in which explicit focus marking occurs and the focal DEs have the highest degree of salience in our data.[3] Examples from the list are the following:

a: Entities referred to by NPs which are focally marked structurally. In Danish this marking occurs in clefts, existential and topicalised constructions.[4]

b: Entities referred to by NPs that follow focusing adverbs, as in **(1)**.

c: Entities focally marked by the prosody (if this information is available) and/or entities providing the information requested in questions, as in **(2)**.

The hierarchy of verbal complements can model givenness preference in Danish. As in English pronouns have high givenness degree (pronominal chain preference). In addition to salience preferences we found that parallelism can account for numerous uses of Danish anaphors. According to parallelism 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. Inspired by the work of (Kameyama, 1996) we have defined a preference interaction model to be used in resolution. Our model is given in figure 1.[5] The interaction model states that givenness preferences are overridden by focality preference, when in conflict, and that they all are overridden by parallelism. Also in Dan-

Parallelism $\supseteq$ Focality $\supseteq$ Pronominal chain $\supseteq$ Givenness
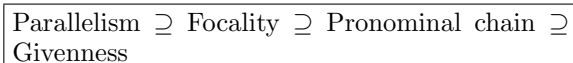
Figure 1: Interaction of preferences

ish demonstrative and personal pronouns refer to entities with different status in the discourse model. Weak (cliticised and unstressed) pronouns usually refer to the most salient entity in the utterance. Strong (stressed and demonstrative) pronouns emphasise or put in contrast the entities they refer to and/or indicate that their antecedents are not the most expected ones.[6] Demonstratives preferentially refer to abstract entities, while personal pronouns preferentially refer to individual entities in ambiguous contexts. All these differences are implemented in DAR.

Approx. half of the APA occurrences in our dialogues refer to entities evoked by larger discourse segments (more turn takings). Thus we follow Eckert and Strube's approach of marking the structure of dialogues and searching for APA antecedents in the right frontier of the discourse tree (Webber, 1991). DAR presupposes different discourse structures for texts and dialogues.

DAR follows the ES00 and PHORA strategy of discriminating between IPAs and APAs by rules looking at the semantic constraints on the predication contexts in which the anaphors occur. DAR relies on many more discriminating rules than ES00. These rules were defined analysing large amounts of data and using the encodings of the Danish PAROLE computational lexicon (Braasch et al., 1998; Navarretta, 1997).

DAR uses language-specific rules to account

---

[3]Many of these constructions are also studied in the Information Structure literature and in some studies on anaphora.

[4]Nominals in clefts are also assigned high salience in e.g. (Sidner, 1983).

[5]The interaction model was defined on the basis of the data and the results of a survey of pronominal uses. Commonsense preferences which override all the other preferences (see inter alia (Hobbs, 1983) are not implemented.

[6]The most frequent Danish third person singular gender pronoun *det* can both be a personal pronoun (corresponding to it) and a demonstrative pronoun (corresponding to this/that). In the latter case it is always stressed.

for Danish APAs. These occur in much more contexts than in English where elliptical constructions or other anaphors such as *too* and *so* are used. Examples of Danish-specific uses of abstract anaphors are given in **(3)** and **(4)**.

(3) *Han var sulten. Det var jeg ikke.* [PID]
(lit. He was hungry. That was I not.)
(My friends were hungry. I wasn't.)

(4) *Han kunne svømme, men det kunne hun ikke.*
(lit. He could swim, but it could she not.)
(He could swim, but she couldn't.)

A language-specific rule recognising APAs is the following: constructions with modal verbs and an object, such as *x skal man* (lit. *x* shall one) (one shall), *x vil man* (lit. *x* will one) (one will).

An example of a rule identifying IPAs is the following: adjectival constructions in which the prepositional complement only subcategorises for concrete entities such as *let for x* (easy for *x*), *fuld af x* (full of *x*).

4 The DAR-algorithm

4.1 Search Space and DE lists

DAR presupposes the discourse structure described by Grosz and Sidner (1986). The minimal discourse unit is the utterance U. Paragraphs correspond to discourse segments in texts. Discourse segments in dialogues were manually marked. The dialogues were structured with *Synchronising Units* (*SU*) according to the definitions in ES00.

The immediate antecedent search space of a pronoun x in utterance U_n is the previous utterance, U_{n-1}. If U_n is the first component in SU_m in dialogues the immediate search space for x is SU_{m-1}. DAR assumes two antecedent domains depending on whether the pronoun has or has not been recognised as an IPA. The antecedent domain for IPAs is first U_{n-1} and then the preceding utterances in the right frontier of the discourse tree searched for in recency order.[7] The antecedent domain for APAs or anaphors which can both be IPAs and APAs is U_{n-1}.

DAR operates on two lists of DEs, the *Ilist* and the *Alist*. The *Ilist* contains the NPs referred to in U_{n-1} ranked according to their degree of salience and enriched with information on gender, number, animacy and other simple semantic types necessary to implement selectional restrictions. In the *Ilist* information

---

[7]The search space in ES00 is the preceding utterance for all pronouns.

> **1** A-list
> **2** within same U, **I** in dialogues,: clause to the left of the clause containing the anaphor
> **3** within previous U (**I**): rightmost main clause and subordinate clauses to its right
> **4** within previous **U**s (**I**s): rightmost complete sentence

Figure 3: The ES00 context ranking

about the grammatical role of nominals is provided and strongly focally marked elements are indicated. The leftmost element in the *Ilist* is the most salient one. Givenness and focality preferences are accounted for in the *Ilist*, as illustrated in figure 2. Focally marked entities are put in front of the list while the remaining DEs are ordered according to verbal complement order. Inside verbal complements nominals are ordered according to their occurrence order as illustrated in the second row of figure 2. The abstract entities which are referred to by an APA in U_{n-1} or SU_{m-1} are encoded in the *Alist*. They are removed from the list after a new utterance (SU in dialogues) has been processed if they have not been mentioned in it. The context ranking for abstract entities is that proposed by Eckert and Strube (2000) and is given in figure 3.

4.2 The Algorithm

DAR consists of two different functions RESOLVEDET and RESOLVEIPA. The former is applied if the actual pronoun x is third person singular neuter, while the latter is applied in all the remaining cases:

if x is *singular & neuter*
    then go to RESOLVEDET(x)
    else go to RESOLVEIPA(x)

RESOLVEIPA takes the IPA x as argument and looks for possible antecedents in the *Ilist* for the preceding U_{n-1} or S_{m-1}, after having applied syntactic constraints and selectional restrictions on the elements of the list. Three different cases are considered: (A) no antecedent has been found in the immediate search space; (B) one antecedent has been found; (C) more antecedents have been found.

If no antecedent has been found (case A), RESOLVEIPA looks for the highest ranked antecedent in recency order in the *Ilist*s of the preceding discourse. If an antecedent is found the algorithm returns it. If no antecedent is

Figure 2: Order of DEs in the *Ilist*

found, x is classified as *inferable*.[8] If one antecedent is found (case B), it is returned. If more candidate antecedents are found (case C), RESOLVEIPA performs tests, implementing the preference interaction model described in section 3, as follows. If U_n and U_{n-1} are parallel[9] and one of the candidate antecedents has the same grammatical role in U_{n-1} as x in U_n, this "parallel" antecedent is marked. In the remaining cases the algorithm marks the highest ranked candidate in the *Ilist*. Pronouns are preferred, unless there are focally marked candidate antecedents. At this point the algorithm individuates the preferred antecedent on the basis of x's type. If x is *weak* the marked candidate proposed in the preceding steps is returned together with the list of the remaining candidate antecedents (possible ambiguity). If x is *strong* the highest ranked candidate antecedent which was not marked in the preceding steps is returned together with the list of candidate antecedents.[10] The approach of marking ambiguities resembles that proposed by Kameyama (1996).

The main structure of the function RESOLVEDET is inspired by ES00. RESOLVEDET tests the pronoun x using the IPA and APA discriminating rules discussed in section 3. If x is IPA, the function RESOLVEIPA-NEU is applied. If x is APA the function RESOLVEAPA is applied. Finally, if the pronoun is neither IPA nor APA, RESOLVEDET looks at its type. If x is strong the algorithm attempts to find an abstract antecedent (RESOLVEAPA), while if it is weak DAR tries to find an individual antecedent (RESOLVEIPA-NEU). RESOLVEIPA-NEU is like RESOLVEIPA except that it returns if no NP antecedents are found in U_{n-1} (case A) so that RESOLVEAPA can be applied.

---

[8]In DAR *inferables* comprise pronouns whose antecedents must be inferred by the context, plural pronouns with complex antecedents and generic uses of *det* (it).

[9]Parallelism is investigated in coordinated, adjacent or explicitly contrasted utterances.

[10]A special rule in DAR is applied to the demonstratives *dette/denne/disse* (this/these) which never corefer with subject candidates.

RESOLVEAPA distinguishes between types of pronoun. If x is *weak*, the preferred antecedent is searched for among the elements indicated in the context ranking, unless it is the object of the verb *gøre* (do), modals, *have* (have) or the abstract subject in copula constructions. In these cases the pronoun is resolved to the VP of the element in the *A-list* or in the context ranking. If x is *strong* RESOLVEAPA attempts to resolve or classify it as *vague* depending on the type of pronoun. This part of the algorithm is specific to Danish and accounts for the fact that different strong pronouns preferentially refer to different abstract entities in the data.

Resolved APAs are inserted into the *Alist*. In case of failure RESOLVEAPA returns so that RESOLVEIPA-NEU can be applied. If both functions fail, the pronoun is classified as *vague*.

4.3 Some Examples

In the following we look at the resolution of example **(2)** from section 3 and the example in **(5)**.

(5): *Du har svært ved at se musemarkøren på skærmen. Hvordan klarer du det?* [EDB]
(You have difficulties seing the mouse-cursor (common-gend) on the screen (common-gend). How do you manage it/this (neuter gender))?

The simplified *Ilist*s and *Alist*s after each utterance has been processed in example **(2)** are given in figure 4. **(2)** contains three SUs. U_2 is an **I/A** thus it belongs to two synchronising units (SU_1 and SU_2). The *Ilist* after U_1 has been processed, contains one element, *din mor* (your mother). In U_2 the personal pronoun *hun* (she) occurs, thus RESOLVEIPA is applied. It resolves *hun* to the compatible NP in the *Ilist*, *din mor*. After U_2 has been processed the *Ilist* contains two elements in this order: the focal marked entity *vores nabo* (our neighbour) and the pronoun *hun* (= din mor). RESOLVEIPA resolves the occurrence of the pronoun *hun* (she) in U_3 to the most salient candidate NP in the *Ilist*, *vores nabo*. Here focal preference overrides pronominal chain preference. The simplified *Ilist*s and *Alist*s after the two utterances in **(5)** have been processed are given in figure 5. After U_1 has been processed there are two com-

SU_1: U_1 (**I**) U_2 (**I/A**):
U_1: *hvem...hvem arbejdede din mor med?*
(with whom... whom did your mother work)
Ilist: [din mor]
Alist:[]

SU_2: U_2 (**I/A**)
U_2: *Hun arbejdede med vores nabo*
(She worked with our neighbour)
Ilist: [vores nabo,hun=din mor]
Alist: []

SU_3: U_3 (**I**)
U_3: *Hun var enke ... havde tre sønner*
(She was a widow... had three sons)
Ilist: [hun=vores nabo,tre sønner]
Alist: []

Figure 4: *Ilist*s and *Alist*s for example (**2**)

U_1: Du har svært ved at se musemarkøren på skærmen.
Ilist: [musemarkøren, skærmen]
Alist:[]

U_2: Hvordan klarer du det?
Ilist:[]
Alist:[det=U_1]

Figure 5: *Ilist*s and *Alist*s for example (**5**)

mon gender singular NPs in the *Ilist*, *muse-markøren* (the mouse cursor) and *skærmen* (the screen). In U_2 the singular neuter gender pronoun *det* (it) occurs, thus RESOLVEDET is applied. The pronoun is neither IPA nor APA according to the discriminating rules. Then RESOLVEDET attempts to find an individual antecedent of the weak pronoun, applying the function RESOLVEIPA-NEU. RESOLVEIPA-NEU fails because the two DEs in the *Ilist* do not agree with the pronoun. Then the function RESOLVEAPA resolves x looking at the context ranking. Being the *Alist* empty, U_1, is proposed as antecedent. The resolved APA is added to the *Alist*.

5 Tests and Evaluation

We have manually tested DAR on randomly chosen texts and dialogues from our collections. The performance of DAR on dialogues has been compared with that of ES00. The function for resolving IPAs (RESOLVEIPA) has similarly been tested on texts, where APAs were excluded. We have compared the obtained results with those obtained by testing BFP (Bren-

nan et al., 1987) and STR98 (Strube, 1998). In all tests the intrasentential anaphors have been manually resolved. Expletive and cataphoric uses of pronouns have been marked and excluded from the tests. Dialogue act units were marked and classified by three persons following the strategy proposed in (Eckert and Strube, 2000). The reliability for the two annotation tasks (κ-statistics (Carletta, 1996)) was of 0.94 and 0.90 respectively. Pronominal anaphors were marked, classified and resolved by two annotators. The κ-statistics for the pronoun classification was 0.86. When the annotators did not agree upon resolution, the pronoun was marked as ambiguous and excluded from evaluation. The results obtained for BFP and STR98 are given in table 1, while the results of DAR's RESOLVEIPA are given in table 2. Because DAR both classifies and resolves anaphors, both precision and recall are given in table 2. Precision indicates the proportion of the resolved pronouns which are correctly resolved, while recall indicates the proportion of all pronouns resolved by humans which are correctly resolved by the algorithm.

The results indicate that RESOLVEIPA performs significantly better than BFP and STR98 on the Danish texts. The better performance of DAR was due to the account of focal and parallelism preferences and of the different reference mechanisms of personal and demonstrative pronouns. Furthermore DAR recognises some generic pronouns and inferable pronouns and excludes them from resolution, but often fails to recognise antecedentless and inferable plural pronouns, because it often finds a plural nominal in the preceding discourse and proposes it as antecedent. The lack of commonsense knowledge explains many incorrectly resolved anaphors. The results of the test of the DAR algo-

algorithm	corr.resolved	res.human	precision
BFP	513	645	79.53
STR98	524	645	81.24

Table 1: Results of BFP and STR98 on texts

corr.res.	res.overall	res.hum.	precis	recall
575	651	645	88.33	89.14

Table 2: Results of RESOLVEIPA on texts

rithm on written texts are in table 3. These results are good compared with the results of the function RESOLVEIPA (table 2). The discriminating rules identify correctly IPAs and

resolution IPA						
Algorithm	correctly resolved	resolved overall	human resolution	precision	recall	f-meaure
ES00	258	411	414	62.77	62.31	62.48
DAR	289	386	414	74.87	68.81	71.71
resolution APA						
Algorithm	correctly resolved	resolved overall	human resolution	precision	recall	f-measure
ES00	179	286	269	62.59	66.54	64.5
DAR	194	277	269	70.04	72.19	69.13

Table 4: Results of ES00 and DAR on dialogues

APAs in the large majority of the cases. Recognition failure often involves pronouns in contexts which are not covered by the discriminating rules. In particular DAR fails to resolve singular neuter gender pronouns with distant antecedents and to identify vague anaphors, because it always "finds" an antecedent in the context ranking. Correct resolution in these cases requires a deep analysis of the context. The

resolution IPA				
corr.res.	res.overall	res.hum.	precis	recall
560	651	645	86.02	86.82
resolution APA				
corr.res.	res.overall	res.hum.	precis	recall
63	87	77	72.41	81.82

Table 3: Results of DAR on texts

results of applying DAR and ES00 on Danish dialogues are reported in table 4.[11] In the last colum the overall performance of the two algorithms is given as f-measure (F) which is defined as $\frac{1}{\alpha \frac{1}{P} + (1-\alpha)\frac{1}{R}}$ where P is precision, R is recall and α is the weight of P and R. We have assigned the same weight to P and R ($\alpha = 0.5$) and thus $F = \frac{2PR}{P+R}$. The results of the tests indicate that DAR resolves IPAs significantly better than ES00 (which uses STR98). The better performance of DAR is also due to the enlarged resolution scope respect to the one used in ES00. DAR correctly resolves more Danish demonstrative pronouns than ES00, because it accounts for language-specific particularities. In general, however, the resolution results for APAs are similar to those obtained for ES00. This is not surprising, because DAR uses the same resolution strategy on these pronouns. DAR performs better on texts than on dialogues. This reflects the more complex nature of dialogues. The results indicate that the IPA/APA discriminat-

ing rules also work well on dialogues. The cases of resolution failure were the same as for the texts. As an experiment we applied DAR on the dialogues without relying on the predefined dialogue structure. In this test the recognition of IPAs and APAs was still good, however the success rate for IPAs was 60.1 % and for APAs was only 39.3%. Many errors were due to the fact that antecedents were searched for in the preceding discourse in linear order and that ungrounded utterances were included in the discourse model.

6 Concluding Remarks

In this paper we presented DAR, an algorithm for resolving IPAs and APAs in Danish texts and dialogues. In DAR differences between the referential characteristics of Danish weak and strong pronouns are accounted for and a novel strategy for resolving individual anaphors is proposed. This strategy combines givenness with focality preferences to model salience and also accounts for parallelism preferences. DAR performs significantly better on IPAs than algorithms which only rely on givenness-based salience models. The strategy and the general assumptions behind DAR should be tested on other languages.

Differing from ES00 and PHORA, DAR has been developed for and tested on both texts and dialogues. DAR extends the ES00 strategy of classifying and resolving (some types of) APAs. The tests of DAR indicate that the ES00's approach of recognising APAs is also promising for texts and other languages than English.

DAR has not been compared with PHORA which is the only abstract anaphora algorithm implemented. We find the algorithm very interesting because it addresses many of the same phenomena, but with different strategies. It would be useful to combine some of these strategies with the approaches proposed in DAR and ES00 to improve the still problematic resolution

---

[11]We extended ES00 with the Danish-specific identification rules before applying it.

of abstract anaphors.

References

A. Braasch, C. Navarretta, and N.H. Sørensen. 1998. Danish lexicon documentation. Technical Report LE-PAROLE. WP3.3-CST, CST.

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

D. Byron and J. Allen. 1998. Resolving Demonstrative Pronouns in the TRAINS93 corpus. In *Proceedings of the Second Colloquium on Discourse Anaphora and Anaphor Resolution (DAARC 2)*, pages 68–81.

D. K. Byron. 2002. Resolving Pronominal Reference to Abstract Entities. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL 2002)*.

J. Carletta. 1996. Assessing agreement on classification tasks. the kappa statistic. *Computational Linguistics*, 22(2):249–254.

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

M. Eckert and M. Strube. 2000. Dialogue acts, synchronising units and anaphora resolution. *Journal of Semantics*, 17:51–89.

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

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

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

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

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

J. R. Hobbs. 1983. Why Is Discourse Coherent? In Fritz Neubauer, editor, *Coherence In Natural-Language Texts*, volume 38 of *Papers in Textlinguistics*, pages 29–70. Helmut Buske Verlag Hamburg.

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

M. Kameyama. 1996. Indefeasible Semantics and Defeasible Pragmatics. In M. Kanazawa, C. Piñon, and H. de Stwart, editors, *Quantifiers, Deduction and Context*, pages 111–138. CSLI, Stanford, CA.

C. Navarretta. 1997. Encoding Danish Verbs in the PAROLE Model. In R. Mitkov, N. Nicolov, and N. Nikolov, editors, *Proceedings of RANLP'97.Recent Advances in Natural Language Processing*, pages 359–363, Tzigov Chark, Bulgaria.

C. Navarretta. 2002a. Combining Information Structure and Centering-based Models of Salience for Resolving Intersentential Pronominal Anaphora. In A. Branco, T. McEnery, and R.Mitkov, editors, *Proceedings of the 4th Discourse Anaphora and Anaphora Resolution Colloqium*, pages 135–140. Ediçoes Colibri.

E. F. Prince. 1981. Toward a taxonomy of given-new information. In P. Cole, editor, *Radical Pragmatics*, pages 223–255. Academic Press.

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

C. Sidner. 1983. Focusing in the Comprehension of Definite Anaphora. In M. Brady and R. Berwick, editors, *Computational Models of Discourse*, pages 267–330. MIT Press.

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

M. Strube and C. Müller. 2003. A machine learning approach to pronoun resolution in spoken dialogue. In *Proceedings of the ACL'03*, pages 168–175.

M. Strube. 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)*, volume II, pages 1251–1257.

B. L. Webber. 1991. Structure and Ostension in the Interpretation of Discourse Deixis. *Natural Language and Cognitive Processes*, 6(2):107–135, January.