Identifying Situation Reference in Danish

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

1 Introduction

This paper deals with the identification of anaphors whose antecedents are verbal phrases or discourse segments in Danish.[1]

These anaphors have been given different names in literature, such as discourse deictics (Levinson 1987, Webber 1991), anaphors to abstract objects (Asher 1993) and situation reference (Fraurud 1992). We follow Fraurud and call them situation anaphors. Situation reference is quite common especially in dialogues, but has seldom been dealt with in computational linguistics.

In this paper we first describe situation reference in Danish (section 2), then we shortly outline Eckert and Strube's algorithm for anaphora resolution a part of which we have modified and extended for identifying Danish situation anaphors (section 3). In section 4 we present our rules for identifying Danish situation reference and we present the results of the manual application of these rules on two dialogues. In section 5 we make some concluding remarks.

2 Danish Situation Reference

Situation anaphors in Danish are third-person neuter gender personal and demonstrative pronouns *det* (it/this/that), *dette* (this), *det her* (this) and *det der* (that). *Dette* is mostly used in written language, while *det her* and *det der* are common in conversations.

We have analysed the occurrences of situation anaphors in a number of dialogues from the collection "Samtale hos Lægen" ("The Talk at the Doctor's"), henceforth **SL**, collected from 1993 to 1995 in the field of psychology of language by researchers at the University of Copenhagen (Duncker & Hermann 1996, Hermann 2000). Furthermore we have analysed situation anaphors in some newspaper articles from **Berlingske Tidende**. As in English situation

[1]The research described has been done under the **Staging** project, funded by the Danish Research Councils.

anaphors can refer to an infinitive, one or more verbal phrases, one or more clauses, a preceding discourse segment or something that can be vaguely inferred from the context. Danish deictics are also used in cases where elliptical constructions are common in English. Some examples of situation reference are the following:

- the anaphor co-refers with an infinitive:

  (1) *At ryge er farligt og det er ogs å dyrt*
  (Smoking is dangerous and it is also expensive)

- the anaphor co-refers with a clause:

  (2) **A:** *Du skal tage en blodprøve*
  (You have to take a blood test)
  B: *Hvorfor det?*
  (Why is that?)
  (**SL**)

- the anaphor is used as the subject complement of *være* (be) and *blive* (become) in answers (or in coordinated successive clauses):

  (3) **A:** *Blev du færdig med opgaven?*
  (Did you finish the task?)
  B: *Ja, det blev jeg*
  (lit. Yes, that was I)
  (Yes, I did)

- the anaphor co-refers with a verb phrase when it is used as the object complement for the verb *have* (have), *gøre* (do) and modal verbs:

  (4) *Alle faldt, men det gjorde jeg ikke*
  (lit. All fell, but that did I not)
  (All fell, but I did not)

- the anaphor co-refers with a clause in constructions with attitude verbs and other verbs which take clausal complements, such as *synes* (think), *tro* (believe) *vide* (know), *sige* (say), *håbe* (hope):

  (5) **A:** *Han falder snart i søvn*
  (He will soon fall asleep)
  B *Det h åber jeg ikke*
  (lit. That hope I not)
  (I hope not)

3 Eckert and Strube's Algorithm

Although situation reference is very common, especially in dialogues, most of the algorithms for resolving pronominal anaphora do not deal with it. An exception is the algorithm proposed by Eckert and Strube (Eckert & Strube 1999*b*, Eckert & Strube 1999*a*), the ES-algorithm henceforth, which was defined for resolving anaphors with individual NP antecedents and with abstract object antecedents. The ES-algorithm is based on rules for discriminating among individual and situation anaphors based on the predicative contexts in which the anaphors occur. Individual anaphors are resolved by a centering-based algorithm (Strube 1998), while some types of situation anaphors are resolved with an algorithm proposed by Eckert and Strube. The athors manually test the approach on selected dialogues and obtain a precision of 63,6% for discourse deictics and 66,2% for individual anaphors.

The algorithm has been adapted to Danish with slightly better results in (Navarretta 2000), but it was found too simplistic for correctly classifying and resolving different types of situation reference. Although we agree, we believe that identifying uses of third-person neuter gender singular personal and demonstrative pronouns as situation anaphors is useful in NLP processing systems and that Eckert and Strube's approach to recognize them from their context is worth pursuing. Thus we have both modified the original rules in Eckert and Strube's algorithm, and added Danish specific rules. The rules are mainly based on the occurrences of situation anaphors in the Danish dialogue collection and corpus of written texts.

4 Identification Rules

In the following we present some of the defeasible preference rules for identifying situation anaphors which we have defined for Danish. We have marked with a star those rules which are simply translations of the rules proposed by Eckert and Strube. defeasible.

- * constructions where a pronoun is equated with an abstract object, e.g., *x er et forslag* (x is a suggestion)
- * copula constructions with adjectives which can only be applied to abstract entities, such as *x er sandt/usandt* (x is true/untrue), *x er rigtigt* (x is correct)
- * arguments of verbs which take S'-complements, e.g., *tro* (believe), *antage* (assume), *sige* (say)
- * anaphoric referent in constructions such as *x er fordi du er holdt op med at ryge* (x is because you have stopped smoking) and *x er p å grund af at du er gravid* (x is because you are pregnant)
- object of gøre (do)

- subject complement with *være* (be) and *blive* (become)
- object of *have* (have) if the verb was not used as a main verb in the previous clause
- object of modal verbs
- in copula constructions where the adjective can both refer to an individual NP and to an abstract object, such as *x er godt* (x is good), *x er dårligt* (x is bad), *elske x* (love x) the anaphor co-refers with an abstract object if the previous clause contains a raising adjective construction (or related constructions where an infinite is the subject)

The latter rule covers cases where the contexts of an anaphor can allow both an individual NP and an abstract object. Consider as illustration the examples (6-a) and (6-b).

(6) a. *Peter boede i et rødt hus. Det hadede han.*
(Peter lived in a red house. He hated it.)
b. *Det er dødsygt at sidde p å et vaskeri. Det hader jeg.*
(It is boring to be in a laundry. I hate it)

The identification rules we have proposed would identify the *det* in example (6-b) as a situation anaphor.

To test the identified rules we have manually marked situation anaphors in two randomly chosen dialogues from the **SL** collection and in a newspaper article. Then we have manually applied the identification rules to the two unmarked dialogues. We have compared the results from human marking and from marking according to the identification rules. In 83 % of the cases the same situation anaphors were identified. Failure cases were mainly anaphors occurring in constructions which allow for both individual and situation reference or anaphors occurring in constructions which we had not identified. The obtained results are encouraging, but it must be noted that we have tested the rules on the same type of dialogue which we also used for identifying the discriminating rules, thus more tests should be done on different dialogue and text types.

5 Concluding Remarks

We have proposed preference rules for identifying situation reference in Danish by modifying and extended the rules for recognizing situation anaphors proposed by Eckert and Strube. We have also presented the results of a first test of these rules, and these results were encouraging, but should be confirmed by more tests on different types of dialogue and text. Although the identified rules are general, they are not yet exhaustive. We believe that the rules can be used in different NLP applications, such as text understanding and dialogue systems

to mark situation anaphors that cannot be resolved by common resolution algorithms which only deal with individual anaphors.

References

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

Duncker, D. & Hermann, J. (1996), 'Patientord og lægeord - særord eller fællesord?', *Månedsskrift for Praktisk Lægegerning - Tidsskrift for Praktiserende Lægers Efteruddannelse* pp. 1019–1030.

Eckert, M. & Strube, M. (1999*a*), Dialogue Acts, Synchronising Units and Anaphora Resolution, *in* J. van Kuppevelt, N. van Leusen, R. van Rooy & H. Zeevat, eds, 'Amstelogue'99 Proceedings - Workshop on the Semantics and Pragmatics of Dialogue'.

Eckert, M. & Strube, M. (1999*b*), Resolving Discourse Deictic Anaphora in Dialogues, *in* 'Proceedings of the EACL-99', pp. 37–44.

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

Hermann, J. (2000), Understandings between doctors and patients - some methodological issues, *in* 'Proceedings ofm the Conference on Medical Interaction, Oct. 18-20', University of Southern Denmark, Odense.

Levinson, S. (1987), 'Pragmatics and the grammar of anaphora: a partial pragmatic reduction of Binding and Control Phenomena', *Journal of Linguistics* **23**(2), 379–434.

Navarretta, C. (2000), Abstract Anaphora Resolution in Danish, *in* L. Dybkjær, K. Hasida & D. Traum, eds, 'Proceedings of 1st SIGdial Workshop on Discourse and Dialogue - Held in conjunction with The 38th Annual Meeting of the ACL', Hong Kong, pp. 56–65.

Strube, M. (1998), Never Look Back: An Alternative to Centering, *in* 'Proceedings of the 36th Meeting of the ACL', Vol. II, Université de Montréal, Montreal, Quebec, Canada, pp. 1251–1257.

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