

The annotation of pronominal abstract anaphora in Danish texts and dialogues

Costanza Navarretta and Sussi Olsen

DAD - report 1

Centre for Language Technology

January, 2009

version final

The project's main aim is to develop a formal model over the use of intersentential pronominal abstract anaphora in Danish.

Danish abstract pronominal anaphora comprise occurrences of the pronouns *det* (it/this/that), *dette* (this), *det her* (this) and *det der* (that) pointing back to verb phrases, predicates in copula constructions, clauses, discourse segments or abstract pronouns (the antecedents). These anaphors refer to abstract entities such as properties, events, states, situations, facts and propositions. An example of abstract anaphor is the following:

og så prøvede jeg så at gå lidt i svømmehallen og **det** prøver jeg sådan ind imellem
(and then I tried to go a little to the swimming pool and I try **that** once in a while)

Samtale med Lægen (Duncker & Hermann 1996)

The pronoun *det* in the example has as antecedent the verb clause *to go a little to the swimming pool* and refers to an activity.

Contents

1	Introduction	2
2	The annotation	3
2.1	The markables	3
2.2	The pronouns and their functions	4
2.3	The type of pronoun	5
2.4	Antecedents	6
2.5	Anaphora and anaphoric relations	7
2.5.1	Anaphoric distance	7
2.5.2	Semantic type of referent	8
2.5.3	Referent	10
2.6	The anaphoric relation	10
2.7	The PALinKa annotation tool	11
3	Annotation examples	12
4	Appendix A: Corpus specificities	16

Chapter 1

Introduction

This report contains the specifications of the annotation of the pronouns *det* (it/this/that) and *dette* (this) in texts and of the pronouns *det* (it/this/that), *det her* (this) and *det der* (that) in dialogues. The focus of the project is on the anaphoric use of the pronouns, i.e. those pronouns whose interpretation depends on linguistic expressions that precede them in discourse.

Furthermore the project focuses on intersentential pronominal anaphora which refer to abstract entities and have clauses, verbal phrases, predicates in copula constructions or discourse segments as their antecedents. For simplicity we call these anaphora *abstract anaphora*. Anaphora which have nominal phrase antecedents are called *individual anaphora*, following Asher (1993).

The format of the annotation is XML and the pronouns are annotated using the PALINKA tool (Orasan 2003)¹.

In the transcriptions of monologues and dialogues a distinction is made between stressed and unstressed pronouns. Stress is transcribed with a comma preceding the letter on which the stress falls. For example the unstressed version of the pronoun *det* is transcribed as *det*, while the stressed version of the pronoun is transcribed *d,et*. Specifications particular to the corpora we annotate can be found in appendix 4.

If the pronouns are preceded by the quantifier *alt* (all), the quantifier is included in the annotation. The manual is organised as follows. In chapter 2 we describe the markables and the attributes used in the DAD project and shortly present how they must be assigned. This description must be supplied with the PALinkA user manual which describes how to use the annotator tool. In chapter 3 we present examples of the annotation of different phenomena in XML-format. Finally in the appendix we discuss corpus specific issues.

¹<http://clg.wlv.ac.uk/projects/PALinkA>".

Chapter 2

The annotation

The DAD annotation scheme is an extension of the MATE/GNOME(co)reference annotation scheme proposed by Poesio (2004). The MATE/GNOME scheme has been extended with some elements and attributes necessary to annotate phenomena related to abstract anaphora.

2.1 The markables

We use the markables proposed in the MATE/GNOME scheme, i.e. **de** marking discourse entities and **seg** marking non nominal referring expressions. The markable **link** codes the relation between referring expressions and their antecedents¹.

To these basic markables we have added the markable **explet** and the markable **abandoned**. The former is used to code pleonastic pronouns in constructions such as *det regner* (it rains), *jeg har det fint* (lit. I have it fine) (I am fine), *det er forbudt at ryge* (it is forbidden to smoke) and *det er godt for mig at toget kom til tiden* (it is good for me that the train arrived in time). The **abandoned** markable is used for pronouns occurring in unfinished and abandoned utterances such as in example 1.

(1)

det er - de er forskellige

(it is - they are different)

We also use a **COMMENT** markable to add comments to all the other markables. The **COMMENT** markable has an attribute **CONTENT** in which comments are typed as strings.

¹This markable is required by the PALinKa tool.

2.2 The pronouns and their functions

The pronouns *det* and *dette* have different functions in discourse. Usually they refer to some entity, but, as discussed above, they can occur in unfinished utterances and the pronoun *det* can be pleonastic. When the pronouns are referring to identifiable entities they are encoded with the **de** markable and their function is marked in the attribute **ATYPE**. The following functions are recognised: *deictic*, *cataphoric*, *abstract anaphoric*, *individual anaphoric*, *textual deictic*, *vague anaphoric*, *vague individual*. In the following we describe each value of the attribute **ATYPE** and the contexts in which they must be used

- *cataphor* (**cataphor**): The pronoun is a cataphor, i.e. it precedes the linguistic expression necessary to its interpretation in discourse as in 2 and 3.
(2)
Det at han kom for sent til mødet, skabte alvorlige problemer for hans kollegaer.
(lit. It that he came too late to the meeting gave serious problems to his colleagues)
(The fact that he came too late to the meeting caused his colleagues serious problems)
(3)
Det der bekymrer mig mest er hans helbred.
(lit. That which preoccupies me mostly is his health)
(What mostly preoccupies me is his health)
- *deictic* (**deictic**): The pronoun refers to something in the physical world as in 4 where the pronoun *det her* (this) is accompanied by a pointing gesture to an object in the world.
(4)
Hvad er det her?
(What is this?)
- *individual* (**indiv**): The pronoun is an individual anaphor, that is its antecedent is a nominal phrase, as in 5.
(5)
*Bordet var for langt, og **det** var også slidt*
(The table was too long and **it** was also shabby)
If a pronoun has an abstract nominal antecedent, such as *situation* (situation) and *problem* (problem), it is classified as an individual anaphor, but the fact that its antecedent is abstract is coded in the **CONTENT** attribute of the **COMMENT** markable, with the string *abstrakt nominal antecedent*.
- *individual vague* (**indiv-vague**): The pronoun is an individual anaphor,

but its nominal antecedent is implicit in the discourse. The implicit antecedent is added to the markable via the **COMMENT** element. The **CONTENT** of this element must start with the string *implicit antecedent*:

- *abstract* (**abstr-ana**): The anaphor is abstract, i.e. has a non nominal antecedent as in 6.

(6)

Leif Hansen vurderer, at skaden er dækket med en indeksreguleret selvrisiko på kr. 5.370,00. Dette vurderes, ud fra at der ved besigtigelsen kunne måles fugt i entreen samt i stuen ved pejsen og i køkkenet.

(Leif Hansen estimates that the damage is covered by an index-linked own risk of 5,370.00 DKr. **That** is estimated from the fact that during the inspection of the house humidity was measured in the hall, in the sitting room near the fireplace and in the kitchen.)

- *abstract vague* (**abstr-vague**): The antecedent of the abstract anaphor is implicit in the discourse.
- *textual deictic*: (**textual-deictic**): The pronoun is a textual deictic, i.e. it refers to the text expression it points back to (Lyons 1977). An example of textual deixis is in 7.

(7)

“Livet er dejligt” - Det gentog han tre gange

(“Life is wonderful”- He repeated that three time);

2.3 The type of pronoun

Because different pronominal types signal different cognitive statuses of the referred entities, see i.a. (Prince 1981, Ariel 1988, Gundel, Hedberg & Zacharski 1993) the individuation of the pronominal type is important when studying anaphora. In English there are three abstract pronouns: *it*, *this* and *that*.

In Danish texts only two pronouns are relevant as abstract anaphors, *det* (it/this/that) whose pronominal type cannot be determined, and the demonstrative pronoun *dette* (this). In spoken language there is a distinction between the personal pronoun *det* (it) which is unstressed and the demonstrative pronoun *det* (this/that) which is stressed. Thus information about stress, when available, must be included in the annotation. In spoken Danish the stressed demonstrative *det* can also occur with two adverbials *det her* (this) and *det der* (that) and also these are annotated. The pronominal type is marked in the attribute **PTYPE** which can take one of the following values *betdet* (stressed det), *ubetdet* (unstressed det), *det* (written det), *detder* (det der), *dether* (det her), *dette* (dette).

2.4 Antecedents

We mark the antecedents of anaphoric pronouns (individual, abstract or textual deictic). Nominal antecedents are marked with the markable *de*, having by default the syntactic type (attribute SYN-TYPE) *NP*. Non-nominal antecedents are marked with the **seg** element.

To study further whether there is a relation between the syntactic type of antecedent and the type of pronoun used we also annotate the syntactic type of non-nominal phrases. We distinguish the following types of non nominal antecedent, see i.a. (Navarretta 2004, Navarretta 2007): *verbal phrases, simple main clauses, matrix clauses, subordinate clauses, complex clauses, discourse segments, other*.

The clausal types we use are inspired by the classification of clauses proposed by Kameyama (1998) in a completely different context.

The exact values of the SYN-TYPE are the following:

- *CL*: a main clause such as *Han ankom for sent* (He arrived too late)
- *MCL*: a matrix clause, i.e. a main clause that is not completely independent, as *han sagde* (he said) in 8.
(8)
Han sagde at han skulle tage afsted
(He said that he had to leave)
- *CCL*: a complex sentence, i.e. coordinated clauses, or main clauses and their subclauses;
- *SCL*: a subordinate finit or infinit clause, such as 9 and 10.
(9)
Hvis hun græder...
(If she cries...)
(10)
For at nå vores mål...
(to reach our objectives...);
- *DS*: larger discourse segments, i.e. more sentences and utterances;
- *VP*: a verbal phrase, such as *tog the book* (took the book);
- *CPR*: a predicate in copula constructions. When the predicates of copula constructions are nominal phrases they are marked with a **de** element;
- *OTHER*: other textual segments which do not correspond to complete clauses or predicate, in cases where an antecedent is not in adjacent discourse. An example of this is in 11, if we want to mark *Stil den bog på*

reolen (Put that book on the shelf). The first element *Stil* is marked with the element **seg** and the attribute tag for the whole construction (a main clause in this case) while the second part of the clause *den bog på reolen* is marked with a new **seg** element and assigned with an *other syn-type* value. The second **seg** markable is linked to the preceding one with a **link** markable of type LTYPE called *disclink* and an attribute called CONNECT.

(11)

Stil, sagde Peter, den bog på reolen...

(Put, Peter said, that book on the shelf...)

In the case where a nominal phrase is not occurring in adjacent text in the data it is coded with two *de* markables, the first **de** is assigned the value *NP* in the SYN-TYPE attribute, while the second **de** markable is assigned the value *part-NP*. The second **de** markable is then connected to the first *de* with a **link** markable of LTYPE *disclink*.

2.5 Anaphora and anaphoric relations

Because the project's focus is on abstract anaphora the markables connected to these pronouns are annotated with extra features that enable to study their use.

More specifically we annotate the anaphoric distance (the distance between anaphor and antecedent) in terms of clausal distance in the markables identifying abstract anaphora and textual deictics. For abstract anaphors we also annotate the semantic type of the referent and the referent.

Because there are cases where anaphors have more than one antecedent/referent, these are also marked. The annotators must choose a preferred antecedent and mark it, but they must also annotate alternative interpretations in the **COMMENT** element. The **CONTENT** value must start with the string *possible antecedent*. In the following we explain how the anaphoric distance, the semantic type of the referent and the referent of abstract anaphors are coded.

2.5.1 Anaphoric distance

The attribute used to encode anaphoric distance is **DIST**. The values of this attribute are integers. The value *0* indicates that the antecedent is in the clause that immediately precedes the clause in which the abstract anaphor occurs; the value *1* indicates that there is a clause in-between the anaphor and its antecedents, while *2* indicates that there are two clauses in-between the abstract anaphor and its antecedent, and so on. Elliptical sentences and utterances where the main

verb is not expressed explicitly count as a clause. Eg. this is often the case in dialogues where utterances can be composed by single feedback words, such as *ja* (yes), *nej* (no) and *tak* (thanks).

2.5.2 Semantic type of referent

The semantic type of the referent of abstract anaphora is annotated because this information is important for identifying the referent, although most work on the automatic resolution of anaphora only focuses on the individuation of the antecedent, see i.a. (Grosz, Joshi & Weinstein 1995).

Many of the semantic types we have used are taken from the middle layer of the hierarchy of abstract objects proposed by Asher (1993) given in 2.1.

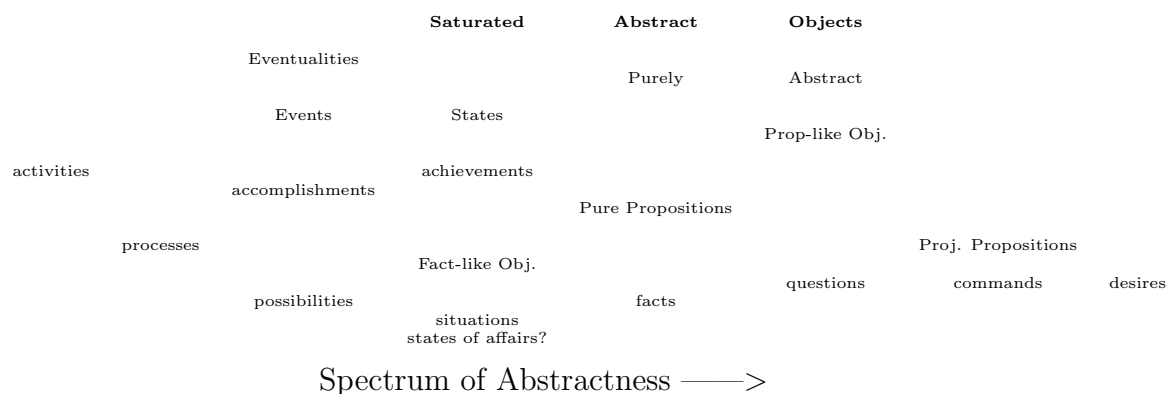


Figure 2.1: Asher's taxonomy of saturated objects

The attribute marking the referential type of the anaphor is called `REF_TYPE` and its acceptable values are the following: *PROP*, *EVENTUALITY*, *FACT-LIKE*, *PROP-L*, *SPEECH-ACT*, *OTHER*. They are used in the following cases:

- *PROP*: The value *PROP* indicates that the referent is of type property. The attribute is often used in cases where the antecedent is the predicate in copula constructions, as in example 12 where the pronoun *det* refers to the property of being very happy.
(12)
Peter var meget glad. Det var jeg også
(lit. Peter was very happy. **That** was I too.)
(Peter was very happy. And so was I.)
- *EVENTUALITY*: The *EVENTUALITY* value is used to mark referents which are states, processes or activities. E.g. this is the case when the anaphor is the subject for verbs such as *ske* (happen), *starte* (start/begin),

slutte (end), *vare* (last, take (time)), or is the object of sense verbs, such as *se* (see), *kigge* (look at), *føle* (sense). The attribute is also used when the anaphor is the object of the verb *gøre* (do) or when the anaphor is said to be an activity, a process or a state as in 13.

(13)

det er en handling/proces/tilstand
(it is an activity/process/state).

- *FACT-LIKE*: The value *FACT-LIKE* marks anaphors who refer to possibilities, situations and facts. Eg. anaphors that occur as objects or subjects of factive verbs such as *vide* (know), *overraske* (surprise), *indrømme* (confess), *tælle* (count), *give mening* (mean), *have betydning for* (have importance for), and for anaphors that are complements for factive adjectives, such as *glad* (happy), *relevant* (relevant). The value can also be used with non-factive verbs when the anaphor refers to possibilities, as in example 14.

(14)

Peter frygtede/forestillede sig at Maria var død
(Peter feared/imagined that Maria was dead).

Finally the value is assigned to anaphors which occur in contexts such as *det er et resultat/et faktum* (it is a result/a factum) and *det er sikkert* (it is sure).

- *EVENT-FACT*: The value *EVENT-FACT* is assigned in cases where the annotator cannot decide from the context whether the referent is an eventuality or a fact. This is often the case when the anaphor refers to a cause of following events/situations². More explicitly this ambiguity can occur when the anaphor is the subject of verbs such as, *tvinge* (force), *forebygge* (prevent), *hindre* (prevent) or occurs in adjectival constructions such as *det er godt/tragisk/betydningsfuld* (it is good/tragic/important).
- *PROP-L*: The value *PROP-L* is assigned to anaphors referring to a propositions, commands, questions, desires. This is often the case when the anaphor is the object of non-factive verbs such as *tro* (believe), *mene* (think), *tvivle* (doubt), *påstå* (affirm), *sige* (say), as in 15 and 16.

(15)

Jeg tror at hun kommer til tiden.
(I believe that she will come in time.)

(16)

Jeg tvivler på at han nåede toget kl. 17.
(I doubt that he caught the train at 17 o'clock.)

The value is also assigned when the anaphor occurs in contexts such as *dette*

²(Vendler 1967) classifies causes, results and consequences as facts, while he classifies effects as eventualities. Other researches, e.g. Peterson (1981), classify causes as eventualities.

er et udsagn/hans mening/et ønske (this is a proposition/his meaning/a desire).

- *SPEECH-ACT*: The value *SPEECH-ACT* is used when the anaphor refers to a speech act, such as a promise, a threat, a lie or similar. An example is in 17.

(17)

A: *Peter er blevet afskediget*

(Peter has been fired)

B: *Det er løgn*

(That's a lie)

2.5.3 Referent

Abstract anaphors are also assigned an attribute REF whose value contains in free words the referent (the attribute is mainly used when comparing the annotations as help for the coders).

2.6 The anaphoric relation

The relation between anaphors and their antecedents are marked with `link` elements which are subordinated to the `de` markable. The `link` element contains a reference to the antecedents' ID's value expressed in the attribute POINT-BACK.³ Anaphora which have more antecedents are marked with more link markables. This is for example the case for split plurals.

The type of relation holding between anaphora and antecedents is given in the LTYPE attribute of the `link` markable. At present we distinguish only two types of anaphoric relation: "identity" in case of coreference (value *ident*) and "non_identity" for all the remaining cases (value *non_ident*).

Individual anaphora, abstract anaphora and textual deictics are all linked to their antecedents.

As explained above, however, the implicit referent of individual vague anaphora is expressed in the CONTENT attribute of the `comment` element. Vague abstract anaphora, instead, contain the attribute REF where the implicit referent can be expressed as a string value.

³In PALinkA this reference is automatically marked when coding the `link` markable.

2.7 The PALinKa annotation tool

In the PALinKa annotation tool the annotation scheme is expressed in the form of a preference file where all markables, attributes and possible values are given. When coding in the tool the annotator only sees the markables declared in the preference file, other markables in the data are not visible. We refer to the PALinKa manual for a description of the tool and its uses.

Chapter 3

Annotation examples

In the following we give examples of the annotation of abstract anaphora and their antecedents in XML.

The first example is from an interview in the Danish television and shows an abstract anaphor (the pronoun *det* which refers to an event).

jeg kunne ikke bare fise ud og gå mig en tur og få noget frisk luft hvis jeg skulle have lyst til d,et (I could not just dash out and take a walk and get some fresh air if I wanted to do that) taken from an interview, is the following:

```
<seg ID="s8" syn-type="VP">
<W>fise</W>
<SPEAKER2 turn="2 33.42044919017667 33.658077079300284"/>
<W>ud</W>
<SPEAKER2 turn="2 33.658077079300284 33.876994111012586"/>
<W>og</W>
<SPEAKER2 turn="2 33.876994111012586 34.07345811383132"/>
<W>gaa</W>
<SPEAKER2 turn="2 34.07345811383132 34.193207601263694"/>
<W>mig</W>
<SPEAKER2 turn="2 34.193207601263694 34.24934017349762"/>
<W>en</W>
<SPEAKER2 turn="2 34.24934017349762 34.58382069664495"/>
<W>tur</W>
<SPEAKER2 turn="2 34.58382069664495 34.922822054414716"/>
<W>og</W>
<SPEAKER2 turn="2 34.922822054414716 35.09496194259875"/>
<W>faa</W>
<SPEAKER2 turn="2 35.09496194259875 35.23281130487392"/>
<W>noget</W>
<SPEAKER2 turn="2 35.23281130487392 35.42179096472813"/>
<W>frisk</W>
```

```

<SPEAKER2 turn="2 35.42179096472813 35.67251645403966"/>
<W>luft</W>
</seg>
<SPEAKER2 turn="2 35.67251645403966 35.79600811295429"/>
<W>hvis</W>
<SPEAKER2 turn="2 35.79600811295429 35.855882856670476"/>
<W>jeg</W>
<SPEAKER2 turn="2 35.855882856670476 35.95879257243267"/>
<W>skulle</W>
<SPEAKER2 turn="2 35.95879257243267 36.09163966005296"/>
<W>have</W>
<SPEAKER2 turn="2 36.09163966005296 36.237584347861166"/>
<W>lyst</W>
<SPEAKER2 turn="2 36.237584347861166 36.35172057807014"/>
<W>til</W>
<SPEAKER2 turn="2 36.35172057807014 36.55386770833042"/>
<de ATYPE="abstr-ana" ID="a1" PTYPE="betdet" dist="0"
ref="fise ud og gaa mig en tur og faa noget frisk luft" ref-type="EVENTUALITY">
<link Ltype="no_ident" POINT-BACK="s8"/>
<W>d,et</W>
</de>
</seg>

```

The second example taken from Pirandello's story *Kapellet* illustrates the annotation of the abstract anaphor *det* when it refers to a proposition.

Stillet sig selv op som Kristus, derinde i kapellet. Er det sandt?
(He had placed himself as Christ inside the chapel. Is this true=)

```

<seg ID="s103" syn-type="SCL">
<W id="w6.297.1" lemma="stille" pos="V_PARTC_PAST">Stillet</W>
<W id="w6.297.2" lemma="sig" pos="PRON_PERS">sig</W>
<W id="w6.297.3" lemma="selv" pos="PRON_DEMO">selv</W>
<W id="w6.297.4" lemma="op" pos="ADV">op</W>
<W id="w6.297.5" lemma="som" pos="UNIK">som</W>
<W id="w6.297.6" lemma="kristus" pos="EGEN">Kristus</W>
<W id="w6.297.7" lemma="," pos="TEGN">,</W>
<W id="w6.297.8" lemma="derinde" pos="ADV">derinde</W>
<W id="w6.297.9" lemma="i" pos="PRAEP">i</W>
<W id="w6.297.10" lemma="kapel" pos="N_DEF_SING">kapellet</W>
</seg>
<W id="w6.297.11" lemma="." pos="SENT">.</W>
</S>
<S id="s6.298">

```

```

    <W id="w6.298.1" lemma="vaere" pos="V_PRES">Er</W>
    <de ATYPE="abstr-ana" ID="a102" PTYPE="det" dist="0"
      ref="udsagnet at han har stillet sig op som Kristus"
      ref-type="PROP-L">
<link Ltype="no_ident" POINT-BACK="s103"/>
<W id="w6.298.2" lemma="den" pos="PRON_DEMO">det</W>
  </de>
  <W id="w6.298.3" lemma="sand" pos="ADJ">sandt</W>
  <W id="w6.298.4" lemma="?" pos="SENT">?</W>
</S>

```

In the third example from a court order the abstract anaphor *dette* is referring to an entity of type FACT-LIKE.

I skrivelse af 16/6 2005 til klageren anførte selskabet bl.a.: "De tre billeder af de to hæveskydedøre viser tydeligt at træet er delamineret. Dette ses ved de vandrette revner i bunden af konstruktionerne..."

(In the letter from 16/6 2005 to the compliant the society stated i.a.: "The three pictures of the two sliding doors show clearly that the tree is delaminated. This can be seen from the horizontal fissures in the basis of the constructions...)

```

<S id="s0.19">
<W id="w0.19.1" lemma="i" pos="PRAEP">I</W>
<W id="w0.19.2" lemma="skrivelse" pos="N_INDEF_SING">skrivelse</W>
<W id="w0.19.3" lemma="af" pos="PRAEP">af</W>
<W id="w0.19.4" lemma="16/6" pos="NUM">16/6</W>
<W id="w0.19.5" lemma="2005" pos="NUM">2005</W>
<W id="w0.19.6" lemma="til" pos="PRAEP">til</W>
<W id="w0.19.7" lemma="klager" pos="N_DEF_SING">klageren</W>
<W id="w0.19.8" lemma="anfoere" pos="V_PARTC_PAST">anfoerte</W>
<W id="w0.19.9" lemma="selskab" pos="N_DEF_SING">selskabet</W>
<W id="w0.19.10" lemma="bl.a." pos="ADV">bl.a.</W>
<W id="w0.19.11" lemma=":" pos="TEGN">:</W>
<W id="w0.19.12" lemma="&quot;" pos="TEGN">&quot;</W>
<W id="w0.19.13" lemma="de" pos="PRON_DEMO">De</W>
<W id="w0.19.14" lemma="tre" pos="NUM">tre</W>
<W id="w0.19.15" lemma="billede" pos="N_INDEF_PLU">billeder</W>
<W id="w0.19.16" lemma="af" pos="PRAEP">af</W>
<W id="w0.19.17" lemma="den" pos="PRON_DEMO">de</W>
<W id="w0.19.18" lemma="to" pos="NUM">to</W>
<W id="w0.19.19" lemma="haeveskydedoere" pos="V_INF">
haeveskydedoere</W>
<W id="w0.19.20" lemma="vise" pos="V_PRES">viser</W>
<W id="w0.19.21" lemma="tydelig" pos="ADJ">tydeligt</W>

```



```

<seg ID="s5" syn-type="SCL">
<W id="w0.19.22" lemma="at" pos="UKONJ">at</W>
<W id="w0.19.23" lemma="trae" pos="N_DEF_SING">traeet</W>
<W id="w0.19.24" lemma="vaere" pos="V_PRESENT">er</W>
<W id="w0.19.25" lemma="delaminere" pos="V_PARTC_PAST">
delamineret</W>
</seg>
<W id="w0.19.26" lemma="." pos="TEGN">
.</W>
</S>
<S id="s0.20">
<de ATYPE="abstr-ana" ID="a4" PTYPE="dette" dist="0"
ref="det faktum at traee er delamineret"
ref-type="FACT-LIKE">
<link Ltype="no_ident" POINT-BACK="s5"/>
<W id="w0.20.1" lemma="denne" pos="PRON_DEMO">Dette</W>
</de>
<W id="w0.20.2" lemma="se" pos="V_PRESENT_PAS">ses</W>
<W id="w0.20.3" lemma="ved" pos="PRAEP">ved</W>
<W id="w0.20.4" lemma="den" pos="PRON_DEMO">de</W>
<W id="w0.20.5" lemma="vandret" pos="ADJ">vandrette</W>
<W id="w0.20.6" lemma="revne" pos="N_INDEF_PLU">revner</W>
<W id="w0.20.7" lemma="i" pos="PRAEP">i</W>
<W id="w0.20.8" lemma="bund" pos="N_DEF_SING">bunden</W>
<W id="w0.20.9" lemma="af" pos="PRAEP">af</W>
<W id="w0.20.10" lemma="konstruktion" pos="N_DEF_PLU">
konstruktionerne</W>
<W id="w0.20.11" lemma="." pos="TEGN">.</W>
</S>

```

Chapter 4

Appendix A: Corpus specificities

In the prosodic transcription of spoken language in the DANPASS project (Grønnum 2006) a distinction is made between stressed and unstressed syllabs. Stress is transcribed with a comma preceding the letter on which the stress falls. For example the unstressed version of the pronoun *det* is transcribed as *det_*, while the stressed version of the pronoun is transcribed *d,et_*. More transcription conventions can be found on the DANPASS project homepage http://www.cphling.dk/ng/danpass_webpage/danpass.htm.

Bibliography

- Ariel, M. (1988), ‘Referring and accessibility’, *Journal of Linguistics* **24**(1), 65–87.
- 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.
- Grønnum, N. (2006), DanPASS - A Danish Phonetically Annotated Spontaneous Speech Corpus, in ‘Proceedings of the 5th International Conference on Language Resources and Evaluation’, Genova.
- Grosz, B., Joshi, A. K. & Weinstein, S. (1995), ‘Centering: A Framework for Modeling the Local Coherence of Discourse’, *Computational Linguistics* **21**(2), 203–225.
- Gundel, J. K., Hedberg, N. & Zacharski, R. (1993), ‘Cognitive status and the form of referring expressions in discourse’, *Language* **69**(2), 274–307.
- Kameyama, M. (1998), Intrasentential centering: A case study., in M. Walker, A. Joshi & E. Prince, eds, ‘Centering Theory in Discourse’, Oxford University Press, Oxford, U.K., pp. 89–112.
- Lyons, J. (1977), *Semantics*, Vol. I-II, Cambridge University Press.
- Navarretta, C. (2004), The main reference mechanisms of danish demonstrative pronominal anaphors, in A. Branco, T. McEnery & R. Mitkov, eds, ‘Proceedings of DAARC-2004- 5th Discourse Anaphora and Anaphora Resolution Colloquium’, Ediçoes Colibri, Furnal, S.Miguel, Portugal, pp. 115–120.
- Navarretta, C. (2007), A contrastive analysis of the use of abstract anaphora, in A. Branco, T. McEnery, R. Mitkov & F. Silva, eds, ‘In Proceedings of

DAARC-2007 - 6th Discourse Anaphora and Anaphora Resolution Colloquium', Centro de Linguística da Universidade do Porto, Lagos, Portugal, pp. 103–109.

Orasan, C. (2003), PALinkA: a highly customizable tool for discourse annotation, *in* 'Proceedings of the 4th SIGdial Workshop on Discourse and Dialogue', Sapporo, pp. 39–43.

Peterson, P. L. (1981), 'What causes effects?', *Philosophical Studies* **39**, 107–139.

Poesio, M. (2004), The MATE/GNOME Proposals for Anaphoric Annotation, *in* 'Proceedings of 5th SIGDIAL', Boston, pp. 154–162.

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

Vendler, Z. (1967), *Linguistics in Philosophy*, Cornell University Press, Ithaca.