

# Danish Verb Phrase Anaphora Database Documentation and Instructions

September 7, 2007

## 1 The database

The Danish Verb Phrase Anaphora database is a searchable XML database (XML v. 1.0), which contains examples of Verb Phrase Ellipsis, Verb Phrase Pronominalization, Verb Phrase Topicalization, as well as other, similar examples. It is transformed into HTML using XSLT (v. 1.0). We use UTF-8 character encoding.

## 2 Searching the database

The database can be searched using the online search form at

<http://linguistics.berkeley.edu/~danish/search.html>

Enter the search criteria you are interested in and click the “Search” button. Search criterion can be combined, but be aware that certain combinations will return null results. Upon searching you will receive a list of entries that meet the criteria you entered. To view all of the entries in the database, leave all of the search parameters blank. On the initial display page you will only see the data, gloss, translation, and context (Danish and English) for each entry. To see the entire annotation for an entry, click on the “Display full entry” link.

### 3 A sample entry

```
<sentence>
  <id>P1</id>
  <data>Man ved ikke, hvordan det er at få børn, før man får dem, og
man ved ikke hvordan det er at miste dem, før man gør det.</data>
  <gloss>one know.PRES not how it be.PRES to get children before one
gets them and one know.PRES not how it be.PRES to lose them before one
does DET</gloss>
  <translation>One doesn't know what it is like to have children until
one has them, and one doesn't know what it is like to lose them until one
does.</translation>
  <contextd></contextd>
  <contexte></contexte>
  <construction>VPP</construction>
  <opposite></opposite>
  <fronting>no</fronting>
  <OS>not possible</OS>
  <VPfiniteness></VPfiniteness>
  <remnants>no</remnants>
  <licensor>
    <aux>gøre</aux>
    <tense>present</tense>
    <multiple-aux>no</multiple-aux>
  </licensor>
  <embedded>embedded non-V2</embedded>
  <source>Winther, Henrik. "Efterladt." I Weekend Avisen. March 17-
23, 2006. p4.</source>
  <genre>written</genre>
  <rhet-rel></rhet-rel>
  <negation>no</negation>
  <comments></comments>
</sentence>
```

### 4 Explanation of nodes and their attributes

<sentence>: The sentence node defines an entry in the database. Each entry contains, and is annotated for, one instance of VP anaphora.

<**id**>: Each entry is given an ID number, which is mainly for internal use. ID numbers are assigned by construction type (see below for a description of the construction types), and the IDs consist of a construction code and sequence number. The following construction codes are used:

**P**=VPP  
**E**=VPE  
**T**=VPT  
**PL**=Pronominalization-like  
**EL**=Ellipsis-like

<**data**>: The data node contains the sentence(s) that contains the instance of VP anaphora and its antecedent. The text in this node can be searched using the “Danish text contains” field in the search form.

<**gloss**>: The gloss node contains a word-by-word morpheme gloss of the text in the data node. The text in the gloss node can be searched using the “Gloss contains” field in the search form. The following abbreviations are used in glossing:

DEF = definite  
DET = the VP proform *det*  
DP = discourse particle  
PAST = past  
POSS = possessive  
PRES = present  
REFL = reflexive

<**translation**>: The translation node contains a free translation of the Danish text in the data node. The text in this node can be searched using the “English translation contains” field in the search form.

<**contextd**>: The contextd node contains the text preceding and/or following the sentence(s) containing the instance of VP anaphora. The notation ‘[...]’ indicates the location in the context where the text in the data goes.

<**contexte**>: The contexte node contains a free translation of the Danish text in the contextd node.

<**construction**>: The construction node contains a code for the type of VP anaphora exhibited in the data. This node is searched using the “Construction” menu in the search form. The following codes are used:

- *VPP*: Verb Phrase Pronominalization
- *VPE*: Verb Phrase Ellipsis
- *VPT*: Verb Phrase Topicalization
- *Pronominalization-like*: VP anaphors that are similar to, but not the same as, VPP
- *Ellipsis-like*: VP anaphors that are similar to, but not the same as, VPE
- *Other*: Anything that cannot be categorized as one of the other construction types

<**opposite**>: The opposite node is only filled in for VPP and VPE examples, and indicates whether, for a given example, the other construction is also possible. For example, in a VPP entry, this node indicates whether a VPE version of the sentence is also possible. This node is searched using the “Opposite Construction Possible” menu in the search form. This node has the following attributes:

- *yes*: The other construction is possible
- *no*: The other construction is not possible

<**fronting**>: This node is only filled in for VPP examples, and indicates whether the VP proform, *det* has been fronted to the beginning of the clause or not. This node is searched using the “Fronting” menu in the search form. This node has the following attributes:

- *yes*: The proform has been fronted to clause initial position
- *no*: The proform has not been fronted

<**OS**>: This node is only filled in for VPP examples in which the proform has not been fronted and indicates whether the proform has undergone Object Shift. This node is not searchable. This node has the following attributes:

- *yes*: The conditions on Object Shift are met, and the proform has undergone Object Shift
- *no*: The conditions of Object Shift are met, but the proform has not undergone Object Shift
- *not possible*: The conditions on Object Shift are not met, and therefore, the proform has not shifted
- *can't tell*: The conditions on Object Shift are met, but it isn't possible to tell whether the proform has shifted or not because the clause does not contain any material of the kind that objects shift across (i.e. adverbs or negation)

<**VPfiniteness**>: This node is only filled in for VPT examples and indicates whether the fronted VP is finite or nonfinite. This node is searched using the “VP Finiteness” menu in the search form. This node has the following attributes:

- *tensed*: The fronted VP is finite
- *tenseless*: The fronted NP is nonfinite

<**remnants**>: The remnants node indicates whether or not there have been any VP-internal adverbials “left behind” by the anaphoric process. This node is not searchable and has the following attributes.

- *yes*: There is at least one remnant
- *no*: There are no remnants

<**licensor**>: This node contains information about the auxiliary that licenses the anaphora (i.e. the auxiliary immediately preceding the base position of the VP anaphor). It has three dependent nodes, which are not searchable.

1. <**aux**>: This node gives the identity of the auxiliary that licenses the anaphora
  - *burde* ‘ought’
  - *kunne* ‘can’
  - *mätte* ‘may/must’

- *skulle* ‘must/shall’
- *turde* ‘dare’
- *have* ‘have’
- *være* ‘be’
- *gøre* support verb
- *blive* passive auxiliary, lit. ‘become’

2. <**tense**>: This node indicates the tense of the licensing auxiliary

- *present*: The auxiliary is in the present tense
- *past*: The auxiliary is in the past tense
- *past-part*: The auxiliary is in the past participle form
- *infinitive*: The auxiliary is in the infinitive form

3. <**multiple-aux**>: This node indicates whether there is more than one auxiliary in the clause containing the anaphor, or whether the licensing auxiliary is the only auxiliary.

- *yes*: There is more than one auxiliary in the clause containing the target of anaphora
- *no*: There is only one auxiliary (the licenser) in the clause containing the target of anaphora

<**embedded**>: This node indicates whether the anaphor is in an embedded or a main clause and, if it’s in an embedded clause, what type of embedded clause it is. This node is searched using the “Embedded” menu in the search form and has the following attributes:

- *nonembedded*: The anaphor is in a main clause
- *embedded V2*: The anaphor is in an embedded clause that requires the verb to be in the second position in the clause
- *embedded non-V2*: The anaphor is in an embedded clause, in which the verb stays in its base position

<**source**>: This node contains information about where the data was collected. It is not searchable.

<**genre**>: This node indicates the type of data in the entry. It is searchable using the “Genre” menu in the search form. This node contains the following attributes:

- *oral*
- *written*

<**rhel-rel**>: This node indicates the rhetorical relation between the antecedent and the target clauses. It is searchable using the “Rhetorical Relation” menu in the search form and has the following attributes:

- *parallel*: The antecedent and target clause stand in a relation that highlights similarities between the entities or events represented in the clauses
- *contrast*: The antecedent and target clause stand in a relation that highlights differences between the entities or events represented in the clauses
- *cause*: The antecedent and target clause stand in a relation, in which an entity or event expressed in one of the clauses brings about the event expressed in the other clause

<**negation**>: This node indicates whether the clause containing the anaphor has been negated. It is searchable using the “Negation” menu in the search form and has the following attributes:

- *yes*: The clause containing the anaphor is negated
- *no*: The clause containing the anaphor is not negated

<**comments**>: This node contains any additional information, interesting tidbits, or notes on incomplete or uncertain information in the entry. It is not searchable.