

# Danish Verb Phrase Anaphora Database Documentation and Instructions

April 10, 2009

## 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 constructions. 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>  
 <actual>no</actual>  
 <status>obligatory</status>  
 </fronting>  
 <OS>not possible</OS>  
 <VPfiniteness></VPfiniteness>  
 <remnants>no</remnants>  
 <licensor>  
 <aux>gøre</aux>  
 <tense>present</tense>  
 <multiple-aux>no</multiple-aux>  
 <secondary-gøre></secondary-gøre>  
 </licensor>  
 <agentivity>  
 <control>no</control>  
 <pi>no</pi>  
 </agentivity>  
 <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  
**L** = VPLD  
**R** = VPR  
**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:

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

SUP = superlative

<**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
- *VPLD*: Verb Phrase Left Dislocation (i.e. left dislocation of a VP with clause-internal resumption by the proform *det*)
- *VPR*: Verb Phrase Relativization (i.e. a non-restrictive relative clause on a VP)
- *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 examples of VPP, VPLD, and PL and contains information about the position of the proform. It has two dependent nodes, which are searchable:

1. <**actual**>: This node indicates whether the proform is fronted to clause-initial position and has the following attributes:
  - *yes*: The proform is fronted.
  - *no*: The proform is not fronted.
2. <**status**>: This node indicates whether the fronting or non-fronting of the proform, as indicated by the value in the <**actual**> field, is obligatory. It has the following attributes:
  - *obligatory*: The actual position of the proform (fronted vs. unfronted) is the only one possible.
  - *optional*: The actual position of the proform is not the only possible one; the opposite position (i.e. fronted position if actually not fronted, and unfronted position if actually fronted) is also possible.
  - *unclear*: It is unclear whether the actual position of the proform is the only possible one (either because judgments on opposite position are unclear or because placing the proform in the opposite position requires non-trivial concomitant changes to the string)

<**OS**>: This node is only filled in for VPP, and VPLP 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 and VPLD 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 certain types of material is “left behind” by the anaphoric process, in particular PPs and subordinate clauses. Medial adverbs and negation are not counted as remnants. 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 four dependent nodes, which are 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-perf* ‘have’ (perfect auxiliary)
  - *have-poss* ‘have’ (main verb)

- *være-perf* ‘be’ (perfect auxiliary)
  - *være-cop* ‘be’ (copula)
  - *gøre* support verb
  - *blive-passive* passive auxiliary, lit. ‘become’
  - *blive-inchoative* ‘become’ (inchoative verb)
  - *other* includes non-traditional auxiliaries like *komme* (come) and various periphrastic expression with *være* and *blive* which express aspectual distinctions.
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
4. <**secondary-gøre**>: This node indicates the status of secondary *gøre*, i.e. the use of the support verb *gøre* in the presence of another auxiliary in the target clause. It is not filled in for examples with primary *gøre*, that is examples where *gøre* is the only auxiliary present, nor is it filled in for examples that have the copula, main verb *have*, or inchoative *blive* as licenser, as secondary *gøre* is categorically excluded with these.
- *obligatory*: secondary *gøre* is present and obligatory
  - *opt-no-sem-change*: secondary *gøre* is present, but could be left out without change of meaning
  - *opt-sem-change*: secondary *gøre* is present, but could be left out, though with a change of meaning

- *poss-no-sem-change*: secondary *gøre* is not present, but could be included without any change of meaning
- *poss-sem-change*: secondary *gøre* is not present, but could be included, though with a change of meaning
- *impossible*: secondary *gøre* is not present and could not be included

<**agentivity**>: This node contains information about the agentivity of the anaphoric and/or displaced VP. It has two dependent nodes, which are searchable.

1. <**control**>: This node gives information as to whether the referent of the subject is in control of the eventuality denoted by the VP.
  - *yes*: The referent of the subject is in control of the eventuality
  - *no*: The referent of the subject is not in control of the eventuality
2. <**pi**>: This node indicates whether the referent of the subject performs or instigates the eventuality denoted by the VP.
  - *yes*: The referent of the subject performs or instigates the eventuality denoted by the VP
  - *no*: The referent of the subject does not perform or instigate the eventuality denoted by the VP

<**embedded**>: This node indicates whether the anaphor originates (but not necessarily appears) 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 originates in a main clause
- *embedded V2*: The anaphor originates in an embedded clause that requires the verb to be in the second position in the clause
- *embedded non-V2*: The anaphor originates 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:

- *spoken*
- *written*
- *quoted*

<**rhet-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.