

# HeliPaD

## User's manual and annotation guide

**Citation:** Walkden, George. 2015. *HeliPaD: the Heliand Parsed Database*. Version 0.9. <http://www.chlg.ac.uk/helipad/>

You can also cite the following article: Walkden, George. 2016. [The HeliPaD: a parsed corpus of Old Saxon](#). *International Journal of Corpus Linguistics* 21(4): 559-571.

This manual is based on the manuals for the Penn Historical Corpora of English, to be accessed from [here](#).

The corpus is released under a [CC-BY 4.0](#) license.

To download the corpus, click [here](#).



## Short table of contents

This version of the manual is a hasty copy-and-paste job from the much better version available online at <http://www.chlg.ac.uk/helipad/>. Most links throughout the document refer to that version; this short table of contents is the exception. A future version of the corpus ought to have a more user-friendly manual.

1. Introduction to the corpus .....	4
2. List of tags and empty categories.....	6
3. Treatment of individual words.....	12
4. Morphological annotation.....	14
5. Syntactic annotation.....	21
6. HeliPaD and the Penn Historical Corpora .....	86
7. Known issues .....	90

## Acknowledgements

George would like to thank the following for their support and advice: Kristin Bech, Kersti Börjars, Tine Breban, Anne Breitbarth, David Denison, Tonya Kim Dewey, Aaron Ecay, Melissa Farasyn, Anthony Kroch, Caitlin Light, Svetlana Petrova, Susan Pintzuk, Ann Taylor, Joel Wallenberg, Sheila Watts, David Willis, and Richard Zimmermann. Particular thanks go to Sheila Watts who corrected the POS-tagging on the last thousand lines.

## o. Table of contents (long)

- [HeliPaD home](#)
- [Introduction to the corpus](#)
  - [The language](#)
  - [The text](#)
  - [The corpus](#)
  - [Orthography](#)
  - [Textual and metrical annotation](#)
  - [Lemmatization](#)
  - [Tokenization](#)
  - [HeliPaD vs. DDD](#)
- [List of tags and empty categories](#)
  - [POS tags](#)
  - [Additional attributes](#)
    - [Case](#)
    - [Person](#)
    - [Number](#)
  - [Syntactic tags](#)
    - [Basic syntactic tags](#)
    - [Extended syntactic tags](#)
  - [Empty categories](#)
- [Treatment of individual words](#)
- [POS annotation](#)
  - [Verbs](#)
    - [Modal verbs \(MD, etc.\)](#)
    - [Wita \(UTP\)](#)
    - [Have, be and become \(HV, BE, RD, etc.\)](#)
    - [Lexical verbs \(VB, etc.\)](#)
    - [Particles, prefixes, clitics \(RP, GE, NEG\)](#)
    - [To-infinitives \(TO\)](#)
  - [Nominal words](#)
    - [Nouns \(N, NPR\)](#)
    - [Pronouns \(PRO, PRO\\$, MAN\)](#)
    - [Adjectives \(ADJ, ADJR, ADJS\)](#)
    - [Quantifiers \(Q, QR, QS\)](#)
    - [Numerals \(NUM\)](#)
    - [Determiners \(D\)](#)
  - [Other words](#)
    - [Adverbs \(ADV, ADVR, ADVS, ALSO\)](#)
    - [Prepositions \(P\)](#)
    - [Interjections \(INTJ\)](#)
    - [Complementizers and conjunctions \(C, CONJ\)](#)
    - [Foreign words \(FW\)](#)
    - [Wh-words](#)
  - [Non-words](#)
    - [Punctuation \(, . ' "\)](#)
    - [Metalinguistic information \(CODE\)](#)

- Syntactic annotation
  - General issues
    - Phrase structure
    - Heads, modifiers and complements
    - Sentence fragments (FRAG) and restarts
    - Text problems (X)
    - Direct speech (-SPE and QTP)
    - Left-dislocation (-LFD) and resumption (-RSP)
    - Appositives and parentheticals (-PRN)
    - Empty categories
    - Expletive constructions
    - Conjunction
    - Elision
    - Negation
  - Clauses
    - Clause structure
    - Clausal extended labels
    - Matrix clauses (IP-MAT)
    - Subordinate clauses
    - Adverbial clauses (CP-ADV)
    - That-clauses (CP-THT)
    - Degree complements (CP-DEG)
    - Comparative clauses (CP-CMP)
    - Direct and indirect questions (CP-QUE)
    - Relative clauses (CP-REL)
    - Free relative clauses (CP-FRL)
    - Infinitival clauses (IP-INF, IP-INF-NCO)
    - Small clauses (IP-SMC)
  - Nominals
    - Noun Phrases (NP)
    - Noun phrase extended labels
    - Subjects
    - Non-subject arguments
    - Adjuncts
    - Case attraction
    - Adjective Phrases (ADJP)
    - Quantifier Phrases (QP)
    - Number Phrases (NUMP)
  - Other constituents
    - Adverb Phrases (ADVP)
    - Prepositional Phrases (PP)
    - Interjection Phrases (INTJP)
    - Participle Phrases (PTP)
    - Foreign phrases (LATIN)
    - Wh-phrases (W\*P)
- HeliPaD and the Penn Historical Corpora
  - Introduction
  - General differences
  - Morphological differences
    - Major differences
    - Minor differences
  - Syntactic differences
    - Major, following Penn
    - Major, different policy
    - Minor differences
- Known issues
  - General
  - Morphology
  - Syntax
  - Specific tokens

# 1. Introduction to the corpus

- The language
- The text
- The corpus
- Orthography
- Textual and metrical annotation
- Lemmatization
- Tokenization
- HeliPaD vs. DDD

## The language

Old Saxon (also known as Old Low German) is a West Germanic language spoken in the area of what is now northern Germany before 1100 AD. It is usually thought to be the ancestor of the Middle Low German language, though the extent to which there is continuity between the language represented in the extant Old Saxon texts and that represented in Middle Low German texts is a matter of debate. Old Saxon is transmitted in two main texts: the *Heliand* (which represents the vast majority of attested Old Saxon), and a verse translation of *Genesis*. In addition, there are a number of shorter texts of no more than a few paragraphs each, as well as a number of glosses.

## The text

The *Heliand* is a gospel harmony written in alliterative verse, and a very loose translation of the Latin *Diatessaron*. In total, 5,983 lines have been preserved, in six manuscripts: C (Cotton), M (Monacensis), S (Straubing), V (Vatican), P (Prague), and L (Leipzig). The S, V, P and L manuscripts are extremely limited in extent, and none of them contains a continuous stretch of more than a hundred lines. The M and C manuscripts are the main witnesses to the text. While the M manuscript contains a number of gaps, the C manuscript (Cotton Caligula A VII, British Library) is complete up to line 5,968. The text is divided into 71 sections, called *fitts*.

There exist two main editions of the *Heliand*: Sievers (1878), a broadly diplomatic edition of manuscripts C and M, and Behaghel (1903 and subsequent editions), the standard critical edition.

## The corpus

This corpus contains all 5,968 lines of the C manuscript of the *Heliand*, using the Sievers (1878) edition. Compared to the standard Behaghel critical edition, this one has the advantages for linguistic research that a) it does not conflate the different forms found in different manuscripts, b) it is not as heavily emended, and c) it is now **in the public domain**.

The corpus is a UTF-8 plain text file designed to be searched using the program **CorpusSearch 2**, with the standard extension `.psd`, broadly following the format of the **Penn Corpora of Historical English** and related projects (**IcePaHC**, **Early New High German Parsed Corpus**, **MCVF**). It is annotated on a number of levels:

- Textual and metrical (page in manuscript, page in edition, line number, caesura)
- Lemmatization
- Parts of speech and morphology
- Syntactic parsing

The total size of the corpus is 46,067 words (not including punctuation and code).

## Orthography

The corpus character encoding is UTF-8, and contains certain special characters such as barred b and d. Word forms are kept as they are in the Sievers edition. Where words have been broken up to facilitate parsing, the site of the break is marked with a dollar sign (\$).

## Textual and metrical annotation

Textual and metrical annotation is POS-tagged as CODE and contained within angle brackets. The order of precedence of these elements is as they appear below.

- Sievers edition page: e.g. P\_7
- Manuscript page: e.g. MS\_5a
- Fitt: e.g. F\_1
- Line: e.g. R\_1
- Caesura (half-line break): C
- Other comments (mostly omissions): e.g. COM:OMISSION

## Lemmatization

A significant difference between the Penn Corpora of Historical English and the HeliPaD (and a property that the HeliPaD shares with the IcePaHC) is that the HeliPaD is lemmatized. The lemma is given after the word form and separated by a hyphen: thus, for the second person singular present indicative of the verb "to be" (*wesan*), what is found in the corpus is *bist-wesan*.

Lemmas are based in form on Köbler's [freely-available Old Saxon dictionary](#), minus length markings. To search a word when you don't know its lemma, the easiest way is to look it up in Köbler. (Note that my assignment of forms to lemmas is not always the same as Köbler's.)

Some words are, unfortunately, indistinguishable by lemma. A small minority of these are also indistinguishable by POS-tag: these include *bord* "edge" and *bord* "shield", and *ger* "year" and *ger* "spear".

In compounded words (joined by a plus sign, +), only the head of the compound is lemmatized. In practice, these are instances of prefixation with GE+ or NEG+ and can be identified morphologically.

## Tokenization

A token is, broadly speaking, a main verb and everything that belongs with it. In many cases, it will be a "sentence", in pretheoretical terms. The main exception is when two independent clauses with finite verbs are conjoined, in which case these are treated as separate tokens.

The token is enclosed in brackets, and consists of a parse followed by a token ID, each of which is itself enclosed in brackets. The ID takes the form **OSHeliandC.foo.bar**, where **foo** is simply a sequential number starting at 1 and **bar** is the range of lines spanned by the token. For instance, **OSHeliandC.265.502-503** is the ID for a token that starts on line 502 and ends on line 503 and is the 265th token in total.

## HeliPaD vs. DDD

The corpus overlaps with the version of the *Heliand* produced as part of the [Referenzkorpus Altdeutsch \(DDD\)](#), but there are a number of differences. The DDD version is based on the Behaghel edition and only contains very shallow parsing. Unlike this corpus, it also contains annotation for alliteration, and indicates stem class of nouns and verbs. The two resources are thus to some extent complementary.

## 2. List of tags and empty categories

- POS tags
- Additional attributes
  - Case
  - Person
  - Number
- Syntactic tags
  - Basic syntactic tags
  - Extended syntactic tags
- Empty categories

### POS tags

A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z

.	sentence-final punctuation
,	sentence-internal punctuation
'	quotation mark
"	double quotation mark
+	joins constituent morphemes in compounds
<b>A</b>	
ADJ	adjective
ADJR	adjective, comparative
ADJS	adjective, superlative
ADV	adverb
ADVR	adverb, comparative
ADVS	adverb, superlative
ALSO	the word <i>ok</i>
<b>B</b>	
BE	<i>wesan</i> , infinitive
BEDI	<i>wesan</i> , past indicative
BEDS	<i>wesan</i> , past subjunctive
BEI	<i>wesan</i> , imperative
BEPI	<i>wesan</i> , present indicative
BEPS	<i>wesan</i> , present subjunctive
BG	<i>wesan</i> , present participle (uninflected)
BGI	<i>wesan</i> , present participle (inflected)
BN	<i>wesan</i> , past participle (uninflected)
BNI	<i>wesan</i> , past participle (inflected)
<b>C</b>	
C	complementizer
<b>CODE</b>	non-text material (e.g., page or line number, caesura)
CONJ	coordinating conjunction
<b>D</b>	
D	determiner
<b>E</b>	
<b>F</b>	
FW	foreign word
<b>G</b>	
GE	the prefix <i>ge-</i>

<b>H</b>	
HG	<i>hebbian</i> , present participle (uninflected)
HGI	<i>hebbian</i> , present participle (inflected)
HN	<i>hebbian</i> , past participle (uninflected)
HNI	<i>hebbian</i> , past participle (inflected)
HV	<i>hebbian</i> , infinitive
HVDI	<i>hebbian</i> , past indicative
HVDS	<i>hebbian</i> , past subjunctive
HVI	<i>hebbian</i> , imperative
HVPI	<i>hebbian</i> , present indicative
HVPS	<i>hebbian</i> , present subjunctive
<b>I</b>	
INTJ	interjection
<b>J</b>	
<b>K</b>	
<b>L</b>	
<b>M</b>	
MAN	indefinite <i>man</i>
MD	modal verb, infinitive
MDDI	modal, past indicative
MDDS	modal, past subjunctive
MDI	modal, imperative
MDPI	modal, present indicative
MDPS	modal, present subjunctive
MG	modal, present participle (uninflected)
MGI	modal, present participle (inflected)
MN	modal, past participle (uninflected)
MNI	modal, past participle (inflected)
<b>N</b>	
N	common noun
NEG	negation
NPR	proper noun
NUM	numeral
<b>O</b>	
<b>P</b>	
P	preposition
PRO	personal pronoun
PRO\$	possessive pronoun
<b>Q</b>	
Q	quantifier
QR	quantifier, comparative (MORE, LESS)
QS	quantifier, superlative (MOST, LEAST)
<b>R</b>	
RD	<i>werthan</i> , infinitive
RDDI	<i>werthan</i> , past indicative
RDDS	<i>werthan</i> , past subjunctive
RDI	<i>werthan</i> , imperative
RDPI	<i>werthan</i> , present indicative
RDPS	<i>werthan</i> , present subjunctive
RG	<i>werthan</i> , present participle (uninflected)

RGI	<i>werthan</i> , present participle (inflected)
RN	<i>werthan</i> , past participle (uninflected)
RNI	<i>werthan</i> , past participle (inflected)
RP	adverbial particle (note that RP can adjoin to verbs)
<b>S</b>	
<b>T</b>	
TO	infinitival <i>to</i>
<b>U</b>	
UTP	the word <i>uton</i>
<b>V</b>	
VB	infinitive, verbs other than BE, HV, RD, modals
VBDI	past indicative
VBDS	past subjunctive
VBI	imperative
VBPI	present indicative
VBPS	present subjunctive
VG	present participle (uninflected)
VGI	present participle (inflected)
VN	past participle (uninflected)
VNI	past participle (inflected)
<b>W</b>	
WADJ	<i>wh</i> -adjective
WADV	<i>wh</i> -adverb
WPRO	<i>wh</i> -pronoun
WPRO\$	possessive <i>wh</i> -pronoun
WQ	<i>hwethar</i> (as polar-question-introducer)
<b>X</b>	
<b>Y</b>	
<b>Z</b>	

## Additional attributes

The HeliPaD differs from the Penn historical corpora of English in making more extensive use of attributes separated by a delimiter, <sup>^</sup>. Where a part of speech has more than one of the three attributes, they occur in the order given below (case, person, number). So a third person singular personal pronoun in the dative case would be PRO<sup>^</sup>D<sup>^</sup>3<sup>^</sup>SG.

Note that all elements that can receive attributes must receive them. This is the case even for elements that are formally invariant (e.g. PRO\$ *his*, PRO\$ *iru*, Q *filo*). Participles with attributes are marked with an I (e.g. VGI, VNI); those with no morphological marking have no I (e.g. VG, VN).

### Case

The possible attribute values for case are <sup>^</sup>N, <sup>^</sup>A, <sup>^</sup>G, <sup>^</sup>D and <sup>^</sup>I. Our approach differs from that of the YCOE in that we label all relevant parts of speech for case, not just morphologically unambiguous instances.

The following parts of speech are labelled for case:

- nouns (N, NPR)
- attributive adjectives (ADJ, ADJR, ADJS, WADJ)
- inflected participles (VNI, VGI, etc.)
- quantifiers (Q, QR, QS)
- determiners (D)



- numbers (NUM)
- pronouns (MAN, PRO, PRO\$, WPRO, WPRO\$)

The case of PRO\$ is the case it receives from context (so not always genitive).

### *Person*

The possible attribute values for person are ^1, ^2 and ^3.

The following parts of speech are labelled for person:

- pronouns (MAN, PRO, PRO\$)
- finite imperative verbs (BEI, HVI, MDI, RDI, VBI)
- finite past indicative verbs (BEDI, HVDI, MDDI, RDDI, VBDI)
- finite past subjunctive verbs (BEDS, HVDS, MDDS, RDDS, VBDS)
- finite present indicative verbs (BEPI, HVPI, MDPI, RDPI, VBPI)
- finite present subjunctive verbs (BEPS, HVPS, MDPS, RDPS, VBPS)

### *Number*

The possible attribute values for number are ^SG, ^PL and ^DU (the latter in Old Saxon only). Our approach differs from the Penn historical corpora of English in that, for consistency, we mark number on nouns as an attribute rather than using the additional tags NS and NPRS.

The following parts of speech are labelled for number:

- nouns (N, NPR)
- adjectives (ADJ, ADJR, ADJS, WADJ)
- inflected participles (VNI, VGI, etc.)
- quantifiers (Q, QR, QS)
- determiners (D)
- pronouns (MAN, PRO, PRO\$)
- finite imperative verbs (BEI, HVI, MDI, RDI, VBI)
- finite past indicative verbs (BEDI, HVDI, MDDI, RDDI, VBDI)
- finite past subjunctive verbs (BEDS, HVDS, MDDS, RDDS, VBDS)
- finite indicative verbs (BEPI, HVPI, MDPI, RDPI, VBPI)
- finite subjunctive verbs (BEPS, HVPS, MDPS, RDPS, VBPS)

The number of PRO\$ is the number it receives from context, not its inherent value. So, for example, *usa* "our" will be treated as singular where it occurs with a singular noun.

## **Syntactic tags**

### *Basic syntactic tags*

Clausal:

- IP: basic clause, including verb
- IPX: incomplete IP
- CP: clausal shell for subordinate and non-declarative clauses
- CPX: incomplete CP

Nominal:

- NP: noun phrase
- WNP: *wh*- noun phrase

- NUMP: number phrase
- QP: quantifier phrase
- WQP: *wh*- quantifier phrase
- ADJP: adjectival phrase
- WADJP: *wh*- adjectival phrase
- PP: prepositional phrase
- WPP: *wh*- prepositional phrase

Other:

- FRAG: sentence fragment
- PTP: participial phrase
- QTP: quotative phrase (quoted non-sentential sequences)
- ADVP: adverbial phrase
- WADVP: *wh*- adverbial phrase
- CONJP: conjunction phrase
- INTJP: interjection phrase
- X: incomplete phrase due to text problems

### *Extended syntactic tags*

For nominals:

- Arguments:
  - -SBJ: subject
  - -OB1: direct or only object
  - -OB2: indirect or second object
  - -PRD: predicate (also for ADJPs and PTPs)
- Non-arguments:
  - -ADT: adjunct (also for ADJPs and PTPs)
  - -LOC: locative
  - -TMP: temporal
  - -VOC: vocative
  - -POS: possessor (only one immediately dominated by each NP)
- -RFL: reflexive (not mutually exclusive with above)

For IPs:

- -MAT: main clause
- -SUB: subordinate clause, inc. direct questions
- -INF: infinitival clause
- -INF-NCO: non-complement infinitival clause
- -SMC: small clause
- -SPE: direct speech (not mutually exclusive with above)

For CPs:

- -REL: relative clause
- -FRL: free relative clause
- -THT: *that*-clause
- -ADV: adverbial clause
- -CMP: comparative clause
- -QUE: interrogative clause
- -DEG: degree clause
- -SPE: direct speech (not mutually exclusive with above)

For ADVPs:

- -LOC: locative
- -DIR: directional
- -TMP: temporal

Miscellaneous:

- -PRN: parenthetical
- -LFD: left-dislocated (with later -RSP)
- -RSP: resumptive (usually with earlier -LFD)
- -X: expletive subject or associate
- -1, -2, -3 etc.: indices indicating movement
- -0, -00, -00 etc.: indices indicating antecedents in instances of ellipsis
- =0, =00, =00 etc.: indices indicating ellipsis sites

Order of extended tags where they co-occur: clausal (except -SPE), nominal (except -RFL), -RFL, -PRN/-LFD/-RSP, -SPE, -X/-1/-2/-3/-0/-00/-000/=0/=00/=000 etc.

## Empty categories

- \*con\*: subject elided under co-ordination
- \*exp\*: null expletive subject (may be co-referential with a clausal element)
- \*arb\*: arbitrary subject in infinitivals
- \*pro\*: null subject, none of the above (usually referential)
- \*T\*: trace of *wh*-movement, always coindexed
- \*ICH\*: trace of other movement, always coindexed
- \*: other empty category (e.g. in *tough*-movement)

### 3. Treatment of individual words

A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z

#### A

*al* Usually Q. Sometimes ADV.  
*al so* CP-ADV. *al* is ADV with untraced ADVP in SpecCP; *so* is C.

#### B

*belgan* VB. May be an accusative subject verb.  
*beter* and *best* ADJR/ADVR and ADJS/ADVS respectively. Lemmatized under *bet*.  
*biddian* VB. Often occurs with an accusative object and a genitive object. In this structure, the genitive object is OB2.  
*biforan* ADV. May head either ADVP-TMP or ADVP-LOC.  
*bihwi* WADV.  
*butan* P, sometimes with clausal complement.

#### C

#### E

*ef* C.  
*en* NUM, even when it seems to be an indefinite determiner or has a focus/"alone" reading.  
*eowiht* N.  
*erist* ADJS or ADVS.

#### F

*fragon* VB. Often occurs with an accusative object and a genitive object. In this structure, the genitive object is OB2.

#### G

*gihwe* GE+WPRO. Is not dominated by a WNP.  
*gihwilik* GE+WADJ. Is not dominated by a WNP.

#### H

*harm* Usually a predicate co-occurring with an extraposed clausal subject.  
*hreuwan* VB. May be an accusative subject verb.  
*hwand* C.  
*hwat* Lemma always *hwe*. WPRO normally; WADV when used as an "interjection", dominated by INTJP.  
*hwethar* WQ when introducing yes-no questions; WADJ when meaning "which of two".  
*hwi* Lemma always *hwe*. WPRO when it occurs as the complement of a preposition. WADV otherwise.  
*hwilik* WADJ.

#### J

*ja* INTJ ("yes") or CONJ ("both"/"and").

#### L

*lerian* VB. Unusually, may take two accusative objects. The person or people being taught are always OB2, even when accusative.  
*likon* VB. Dative subject verb.  
*lustian* VB. Accusative subject verb. Other argument typically genitive.  
*luttill* ADJ.

#### M

*man* MAN when it occurs unmodified as singular subject. N otherwise.  
*manon* VB. Often occurs with an accusative object and a genitive object. In this structure, the genitive object is OB2.  
*mer, mero* and *mest* QR, QR, QS (lemma *mer* for the latter).

<i>mikil</i>	ADJ.
<b>N</b>	
<i>ne</i>	NEG alone as a negator. NEG+CONJ when it is a conjunction. NEG+C when it is a complementizer.
<i>ne si</i>	CP-ADV with no C and IP-SUB.
<i>nek</i>	NEG+CONJ.
<i>neo</i>	NEG+ADV (lemma <i>eo</i> ).
<i>neowiht</i>	NEG+N (lemma <i>ewiht</i> ).
<i>nen</i>	INTJ.
<i>newa</i>	P or C depending on its complement.
<i>newan</i>	P, sometimes with clausal complement.
<i>nigen</i>	NEG+Q.
<i>niud</i>	Usually a predicate co-occurring with an extraposed clausal subject.
<b>O</b>	
<i>ok</i>	ALSO.
<b>R</b>	
<i>reht so</i>	CP-ADV. <i>reht</i> is ADV with untraced ADVP in SpecCP; <i>so</i> is C.
<i>risan</i>	VB. May be a dative subject verb.
<b>S</b>	
<i>scin</i>	Usually a predicate co-occurring with an extraposed clausal subject.
<i>self</i>	ADJ.
<i>sia</i>	Always lemmatized under <i>he</i> or <i>siu</i> , even when plural.
<i>sinu</i>	INTJ.
<i>so</i>	C in comparative contexts, ADV otherwise, often modifying another element.
<i>spanan</i>	VB. May be an accusative subject verb. Other argument genitive or nominative.
<i>sulik</i>	ADJ.
<b>T</b>	
<i>te</i>	Usually P, but can be ADV when it means "too (much)".
<i>than</i>	C in comparative contexts, ADV otherwise - either heading ADVP-TMP or modifying another element.
<i>tharf</i>	Usually a predicate co-occurring with an extraposed clausal subject.
<i>that</i>	Normally D (lemma <i>the</i> ), including in relative clauses, but can be C (lemma <i>that</i> ) when heading a <i>that</i> -clause or degree clause.
<i>the</i>	Most usually D, or C when it is an invariant complementizer, or CONJ when it is a conjunction.
<i>these</i>	D.
<i>thunkian</i>	VB. Dative subject verb. Often occurs with a genitive object. In this structure, the genitive object is OB2.
<i>thurstian</i>	VB. Accusative subject verb.
<b>U</b>	
<i>und</i>	Usually P. May be ADV in the ADV-TMP <i>und er</i> .
<i>untat</i>	C.
<b>W</b>	
<i>wela</i>	INTJ.
<i>wola</i>	INTJ.
<i>wiht</i>	N.

## 4. Morphological annotation

- Verbs
  - Modal verbs (MD, etc.)
  - *Wita* (UTP)
  - *Have, be* and *become* (HV, BE, RD, etc.)
  - Lexical verbs (VB, etc.)
  - Particles, prefixes, clitics (RP, GE, NEG)
  - *To*-infinitives (TO)
- Nominal words
  - Nouns (N, NPR)
  - Pronouns (PRO, PRO\$, MAN)
  - Adjectives (ADJ, ADJR, ADJS)
  - Quantifiers (Q, QR, QS)
  - Numerals (NUM)
  - Determiners (D)
- Other words
  - Adverbs (ADV, ADVR, ADVS, ALSO)
  - Prepositions (P)
  - Interjections (INTJ)
  - Complementizers and conjunctions (C, CONJ)
  - Foreign words (FW)
  - *Wh*-words
- Non-words
  - Punctuation ( , . ' ")
  - Metalinguistic information (CODE)

☒ Where the HeliPaD follows other Penn historical corpora, the text will be marked like this.

☒ Where the HeliPaD does its own thing, the text will be marked like this.

There's also a [page](#) with a summary of these differences.

See also the [List of tags and empty categories](#).

### Verbs

☒ All finite verbs have person and number marked as attributes. See [Additional attributes](#).

☒ The ambiguity tags VBP and VBD etc., for formally ambiguous indicative/subjunctive/imperative verbs, are not used. (Verb form classification follows Köbler.)

☒ The tag AX\*, for auxiliary verbs, is not used.

☒ Participles are divided into inflected (VGI, VNI, etc.) and uninflected (VG, VN, etc.) tags. The inflected tags take nominal attributes.

#### *Modal verbs (MD, etc.)*

- MD: infinitive
- MDI: imperative
- MDPI: present indicative
- MDPS: present subjunctive
- MDDI: past indicative
- MDDS: past subjunctive

- MG: present participle, uninflected
- MGI: present participle, inflected (not used)
- MN: past participle, uninflected (not used)
- MNI: past participle, inflected (not used)

Modals are a closed class including (always and) only the following: *kunnan*, *motan*, *mugan*, *skulan*, *thurvan*, *willian*.

### Wita (UTP)

*Wita*, which is used to introduce hortative clauses ("let us..."), is tagged UTP.

### Have, be and become (HV, BE, RD, etc.)

- HV: infinitive
  - HVI: imperative
  - HVPI: present indicative
  - HVPS: present subjunctive
  - HVDI: past indicative
  - HVDS: past subjunctive
  - HG: present participle, uninflected (not used)
  - HGI: present participle, inflected (not used)
  - HN: past participle, uninflected (not used)
  - HNI: past participle, inflected (not used)
- 
- BE: infinitive
  - BEI: imperative
  - BEPI: present indicative
  - BEPS: present subjunctive
  - BEDI: past indicative
  - BEDS: past subjunctive
  - BG: present participle, uninflected (not used)
  - BGI: present participle, inflected (not used)
  - BN: past participle, uninflected (not used)
  - BNI: past participle, inflected (not used)
- 
- RD: infinitive
  - RDI: imperative (not used)
  - RDPI: present indicative
  - RDPS: present subjunctive
  - RDDI: past indicative
  - RDDS: past subjunctive
  - RG: present participle, uninflected (not used)
  - RGI: present participle, inflected (not used)
  - RN: past participle, uninflected
  - RNI: past participle, inflected (not used)

There is a one-to-one mapping between the lemma *hebbian* and the tags HV\*/HG\*/HN\*, between *wesan* and BE\*/BG\*/BN\*, and between *werthan* and RD\*/RG\*/RN\*.

☐ Forms of *werthan* are tagged RD\*/RG\*/RN\*, as in the IcePaHC and ENHG Parsed Corpus.

### Lexical verbs (VB, etc.)

- VB: infinitive
- VBI: imperative

- VBPI: present indicative
- VBPS: present subjunctive
- VBDI: past indicative
- VBDS: past subjunctive
- VG: present participle, uninflected
- VGI: present participle, inflected
- VN: past participle, uninflected
- VNI: past participle, inflected

All remaining verbs are labelled VB\*/VG\*/VN\*.

### *Particles, prefixes, clitics (RP, GE, NEG)*

☐ The tag RP is closed class, and used for the particles *an*, *to*, *up* and *ut*. It does not occur prefixed to verbs as in the YCOE.

☐ The tag GE has a one-to-one mapping with the prefix *gi-*. It never occurs independently, but always prefixed/cliticized to a verbal form. Nominal *gi-* is not tagged in this way, except in the context of *gihwilik* and *gihwe*.

The negative particle *ne* (NEG) can occur either independently or prefixed/cliticized to a verbal, adverbial, or nominal form, or to a conjunction.

Clitic forms (separated by a plus sign) are not reflected in the **lemmatization**.

### *To-infinitives (TO)*

The tag TO is used for forms of *to* when co-occurring with inflected infinitives.

☐ Inflected infinitives are not given special treatment, unlike in the YCOE. They can always be retrieved due to their co-occurrence with TO within an IP-INF.

## **Nominal words**

☐ All nominal words have case and number marked as attributes. Pronouns also have person marked as an attribute. See **Additional attributes**. Attribute annotation follows Köbler. Where in doubt, nominative has been preferred over accusative over dative over genitive over instrumental.

NP-internal agreement is forced wherever possible, thus allowing ambiguous elements to be tagged (the HeliPaD's approach to attributes is "maximalist"). The main exception to this is with instrumental elements, which often co-occur with formally dative elements. This is not treated as a case clash, and instrumental is only preferred where unambiguous.

### *Nouns (N, NPR)*

All singular, plural, collective, and compound nouns are tagged as N. See the syntactic manual on **Noun Phrases** for details of compounding.

☐ Proper nouns are tagged as NPR (not NR as in the YCOE).

### *Pronouns (PRO, PRO\$, MAN)*

MAN is used for singular, unmodified *man* subjects.



PRO is used for personal pronouns. They can be used as reflexives in the HeliPaD, but this is indicated **at phrase level**. PRO is a closed class consisting of the following lemmas: *ik, wit, we, thu, git, gi, he, siu, it*.

Subject pronouns enclitic to verbs are always separated out for the purposes of parsing. This is done using the dollar sign (\$) and without adding or removing segments.

PRO\$ is used for possessive "pronouns", and is also a closed class. Like other nominal categories, it is annotated for case and number, and also for person. The third person forms *is* and *iru* are never inflected, but receive attributes anyway, in agreement with other elements in their phrase. (When there are no such other elements these are treated as genitive pronoun forms.) The other forms - *min, unka, usa, thin, inka, iuwa*, and the reflexive *sin* - are formally inflected and this is reflected in the annotation for attributes.

### **Adjectives (ADJ, ADJR, ADJS)**

ADJR and ADJS are used for comparative and superlative adjectives respectively. All other adjectives are ADJ.

☐ Unlike in the YCOE, where a weak adjective is used nominally (i.e. without a noun head), it normally retains its adjectival tag.

Ordinal numbers, including *othar*, are also tagged ADJ (and not NUM). *erist* is treated as superlative, and may also be tagged ADVS when it is a temporal adverb.

☐ The adjectives *mikil* and *luttil* are tagged as adjectives, even when they are clearly quantifiers. Cognates in the YCOE and other Penn corpora are treated in the exact opposite way.

*Self* and *sulik* are tagged as ADJ.

☐ Annotation as an inflected participle (VGI, VNI, etc.) is always preferred to annotation as an adjective, if possible.

### **Quantifiers (Q, QR, QS)**

Quantifiers are a closed class, and include all and only the following: *al, bethia, enhwilik, enig, filo, manag, sum*. These elements have a tendency not to inflect (and *filo* never does). When they do inflect, they do so in a similar way to adjectives, but are never tagged as such. They always bear case and number attributes in the annotation. *al* can also be an adverb on occasion.

☐ *Wh*-indefinites such as *hwilik* are always tagged as W\* and not as Q, regardless of their syntactic role, which is disambiguated at phrasal level.

*nigen* is treated as NEG+Q, and *neowiht* as NEG+N.

☐ Some apparently quantificational elements such as *wiht, eowiht* etc. are treated as nouns in the HeliPaD rather than as quantifiers as the corresponding items are in the YCOE.

### **Numerals (NUM)**

Numerals, when cardinal, are tagged NUM; this is in principle a closed class. *half* is treated as a numeral. *en* is treated as a numeral, even when it seems to be an indefinite determiner, means "alone" and/or has a focus reading.

### **Determiners (D)**

Determiners are a closed class which includes *the* (in many forms), the distal demonstrative/article, and *these*, the proximal demonstrative.

## Other words

☐ The FP (focus particle) and XX (problematic word) tags are not used in the HeliPaD, mainly since there is no call for them in the current material.

### *Adverbs (ADV, ADVR, ADVS, ALSO)*

☐ Adverbs do not bear extended tags ^T, ^L and ^D for temporal, locative and directional, as they do in the YCOE. This information is retrievable from the **phrasal extended label** and from the lemma.

The following adverbs always head ADVP-

TMP: *aftar, eft, eo, er, erist, forn, get, ju, hald, hindag, hiudu, hwanna, lang, lango, noh, nu, oft, san, sana, simbla, simblon, sith, sithor, sniumo, tho, und.*

The following adverbs always head ADVP-

DIR: *angepin, ellior, fer, ferran, forana, herod, herodwardes, hinan, in, nithana, nithar, north, ostana, os tar, tegegnas, thanan, tharod, thurh, towardes, westan, westar, witharwardes.*

The following adverbs always head ADVP-

LOC: *bihindan, biovan, foran, her, innan, nithara, sundar, thar, uppan, up, uta, wido.*

The following adverbs may head either ADVP-TMP or ADVP-DIR: *forth, forthward, furthor.*

The following adverbs may head either ADVP-DIR or ADVP-LOC: *hoho, nah, ostan, ovana.*

*biforan* may be temporal or locative. *than* may be temporal or atemporal. *neo* (lemma *eo*) is tagged NEG+ADV.

☐ The word *ok* is tagged ALSO (the cognate is ADV in the YCOE). ALSO does not head a phrase, may modify adjectives, and often co-occurs with conjunctions within a CONJP.

Some words function as both adverbs and prepositions. An analysis as a (stranded) preposition is always preferred when a possible complement is present.

### *Prepositions (P)*

The P tag is always and only used for prepositions with a complement. Otherwise, these elements are labelled ADV or RP.

☐ Subordinators are not tagged as P in the HeliPaD. They are treated as either **complementizers** or **adverbs**.

The following words may be tagged

P: *af, aftar, an, and, angegin, ano, at, bi, biforan, butan, er, fan, farutar, for, fram, furi, in, innan, inne, m id, newa, newan, ovar, te, thurh, to, twisk, um, umbi, und, undar, up, uppa, uppan, uta, with, withar.*

*butan, newa* and *newan* can take a *that*-clause complement, though *newa* is more usually C.

### *Interjections (INTJ)*

INTJ is a small closed class only used when no other analysis is available. The HeliPaD contains the following interjections: *ja, nen, sinu, wela, wola*. "Interjectional" *hwat* (lemma *hwe*) is WADV and dominated by an INTJP.

### Complementizers and conjunctions (C, CONJ)

The following may be tagged as **co-ordinating conjunctions**: *ak, eftha, endi, ge, ja, jak, noh, the*.

*ne* and *nek*, when used as conjunctions, are tagged NEG+CONJ.

☐ The word *ok* is tagged ALSO (the cognate is ADV in the YCOE). ALSO does not head a phrase, may modify adjectives, and often co-occurs with conjunctions within a CONJP.

☐ **Subordinating conjunctions** are treated as **adverbs** in **adverb phrases** if they are homophonous with adverbs, as they usually are, and never as prepositions.

The genuine complementizer tag (C) is limited to *ef, hwand, newa, so, than* in comparatives, *that, the*, and *untat*. In addition, *ne* when used as a complementizer is tagged NEG+C. Most of the time, in the HeliPaD, C is null.

### Foreign words (FW)

Unintegrated foreign words, always Latin in the HeliPaD, are labelled FW.

### Wh-words

All *wh*-words are closed class. The following tags exist: WADJ, WADV, WPRO, WPRO\$ (not used), WQ.

☐ Morphologically *wh*- elements are tagged using the W\* tags even when they are not part of an extraction structure. Since Old Saxon is particularly flexible in using *wh*-words as indefinites, this is quite important.

WADJ is used for *hwilik*, which may head a (W)ADJP, or may be part of a (W)NP (when it is not the head). WADJ takes the attributes of an **adjective**. In non-extraction structures, *gihwilik* occurs and is tagged GE+WADJ. The tag WADJ is also used for *hwethar* when it means "which of two".

WADV is used for *bihwi*, *hwan* (projecting (W)ADVP-TMP), *hwanan* (projecting (W)ADVP-DIR), *hwar* (projecting (W)ADVP-LOC), *hwarod* (projecting (W)ADVP-DIR), and *hwo*. It is also used for "interjectional" *hwat* (lemma *hwe*), which is dominated by INTJP.

The instrumental form *hwi* (lemma *hwe*) meaning "why", when used alone to form a question, is treated as WADV as in the YCOE. When it occurs with a preposition (e.g. *te hwi*), it is treated as WPRO. The fixed combination *bihwi* is always WADV.

WPRO behaves like other **pronouns** in terms of its morphology and attributes, though does not take person or number. It is used for the generic *wh*-element *hwe*, except for "interjectional" *hwat*. In non-extraction structures, *gihwe* occurs and is tagged GE+WPRO.

WPRO\$; is the *wh*-counterpart of PRO\$, and could in principle be used for non-head instances of the genitive *hwes*. No instances have been found in the HeliPaD.

When it introduces a yes-no question, *hwethar* receives the tag WQ, which takes no attributes.

## Non-words

### Punctuation (, . ' ")

Editorial speech marks, either single or double, are tagged as themselves (' or ").

All other punctuation is either tagged as . (if it is token-final, modulo speech marks and CODE elements) or as , (otherwise). Remember that CorpusSearch ignores punctuation by default.

### *Metalinguistic information (CODE)*

The tag CODE is used for the following things (which, when they occur, occur in the following order):

- Sievers edition page: e.g. P\_7
- Manuscript page: e.g. MS\_5a
- Fitt: e.g. F\_1
- Line: e.g. R\_1
- Caesura (half-line break): C
- Other comments (mostly omissions): e.g. COM:OMISSION

## 5. Syntactic annotation

- General issues
  - Phrase structure
  - Heads, modifiers and complements
  - Sentence fragments (FRAG) and restarts
  - Text problems (X)
  - Direct speech (-SPE and QTP)
  - Left-dislocation (-LFD) and resumption (-RSP)
  - Appositives and parentheticals (-PRN)
  - Empty categories
  - Expletive constructions
  - Conjunction
  - Elision
  - Negation
- Clauses
  - Clause structure
  - Clausal extended labels
  - Matrix clauses (IP-MAT)
  - Subordinate clauses
  - Adverbial clauses (CP-ADV)
  - *That*-clauses (CP-THT)
  - Degree complements (CP-DEG)
  - Comparative clauses (CP-CMP)
  - Direct and indirect questions (CP-QUE)
  - Relative clauses (CP-REL)
  - Free relative clauses (CP-FRL)
  - Infinitival clauses (IP-INF, IP-INF-NCO)
  - Small clauses (IP-SMC)
- Nominals
  - Noun Phrases (NP)
  - Noun phrase extended labels
  - Subjects
  - Non-subject arguments
  - Adjuncts
  - Case attraction
  - Adjective Phrases (ADJP)
  - Quantifier Phrases (QP)
  - Number Phrases (NUMP)
- Other constituents
  - Adverb Phrases (ADVP)
  - Prepositional Phrases (PP)
  - Interjection Phrases (INTJP)
  - Participle Phrases (PTP)
  - Foreign phrases (LATIN)
  - *Wh*-phrases (W\*P)

### Preliminaries

The HeliPaD parsing closely follows the principles of the **York-Toronto-Helsinki Parsed Corpus of Old English Prose (YCOE)**, and this manual broadly follows the structure of that one. On occasion, my practice deviates from that of the YCOE and follows that of the **Icelandic Parsed Historical Corpus (IcePaHC)** or the **other Penn historical corpora** instead, or adopts a different solution entirely.

☐ Where the HeliPaD follows other Penn historical corpora, the text will be marked like this.

☐ Where the HeliPaD does its own thing, the text will be marked like this.

There's also [a page with a summary of these differences](#).

Examples used in this manual are from HeliPaD version 0.9. You may also be interested in the principles of [tokenization](#), [POS-tagging](#), or the [list of tags and categories](#).

## General issues

### *Phrase structure*

The annotation scheme used for the Penn historical corpora and for the HeliPaD uses labelled bracketing, which is easily translatable into tree notation. The trees thus derived, however, are substantially flatter than those usually found in GB or Minimalist syntactic theory: multiple branching is permitted, and many phrase types - especially clauses - end up with quite flat structure. This is not intended as a theoretical commitment, but rather is done for ease of annotation and retrieval.

In particular, there is no VP in the Penn historical corpora, or the HeliPaD. The verb and all its objects and adjuncts are immediately dominated by IP and are sisters there. Here's a simple example:

```
( (IP-MAT-SPE (CODE <C>)
  (ADVP-TMP (ADV Thuo-tho))
  (VBDI^3^PL slogun-slahan)
  (ADVP-LOC (ADV thar-thar))
  (ADVP-TMP (ADV eft-eft))
  (NP-SBJ (N^N^PL crud-krud))
  (PP (P an-an)
      (NP (N^A^SG gimang-gimang)))
  (. , - ,))
  (ID OSHeliandC.1291.2409))
```

All phrase types (except IP and CP, [on which more here](#)) have essentially the same structure in the corpus. The key points are the following:

- Intermediate phrase levels such as bar-levels are not used. Phrasal nodes (e.g. NP, ADVP, NUMP) always immediately dominate the head category (e.g. N, ADV, NUM).
- Structure is endocentric: heads project a corresponding phrasal node. This is not exceptionless: some heads, such as verbs, as well as determiners (D) and particles (RP), never project a phrase. Single-word modifiers also do not project.
- Complements, whether single-word or multi-word, always project. This includes possessive genitive phrases, but not possessive pronouns.

☐ In the HeliPaD, single-word modifiers do not project a phrase, even when they follow the head.

### *Heads, modifiers and complements*

Since there are no bar levels, both modifiers and complements are in a sisterhood relation to the **head**, and the phrase immediately dominates all three. Phrases (except IP) are, in general, the same category as their head (e.g. NP dominates N).

In some cases, phrases are not headed in this way. In some cases, this is because the head categories are viewed as subclasses of one another. For instance, pronouns (PRO) and proper nouns (NPR) are seen as a subcategory of N:

```
(NP-SBJ (PRO^N^3^PL sie-he))
```

(OSHeliandC.4.12-15)

In other cases, the absence of a head may be due to elision, which is not represented directly in the corpus as a process. The annotation is neutral as to which of these possibilities is the correct one in any given case: the user is free to judge.

It is in **nominals** where the greatest discrepancy between head label and phrase label can be found. In addition to nouns (N, NPR), the following categories may serve as "heads":

- pronouns (PRO, PRO\$, MAN)
- determiners (D)
- adjectives (ADJ, ADJR, ADJS)
- quantifiers (Q, QR, QS)
- numbers/numerals (NUM)
- inflected participles (VNI, VGI, etc.)

The default treatment of pronouns and determiners is as heading the phrases that immediately dominate them, whereas the other categories listed above are normally treated as modifiers. As such, these other categories may project their own phrase level (e.g. ADJP) when they are themselves modified. When they are not modified, an NP may dominate two potential heads:

```
(NP-SBJ (PRO^N^3^SG it-it)
        (Q^N^SG all-al))
```

(OSHeliandC.1303.2428-2430)

```
(NP-SBJ (D^N^PL thiu-the)
        (NUM^N^PL fiui-fif))
```

(OSHeliandC.12.47)

☐ If an NP immediately dominates only a "modifier" in this sense, the modifier is treated as the head, and the extra level omitted. For instance, in the following example, no QP is present:

```
(NP-POS (ADV so-so)
        (Q^G^PL managaro-manag))
```

(OSHeliandC.974.1862-1865)

☐ ADJPs can be headed only by adjectives, inflected participles, and possessive pronouns, and not by participle phrases or by quantifiers as in the YCOE.

PPs are always headed by prepositions, ADVPs are always headed by adverbs, and QPs are always headed by quantifiers. NUMPs may contain more than one NUM, in which case these together are treated as the head, much as in compounds.

**Modifiers**, like complements, are daughters of the phrasal node and sisters to the head. Modifiers only project a phrase level when they are multi-word and/or themselves modified. The following examples show a construction with multiple adjectives, then one with a multi-word modifier:

```
(NP-SBJ-PRN (ADJ^N^SG mari-mari)
            (ADJ^N^SG mahtig-mahtig)
            (NPR^N^SG Crist-Krist))
```

(OSHeliandC.1384.2574-2577)

```
(NP (N^D^PL uuordon-word)
    (QP (ADV so-so)
        (Q^D^PL filo-filo)))
```

(OSHeliandC.1001.1923-1926)

☐ When modifiers are separated from a head with which they agree, these are traced to the head, as with *alla* below. Unlike in the YCOE, this is the case regardless of whether they are case-marked.

```
( (IP-MAT-SPE (CODE <C>)
  (INTJP (WADV Huat-hwe))
  (, , -, )
  (NP-SBJ (PRO^N^1^PL uui-we)
    (QP *ICH*-1))
  (NP-OB1 (D^A^SG thia-the))
  (ADVP-LOC (ADV hier-her))
  (VBPI^1^PL uuitun-witan)
  (QP-1 (Q^N^PL alla-al))
  (. , -, ))
  (ID OSHeliandC.1426.2654))
```

**Complements**, unlike modifiers, always have a phrase label. This includes single-word complements. Possessor NPs are treated as complements in all instances, but possessive pronouns are usually treated as modifiers.

```
(NP-SBJ (D^N^SG that-the)
  (NP-POS (N^G^PL manno-man))
  (N^N^SG folc-folk))
```

(OSHeliandC.1430.2658)

```
(NP-OB1 (PRO$^A^3^SG is-is)
  (N^A^SG gibodscipi-gibodskepi))
```

(OSHeliandC.1433.2659-2660)

NP arguments (including complements) are distinguished from adjuncts at the IP level, but there is an irreducible element of subjectivity to this task, and the corpus user interested in argument structure should check these distinctions carefully. No attempt has been made to do the same for PPs.

☐ Within other phrases, excluding IPs, PTPs and CPs but including noun phrases, NPs are indicated as possessive (NP-POS) if this is their function (as it usually is within NPs themselves), and unmarked otherwise (as is always the case within PPs). This differs from the approach taken in the YCOE, which labels constituents for case. The default treatment of NPs is to attach them as high as possible in the structure: this means treating constituents as NP-ADT rather than arguments of a non-verbal element.

### *Sentence fragments (FRAG) and restarts*

The exocentric phrasal category FRAG is used for fragmentary clauses or utterances that were written as such in the text, not for those that have been distorted by **textual problems** or omission.

In HeliPaD 0.9, FRAG is only used for restarts, i.e. for a clause that breaks off and then begins again. The incomplete beginning is included within the FRAG as an **IPX-** with no equal-sign index, and the FRAG itself is included within the complete clause.

```
( (IP-MAT (CODE <R_507>)
  (FRAG (IPX-MAT (NP-SBJ (PRO^N^3^SG Siu-siu))
    (MDDI^3^SG muosta-motan)
    (PP (P after-aftar)
      (NP (PRO$^D^3^SG ira-iru)
        (N^D^SG magathedi-magathhed)))
    (, , -, )))
  (CODE <C>)
```



```

(CP-ADV (ADVP-TMP (ADV sithor-sithor))
  (C 0)
  (IP-SUB (NP-SBJ (PRO^N^3^SG siu-siu))
    (NP-PRD (N^G^SG mannes-man))
    (RDDI^3^SG uuarth-werthan)
    (CODE <R_508>)
    (NP-PRD-PRN (N^G^SG erlas-erl))
    (PP (P an-an)
      (NP (N^D^SG ehti-eht)))
    (CODE <C>)
    (NP-SBJ-PRN (ADJ^N^SG edili-ethili)
      (N^N^SG thiorna-thiorna))))
(, ,-,)
(CODE <R_509>)
(ADVP (ADV so-so))
(MDDI^3^SG muosta-motan)
(NP-SBJ (PRO^N^3^SG siu-siu))
(PP (P mid-mid)
  (NP (PRO$^D^3^SG iro-iru)
    (N^D^SG brudigumen-brudigumo)))
(CODE <C>)
(NP-OB1 (N^G^PL bodlu-bodal))
(GE+VB giuualdan-waldan)
(CODE <R_510>)
(NP-TMP (NUM^A^PL sibun-sivun)
  (N^A^PL uuintar-wintar))
(ADVP (ADV saman-saman))
(. .-.))
(ID OSHeliandC.270.507-510))

```

In subordinate clauses, the subordinator is usually repeated.

```

( (IP-MAT-SPE (CODE <C>)
  (CONJ endi-endi)
  (NP-SBJ-X *exp*)
  (NP-OB2 (PRO^D^1^PL us-we))
  (BEPI^3^SG is-wesan)
  (NP-ADT (N^D^PL firinon-firina))
  (NP-PRD (N^N^SG tharf-tharf))
  (CODE <R_2429>)
  (CP-THT-SPE-X (FRAG (CP-THT-SPE (C that-that)
    (IPX-SUB-SPE (NP-SBJ
      (PRO^N^1^PL uui-we))
      ...
      (, ,-,)
      (CODE <R_2430>)
      (C that-that)
      (IP-SUB-SPE (NP-SBJ (PRO^N^1^PL uui-we))
        (NP-OB1 (PRO^A^3^SG it-it))
        (PP (P an-an)
          (NP (D^D^SG theson-these)
            (N^D^SG lande-land))))
        (PP (P at-at)
          (NP (PRO^D^2^SG thi-thu))))
        (CODE <C>)
        (VB linon-linon)
        (MDPS^1^PL muotin-motan))))
      (. .-.))
      (' ' -'))
  (ID OSHeliandC.1303.2428-2430))

```

### *Text problems (X)*

The exocentric category X is used for ungrammatical structures that arise due to text problems (typically manuscript problems), usually alongside a comment that indicates what the problem is.

```
( (X (CODE <C>)
  (CODE <COM:OMISSION>)
  (NP-SBJ-X *exp*)
  (ADJP-PRD (ADJ^A^SG cuth-kuth))
  (CODE <R_5891>)
  (NP-OB2 (D^D^PL them-the)
    (N^D^PL liudon-liudi))
  (PP (P after-aftar)
    (NP (D^D^SG them-the)
      (N^D^SG lande-land)))
  (CODE <C>)
  (CP-THT-X (C that-that)
    (IP-SUB (NP-SBJ (PRO^N^3^PL sia-he)
      (NP-OB1 (ADJ^A^PL sulica-sulik)
        (N^A^PL lugina-lugina))
      (MDDI^3^PL uuoldun-willian)
      (CODE <R_5892>)
      (VB ahebbian-ahebbian)
      (PP (P be-bi)
        (NP (D^A^SG than-the)
          (ADJ^A^SG helagan-helag)
          (N^A^SG drohtin-drohtin))))))
    (. .-.))
  (ID OSHeliandC.3489.5890-5892))
```

### *Direct speech (-SPE and QTP)*

Direct speech is marked in the corpus by adding the label extension -SPE to any IPs and CPs that are spoken. Any IP or CP dominated by an IP or CP labelled -SPE is also labelled -SPE.

```
( (IP-MAT-SPE (CODE <R_118>)
  (NP-SBJ (PRO$^N^2^SG thin-thin)
    (N^N^SG theonost-thionost))
  (BEPI^3^SG is-wesan)
  (NP-OB2 (PRO^D^3^SG im-he))
  (CODE <MS_8a>)
  (PP (P an-an)
    (NP (N^D^SG thanke-thank)))
  (, ,-,)
  (CODE <C>)
  (CP-THT-SPE (C that-that)
    (IP-SUB-SPE (NP-SBJ (PRO^N^2^SG thu-thu)
      (NP-OB1 (ADJ^A^SG sulica-sulik)
        (N^A^SG githaht-githaht))
      (HVPI^2^SG habes-hebbian)
      (CODE <R_119>)
      (PP (P an-an)
        (NP (NP-POS (PRO^G^3^SG is-he)
          (NUM^G^SG enes-en))
          (N^A^SG craft-kraft))))))
    (. .-.))
  (ID OSHeliandC.48.118-119))
```

Non-clausal sequences of direct speech are labelled as QTP, an exocentric phrasal category.

```

( (IP-MAT (CODE <C>)
  (CONJ endi-endi)
  (NP-SBJ *con*)
  (NP-OB2 (D^D^SG them-the)
    (N^D^SG alouualden-alowaldo)
    (CODE <R_2843>)
    (ADJ^D^SG selbem-self))
  (VBDI^3^SG sagda-seggian)
  (, , -, )
  (CODE <C>)
  (CP-THT (C that-that)
    (IP-SUB (NP-SBJ (PRO^N^3^PL sia-he))
      (PP (P an-an)
        (NP (PRO$^D^3^SG iro-iru)
          (N^D^SG gisithie-gisithi)))
      (NP-OB1 (ADV than-than)
        (QR^A^SG mer-mer)
        (CODE <R_2844>)
        (NP-POS (ADJ^G^SG garoes-garu)))
      (NEG ni-ne)
      (HVDS^3^PL habdin-hebbian)
      (CODE <C>)
      (' ' - '))
    (QTP (PP (P nouan-newan)
      (NP (NP (ADJ^A^PL gerstin-girstin)
        (N^A^PL bruod-brod)
        (CODE <R_2845>)
        (NUM^A^PL fiui-fif))
      (CONJP *ICH*-1)))
      (PP (P an-an)
        (NP (PRO$^D^1^SG usero-usa)
          (N^D^SG ferde-fard)))
      (CODE <C>)
      (CONJP-1 (CONJ endi-endi)
        (NP (N^A^PL fiscos-fisk)
          (NUM^A^PL tuena-twene))))))
  (. .-.))
(ID OSHeliandC.1544.2842-2845))

```

For the *Heliand*, the fact that it was originally an oral epic does not mean that all clauses are annotated as -SPE! Only those which are spoken by characters in the story are marked with -SPE or QTP.

☐ In the YCOE, the first independent clause following a verb of saying is included in the parse as the complement of the verb of saying, whereas later independent clauses are treated as separate tokens. In the HeliPaD, these clauses are always treated as independent tokens.

### ***Left-dislocation (-LFD) and resumption (-RSP)***

Constituents in the left periphery of a clause that are later resumed by a coreferential phrase are given the extended label -LFD.

☐ In the HeliPaD, unlike in the YCOE, the label -LFD is used systematically with all clausal categories, including CP-ADV (e.g. in *if ... then* constructions).

```

( (IP-MAT-SPE (CODE <C>)
  (CP-ADV-LFD-SPE (ADVP (ADV thoh-thoh))
    (C 0)
    (IP-SUB-SPE (NP-SBJ-1 (PRO^N^3^SG hie-he))
      (ADJP (ADV so-so)
        (ADJ^N^SG othi-othi)

```

```

(IP-INF-SPE *ICH*-1))
(NEG ni-ne)
(BEPS^3^SG si-wesan)
(CODE <R_1787>)
(NP-OB2 (N^D^PL firihon-firihos))
(IP-INF-SPE-1 (NP-OB1 *-1)
              (TO te-te)
              (VB faranne-
faran))))
(, , -, )
(CODE <C>)
(ADVP-RSP (ADV thoh-thoh))
(MDPI^3^SG scal-skulan)
(NP-SBJ (PRO^N^3^SG hie-he))
(PP (P te-te)
    (NP (N^D^SG frumu-fruma)))
(RD uerthan-werthan)
(CODE <R_1788>)
(CP-FRL-OB2-SPE (WNP-SBJ-2 (ADV so-so)
                          (WPRO^N hue-hwe))
                (C so-so)
                (IP-SUB-SPE (NP-SBJ *T*-2)
                          (NP-OB1 (PRO^A^3^SG ina-he))
                          (VBPI^3^SG thurugengit-
thurhgangan)))
(, , -, )
(ID OSHeliandC.940.1786-1788))

```

Resumptive elements (-RSP) are used either to match -LFD elements, as in the example above, or on a resumptive pronoun in an unbounded dependency context where a trace would be expected, as in the example below.

```

(NP-PRD (D^N^SG that-the)
        (N^N^SG barn-barn)
        (NP-POS (NPR^G^SG godes-god)))
(CODE <R_912>)
(' ' -')
(CP-REL-SPE (WNP-SBJ-1 (D^N^SG that-that))
            (C 0)
            (IP-SUB-SPE (ADVP-LOC (ADV hier-her))
                      (ADVP-TMP (ADV lango-lango)
                                (ADV iu-ju))
                      (' ' -')
                      (IP-MAT-PRN (VBDI^3^PL quathun-kwethan)
                                (NP-SBJ (PRO^N^3^PL sea-he))))
            (, , -, )
            (CODE <C>)
            (' ' -')
            (NP-SBJ (N^N^PL liudi-liudi))
            (VBDI^3^PL sagdun-seggian)
            (CODE <R_913>)
            (NP-SBJ-PRN (N^N^PL ueros-wer))
            (ADVP (ADV uuarlico-warliko))
            (, , -, )
            (CODE <C>)
            (CP-THT-SPE (C that-that)
                      (IP-SUB-SPE (NP-SBJ-RSP-1 (PRO^N^3^SG
hie-he))
                                (MDDS^3^SG scoldi-skulan)
                                ...
                                (VB cuman-kuman))))))

```

(OSHeliandC.489.911-913)

### *Appositives and parentheticals (-PRN)*

Appositive or parenthetical constituents are given the extended label -PRN, which may occur with almost any phrase type. Appositive nominals are particularly common in the *Heliand*.

▣ In the YCOE, nominal appositive constituents are either contained within, or indexed to, the constituent to which they are in apposition. In the HeliPaD, they are instead treated as sisters to that constituent (or indexed to a sisterhood position). This enables a much less cluttered clausal representation, and the intended apposition relations are usually very clear semantically.

```
( (IP-MAT-SPE (CODE <C>)
  (VBPI^3^SG farit-faran)
  (NP-OB1 (PRO^A^3^SG ina-he))
  (NP-SBJ (NP-POS (N^G^SG uuerodes-werod))
    (ADJ^N^SG liut-lut))
  ( , , -, )
  (CODE <R_1783>)
  (NP-SBJ-PRN (ADJ^N^SG faho-faho)
    (N^N^SG folcscepi-folkskepi))
  (. :-:))
(ID OSHeliandC.937.1782-1783))
```

Clauses that are in apposition are contained within or indexed to the constituent to which they are in apposition.

```
( (IP-MAT-SPE (CODE <C>)
  (PP (P For-for)
    (CODE <MS_53b>)
    (NP (D^I^SG thiu-the)
      (CP-THT-PRN-SPE *ICH*-1)))
  (NP-SBJ (PRO^N^2^PL gi-gi))
  (VB sorgon-sorgon)
  (MDPI^2^PL sculun-skulan)
  (CODE <R_1881>)
  (CP-THT-PRN-SPE-1 (C that-that)
    (IP-SUB-SPE (NP-OB2 (PRO^D^2^PL iu-gi))
      (NP-SBJ (D^N^PL thia-the)
        (N^N^PL man-man))
      (NEG ni-ne)
      (MDPS^3^PL mugin-mugan)
      (CODE <C>)
      (NP-OB1 (N^A^PL muodgithahti-
modgithaht))
      ( , , -, )
      (CODE <R_1882>)
      (NP-OB1-PRN (N^A^SG uuilleon-
willio))
      (VB auuendan-awendan)))
  (. .-.))
(ID OSHeliandC.980.1880-1882))
```

Parenthetical clauses are also contained within the clause to which they are parenthetical. In most cases, these involve the word *quathie*:

```
( (CP-QUE-SPE (CODE <R_2025>)
  ( ' ' - ' )
  (WNP-SBJ-1 (WPRO^N Huat-hwe))
  (IP-SUB-SPE (NP-SBJ *T*-1)
```

```

(BEPI^3^SG ist-wesan)
(NP-OB2 (NP-OB2 (PRO^D^1^SG mi-ik))
        (CONJP (CONJ endi-endi)
                (NP-OB2 (PRO^D^2^SG thi-thu))))
(' '-')
(IP-MAT-PRN (VBDI^3^SG quat$-kwethan)
            (NP-SBJ (PRO^N^3^SG $hie-he)))
(, , -, )
(CODE <C>)
(' '-')
(PP (P umbi-umbi)
    (NP (NP-POS (D^G^PL thesaro-these)
                (N^G^PL manno-man))
        (N^A^SG lith-lith)))
(, , -, )
(CODE <R_2026>)
(PP (P umbi-umbi)
    (NP (NP-POS (D^G^SG theses-these)
                (N^G^SG uuerodes-werod))
        (N^A^SG uuin-win))))
(. ?-?)
(ID OSHeliandC.1047.2025-2026))

```

-PRN IPs are also used in the annotation of **right node raising**.

### *Empty categories*

There are seven types of empty category in the HeliPaD:

- \*con\*: subject elided under co-ordination
- \*exp\*: null expletive subject (may be co-referential with a clausal element)
- \*arb\*: arbitrary subject in infinitivals
- \*pro\*: null subject, none of the above (usually referential)
- \*T\*: trace of *wh*-movement, always coindexed
- \*ICH\*: trace of other movement, always coindexed
- \*: other empty category (e.g. in *tough*-movement)

In addition, empty complementizers and *wh*-operators are given a notional phonological content of *o*.

Since all complete finite clauses and small clauses are required to have a subject, an **empty subject** is added if there is no overt one. Non-finite clauses (IP-INF) are not required to have a subject in the annotation, and in cases of raising-to-subject or subject control they will not have one. For expletive subjects, see **Expletive constructions**.

In cases of conjunction reduction, the elided subject is \*con\* in all clause types. This category can only be used when the subject is coreferential and identical in number; otherwise, \*pro\* is used. A conjunction itself is not necessary for \*con\* to be used: see below for examples both with and without conjunction.

```

( (IP-MAT (CODE <C>)
      (CONJ endi-endi)
      (NP-SBJ (PRO^N^3^SG hie-he)
              (ADJP *ICH*-1))
      (ADVP-LOC (ADV thar-thar))
      (PP (P mid-mid)
          (NP (PRO$^D^3^PL is-is)
              (N^D^PL fingron-fingar))))
      (ADVP-TMP (ADV thuo-tho))
      (CODE <R_2042>)
      (VBDI^3^SG segnoda-segnon)

```

```

(ADJP-1 (ADJ^N^SG selbo-self))
(CODE <C>)
(NP-OB2 (PRO$^D^3^PL sinon-sin)
        (N^D^PL handon-hand))
(. ,-,))
(ID OSHeliandC.1055.2041-2042))

( (IP-MAT (CODE <R_2043>)
        (NP-SBJ *con*)
        (VBDI^3^SG uuarahta-wirkian)
        (NP-OB1 (PRO^A^3^SG it-it))
        (PP (P te-te)
            (NP (N^D^SG uuine-win))))
  (. ,-,))
(ID OSHeliandC.1056.2043))

( (IP-MAT (CODE <C>)
        (CONJ endi-endi)
        (NP-SBJ *con*)
        (VBDI^3^SG hiet-hetan)
        (IP-INF (IP-INF (NP-SBJ *arb*)
                        (NP-OB1 (PRO^G^3^SG is-it))
                        (PP (P an-an)
                            (NP (NUM^A^SG enn-en)
                                (N^A^SG uuegi-wegi))))
                        (VB hladan-hladan))
        (, ,-,))
  (CODE <R_2044>)
  (CONJP (IP-INF (NP-SBJ *arb*)
                (VB scepian-skeppian)
                (PP (P mid-mid)
                    (NP (NUM^D^SG enaro-en)
                        (N^D^SG scalun-skala))))))
  (. ,-,))
(ID OSHeliandC.1057.2043-2044))

```

▣ Arbitrary PRO in ECM infinitives is indicated by \*arb\*, as it is in the PPCME2 and IcePaHC (but not the YCOE); see the above example.

Where \*con\*, \*exp\* and \*arb\* are all impossible, \*pro\* is used. This is not a commitment to a genuine *pro*-drop analysis; rather, \*pro\* is a dustbin category for empty subjects, some of which may represent genuine referential null subjects, others scribal error. Cases of conjunction reduction with number mismatch are also treated as \*pro\*. All empty subjects, whether \*pro\* or not, appear as early in the clause as possible after such elements as *wh*-traces, conjunctions, vocatives, interjections, and left-dislocated items.

```

( (IP-MAT (CODE <C>)
        (NEG Ne-ne)
        (BEDI^3^SG uuas-wesan)
        (ADVP-TMP (ADV io-eo))
        (NP-POS-1 (NPR^G^PL Iudeo-Judeo))
        (ADVP (ADV bethiu-bethiu))
        (CODE <R_2361>)
        (NP-SBJ (NP-POS *ICH*-1)
                (NP-POS-PRN (N^G^SG lethes-leth)
                            (N^G^SG liudscipes-liudskepi))
                (CODE <C>)
                (N^N^SG giloĥo-gilovo))
        (ADJP-PRD (D^I^SG thiu-the)
                 (ADJR^N^SG bettera-bet))
        (CODE <R_2362>)
        (PP (P an-an)

```

```

      (NP (D^A^SG thena-the)
          (ADJ^A^SG helagon-helag)
          (NPR^A^SG Crist-Krist)))
    (. ,-,))
  (ID OSHeliandC.1258.2360-2362))

( (IP-MAT (CODE <C>)
  (CONJ ac-ak)
  (NP-SBJ *pro*)
  (HVDI^3^PL habdun-hebbian)
  (NP-OB2 (PRO^D^3^SG im-he))
  (NP-OB1 (ADJ^A^SG hardon-hard)
           (N^A^SG muod-mod))
  (, ,-,)
  (CODE <R_2363>)
  (NP-OB1-PRN (ADJP (ADV suitho-switho)
                   (ADJ^A^SG starcan-stark))
              (N^A^SG strid-strid))
  (. ,-,))
  (ID OSHeliandC.1259.2362-2363))

```

\*pro\* is also used in examples where a subjunctive is used with imperative force.

```

( (IP-MAT (CODE <R_3912>)
  (CP-FRL-LFD (WNP-SBJ-1 (ADV so-so)
                        (WPRO^N hue-hwe))
              (C so-so)
              (IP-SUB (NP-SBJ *T*-1)
                      (ADVP-LOC (ADV thar-thar))
                      (PP (P mid-mid)
                          (NP (N^I^SG thurstu-thurst))))
                    (CODE <C>)
                    (VN bithuungan-bithwingan)
                    (BEDS^3^SG uuari-wesan)))
  (, ,-,)
  (CODE <R_3913>)
  (' '-')
  (QTP (NP-SBJ-RSP *pro*)
       (ADVP (ADV so-so))
       (VBPS^3^SG gange-gangan)
       (NP-OB2-RFL (PRO^D^3^SG im-he))
       (ADVP-DIR (ADV herod-herod))
       (IP-INF-SPE (VB drincan-drinkan)
                  (PP (P te-te)
                      (NP (PRO^D^1^SG mi-ik)))
                  (NP-OB1 *ICH*-2))
       (' '-')
       (IP-MAT-PRN (VBDI^3^SG quat$-kwethan)
                  (NP-SBJ (PRO^N^3^SG $hie-he)))
       (, ,-,)
       (CODE <C>)
       (' '-')
       (NP-TMP (NP-POS (N^G^PL dago-dag)
                    (GE+WADJ^G^SG gihuilikes-gihwilik))
              (CODE <R_3914>)
              (NP-OB1-2 (ADJ^A^PL suotian-swoti)
                       (N^A^PL brunnon-brunno)))
       (. .-.))
  (ID OSHeliandC.2262.3912-3914))

```



Most complete CPs must have a **complementizer** (C), which is represented as phonologically *o* when not overtly present. Direct questions, yes/no-questions, and V1 conditionals do not require a C.

```
( (IP-MAT (CODE <C>)
  (ADVP (ADVR Uuirss-wirs))
  (BEPI^3^SG is-wesan)
  (NP-OB2 (D^D^PL them-the)
    (ADJP (ADJ^D^PL odron-othar)))
  (, , -, )
  (CODE <R_1348>)
  (ADJP-PRD (ADJ^N^SG gibidig-gividig))
  (NP-SBJ (ADJR^N^SG grimmera-grim)
    (N^N^SG thing-thing))
  (CODE <MS_39b>)
  (CODE <C>)
  (NP-OB2-PRN (D^D^PL them-the)
    (CP-REL-SPE (WNP-SBJ-1 (D^N^PL thia-the)
      (C 0)
      (IP-SUB-SPE (NP-SBJ *T*-1)
        (ADVP-LOC (ADV hier-her))
        (NP-OB1 (N^A^SG guod-god))
        (VBPI^3^PL egun-egan)
        (, , -, )
        (CODE <R_1349>)
        (NP-OB1-PRN (ADJ^A^SG uuidon-
          wid)
            (N^A^SG
            uuerolduuelon-weroldwelo))))))
    (. :-:))
  (ID OSHeliandC.737.1347-1349))

( (CP-QUE-SPE (CODE <R_923>)
  (IP-SUB-SPE (BEPI^2^SG Bist-wesan)
    (NP-SBJ (PRO^N^2^SG thu-thu))
    (NP-PRD (Q^N^SG enig-enig)
      (NP-POS (D^G^PL thero-the)
        (CODE <C>)
        (CP-REL-SPE (WNP-SBJ-1 (D^N^SG
          thi-the))
            (C 0)
            (IP-SUB-SPE (NP-SBJ
              *T*-1)
                (ADVP-LOC
                (ADV hier-her))
                (ADVP-TMP
                (ADV er-er))
                (BEDS^3^SG uuari-wesan))))))
    (CODE <R_924>)
    (NP-POS-PRN (ADJ^G^PL uuissaro-wis)
      (N^G^PL uuarsagono-
      warsago))))
  (. ?-?))
  (ID OSHeliandC.498.923-924))
```

☐ In CP-CMP, CP-REL, CP-FRL and *wh*-CP-QUE, ***wh*-operators** must be present, and are represented as phonologically *o* when not overtly present. The clause from which the *wh*-operator is extracted must contain a trace of the same category, or else a resumptive. Unlike in the YCOE, this trace is marked with all the same extended labels as the operator itself. The extended labels of *wh*-operators are identical in every way to those of overt constituents.

```

( (IP-MAT-SPE (CODE <C>)
  (VBI^2^SG Sagi-seggian)
  (NP-OB2 (PRO^D^1^PL us-we))
  (CP-QUE-SPE (WNP-PRD-1 (WPRO^N huat-hwe)
    (NP-POS *ICH*-2))
    (C 0)
    (IP-SUB-SPE (NP-PRD *T*-1)
      (NP-SBJ (PRO^N^2^SG thu-thu))
      (NP-POS-2 (N^G^PL manno-man))
      (BEPS^2^SG sis-wesan)))
    (. !-!))
  (ID OSHeliandC.497.922))

( (IP-MAT (CODE <C>)
  (NP-SBJ *con*)
  (VBDI^3^SG quat-kwethan)
  (CP-THT (C that-that)
    (IP-SUB (NP-SBJ (D^N^PL thia-the)
      (CP-REL *ICH*-1))
      (ADJP-PRD (ADJ^N^PL saliga-salig))
      (BEDS^3^PL uuarin-wesan)
      (CODE <R_1301>)
      (NP-SBJ-PRN (N^N^PL mann-man))
      (PP (P an-an)
        (NP (D^D^SG thesaro-these)
          (N^D^SG middilgardun-middilgarda)))
      (CODE <C>)
      (CP-REL-1 (WNP-SBJ-2 (D^N^PL thia-the))
        (C 0)
        (IP-SUB (NP-SBJ *T*-2)
          (ADVP-LOC (ADV her-her))
          (PP (P an-an)
            (NP (PRO$^D^3^SG iro-iru)
              (N^D^SG muode-mod)))
            (BEDS^3^PL uuarin-wesan)
            (CODE <R_1302>)
            (ADJP-PRD (ADJ^N^PL arma-arm))
            (PP (P thuru-thurh)
              (NP (N^A^SG odmuodig-
othmodi))))))))))
  (. :-:))
  (ID OSHeliandC.712.1300-1302))

```

**Scrambling and extraposition** are indicated by \*ICH\* rather than \*T\*, and are only represented if the movement in question has taken the moved constituent out of its source category (that is, structurally vacuous "linear" scrambling and extraposition do not exist in this annotation scheme). Traces are always as early as possible in the source constituent if the movement is upward/to the left, and as late as possible if the movement is downward/to the right. With non-finite clauses, elements are traced to the position of the non-finite verb, which is taken as the position of the clause.

```

( (IP-MAT (CODE <C>)
  (ADVP-TMP (ADV Than-than))
  (MDPI^2^PL motun-motan)
  (NP-SBJ (PRO^N^2^PL gi-gi))
  (NP-OB1 (D^A^SG thia-the)
    (N^A^SG fruma-fruma)
    (CP-THT-SPE *ICH*-1))
  (VB egan-egan)
  (CODE <R_1461>)
  (CP-THT-SPE-1 (C that-that)
    (IP-SUB-SPE (NP-SBJ (PRO^N^2^PL gi-gi))

```

```

(MDPI^2^PL muotun-motan)
(VB hetan-hetan)
(CODE <C>)
(NP-PRD (NP-POS (N^G^SG hebancuninges-
hevankuning))
(N^N^PL suni-sunu))
(, ,-,)
(CODE <R_1462>)
(NP-PRD-PRN (PRO$^N^3^PL is-is)
(ADJ^N^PL blithi-blithi)
(N^N^PL barn-barn))))
(. .-.))
(ID OSHeliandC.787.1460-1462))

```

An index number on the scrambled/extrapolated constituent indicates that it does not belong to (i.e. originate in) the phrase that immediately dominates it.

```

( (IP-MAT-SPE (CODE <C>)
(, ---)
(NP-SBJ (PRO^N^2^PL gi-gi))
(VBDI^2^PL hordun-horian)
(NP-OB1-1 (PRO^A^3^SG it-it))
(ADVP-TMP (ADV oft-oft))
(IP-INF-SPE (NP-OB1 *ICH*-1)
(VB sprekean-sprekan)
(CODE <P_102>)
(CODE <R_1433>)
(NP-ADT (N^D^PL uuordun-word))
(NP-SBJ (ADJ^A^PL uuisa-wis)
(N^A^PL man-man)))
(, ---)
(. :-:))
(ID OSHeliandC.776.1432-1433))

```

☞ **Raising to subject** is not usually explicitly represented in the HeliPaD.

The exception is **tough-movement**, which is represented as in the following example, using a default empty category:

```

(CP-ADV-LFD-SPE (ADVP (ADV thoh-thoh))
(C 0)
(IP-SUB-SPE (NP-SBJ-1 (PRO^N^3^SG hie-he))
(ADJP (ADV so-so)
(ADJ^N^SG othi-othi)
(IP-INF-SPE *ICH*-1))
(NEG ni-ne)
(BEPS^3^SG si-wesan)
(CODE <R_1787>)
(NP-OB2 (N^D^PL firihon-firihos))
(IP-INF-SPE-1 (NP-OB1 *-1)
(TO te-te)
(VB faranne-faran))))
(OSHeliandC.940.1786-1788)

```

### *Expletive constructions*

Among overt elements, only the pronoun *it* is treated as an expletive; similar constructions involving *that* are treated as instances of **apposition**. Expletive coindexing involves a final -X, as in the YCOE (though it is upper case, for technical reasons). Extrapolated clausal subjects (*that*-clauses, infinitives

or questions) are coindexed to an overt (*it*) or null (\*exp\*) expletive. These extraposed clausal subjects occur in various constructions.

```
( (IP-MAT (CODE <C>)
  (NP-SBJ-X *exp*)
  (NP-OB2 (PRO^D^3^SG Im-he))
  (BEDI^3^SG uuas-wesan)
  (ADVP-TMP (ADV thuo-tho))
  (NP-PRD (N^N^SG uuilleo-willio)
    (ADJ^N^SG mikil-mikil))
  (CODE <R_872>)
  (CP-THT-X (C that-that)
    (IP-SUB (NP-SBJ (PRO^N^3^SG hie-he))
      (PP (P fan-fan)
        (NP (ADJ^D^PL solicon-sulik)
          (N^D^PL saldon-salitha)))
      (CODE <C>)
      (VB seggean-seggian)
      (MDDS^3^SG muosti-motan)))
    (. .-.))
  (ID OSHeliandC.470.871-872))

( (IP-MAT (CODE <C>)
  (NP-SBJ-X *exp*)
  (BEDI^3^SG uuas-wesan)
  (NP-OB2 (PRO^D^3^PL im-he))
  (NP-PRD (NP-POS (PRO$^G^3^SG is-is)
    (N^G^SG huldi-huldi))
    (NP-POS-PRN (N^G^PL helpono-helpa))
    (N^N^SG tharf-tharf))
  (CODE <R_1188>)
  (IP-INF-X (TO te-te)
    (GE+VB githiononne-thionon))
  (. :-:))
  (ID OSHeliandC.650.1187-1188))

( (IP-MAT (CODE <C>)
  (NP-SBJ-X *exp*)
  (BEDI^3^SG uuas-wesan)
  (NP-OB2 (PRO^D^3^PL im-he))
  (NP-PRD (N^N^SG niud-niud)
    (ADJ^N^SG mikil-mikil))
  (CODE <R_183>)
  (CP-QUE-X (WNP-OB1-1 (WPRO^A huat-hwe)
    (NP-POS *ICH*-2))
    (C 0)
    (IP-SUB (IP-SUB-0 (NP-OB1 *T*-1)
      (NP-SBJ (PRO^N^3^SG he-he))
      (NP-OB2 (PRO^D^3^PL im-he))
      (NP-POS-2 (ADJ^G^SG suodlicas-
sothlik))
      (CODE <C>)
      (VB seggean-seggian)
      (MDDS^3^SG uueldi-willian))
      (, ,-,)
      (CODE <R_184>)
      (CONJP (IPX-SUB=0 (VB uuisan-wisian)
        (PP (P te-te)
          (NP (N^D^PL uuarun-
war))))))
      (. .-.))
```

(ID OSHeliandC.86.182-184))

Other expletive subjects occur in impersonal constructions, and are not coindexed.

```
( (IP-MAT (CODE <C>)
  (NP-SBJ *exp*)
  (ADVP-TMP (ADV Thuo-tho))
  (RDDI^3^SG uuarth-werthan)
  (NP-OB2 (Q^D^PL managon-manag))
  (ADJP-PRD (ADJ^N^SG cuth-kuth))
  (CODE <R_387>)
  (PP (P ober-ovar)
    (NP (D^A^SG thesa-these)
      (ADJ^A^SG uuidun-wid)
      (N^A^SG uuerold-werold)))
  (. , -,))
(ID OSHeliandC.213.386-387))

( (IP-MAT-SPE (CODE <R_418>)
  (' ' -')
  (NP-SBJ *exp*)
  (NP-PRD (NP-PRD (N^N^SG Diuritha-diuritha))
    (CONJP *ICH*-1))
  (BEPS^3^SG si-wesan)
  (ADVP-TMP (ADV nu-nu))
  (' ' -')
  (IP-MAT-PRN (VBDI^3^PL quathun-kwethan)
    (NP-SBJ (PRO^N^3^PL sia-he)))
  (, , -,)
  (CODE <C>)
  (' ' -')
  (NP-OB2 (N^D^SG drohtine-drohtin)
    (ADJ^D^SG selbon-self))
  (CODE <R_419>)
  (PP (P an-an)
    (NP (D^D^SG thiem-the)
      (ADJS^D^SG hohoston-hoh)
      (CODE <C>)
      (NP-POS (N^G^PL himilo-himil))
      (N^D^SG rikea-riki)))
  (, , -,)
  (CODE <R_420>)
  (CONJP-1 (CONJ endi-endi)
    (NP-PRD (N^N^SG fridu-frithu)))
  (PP (P an-an)
    (NP (N^D^SG erdu-ertha)))
  (CODE <C>)
  (NP-OB2-PRN (NP-POS (N^G^PL firio-firihos))
    (N^D^PL barnon-barn))
  (, , -,)
  (CODE <R_421>)
  (NP-OB2-PRN (ADJ^D^PL guoduuilligon-godwillig)
    (N^D^PL gumon-gumo)
  (, , -,)
  (CODE <C>)
  (CP-REL (WNP-SBJ-2 (D^N^PL thie-the)
    (C thia-the)
    (IP-SUB (NP-SBJ *T*-2)
      (NP-OB1 (NPR^A^SG god-god))
      (VBPI^3^PL ankenneat-andkennian)
      (CODE <R_422>))
```

```

hluttar)
        (PP (P thuru-thurh)
            (NP (ADJ^A^SG hluttran-
                (N^A^SG hugi-hugi))))))
        (. .-..)
        (' '-'))
(ID OSHeliandC.229.418-422))

```

### Conjunction

Conjunction of everything other than matrix clauses with coreferential subjects is represented using CONJP, a phrasal category, usually endocentric with CONJ as its head. The basic structure is as follows:

```

(XP1 (XP2 first_conjunct)
     (CONJP (CONJ conjunction)
            (XP3 second_conjunct)))

```

CONJP is sister to the first conjunct, with the second conjunct acting as complement within the CONJP (following the CONJ itself, if there is one).

```

(PP (PP (P mid-mid)
        (NP (N^D^PL uuordun-word)))
    (CONJP (CONJ endi-endi)
           (PP (P mid-mid)
               (NP (N^D^PL uuercun-werk)))))

```

(OSHelianC.1.1-5)

Third and subsequent conjuncts are also within CONJPs and sisters to the first conjunct.

```

(PP (PP (P mid-mid)
        (NP (ADJ^I^SG leohtu-lioht)
            (N^I^SG hugi-hugi)))
    (CODE <C>)
    (CONJP (CONJ endi-endi)
           (PP (P mid-mid)
               (NP (N^D^SG gilobon-gilovo)
                   (ADJ^D^SG guodon-god))))
    (CODE <P_24>)
    (CODE <R_291>)
    (CONJP (CONJ endi-endi)
           (PP (P mid-mid)
               (NP (ADJ^D^PL hluttron-hluttar)
                   (N^D^PL treuon-treuwa)))))

```

(OSHelianC.160.288-291)

In cases of "balanced" conjunction, i.e. where a conjunction precedes the first conjunct, it is included under the root node.

```

(NP-OB1-PRN (CONJ ge-ge)
            (NP-OB1-PRN (N^A^SG hosc-hosk))
            (CONJP (CONJ gi-ge)
                  (NP-OB1-PRN (N^A^SG harmquidi-harmkwidi))))

```

(OSHelianC.987.1894-1896)

▣ The HeliPaD's approach to conjunction differs in three important ways from that of the YCOE and other Penn corpora. First, single-word conjuncts are treated in exactly the same way as other conjuncts. Secondly,

any extended labels borne by the root node are inherited by the two conjuncts. Both these points are illustrated by the above example. Thirdly, shared pre- and post-head modifiers are simply included under the root node, as illustrated by the below:

```
(NP-OB1 (NP-POS (NP-POS (Q^G^PL allaro-al)
                        (N^G^PL manno-man))
          (GE+WPRO^G gihues-gihwe))
  (CODE <C>)
  (NP-OB1 (N^A^PL muodgithahti-modgithaht))
  (, , -, )
  (CODE <R_1926>)
  (CONJP (NP-OB1 (N^A^SG uuord-word)))
  (CONJP (CONJ endi-endi)
        (NP-OB1 (N^A^SG uuilleon-willio))))
```

(OSHeliandC.1001.1923-1926)

```
(NP-OB1-PRN (NP-OB1-PRN (N^A^PL scenkion-skenkio))
  (CONJP (CONJ endi-endi)
        (NP-OB1-PRN (N^A^PL scapuuardos-skapward)))
  (, , -, )
  (CODE <C>)
  (CP-REL (WNP-SBJ-1 (D^N^PL thia-the))
    (C 0)
    (IP-SUB (NP-SBJ *T*-1)
      (ADVP-LOC (ADV thar-thar))
      (MDDI^3^PL scoldun-skulan)
      (NP-OB2 (D^D^SG thera-the)
        (N^D^SG scola-skola))
      (CODE <MS_57b>)
      (VB thionon-thionon))))
```

(OSHeliandC.1051.2032-2036)

When unlike constituents are conjoined, the dominating phrase will be the same as the first conjunct.

```
(QP-2 (QP (Q^A^SG manag-manag))
  (CONJP (CONJ endi-endi)
        (ADJP (ADJ^A^SG mislic-mislik))))
```

(OSHeliandC.985.1889-1892)

Finally, focus-sensitive adverbs and ADVPs such as *ok* (tagged ALSO), *so self* and *so sama* may be treated as part of a CONJP, whether preceding or following the conjunct.

```
(CONJP (CONJ endi-endi)
  (ALSO oc-ok)
  (NP-OB1 (D^A^SG that-the)
    (N^A^SG barn-barn)
    (NP-POS (NPR^G^SG godes-god))))
```

(OSHeliandC.257.478-480)

```
(PP (PP (P umbi-umbi)
  (NP (NP-POS (D^G^PL thesaro-these)
    (N^G^PL liudio-liudi))
    (N^A^SG lif-lif)))
  (CODE <C>)
  (CONJP (CONJ endi-endi)
```

```
(ALSO oc-ok)
(PP (P umbi-umbi)
  (NP (D^A^SG thit-these)
    (N^A^SG land-land)))
(ADVP (ADV so-so)
  (ADV samo-sama))))))
```

(OSHeliandC.2163.3755-3757)

### *Elision*

Noun phrase elision is never indicated in the HeliPaD, but clauses may be complete (IP-, CP-) or incomplete (IPX-, CPX-). Elision is the usual source of such incomplete clauses; **textual problems** are treated differently, using the X phrase. Incomplete clauses arising from elision have equal-sign coindexing: the incomplete clause will bear =0 (or =00, =000 etc.), while the clause it is patterned on will be marked -0 (or -00, -000 etc.).

```
( (IP-MAT-SPE=0 (CODE <C>)
  (CONJ endi-endi)
  (VBPI^3^SG gangat-gangan)
  (NP-OB2 (PRO^D^3^SG im-he))
  (NP-SBJ (N^N^SG diuball-diuval))
  (ADVP-DIR (ADV ferr-fer))
  (, , -, )
  (CODE <R_2481>)
  (NP-SBJ-PRN (ADJ^N^PL uuretha-wreth)
    (CODE <MS_69b>)
    (N^N^PL uuihti-wiht))
  (CODE <C>)
  (IPX-MAT-PRN-SPE=0 (CONJ endi-endi)
    (NP-SBJ (D^N^SG thie-the)
      (N^N^SG uuard-ward)
      (NP-POS (NPR^G^SG godes-god)))
    (CODE <R_2482>)
    (ADVP-DIR (ADVR nahor-nah)
      (ADJ^I^SG mikilu-mikil)))
  (CODE <C>)
  ...
  (. ; -;))
(ID OSHeliandC.1328.2480-2485))
```

**VP conjunction** is not distinguished from IP conjunction due to the lack of a VP node. It is treated as elision, using an IPX with equal-sign coindexing.

```
( (IP-MAT-SPE (CODE <C>)
  (IP-MAT-SPE=0 (ADVP-TMP (ADV Than-than))
    (MDPI^2^SG scalt-skulan)
    (NP-SBJ (PRO^N^2^SG thu-thu))
    (ADVP-TMP (ADV eft-eft))
    (NP-OB1 (uuord-word N^A^SG))
    (VB sprekan-sprekan))
  (, , -, )
  (CODE <R_169>)
  (CONJP (IPX-MAT-SPE=0 (HV hebban-hebbian)
    (NP-OB1-PRN (NP-POS (PRO$^G^2^SG
thinera-thin)
      (N^G^SG stemna-
stemna))
      (N^A^SG giuuald-giwald))))))
  (. :-:))
```



(ID OSHeliandC.78.168-169))

In **right-node raising** the second conjunct is placed in a **parenthetical** IPX, itself at IP level, without a CONJP.

```
( (IP-MAT-0 (CODE <R_1075>)
  (ADVP-TMP (ADV Thuo-tho))
  (VBDI^3^SG bigan-biginnan)
  (ADVP-TMP (ADV eft-eft))
  (VB niuson-niuson)
  (CODE <C>)
  (IPX-MAT-PRN=0 (CONJ endi-endi)
    (ADVP-DIR (ADVR nahor-nah))
    (VBDI^3^SG gieng-gangan))
  (CODE <R_1076>)
  (NP-SBJ (ADJ^N^SG unhiuri-unhiuri)
    (N^N^SG fiond-fiund))
  (CODE <C>)
  (NP-TMP (ADJ^I^SG othar-othar)
    (N^I^SG sithu-sith))
  (. , -,))
(ID OSHeliandC.586.1075-1076))
```

Unindexed IPX is used in certain types of **comparative** (with CPX) and in **restarts**.

CPs can also be incomplete (CPX) in certain types of **comparative**, and in *that*-clauses with an elided complementizer.

### Negation

The sentential negator in Old Saxon is the particle *ne* (NEG), placed at IP level, usually directly preverbal. The negator *ne* plus *eowiht* is always parsed as a nominal constituent, and hence as involving constituent negation, regardless of whether or not it is inflected. In constituent negation, NEG is contained within the negated constituent, and often cliticized onto the following word (in the manuscript perhaps even more often than in the Sievers edition).

```
( (IP-MAT-SPE (CODE <C>)
  (NEG Ni-ne)
  (HVI^2^SG hãbi-hebbian)
  (NP-SBJ (PRO^N^2^SG thu-thu))
  (NP-OB1 (ADJ^A^SG uuekean-wek)
    (N^A^SG hugi-hugi))
  (. , -,))
(ID OSHeliandC.139.262))

( (IP-MAT (CODE <C>)
  (NP-SBJ *con*)
  (NEG ni-ne)
  (VBDI^3^SG antuuordida-andwordian)
  (NP-OB1 (NEG+N^A^SG niouuiht-eowiht))
  (CODE <R_5383>)
  (PP (P uuiđ-with)
    (NP (PRO$^A^3^PL iro-he)
      (ADJ^A^PL uurethun-wreth)
      (N^A^PL uuord-word)))
  (. :-:))
(ID OSHeliandC.3187.5382-5383))
```

## Clauses

### Clause structure

As described under **phrase structure**, the clause structure representations in the HeliPaD are quite flat due to the lack of a VP node. In addition, there are a number of word-level categories that may be immediately dominated by the IP with no phrasal node intervening. These are:

- All verbs, finite and non-finite
- Particles (RP, ALSO)
- Sentential conjunctions (CONJ)
- Negation (NEG)
- Single-word interjections (INTJ)
- The word *wita* (UTP)

```
( (IP-MAT-SPE (CODE <C>)
  (CONJ endi-endi)
  (NP-SBJ *con*)
  (ADVP-TMP (ADV eft-ef))
  (RP upp-up)
  (GE+VBDI^3^SG gigeng-gangan)
  (. ,-,))
  (ID OSHeliandC.1288.2408))

( (IP-MAT-SPE (CODE <C>)
  (CONJ ac-ak)
  (IP-MAT-SPE-0 (UTP uuita-wita)
    (NP-1 (PRO^D^3^SG im-he))
    (VB uuonian-wonon)
    (PP (NP *ICH*-1)
      (P mid-mid))))
  (. ,-,)
  (CODE <R_3996>)
  (CONJP (IPX-MAT-SPE=0 (VB thuoloian-tholon)
    (PP (P mid-mid)
      (NP (PRO$^D^1^SG usson-usa)
        (N^D^SG thiodne-thiodan))))))
  (. :-:))
  (ID OSHeliandC.2317.3995-3996))
```

Any complete finite clause (IP-MAT\*, IP-SUB\*) will contain at least a subject and a finite verb. See **Subjects** for details of the former.

There are also two types of non-finite IPs: infinitives (IP-INF\*) and small clauses (IP-SMC). The former are headed by a non-finite verb, and do not necessarily have a subject (though they may). The latter always contain a subject and a non-verbal predicate or participial verb form. In addition, **participial phrases** behave like clauses in some respects.

The structure of CPs is discussed under **Subordinate clauses**.

### Clausal extended labels

Clausal phrases (IPs, CPs) must have one of the following extended labels:

For IPs:

- -MAT: main clause
- -SUB: subordinate clause, inc. direct questions

- -INF: infinitival clause
- -INF-NCO: non-complement infinitival clause
- -SMC: small clause

☐ For IPs, the extended label -SUB-CON is not used in the HeliPaD, as it is redundant and all instances of conjoined subordinate clauses can be retrieved in other ways. The label -ABS (for infinitival absolutes) is also not used.

For CPs:

- -REL: relative clause
- -FRL: free relative clause
- -THT: *that*-clause
- -ADV: adverbial clause
- -CMP: comparative clause
- -QUE: interrogative clause
- -DEG: degree clause

☐ For CPs, the HeliPaD does not include -CAR (clause-adjoined relatives), -CLF (clefts), -EXL (exclamatives), and -EOP (gapped infinitival relative/purpose clauses), as these structures are essentially not found in the *Heliand*.

Following these functional extended labels, an IP or a CP may additionally bear one or more of the following, in that order:

- A **nominal extended label** (free relative clauses only)
- -PRN (**parenthetical**) or -LFD (**left-dislocated**)
- -SPE (**direct speech**)
- An index number (e.g. -0, -1) or -X

```
(CP-FRL-OB1-PRN-SPE (WNP-SBJ-1 (ADV so-so)
                                (WPRO^N huat-hwe))
 (C so-so)
 (IP-SUB-SPE (NP-SBJ *T*-1)
             (NP-OB2 (PRO^D^2^SG thi-thu))
             (ADJP-PRD (ADJ^N^SG gibidig-gividig))
             (ADVP-DIR (ADV forth-forth))
             (CODE <R_3379>)
             (RD uuerthan-werthan)
             (MDDS^3^SG scoldi-skulan)))
```

(OSHeliandC.1918.3377-3379)

### ***Matrix clauses (IP-MAT)***

The label for matrix clauses is IP-MAT. Almost all tokens are IP-MAT at the highest level.

```
( (IP-MAT (CODE <C>)
         (ADVP-TMP (ADV Thuo-tho))
         (VBDI^3^SG bigan-biginnan)
         (NP-OB2-RFL (PRO^D^3^SG im-he))
         (NP-SBJ (D^N^SG the-the)
                (ADJ^N^SG uuiso-wis)
                (N^N^SG mann-man))
         (, , -, )
         (CODE <R_313>)
         (NP-SBJ-PRN (ADJP (ADV suitho-switho)
                          (ADJ^N^SG guod-god))
```

```

(N^N^SG gumo-gumo))
(CODE <C>)
(NP-SBJ-PRN (NPR^N^SG Ioseph-Joseph))
(PP (P an-an)
  (NP (PRO$^D^3^SG is-is)
    (N^D^SG muode-mod)))
(CODE <R_314>)
(VB thenkean-thenkian)
(NP-OB1 (D^G^PL thero-the)
  (N^G^PL thingo-thing))
(, ,-,)
(CODE <C>)
(CP-QUE (WADVP-1 (WADV huo-hwo))
  (C 0)
  (IP-SUB (ADVP *T*-1)
    (NP-SBJ (PRO^N^3^SG hie-he))
    (NP-OB1 (D^A^SG thea-the)
      (N^A^SG thiornun-thiorna))
    (ADVP-TMP (ADV tho-tho))
    (CODE <R_315>)
    (NP-ADT (N^D^PL listion-list))
    (VBDS^3^SG forlieti-farlatan)))
(. .-.))
(ID OSHeliandC.174.312-315))

```

```

( (IP-MAT (CODE <R_334>)
  (NP-SBJ *con*)
  (VBDI^3^SG bisorogoda-bisorgon)
  (NP-OB1 (PRO^A^3^SG sea-siu))
  (PP (P an-an)
    (NP (PRO$^D^3^SG is-is)
      (N^D^SG gisitha-gisith)))
  (. ,-,))
  (ID OSHeliandC.190.334))

```

Imperative clauses are also IP-MAT, but are distinguished by having an imperative verb form (e.g. VBI) and no requirement for a subject.

```

( (IP-MAT-SPE (CODE <C>)
  (NEG Ni-ne)
  (VBI^2^PL latat-latan)
  (IP-INF (NP-SBJ (PRO$^A^2^SG iuuan-iuwa)
    (N^A^SG hugi-hugi))
    (VB tuiflean-twiflian))
  (. !-!)
  (' '-'))
  (ID OSHeliandC.514.948))

```

```

( (IP-MAT-SPE (CODE <C>)
  (HVI^2^PL Hebbeat-hebbian)
  (NP-OB1 (PRO$^A^2^SG iuuan-iuwa)
    (N^A^SG uuilleon-willio))
  (ADVP-DIR (ADV tharod-tharod))
  (, ,-,)
  (CODE <R_944>)
  (NP-VOC (N^N^PL liudi-liudi))
  (NP-OB1-PRN (PRO$^A^2^SG iuuan-iuwa)
    (N^A^SG giloḃon-gilovo))
  (. :-:))
  (ID OSHeliandC.512.943-944))

```

```
( (IP-MAT-SPE (CODE <R_1067>)
  (GE+VBI^2^SG Giheli-helian)
  (NP-OB1 (PRO$^A^2^SG thinan-thin)
    (N^A^SG hungar-hungar))
  (. .-.))
  (' '-'))
(ID OSHeliandC.580.1067))
```

Clauses containing *wita* (UTP) as main verb are labelled IP-MAT, and like imperatives have no requirement for a subject to be present.

```
( (IP-MAT-SPE (CODE <C>)
  (UTP Uuita-wita)
  (VB kiesan-kiosan)
  (NP-OB2 (PRO^D^3^SG im-he))
  (NP-OB1 (ADJ^A^SG oðerna-othar)
    (CODE <R_224>)
    (ADJ^A^SG niudsamana-niudsam)
    (N^A^SG namon-namo))
  (. :-:))
(ID OSHeliandC.113.223-224))
```

### Subordinate clauses

As a rule, subordinate clauses are CPs, at least when they are finite. All complete CPs are **labelled for type** and contain an IP-SUB or IPX-SUB complement. In *wh*-questions, relatives, free relatives, comparatives, etc., they will also contain a *wh*-phrase (overt or empty). All CPs contain a C head in the HeliPaD, except direct questions and conditional inversion.

☐ The HeliPaD does not share the Penn corpora's strange fetish for treating subordinating conjunctions as prepositions.

```
(CP-CMP (WADVP-TMP-2 (ADV than-than)
  (ADV lang-lang))
  (C the-the)
  (IP-SUB (ADVP-TMP *T*-2)
    (NP-OB1 (PRO^A^3^SG sia-siu))
    (NP-SBJ (NP-POS (N^G^PL firio-firihos)
      (N^N^PL barn-barn))
    (CODE <R_4455>)
    (VB ardon-ardon)
    (MDPI^3^PL muotun-motan)))
```

(OSHeliandC.2580.4452-4456)

The position in front of the *wh*-word or complementizer ("SpecCP", though no theoretical significance should be attached to this term) can host elements such as adverbs, left-dislocated constituents, etc.

```
(CP-ADV-SPE (C huand-hwand)
  (IP-SUB-SPE (CP-FRL-LFD-SPE (WNP-SBJ-1 (ADV so-so)
    (WPRO^N hue-hwe))
  (C so-so)
  (IP-SUB-SPE (NP-SBJ *T*-1)
    (NP-OB1 (NP-POS (N^G^PL
  uuapno-wapan))
    (N^A^SG nith-
  nith))
  (, , - ,)
  (CODE <R_4897>)
```

```

(NP-OB1-PRN (ADJ^A^SG
grimman-grim)
(N^A^SG
gerheti-gerheti))
(MDPI^3^SG uuil-
willian)
(CODE <C>)
(ADVP (ADV gerno-gern))
frummian)))
(VB frummean-
(, , -, )
(CODE <R_4898>)
(IP-SUB-SPE (NP-SBJ-RSP (PRO^N^3^SG hie-he))
(VBPI^3^SG sueltid-sweltan)
(NP-OB2-RFL (PRO^D^3^SG im-he))
(ADVP-TMP (ADV oft-oft))
(CODE <C>)
(NP-ADT (NP-POS (N^G^SG suerdes-swerd))
(N^I^PL eggion-eggia)))
(, , -, )
(CODE <R_4899>)
(CONJP (IP-SUB-SPE (NP-SBJ *con*)
(VBPI^3^SG doit-doian)
(NP-OB2-RFL (PRO^D^3^SG im-he))
(ADJP-ADT (ADJ^N^SG drorag-
drorag))))))

```

(OSHeliandC.2885.4895-4899)

In subordinate clauses, elements before a single *that*-complementizer are treated as in SpecCP, and elements between two *that*-complementizers are treated as involving a *that*-clause embedded within another *that*-clause, as in the below.

```

( (IP-MAT-SPE (CODE <R_587>)
(NP-SBJ (PRO^N^3^SG Hie-he))
(VBDI^3^SG quat-kwethan)
(CP-THT-SPE (C that-that)
(CP-THT-SPE (PP (P an-an)
(NP (D^D^SG them-the)
(ADJ^D^SG selbon-self)
(N^D^SG dage-dag)
(CODE <C>)
...
(, , -, )
(CODE <R_589>)
(IP-MAT-PRN-SPE (ADVP (ADV so-so))
(VBDI^3^SG quat-
kwethan)
(NP-SBJ (PRO^N^3^SG
he-he)))
(C that-that)
(IP-SUB-SPE (ADVP-DIR (ADV oстана-
ostana))
(CODE <C>)
(NUMP-2 (NUM^N^SG en-en))
(MDDS^3^SG scoldi-skulan)
(VB scinan-skinan)
(CODE <R_590>)
...
(. .-.))
(ID OSHeliandC.311.587-592))

```

### Adverbial clauses (CP-ADV)

CP-ADV is the label used for adverbial clauses, which may have a variety of forms and functions.

☐ In the HeliPaD, non-argumental *that*-clauses (e.g. with resultative meaning) are still labelled CP-THT and not CP-ADV.

Most subordinating conjunctions are treated as ADVPs in SpecCP along with a null C.

```
(CP-ADV (ADVP-TMP (ADV thuo-tho))
  (C 0)
  (IP-SUB (NP-SBJ (PRO^N^3^PL sia-he))
    (ADVP-TMP (ADV eft-eft))
    (PP (P te-te)
      (NP (N^D^SG hebanuuange-hevanwang)))
    (CODE <R_415>)
    (VBDI^3^PL uundun-windan)
    (PP (P thuru-thurh)
      (NP (D^A^PL thiu-the)
        (N^A^PL uuolkan-wolkan))))))
```

(OSHeliandC.227.414-415)

```
(CP-ADV (ADVP-LOC (ADV thar-thar))
  (C 0)
  (IP-SUB (NP-SBJ (PRO^N^3^SG hie-he))
    (PP (P an-an)
      (NP (PRO$^D^3^SG is-is)
        (N^D^SG rikie-riki)))
    (VBDI^3^SG sat-sittian)))
```

(OSHeliandC.384.715-718)

Such clauses are treated as relative (CP-REL) if, and only if, there is a coreferential adverb preceding it in the clause.

```
( (IP-MAT (CODE <R_1136>)
  (NP-SBJ *con*)
  (VBDI^3^SG fuor-faran)
  (NP-OB2-RFL (PRO^D^3^SG im-he))
  (PP (P te-te)
    (NP (D^D^PL them-the)
      (N^D^PL friundun-friund)
      (CODE <C>)
      (CP-REL (WADVP-LOC-1 (ADV thar-thar))
        (C 0)
        (IP-SUB (IP-SUB-0 (ADVP-LOC *T*-1)
          (NP-SBJ (PRO^N^3^SG hie-he))
          (VN afuodid-afodian)
          (BEDI^3^SG uuas-wesan))
          (, , -, )
          (CODE <R_1137>)
          (CONJP (IPX-SUB=0 (ADVP (ADV tirlico-
            tirliko))
              (VN atogan-
                atiohan))))))
        (. , -, ))
    (ID OSHeliandC.620.1136-1137))
```

A limited number of subordinators, mostly those that are not homophonous with adverbs, are treated as C elements (*gif, huand, ne* (NEG+C), *newa, so, than* in comparatives, *that, the, untat*).

```
(CP-ADV (C unthat-untat)
  (IP-SUB (NP-SBJ (PRO^N^3^SG hie-he))
    (PP (P te-te)
      (NP (N^D^SG middean-middia)))
    (VBDI^3^SG quam-kuman)
    (CODE <R_2241>)
    (NP-SBJ-PRN (N^N^SG uualdand-waldand))
    (PP (P mid-mid)
      (NP (PRO$^I^3^SG is-is)
        (N^I^SG uuerodu-werod))))))
```

(OSHeliandC.1168.2239-2241)

```
(CP-ADV (C huand-hwand)
  (IP-SUB (ADVP-LOC (ADV thar-thar))
    (BEDI^3^SG uuas-wesan)
    (NP-POS-1 (N^G^PL gumono-gumo))
    (GE+VN gitald-tellian)
    (CODE <R_2871>)
    (PP (P ano-ano)
      (NP (NP (N^A^SG uuiḃ-wif))
        (CONJP (CONJ endi-endi)
          (NP (N^A^SG kind-kind))))))
    (CODE <C>)
    (NP-SBJ-PRN (N^N^PL uueros-wer))
    (ADVP (ADV atsamma-atsamne))
    (CODE <R_2872>)
    (NP-SBJ (NP-POS *ICH*-1)
      (NUM^N^PL fif-fif)
      (NUM^N^PL thusundig-thusundig))))
```

(OSHeliandC.1566.2869-2872)

As in the YCOE, *so*-clauses are treated as CP-ADV and not CP-CMP when there is no obvious extraction or comparison involved, particularly in the combinations *reht so* and *al so* - but this can be a tricky distinction to make.

```
( (IP-MAT-SPE (CODE <R_155>)
  (BEPI^3^PL sind-wesan)
  (NP-SBJ (PRO$^N^1^PL unca-unka)
    (N^N^PL andbari-andbari))
  (CODE <C>)
  (ADJP-PRD (ADJR^N^PL odarlicron-otharlik))
  (, , -, )
  (CODE <R_156>)
  (NP-SBJ-PRN (NP-SBJ-PRN (N^N^SG muod-mod))
    (CONJP (CONJ endi-endi)
      (NP-SBJ-PRN (N^N^SG megincraft-
meginkraft))))
  (, , -, )
  (CODE <C>)
  (CP-ADV-SPE (C so-so)
    (IP-SUB-SPE (NP-SBJ (PRO^N^1^DU uuit-wit))
      (ADVP-TMP (ADV iu-ju))
      (NP-TMP (QP (ADV so-so)
        (Q^A^SG managan-manag))
        (N^A^SG dag-dag))
```



```

(CODE <R_157>)
(BEDI^1^PL uuarun-wesan)
(PP (P an-an)
      (NP (D^D^SG thesaro-these)
            (CODE <MS_9a>)
            (NP-OB2 (N^D^SG uueroldi-
werold))))))
      (. :-:))
(ID OSHeliandC.74.155-157))

( (IP-MAT (CODE <C>)
      (ADVP-TMP (ADV Thuo-tho))
      (HVDI^3^SG habda-hebbian)
      (NP-SBJ (PRO^N^3^SG hie-he))
      (NP-OB1 (NP-POS (PRO$^G^1^SG usas-usa)
                     (N^G^SG uualdandes-waldand))
              (CODE <R_191>)
              (N^A^SG geld-geld))
      (GE+VN gilestid-lestian)
      (, , -, )
      (CODE <C>)
      (CP-ADV (ADVP (ADV all-al))
              (C so-so)
              (IP-SUB (NP-SBJ (PRO$^N^3^SG is-is)
                              (N^N^SG gigengi-gigengi))
                      (BEDI^3^SG uuas-wesan)
                      (CODE <R_192>)
                      (GE+VN gimarcod-markon)
                      (PP (P mid-mid)
                          (NP (N^D^PL mannon-man))))))
      (. .-.))
(ID OSHeliandC.89.190-192))

```

▣ Purpose clauses headed by a genitive demonstrative are not treated as CP-ADV but as CP-FRL-ADT.

V1 conditionals, usually *unless*-clauses, are CP-ADV with no C.

```

(CP-ADV (IP-SUB (NP-SBJ-X *exp*)
              (NEG ni-ne)
              (BEDS^3^SG uuari-wesan)
              (CP-THT-X (C that-that)
                        (IP-SUB (NP-SBJ (PRO^N^3^SG it-it))
                                (NP-PRD (N^N^SG gibod-gibod)
                                        (NP-POS (NPR^G^SG godes-god)
                                                (CODE <R_206>)
                                                (ADJ^G^SG selbes-self))))
                        (BEDS^3^SG uuari-wesan))))))

```

(OSHeliandC.101.203-206)

### **That-clauses (CP-THT)**

▣ The HeliPaD is more liberal than the YCOE in its use of CP-THT, which is used broadly for any clause introduced by *that*, as well as some in the scope of negation introduced by *ne*, unless they are instances of **CP-DEG**. CP-THT is thus essentially a formal rather than functional label. CP-THT does not need to be the complement of a verb or adjective, and is usually unindexed at IP level.

```

(CP-THT (NEG+C ne-ne)
      (IP-SUB (NP-SBJ (PRO^N^3^SG siu-siu))
              (NP-OB1 (PRO^A^3^SG ina-he))

```

```

(ADVP-TMP (ADV simla-simbla))
(CODE <MS_16b>)
(ADVP-DIR (ADV tharod-tharod))
(CODE <R_457>)
(PP (P te-te)
    (NP (D^D^SG them-the)
        (NP-POS (NPR^G^SG godes-god))
        (N^D^SG uuihe-wih)))
(CODE <C>)
(VB forgeþan-fargevan)
(MDDI^3^SG scolda-skulan))

```

(OSHelilandC.243.453-457)

```

( (IP-MAT (CODE <C>)
  (ADVP-TMP (ADV Thuo-tho))
  (GE+VBIDI^1^SG gifragen-fregnan)
  (NP-SBJ (PRO^N^1^SG ik-ik))
  (CP-THT (C that-that)
    (IP-SUB (NP-OB2 (PRO^D^3^SG iro-siu))
      (ADVP-LOC (ADV thar-thar))
      (NP-SBJ (N^N^SG sorga-sorga))
      (GE+VBIDI^3^SG gistuod-standan)))
  (, , -, )
  (CODE <R_511>)
  (CP-THT-PRN (C that-that)
    (IP-SUB (NP-OB1 (PRO^A^3^SG sia-siu))
      (NP-SBJ (D^N^SG thiu-the)
        (ADJ^N^SG mikila-mikil)
        (N^N^SG maht-maht)
        (CODE <C>)
        (NP-POS (N^G^SG metodes-metod)))
      (VBIDI^3^SG tedelda-tedelian)
      (, , -, )
      (CODE <R_512>)
      (NP-SBJ-PRN (ADJ^N^PL uureth-wreth)
        (N^N^PL uurðigiscapu-
wurdigiskap))))
  (. .-.))
  (ID OSHelilandC.271.510-512))

```

*That*-clauses may be subjects, as described in **Expletive constructions**.

*That*-clauses are quite often parenthetical to an overt demonstrative or object pronoun:

```

( (IP-MAT (CODE <C>)
  (BEDI^3^SG Uuas-wesan)
  (NP-SBJ (D^N^SG that-the)
    (CP-THT-PRN *ICH*-1))
  (PP (PP (P an-an)
    (NP (PRO$^D^3^PL is-is)
      (N^D^PL uuordon-word))))
    (CONJP *ICH*-2))
  (ADJP-PRD (ADJ^N^SG scin-skin))
  (CODE <R_1208>)
  (CONJP-2 (CONJ iac-jak)
    (PP (P an-an)
      (NP (PRO$^D^3^PL is-is)
        (N^D^PL dadeon-dad))))
  (ADVP (ADV so-so)

```

```

                                (ADV samo-sama)))
(CODE <C>)
(CP-THT-PRN-1 (C that-that)
              (IP-SUB (IP-SUB (NP-SBJ (PRO^N^3^SG hie-he))
                                (NP-PRD (N^N^SG drohtin-drohtin))
                                (BEDI^3^SG uuas-wesan)
                                (, , -, )
                                (CODE <P_86>)
                                (CODE <R_1209>)
                                (NP-PRD-PRN (ADJ^N^SG himilisk-
himilisk)
                                                (N^N^SG herro-herro))))
                                (CODE <C>)
                                (CONJP (CONJ endi-endi)
                                        (IP-SUB (NP-SBJ *con*)
                                                (PP (P te-te)
                                                    (NP (N^D^SG helpa-
helpa))))
                                        (VBDI^3^SG quam-kuman)
                                        (CODE <R_1210>)
                                        (PP (P an-an)
                                            (NP (D^A^SG thesan-
these)
                                                (N^A^SG middilgard-
middilgard))))
                                        (CODE <C>)
                                        (NP-OB2 (NP-POS (N^G^PL
manno-man))
                                                (N^D^PL barnon-
barn))
                                        (, , -, )
                                        (CODE <R_1211>)
                                        (NP-OB2-PRN (N^D^PL liudeon-
liudi))
                                        (PP (P te-te)
                                            (NP (D^D^SG theson-
these)
                                                (N^D^SG liohte-
lioht))))))
                                (. .-.))
                                (ID OSHeliandC.663.1207-1211))
( (IP-MAT (CODE <C>)
        (NP-SBJ (PRO^N^3^SG Hie-he))
        (GE+VBDI^3^SG gideda-don)
        (NP-SBJ-1 (PRO^A^3^SG it-it)
                  (CP-THT-PRN *ICH*-2))
        (ADVP-TMP (ADV san-san))
        (PP (P after-aftar)
            (NP (D^D^SG thiu-the)))
        (CODE <R_996>)
        (IP-SMC (NP-SBJ *ICH*-1)
                (NP-OB2 (N^D^PL mannon-man))
                (ADJP-PRD (ADJ^A^SG mari-mari)))
        (, , -, )
        (CODE <C>)
        (CP-THT-PRN-2 (C that-that)
                      (IP-SUB (NP-SBJ (PRO^N^3^PL sia-he))
                              (ADVP-LOC (ADV thar-thar))
                              (NP-OB1 (ADJ^A^SG mahtina-mahtig)
                                      (CODE <R_997>))

```

```

(N^A^SG herron-herro))
(HVDI^3^PL haʒdun-hebbian)))
(. :-:))
(ID OSHeliandC.547.995-997))

( (IP-MAT (CODE <C>)
(NP-SBJ *con*)
(BEDI^3^PL uuarun-wesan)
(NP-OB2-RFL (PRO^D^3^PL im-he))
(ADVP-LOC (ADV thar-thar))
(GE+VNI^A^PL gifarana-faran)
(PP (P te-te)
(NP (D^I^SG thiu-the)
(CODE <R_1229>)
(CP-THT-PRN (C that-that)
(IP-SUB (NP-SBJ (PRO^N^3^PL sia-he))
(NP-OB1 (NP-POS (PRO$^G^1^SG uses-
usa)
(N^G^SG drohtines-
drohtin))
(CODE <C>)
(NP-OB1 (N^G^PL dadeo-dad))
(NP-OB1 (CONJ endi-endi)
(NP (N^G^PL uuordo-
word))))))
(CODE <R_1230>)
(VB faran-faron)
(MDDI^3^PL uuoldun-willian))))))
(. ,-,))
(ID OSHeliandC.668.1228-1230))

```

In some **comparative** contexts, the complementizer may be elided, in which case CPX-THT is used.

```

( (IP-MAT-SPE-0 (CODE <C>)
(ADVP (ADV So-so))
(MDPI^3^SG mag-mugan)
(NP-SBJ (D^N^SG that-the)
(CP-THT-PRN-SPE *ICH*-1))
(PP (P an-an)
(NP (PRO$^D^3^SG is-is)
(N^D^SG hugi-hugi)))
(CODE <MS_49a>)
(NP-PRD (QR^N^SG mera-mero)
(CP-CMP-SPE *ICH*-2))
(CODE <P_122>)
(CODE <R_1712>)
(PP (P an-an)
(NP (D^D^SG thesaro-these)
(N^D^SG middilgard-middilgard)))
(CODE <C>)
(NP-OB2 (NP-POS (N^G^PL manno-man))
(GE+WADJ^D^SG gihuilicon-gihwilik))
(CODE <R_1713>)
(BE uesan-wesan)
(PP (P an-an)
(NP (D^D^SG thesaro-these)
(N^D^SG uueroldi-werold)))
(CODE <C>)
(CP-THT-PRN-SPE-1 (C that-that)
(IP-SUB (NP-SBJ (PRO^N^3^SG hie-he))
(ADVP-LOC (ADV hier-her))

```

```

(NP-OBJ (N^G^SG uuammes-wam))
(GE+VBPI^3^SG giduot-don)))
(, , -, )
(CODE <R_1714>)
(CP-CMP-SPE-2 (WADJP-PRD-3 0)
(C than-than)
(IPX-SUB-SPE=0 (ADJP-PRD *T*-3)
(CPX-THT-SPE (IP-SUB-SPE ...
(. .-.))
(ID OSHeliandC.905.1711-1716))

```

### *Degree complements (CP-DEG)*

Degree complements occur with the adverb *so*, and are traced to this constituent. In some instances they can be difficult to distinguish from purpose clauses.

```

( (IP-MAT (CODE <C>)
(ADVP-TMP (ADV Thann-than))
(ADVP (ADV thoh-thoh))
(GE+VBDI^3^SG gitruoda-truon)
(ADVP (ADV so-so)
(ADV uuel-wel)
(CP-DEG *ICH*-1))
(CODE <P_144>)
(CODE <R_2029>)
(PP (P an-an)
(NP (PRO$^D^3^PL iru-iru)
(N^D^PL hugisceftion-hugiskaft)))
(CODE <C>)
(NP-SBJ (ADJ^N^SG helag-helag)
(N^N^SG thiorna-thiorna))
(, , -, )
(CODE <R_2030>)
(CP-DEG-1 (C that-that)
(IP-SUB (NP-OBJ (PRO^G^3^SG is-it))
(PP (P after-aftar)
(NP (D^D^PL them-the)
(N^D^PL uuordon-word)))
(CODE <C>)
(NP-SBJ (NP-POS (N^G^SG uualdandes-waldand))
(N^N^SG barn-barn))
(, , -, )
(CODE <R_2031>)
(NP-SBJ-PRN (NP-POS (VGI^G^PL helandero-
helian))
(ADJS^N^SG best-bet))
(CODE <C>)
(VB helpan-helpan)
(MDDS^3^SG uueldi-willian))
(. .-.))
(ID OSHeliandC.1050.2028-2031))

```

### *Comparative clauses (CP-CMP)*

Comparative clauses are one of the most difficult things to annotate properly. Because of elision, the CPX and IPX ("incomplete") labels get a lot of use here.

The CPX-CMP label is used when the *wh*-operator and complementizer can be overtly represented in the annotation. This is usually because the category or position of the trace is not straightforwardly recoverable, or else because large-scale absence of structure makes reconstruction impossible.

```

( (IP-MAT-SPE (CODE <C>)
  (ADVP-TMP (ADV Nu-nu))
  (NP-SBJ (PRO^N^1^SG ik-ik))
  (NP-OB1 (PRO^A^2^PL iu-gi))
  (VB sendean-sendian)
  (MDPI^1^SG scal-skulan)
  (CODE <R_1874>)
  (PP (P after-aftar)
    (NP (D^D^SG theson-these)
      (N^D^SG landscepie-landskepi)))
  (CODE <C>)
  (CPX-CMP-SPE (C so-so)
    (IPX-SUB-SPE (NP-SBJ (N^N^SG lamb-lamb))
      (PP (P under-undar)
        (NP (N^A^PL uuluos-wulf))))))
  (. :-:))
(ID OSHeliandC.977.1873-1874))

```

CP-CMP dominates IP-SUB only if nothing other than the extracted category has been elided. In cases where more structure has been elided, IPX-SUB is used. Equal-sign coindexing (=0 with a corresponding -o) shows where the model for the elided structure is to be understood.

```

( (IP-MAT-SPE-0 (CODE <C>)
  (BEI^2^PL Uuesat-wesan)
  (NP-OB2-RFL (PRO^D^2^PL iu-gi))
  (ADJP-PRD (ADV so-so)
    (ADJ^N^PL giuuara-giwar)
    (CP-CMP-SPE *ICH*-1))
  (PP (P uuidar-withar)
    (NP (D^I^SG thi-u-the)))
  (, , -, )
  (CODE <R_1883>)
  (PP (P uuid-with)
    (NP (PRO^D^3^PL iro-iru)
      (N^D^PL fecnon-fekan)))
  (ADVP-LOC (ADV thar-thar))
  (CODE <C>)
  (CP-CMP-SPE-1 (WADJP-PRD-2 0)
    (C so-so)
    (IPX-SUB-SPE=0 (ADJP-PRD *T*-2)
      (NP-SBJ (MAN^N^3^SG man-man))
      (PP (P uuidar-withar)
        (NP (N^D^PL fiondon-
fiund))))
      (MDPI^3^SG scal-skulan)))
  (. .-.))
(ID OSHeliandC.981.1882-1883))

```

The most usual types of CP(X)-CMP involve a head *so*, with an accompanying *so* or *sulik* X constituent (as in the above), *than* along with comparative morphology, or *than X the* (as in the below).

```

( (IP-MAT (CODE <C>)
  (NP-SBJ *con*)
  (VBDI^3^PL quathun-kwethan)
  (CP-THT (C that-that)
    (IP-SUB-0 (NP-OB2 (PRO^D^3^SG im-he))
      (NP-SBJ (N^N^SG uualdand-waldand)
        (ADJ^N^SG self-self))
      (, , -, )
      (CODE <R_2337>))

```

```

(NP-SBJ-PRN (NPR^N^SG god-god)
              (ADJ^N^SG alomahtig-alomahtig))
(CODE <C>)
(VN forgeban-fargevan)
(HVDS^3^SG habdi-hebbian)
(CODE <R_2338>)
(NP-OB1 (QR^A^PL merun-mero)
         (N^A^PL mahti-maht)
         (CODE <C>)
         (CP-CMP (WNP-OB1-1 0)
                  (C than-than)
                  (IPX-SUB=0 (NP-OB1 *T*-1)
                              (ADVP (ADV elcor-
elkor))
                              (NP-OB2 (Q^D^SG
enigon-enig)
                              (NP-POS
(N^G^SG mannes-man))
                              (N^D^SG
sunie-sunu))))))
(, , -, )
(CODE <R_2339>)
(NP-OB1-PRN (NP-OB1-PRN (N^A^SG craft-kraft))
            (CONJP (CONJ endi-endi)
                   (NP-OB1-PRN (N^A^PL cunsti-kunst))))
(. ; - ;))
(ID OSHeliandC.1243.2336-2339))

( (IP-MAT (CODE <C>)
         (NP-SBJ *con*)
         (GE+VBPI^3^SG gihugit-huggian)
         (CP-QUE (WNP-OB1-1 (WPRO^A huat-hwe)
                           (NP-POS *ICH*-2))
                 (C 0)
                 (IP-SUB (NP-OB1 *T*-1)
                          (NP-SBJ (PRO^N^3^SG hie-he)
                                   (ADJ^N^SG selbo-self))
                          (GE+VBDI^3^SG gifrumida-frummian)
                          (CODE <R_3497>)
                          (NP-POS-2 (ADJ^G^SG grimmes-grim))
                          (ADVP-TMP (ADV thann-than)
                                    (ADV lang-lang)
                                    (CP-CMP (WADV-TMP-3 0)
                                             (C the-the)
                                             (IP-SUB (ADVP-TMP *T*-3)
                                                      (NP-SBJ (PRO^N^3^SG
hie-he))
                                                      (MDDI^3^SG muosta-
motan)
                                                      (CODE <MS_98b>)
                                                      (NP-OB1 (PRO$^G^3^SG
is-is)
                                                      (N^G^SG
iuguthi-juguth))
                                                      (VB niotan-niotan))))))
         (. ; - ;))
         (ID OSHeliandC.2000.3496-3497))

```

For instances where a subordinator may be elided, see the section on *that-clauses*.

### Direct and indirect questions (CP-QUE)

Both direct and indirect questions are CP-QUE dominating an IP-SUB. The difference in annotation is that indirect questions contain a null C, whereas direct questions do not. When a *wh*-phrase is present, it is traced into the subordinate clause in which it belongs.

```
( (CP-QUE-SPE (CODE <C>))
  (' '-'))
(WADVP-1 (WADV Huo-hwo))
(IP-SUB-SPE (ADVP *T*-1)
  (MDPI^3^SG mag-mugan)
  (NP-SBJ (D^N^SG that-the))
  (GE+RD giuuerthan-werthan)
  (ADVP (ADV so-so))
  (' '-'))
(IP-MAT-PRN (VBDI^3^SG quat$-kwethan)
  (NP-SBJ (PRO^N^3^SG $hie-he)))
(, ,-,)
(CODE <R_142>)
(' '-'))
(ADVP-TMP (ADV after-aftar))
(PP (P an-an)
  (NP (N^D^SG aldre-aldar))))
(. ?-?))
(ID OSHeliandC.65.141-142))

( (IP-MAT (CODE <R_210>)
  (NP-SBJ *con*)
  (VBDI^3^SG fragode-fragon)
  (ADVP (ADV niudlico-niudliko))
  (CODE <C>)
  (CP-QUE (WNP-PRD-1 (WPRO^N huat-hwe))
    (C 0)
    (IP-SUB (NP-PRD *T*-1)
      (NP-SBJ (PRO$^N^3^SG is-is)
        (N^N^SG namo-namo))
      (MDDS^3^SG scoldi-skulan)
      (CODE <R_211>)
      (BE uuesan-wesan)
      (PP (P an-an)
        (NP (D^D^SG thesaro-these)
          (N^D^SG uueroldi-werold))))))
  (. :-:))
  (ID OSHeliandC.105.210-211))
```

**Yes-no questions** are annotated as CP-QUE immediately dominating an IP-SUB but no C or *wh*-phrase.

```
( (CP-QUE-SPE (CODE <R_3813>)
  (IP-SUB-SPE (BEPI^3^SG is-wesan)
    (NP-SBJ (PRO^N^3^SG it-it))
    (NP-PRD (N^N^SG reht-reht))))
  (ID OSHeliandC.2196.3813))
```

When it introduces a yes-no question, *hwethar* is labelled WQ. These questions follow the same rules as others: C is present in indirect questions and absent in direct questions.

```
( (CP-QUE-SPE (CODE <C>)
  (' '-'))
  (WQ Hueder-hwethar)
  (IP-SUB-SPE (VBPI^2^PL ledeat-ledian)
```



```

(NP-SBJ (PRO^N^2^PL gi-gi))
(NP-OB1 (VNI^A^SG uundan-windan)
        (N^A^SG gold-gold))
(CODE <R_555>)
(PP (P te-te)
    (NP (N^D^SG geþu-geva)))
(NP-OB2 (WADJ^D^SG huilicon-hwilik)
        (NP-POS (N^G^PL gumono-gumo))))
(. ?-?)
(ID OSHeliandC.291.554-555))

```

```

( (IP-MAT-SPE (CODE <C>)
  (NEG ne-ne)
  (VBI^2^PL ruokeat-rokian)
  (CP-QUE-SPE (WQ hueder-hwethar)
    (C 0)
    (IP-SUB-SPE (NP-SBJ (PRO^N^2^PL gi-gi))
      (NP-OB1 (NP-POS (D^G^SG thes-the))
              (NP-OB1 (Q^A^SG enigan-enig)
                      (N^A^SG thanc-thank))
              (CONJP *ICH*-1))
      (VBPS^2^PL antfahan-andfahan)
      (CODE <R_1542>)
      (CONJP-1 (CONJ eftha-eftha)
              (NP-OB1 (N^A^SG lon-lon)))
      (PP (P an-an)
          (NP (D^D^SG thesaro-these)
              (ADJ^D^SG lehnun-lehni)
              (N^D^SG ueroldi-werold))))))
(. ,-,))
(ID OSHeliandC.818.1541-1542))

```

*Hwethar* may also serve as a *wh*-word meaning "which of two". In this role, it is tagged as WADJ.

```

( (IP-MAT (CODE <R_5409>)
  (ADVP-TMP (ADV Thuo-tho))
  (VBDI^3^SG bigan-biginnan)
  (NP-SBJ (D^N^SG thie-the)
          (N^N^SG heritogo-heritogo))
  (CODE <C>)
  (NP-OB1 (D^A^SG thia-the)
          (N^A^SG heri-heri)
          (NP-POS (NPR^G^PL Iudeono-Judeo)))
  (, ,-,)
  (CODE <R_5410>)
  (NP-OB1-PRN (D^A^SG that-the)
              (N^A^SG folc-folk))
  (VB fragoian-fragon)
  (, ,-,)
  (CODE <C>)
  (CP-ADV (ADVP-LOC (ADV thar-thar))
    (C 0)
    (IP-SUB (NP-SBJ (PRO^N^3^PL sia-he))
      (PP (NP (PRO^D^3^SG im-he))
          (P fora-for))
      (VBDI^3^PL stuodun-standan)))
  (, ,-,)
  (CODE <R_5411>)
  (CP-QUE (WNP-OB1-1 (WADJ^A^SG huederon-hwethar)
                    (NP-POS *ICH*-2))
    (C 0)

```

```

(IP-SUB (IP-SUB-0 (NP-OB1 *T*-1)
                  (NP-SBJ (PRO^N^3^PL sia-he))
                  (NP-POS-2 (D^G^PL thero-the)
                             (NUM^G^PL tueio-twene)
                             (CP-REL-SPE *ICH*-3))
                  (CODE <C>)
                  (VB tuomian-tomian)
                  (MDDS^3^PL uueldin-willian))
(, , -, )
(CODE <R_5412>)
(CONJP (IPX-SUB=0 (NP-OB1 (N^G^SG ferahes-ferah)
                          (VB biddian-biddian)))
(, :-:))
...
(ID OSHeliandC.3203.5409-5413))

```

When *hwat* has an adverbial role ("why" or "how"), it is treated as accusative and part of a WNP-ADT. See also **Interjection Phrases**.

```

( (CP-QUE-SPE (CODE <C>)
              (' '-'))
  (WNP-ADT-1 (WPRO^A Huat-hwe))
  (IP-SUB-SPE (NP-ADT *T*-1)
              (MDPI^2^PL uuelliat-willian)
              (NP-SBJ (PRO^N^2^PL gi-gi))
              (ADJP-2 (PRO$^G^1^PL minero-min))
              (ADV-LOC (ADV hier-her))
              (' '-'))
  (IP-MAT-PRN (VBDI^3^SG quat$-kwethan)
              (NP-SBJ (PRO^N^3^SG $he-he)))
(, , -, )
(CODE <R_3574>)
(' '-'))
(NP-OB1 (ADJP *ICH*-2)
        (N^G^PL helpono-helpa))
(VB biddean-biddian))
(. ?-?)
(' '-'))
(ID OSHeliandC.2061.3573-3574))

```

In some questions, *wh*-doubling may be found. In these cases, the matrix clause *wh*-element is treated as a dummy and coindexed to the lower, "real" version.

```

( (CP-QUE-SPE (CODE <C>)
              (WNP-1 (WPRO^N huat-hwe))
              (IP-SUB-SPE (VBPI^3^PL quethat-kwethan)
                          (NP-SBJ (D^N^PL thesa-these)
                                   (NP-POS (NPR^G^PL Iudeo-Judeo))
                                   (N^N^PL liudi-liudi))
                          (, , -, )
                          (CODE <R_3040>)
                          (NP-SBJ-PRN (ADJ^N^SG mari-mari)
                                       (N^N^SG meginthioda-meginthioda))
                          (, , -, )
                          (CODE <C>)
                          (CP-QUE-SPE (WNP-PRD-1 (WPRO^N huat-hwe)
                                                  (NP-POS *ICH*-2))
                          (C 0)
                          (IP-SUB-SPE (NP-PRD *T*-1)

```

```

ik))
man))
      (NP-SBJ (PRO^N^1^SG ik-
      (NP-POS-2 (N^G^PL manno-
      (BEPS^1^SG si-wesan))))
      (. ?-?)
      (' '-'))
(ID OSHeliandC.1683.3039-3040))

```

### Relative clauses (CP-REL)

All relative clauses have a *wh*-position and a complementizer position. It is rare for neither to be overtly filled. Since *thie* and *the* are ambiguous between invariant C and demonstrative, the preference is to annotate it as a demonstrative if that is formally possible.

```

( (IP-MAT-SPE (CODE <R_1683>)
  (NP-SBJ (N^N^PL liudi-liudi)
    (CP-REL-SPE *ICH*-1))
  (BEPI^3^PL sind-wesan)
  (NP-OB2 (PRO^D^3^SG im-he))
  (ADJP-PRD (ADJR^N^PL leoðrun-liof))
  (ADJP-ADT (ADJ^I^SG mikilo-mikil))
  (CODE <C>)
  (CP-REL-SPE-1 (WNP-OB1-2 (D^A^PL thie-the))
    (C 0)
    (IP-SUB-SPE (NP-OB1 *T*-2)
      (NP-SBJ (PRO^N^3^SG hie-he))
      (NP-OB2-RFL (PRO^D^3^SG im-he))
      (PP (P an-an)
        (NP (D^D^SG them-the)
          (N^D^SG lande-land)))
      (GE+VBDI^3^SG giuuarahtha-wirkian)
      (CODE <R_1684>)
      (NP-SBJ-PRN (N^N^SG uualdand-
waldand))
      (PP (P an-an)
        (NP (N^D^SG uuilleon-willio)
          (PRO$^D^3^SG sinan-sin))))))
  (. .-.))
(ID OSHeliandC.888.1683-1684))

```

```

( (IP-MAT (CODE <C>)
  (CONJ endi-endi)
  (NP-SBJ *con*)
  (NP-OB2 (D^D^PL them-the)
    (N^D^PL mannon-man))
  (VBDI^3^SG sagda-seggian)
  (CODE <R_1296>)
  (NP-ADT (ADJ^D^PL spahon-spah)
    (N^D^PL uuordon-word))
  (CODE <MS_38a>)
  (CODE <C>)
  (NP-OB2-PRN (D^D^PL them-the)
    (CP-REL (WNP-OB1-1 0)
      (C the-the)
      (IP-SUB (NP-OB1 *T*-1)
        (NP-SBJ (PRO^N^3^SG hie-he))
        (PP (P te-te)
          (NP (D^D^SG thero-the)
            (N^D^SG spraco-spraka)))
        (ADVP-DIR (ADV tharod-tharod))

```

```

                                (CODE <P_92>C>)
                                (CODE <R_1297>C>)
                                (NP-SBJ-PRN (NPR^N^SG Crist-Krist)
                                   (N^N^SG alouualdo-
alowaldo))
                                (CODE <C>)
                                (GE+VN gicoran-kiosan)
                                (HVDI^3^SG haḃda-hebbian))))
                                ...
                                (ID OSHeliandC.710.1295-1299))

```

There is no evidence for *that* as an invariant complementizer in the HeliPaD (unlike e.g. in late OE).

Temporal and locative relative clauses behave like other relative clauses. There is usually an overt adverbial *thar* or *than*, but in some cases there may be only a complementizer *the*.

```

(ADVP-DIR (ADV thanan-thanan)
  (CP-REL-SPE (WADV-1 0)
    (C the-the)
    (IP-SUB-SPE (ADVP *T*-1)
      (NP-SBJ (MAN^N man-man))
      (NP-OB1 (PRO^A^2^PL iu-gi))
      (VB antfahan-andfahan)
      (NEG ni-ne)
      (MDPI^3^SG uuili-willian))))

```

(OSHeliandC.1014.1946-1947)

```

(PP (P an-an)
  (NP (D^A^SG that-the)
    (ADJ^A^SG hoha-hoh)
    (N^A^SG hus-hus)
    (, , -, )
    (CODE <C>)
    (CP-REL (WADV-LOC-1 (ADV thar-thar))
      (C 0)
      (IP-SUB (ADVP-LOC *T*-1)
        (NP-SBJ (D^N^SG thiu-the)
          (N^N^SG heri-heri))
        (VBDI^3^SG dranc-drinkan)
        (, , -, )
        (CODE <R_2002>)
        (NP-SBJ-PRN (D^N^PL thia-the)
          (NPR^N^PL Iudeon-Judeo))
        (PP (P an-an)
          (NP (D^D^SG them-the)
            (N^D^SG gestseli-gastseli))))))

```

(OSHeliandC.1030.1999-2002)

```

( (IP-MAT-SPE (CODE <R_2496>C>)
  (ADVP-TMP (ADV Than-than)
    (CP-REL-SPE *ICH*-1))
  (VBPI^3^SG biginnit-biginnan)
  (NP-SBJ (PRO^D^3^SG im-he))
  (VB thunkean-thunkian)
  (, , -, )
  (CODE <C>)
  (CP-REL-SPE-1 (WADV-TMP-2 (ADV than-than))
    (C 0)

```

```

(IP-SUB-SPE (IP-SUB-SPE (ADVP-TMP *T*-2)
(NP-SBJ (PRO^N^3^SG
hie-he))
(P (P under-undar)
(NP (D^D^SG thero-
(N^D^SG thieda-
thiod)))
(VBPI^3^SG sted-stan))
...
(ID OSHeliandC.1336.2496-2498))

```

See also [Case attraction](#).

### *Free relative clauses (CP-FRL)*

Free relatives (CP-FRL) are treated essentially like nominals in terms of their external syntax: they take **nominal extended labels** and thus may be subjects, locatives, predicates, etc. In case of ambiguity between questions and free relatives, the **question** parse is chosen. Relative clauses with no possible antecedent that are headed by a demonstrative are treated as free relatives, as in the first example below.

```

( (IP-MAT-SPE (CODE <R_2582>C>)
(NP-SBJ (PRO^N^1^SG ik-ik)
(ADJ^N^SG selbo-self))
(BEPI^1^SG biun-wesan)
(CP-FRL-PRD-SPE (WNP-SBJ-1 (D^N^SG that-the))
(C 0)
(IP-SUB-SPE (NP-SBJ *T*-1)
(ADVP-LOC (ADV thar-thar))
(VBPI^1^SG saiu-saian)))
(. , -,))
(ID OSHeliandC.1388.2582))

( (IP-MAT (CODE <C>)
(NP-SBJ (D^N^PL thia-the)
(N^N^PL liudi-liudi)
(QP *ICH*-1))
(RDDI^3^PL uurdun-werthan)
(QP-1 (Q^N^PL alla-al))
(CODE <R_2862>)
(ADJP-PRD (ADJ^N^PL sada-sad))
(NP-SBJ-PRN (ADJ^N^SG salig-salig)
(N^N^SG folc-folk))
(, , -,)
(CODE <C>)
(CP-FRL-SBJ-PRN (WNP-SBJ-2 (ADV so-so)
(WPRO^N huat-hwe))
(C so-so)
(IP-SUB (NP-SBJ *T*-2)
(ADVP-LOC (ADV thar-thar))
(GE+VN gisamnod-samnon)
(BEDI^3^SG uuas-wesan)
(CODE <R_2863>)
(P (P fön-fan)
(NP (Q^D^PL allon-al)
(ADJ^D^PL uuidon-wid)
(N^D^PL uegon-weg))))))
(. .-.))
(ID OSHeliandC.1561.2861-2863))

```

```

( (IP-MAT (CODE <C>)
  (NP-SBJ (D^N^SG That-the)
    (N^N^SG uuerod-werod)
    (ADJ^N^SG oder-othar))
  (VBDI^3^SG bed-bidan)
  (CODE <R_104>)
  (PP (P umbi-umbi)
    (NP (D^A^SG thena-the)
      (N^A^SG alah-alah))
    (ADV utan-utan))
  (CODE <C>)
  (NP-SBJ-PRN (NP-POS (NPR^G^PL Hebreo-Ebreo))
    (N^N^PL liudi-liudi))
  (, , -, )
  (CODE <R_105>)
  (CP-FRL-TMP (WADVP-TMP-1 (WADV huan-hwan)
    (ADV er-er))
    (C 0)
    (IP-SUB (ADVP-TMP *T*-1)
      (D^N^SG thie-the)
      (NP-SBJ (ADJ^N^SG fruodo-frod)
        (N^N^SG man-man))
      (CODE <C>)
      (GE+VN gifrumid-frummian)
      (HVDS^3^SG habdi-hebbian)
      (CODE <P_12>)
      (CODE <R_106>)
      (NP-OB1 (NP-POS (N^G^SG uualdandes-waldand))
        (N^A^SG uuilleon-willio))))
  (. .-.))
(ID OSHeliandC.41.103-106))

```

Free relatives with *wh*-elements are mostly of the *so wh-* *so* type. *so wh-* is in the *wh*-position, and the second *so* is in C.

```

( (IP-MAT (CODE <C>)
  (NP-SBJ (D^N^SG That-the)
    (N^N^SG fri-fri))
  (NP-OB1 (Q^A^SG all-al))
  (VBDI^3^SG biheld-bihaldan)
  (CODE <R_436>)
  (PP (P an-an)
    (NP (PRO$^D^3^PL iro-iru)
      (N^D^PL hugisceftion-hugiskaft)))
  (CODE <C>)
  (NP-SBJ-PRN (ADJ^N^SG helag-helag)
    (N^N^SG thiorno-thiorna))
  (, , -, )
  (CODE <R_437>)
  (NP-SBJ-PRN (D^N^SG thiu-the)
    (N^N^SG magat-magath))
  (CODE <MS_16a>)
  (PP (P an-an)
    (NP (PRO$^D^3^SG iru-iru)
      (N^D^SG muode-mod)))
  (, , -, )
  (CODE <C>)
  (CP-FRL-OB1-PRN (WNP-OB1-1 (ADV so-so)
    (WPRO^A huat-hwe))
    (C so-so)
    (IP-SUB (NP-SBJ (PRO^N^3^SG siu-siu))

```

```

(GE+VBDI^3^SG gihorda-horian)
(IP-INF (NP-OB1 *T*-1)
        (NP-SBJ (D^A^PL thia-the)
                (N^A^PL man-man))
        (VB spreca-sprekan)))
(. .-.))
(ID OSHeliandC.236.435-437))

```

### *Infinitival clauses (IP-INF, IP-INF-NCO)*

Infinitival clauses are normally the complements of verbs in the HeliPaD. Only with ECM verbs is a subject present in an infinitival clause.

```

(CP-FRL-OB1-PRN (WNP-OB1-1 (ADV so-so)
                             (WPRO^A huat-hwe))
 (C so-so)
 (IP-SUB (NP-SBJ (PRO^N^3^SG siu-siu))
         (GE+VBDI^3^SG gihorda-horian)
         (IP-INF (NP-OB1 *T*-1)
                 (NP-SBJ (D^A^PL thia-the)
                         (N^A^PL man-man))
                 (VB spreca-sprekan))))

```

(OSHeliandC.236.435-437)

```

( (IP-MAT-SPE (CODE <C>)
              (NP-SBJ (D^N^SG That-the)
                      (IP-INF-PRN-SPE *ICH*-1))
              (RDPI^3^SG uuirdit-werthan)
              (NP-OB2 (PRO^D^2^SG thi-thu))
              (NP-PRD (N^N^SG uuerk-werk)
                     (ADJ^N^SG mikil-mikil))
              (, , -, ,)
              (CODE <P_38>)
              (CODE <R_502>)
              (IP-INF-PRN-SPE-1 (NP-OB1 (N^A^SG thrim-thrim))
                                (TO te-te)
                                (GE+VB githolonne-tholon))
              (. .-.))
              (' ' -'))
(ID OSHeliandC.264.501-502))

```

**Exceptional case marking (ECM)** verbs, also known as **Accusative and infinitive (Acl)** verbs, are lexically restricted to the following:

- Perception verbs:
  - *findan*
  - *horian*
  - *sehan*
- Verbs of permission and commanding:
  - *biddian*
  - *hetan*
  - *latan*
- *witan*

They involve a bare infinitive and always include a subject.

```

( (IP-MAT (CODE <C>)
          (NP-SBJ *con*))

```

```

(GE+VBDI^3^PL gisahun-sehan)
(IP-INF (NP-SBJ (PRO$^A^3^SG iro-iru)
              (N^A^SG boggebon-boggevo))
        (CODE <R_2739>)
        (BE uuesan-wesan)
        (PP (P an-an)
            (NP (N^D^PL uunneon-wunnia))))
(. .-.))
(ID OSHeliandC.1476.2738-2739))

```

```

( (IP-MAT (CODE <C>)
  (NP-SBJ *con*)
  (VBDI^3^SG hiet-hetan)
  (ADVP-TMP (ADV thuo-tho))
  (IP-INF (NP-SBJ (PRO$^A^3^SG is-is)
                  (N^A^SG uuapanberan-wapanberand))
          (CODE <R_2780>)
          (VB gangan-gangan)
          (PP (P fan-fan)
              (NP (D^D^SG them-the)
                  (N^D^SG gastseli-gastseli))))))
(ID OSHeliandC.1503.2779-2780))

```

▣ Arbitrary subjects of such clauses are given as the **empty category** \*arb\*.

In **tough-movement** constructions, the subject is traced to the infinitive, which is contained within the ADJP.

```

( (IP-MAT-SPE (CODE <C>)
  (NP-SBJ-1 *pro*)
  (ADJP-PRD (ADJ^N^SG othi-othi)
            (IP-INF-SPE *ICH*-2))
  (BEPI^3^SG ist-wesan)
  (ADVP-DIR (ADV tharod-tharod))
  (IP-INF-SPE-2 (NP-OB1 *-1)
                (TO te-te)
                (VB faranne-faran))
  (CODE <R_1780>)
  (NP-OB2 (N^D^PL eldibarnon-eldibarn))
  (, , -, )
  (CODE <C>)
  (CP-ADV-SPE (ADVP (ADV thoh-thoh))
              (C 0)
              (IP-SUB-SPE (NP-SBJ (PRO^N^3^SG it-it))
                          (NP-OB2 (PRO^D^3^PL im-he))
                          (PP (P at-at)
                              (NP (D^D^SG them-the)
                                  (N^D^SG endie-endi))))
              (NEG ni-ne)
              (VBPS^3^SG dugi-dugan)))
(. .-.))
(ID OSHeliandC.935.1779-1780))

```

**Non-complement infinitives** (IP-INF-NCO) are those which are not the complements of verbs, nouns, or adjectives, and not the subject of the clause. Most usually, these co-occur with motion verbs.

```

( (IP-MAT (CODE <C>)
  (NP-SBJ *con*)
  (VBDI^3^SG geng-gangan)
  (IP-INF-NCO (PP (P uuiđ-with)

```



```

                (NP (PRO$^A^3^SG iro-iru)
                   (N^A^SG kind-kind)))
        (VB sprekan-sprekan)
        (CODE <R_2019>)
        (PP (P uuiđ-with)
            (NP (PRO$^A^3^SG iro-iru)
                (N^A^SG suno-sunu)
                (ADJ^A^SG selban-self))))
        (. , - ,)
    (ID OSHeliandC.1042.2018-2019))

```

### *Small clauses (IP-SMC)*

Small clauses consist of a subject and a predicate. They are the only type of IP that does not contain a verbal element by default. The subjects of small clauses are almost always in accusative case.

```

    ( (IP-MAT (CODE <C>)
        (ADVP-TMP (ADV Thuo-tho))
        (HVDI^3^SG habda-hebbian)
        (ADVP-TMP (ADV eft-ef))
        (IP-SMC (NP-SBJ (PRO$^A^3^SG is-is)
                       (N^A^SG uuord-word))
                (ADJP-PRD (ADJ^A^SG garo-garu)))
        (CODE <R_2024>)
        (NP-SBJ (ADJ^N^SG mahtig-mahtig)
                (N^N^SG barn-barn)
                (NP-POS (NPR^G^SG godes-god))))
    (ID OSHeliandC.1045.2023-2024))

    ( (IP-MAT-SPE (CODE <C>)
        (ADVP-TMP (ADV than-than))
        (VBPI^2^SG findis-findan)
        (NP-SBJ (PRO^N^2^SG thu-thu))
        (IP-SMC-SPE (ADJP-PRD (ADJ^A^SG gisundan-gisund))
                    (PP (P at-at)
                        (NP (N^D^SG hus-hus))))
        (CODE <R_2151>)
        (NP-SBJ (ADJ^A^SG maguungan-magujung)
                (N^A^SG man-man)))
        (. :-:))
    (ID OSHeliandC.1105.2150-2151))

```

If the subject and predicate are not adjacent, the small clause is taken to be where the predicate is, and the subject is traced to that.

```

    ( (IP-MAT (CODE <C>)
        (NP-SBJ-1 (D^A^SG thena-the))
        (GE+VBDI^3^SG gideda-don)
        (NP-SBJ (D^N^SG thie-the)
                (VGI^N^SG heland-helian)
                (ADJ^N^SG self-self))
        (. , - ,)
        (CODE <R_2355>)
        (NP-SBJ-PRN (NPR^N^SG Crist-Krist))
        (PP (P thuru-thurh)
            (NP (PRO$^A^3^SG is-is)
                (N^A^SG craft-kraft)
                (ADJ^A^SG mikil-mikil))))
        (CODE <C>)
        (IP-SMC (NP-SBJ *ICH*-1)

```

```

                (ADJP-PRD (ADJ^A^SG quican-kwik))
        (PP (P aftar-aftar)
            (NP (N^D^SG doðe-doth)))
        (. , - ,))
(ID OSHeliandC.1252.2354-2355))

```

Small clauses often occur with verbs of naming, with an accusative subject as usual.

## Nominals

### *Noun Phrases (NP)*

The head of a noun phrase is a nominal element, usually a noun (common or proper) or a pronoun. There is no explicit marking of any assumed null noun or pronoun in NPs.

```

(NP-SBJ (PRO^N^3^PL sea-he))
(NP-POS (NPR^G^SG Cristes-Krist))
(NP-OB1 (Q^A^SG all-al))

```

(OSHeliandC.9.32-42)

The word *man*, with its special tag MAN, heads an NP.

```

(NP-SBJ (MAN^N^3^SG man-man))

```

(OSHeliandC.42.106-113)

Noun-noun compounds are tagged as a single N when written together, as in the YCOE, but as separate nouns within the same NP when written separately, as in the case of proper names.

```

(NP-SBJ (NPR^N^SG Simon-Simon)
        (NPR^N^SG Petrus-Petrus))

```

(OSHeliandC.1691.3053-3054)

☒ Titles are never labelled as separate appositive phrases in the HeliPaD, whether modified or not:

```

(NP-SBJ (NPR^N^SG Iohannes-Johannes)
        (CODE <C>)
        (N^N^SG iungro-jungaro)
        (NP-POS (NPR^G^SG Cristes-Krist)))

```

(OSHeliandC.3321.5610-5611)

```

(NP-SBJ (NPR^N^SG Iohannes-Johannes)
        (D^N^SG thie-the)
        (ADJ^N^SG guodo-god))

```

(OSHeliandC.1460.2710-2711)

See the section on **Heads, modifiers and complements** for information on NP-internal modification, and the section on **Appositives and parentheticals** for information on apposition.

☐ NP-COM is not used, and nouns are assumed not to have nominal complements. NP-POS or clause-level adjuncts take over much of the work of this label in the HeliPaD.

Possessive pronouns are labelled PRO\$, and like other inflected nominal categories in the HeliPaD are always given attributes, even in the case of the third person possessives which do not inflect. These items are also found as genitive pronouns in some cases, essentially when they are used alone in a phrase, or modified only by *al*, *en* or *self*.

```
( (IP-MAT (CODE <C>)
  (NP-SBJ (NUM^N^SG En-en)
    (NP-POS *ICH*-1))
  (BEDI^3^SG uuas-wesan)
  (NP-POS-1 (PRO^G^3^PL iro-he))
  (ADVP-TMP (ADV thuo-tho))
  (ADVP-TMP (ADV noh-noh))
  (ADVP-TMP (ADV than-than))
  (CODE <R_47>)
  (NP-OB2 (NP-POS (N^G^PL firio-firihos)
    (N^D^PL barnun-barn))
  (ADVP-LOC (ADV biforan-biforan))
  (. , -,))
  (ID OSHeliandC.11.46-47))

( (IP-MAT-SPE (CODE <R_1632>)
  (CONJ ac-ak)
  (VBI^2^PL mithat-mithan)
  (NP-OB1 (PRO^G^3^SG is-it))
  (PP (P for-for)
    (NP (ADJ^D^PL odron-othar)
      (N^D^PL mannon-man)))
  (. :-:))
  (ID OSHeliandC.864.1632))

( (IP-MAT-SPE (CODE <C>)
  (CONJ endi-endi)
  (NP-SBJ (D^N^SG thie-the)
    (PRO$^N^2^SG iuuua-iuwa)
    (N^N^SG fritho-frithu))
  (VBPI^3^SG huiri-bit-hwervan)
  (CODE <R_1944>)
  (ADVP-TMP (ADV eft-eft))
  (PP (P an-an)
    (NP (NP-POS (PRO^G^2^PL iuuuer-gi)
      (ADJ^G^PL selbaro-self))
    (N^A^SG sith-sith)))
  (. , -,))
  (ID OSHeliandC.1011.1943-1944))
```

NP-internal agreement is forced wherever possible, thus allowing ambiguous elements to be tagged (the HeliPaD's approach to attributes is "maximalist"). The main exception to this is with instrumental elements, which often co-occur with formally dative elements. This is not treated as a case clash, and instrumental is only preferred where unambiguous.

### *Noun phrase extended labels*

☐ In general, nominal extended labels for arguments in the HeliPaD work like they do in the PPCME2 and IcePaHC and not as they do in the YCOE. Hence, not every paragraph in this section will be flagged in green.

At clause level, NPs always either bear a functional label or an index to show where they have been extracted from. The possibilities for nominal extended labels are:

- Arguments:
  - -SBJ: subject
  - -OB1: direct or only object
  - -OB2: indirect or second object
  - -PRD: predicate (also for ADJPs and PTPs)
- Non-arguments:
  - -ADT: adjunct (also for ADJPs and PTPs)
  - -LOC: locative
  - -TMP: temporal
  - -VOC: vocative
  - -POS: possessor (only one immediately dominated by each NP)
- -RFL: reflexive (not mutually exclusive with above)

☐ These replace the YCOE's phrasal case labels.

The two broad classes of NPs in the corpus are arguments and adjuncts. Temporal (-TMP) and locative (-LOC) NPs are a subclass of adjuncts (-ADT), as are vocatives (-VOC) and **left-dislocations** (-LFD). Possessors (-POS) only occur NP-internally or coindexed to an NP-internal position.

Any NP that contains a pronoun coreferential with the subject of the clause is labelled -RFL (reflexive), no matter where it occurs in the clause.

### Subjects

☐ In general, nominal extended labels for arguments in the HeliPaD work like they do in the PPCME2 and IcePaHC and not as they do in the YCOE. Hence, not every paragraph in this section will be flagged in green.

All complete **finite IPs** must have a subject (-SBJ), other than the following:

- Clauses with *wita* (equivalent to *uton* in the YCOE)
- Imperative clauses (though they may have one)
- The outer wrapper of **conjoined** clauses

```
( (IP-MAT-SPE (CODE <C>)
  (UTP Uuita-wita)
  (VB kiesan-kiosan)
  (NP-OB2 (PRO^D^3^SG im-he))
  (NP-OB1 (ADJ^A^SG oderna-othar)
    (CODE <R_224>)
    (ADJ^A^SG niudsamana-niudsam)
    (N^A^SG namon-namo))
  (. :-:))
(ID OSHeliandC.113.223-224))
```

```
( (IP-MAT-SPE (CODE <C>)
  (VBI^2^SG Sagi-seggian)
  (NP-OB2 (PRO^D^1^PL us-we))
  (, , -, )
  (CP-QUE-SPE (WPP-1 (P under-undar)
    (WNP (WADJ^D^SG huillicon-hwilik)
      (NP-POS *ICH*-2)))
  (C 0)
  (IP-SUB-SPE (PP *T*-1)
    (NP-SBJ (PRO^N^3^SG hie-he))
    (BEPS^3^SG si-wesan)
```

```

(NP-POS-2 (D^G^PL thesaro-these)
           (N^G^PL cunnio-kunni))
(VN afuodid-afodian))
      (. .-.))
      (' '-'))
(ID OSHeliandC.322.605))

( (IP-MAT-SPE (CODE <R_263>)
             (NEG ni-ne)
             (VBI^2^SG forohti-forhtian)
             (NP-SBJ (PRO^N^2^SG thu-thu))
             (NP-OB2 (PRO$^D^2^SG thinon-thin)
                     (N^D^SG ferahe-ferah))
             (. .-:))
  (ID OSHeliandC.140.263))

( (IP-MAT (CODE <R_994>)
        (IP-MAT-0 (NP-OB1 (D^A^SG That-the))
                  (MDDI^3^SG muosta-motan)
                  (NP-SBJ (NPR^N^SG Iohannes-Johannes))
                  (ADV-TMP (ADV thuo-tho))
                  (, ,-,))
                  (CODE <C>)
                  (CP-ADV (ADV (ADV all-al))
                          (C so-so)
                          (IP-SUB (NP-OB1 (PRO^A^3^SG it-it))
                                   (NP-SBJ (NPR^N^SG guod-god))
                                   (MDDI^3^SG uuelda-willian)))
                  (, ,-,))
                  (CODE <R_995>)
                  (GE+VB gisehan-sehan))
  (CONJP (CONJ endi-endi)
        (IPX-MAT=0 (GE+VB gihorean-horian)))
  (. .-.))
(ID OSHeliandC.546.994-995))

```

If no overt subject is present in such clauses, an **empty subject** is added.

Subjects are almost always nominative, though they are not required to be. A small selection of verbs may have non-nominative subjects for ease of retrieval. (This selection of verbs is likely to change in future editions of the corpus.)

- *thunkian* (dat.)
- *likon* (dat.)
- *lustian* (acc.)
- *thurstian* (acc.)
- *risan* (dat.) (variable)
- *belgan* (acc.) (variable)
- *spanan* (acc.) (variable)
- *hreuwan* (acc.) (variable)

In **non-finite IPs** there is no requirement to have a subject, on the whole. Overt subjects are marked -SBJ. ECM verbs have a subject; if it is absent, \*arb\* is used. Small clauses (-SMC) are required to have a subject and predicate.

```

( (IP-MAT (CODE <C>)
        (NP-SBJ-X *exp*)
        (BEDI^3^SG uuas-wesan)
        (NP-OB2 (PRO^D^3^PL im-he))
        (NP-PRD (NP-POS (PRO$^G^3^SG is-is))

```

```

                (N^G^SG huldi-huldi))
            (NP-POS-PRN (N^G^PL helpono-helpa))
            (N^N^SG tharf-tharf))
        (CODE <R_1188>)
        (IP-INF-X (TO te-te)
            (GE+VB githiononne-thionon))
        (. :-:))
    (ID OSHeliandC.650.1187-1188))

( (IP-MAT-SPE (CODE <C>)
    (NEG Ni-ne)
    (VBI^2^PL latat-latan)
    (IP-INF (NP-SBJ (PRO$^A^2^SG iuuan-iuwa)
        (N^A^SG hugi-hugi))
        (VB tuiflean-twiflian))
    (. !-!)
    (' '-'))
    (ID OSHeliandC.514.948))

```

```

( (IP-MAT (CODE <C>)
    (ADVP-TMP (ADV Thuo-tho))
    (HVDI^3^SG haþda-hebbian)
    (ADVP-TMP (ADV eft-eft))
    (IP-SMC (NP-SBJ (PRO$^A^3^SG is-is)
        (N^A^SG uuord-word))
        (ADJP-PRD (ADJ^A^SG garo-garu)))
    (CODE <R_2024>)
    (NP-SBJ (ADJ^N^SG mahtig-mahtig)
        (N^N^SG barn-barn)
        (NP-POS (NPR^G^SG godes-god))))
    (ID OSHeliandC.1045.2023-2024))

```

As in the YCOE, and unlike in the PPCME2, presentational *thar* is not specially tagged, and is always part of an ADVP-LOC. Expletive *it* is also not specially marked, except as part of an **expletive construction**, when it is part of an NP-SBJ-X.

```

( (IP-MAT (CODE <C>)
    (NP-SBJ *con*)
    (VBDI^3^SG quat-kwethan)
    (CP-THT (C that-that)
        (IP-SUB (ADVP-LOC (ADV thar-thar))
            (QP-1 (ADV so-so)
                (Q^N^SG filo-filo))
            (BEDS^3^SG uuari-wesan)
            (CODE <R_2833>)
            (NP-SBJ (QP *ICH*-1)
                (NP-POS (N^G^PL manno-man))
                (N^N^SG menigi-menigi))))
    (. :-:))
    (ID OSHeliandC.1537.2832-2833))

```

```

( (IP-MAT (CODE <C>)
    (NP-SBJ *con*)
    (VBDI^3^SG quat-kwethan)
    (CP-THT (C that-that)
        (IP-SUB (NP-SBJ-X (PRO^N^3^SG it-it))
            (NP-OB2 (NPR^D^SG gode-god))
            (BEDS^3^SG uuari-wesan)
            (CODE <R_2712>)
            (NP-OB2-PRN (N^D^SG uualdande-waldand))
            (ADJP-PRD (ADJ^N^SG uuidarmuod-witharmod))

```

```

(CODE <C>)
(CP-THT-X (C that-that)
...
(ID OSHeliandC.1461.2711-2714))

```

Free relative subjects are labelled -SBJ. Extraposed clausal subjects are coindexed with \*exp\* as part of an **expletive construction**. There are no other clausal subjects in the HeliPaD.

```

( (IP-MAT (CODE <R_654>)
  (BEDI^3^PL uuarun-wesan)
  (NP-OB2 (PRO^D^3^PL im-he))
  (NP-PRD (ADJ^N^PL glauua-glau)
    (N^N^PL guomon-gumo))
  ( , , -, )
  (CODE <C>)
  (CP-FRL-SBJ (WNP-SBJ-1 (D^N^PL the-the))
    (C 0)
    (IP-SUB (NP-SBJ *T*-1)
      (NP-OB1 (D^A^PL thea-the)
        (N^A^PL gifa-geva))
      (VBDI^3^PL leddun-ledian)))
  (. .-.))
(ID OSHeliandC.349.654))

```

```

( (IP-MAT (CODE <C>)
  (NP-SBJ-X *exp*)
  (NP-OB2 (PRO^D^3^SG Im-he))
  (BEDI^3^SG uuas-wesan)
  (ADVP-TMP (ADV thuo-tho))
  (NP-PRD (N^N^SG uuilleo-willio)
    (ADJ^N^SG mikil-mikil))
  (CODE <R_872>)
  (CP-THT-X (C that-that)
    (IP-SUB (NP-SBJ (PRO^N^3^SG hie-he))
      (PP (P fan-fan)
        (NP (ADJ^D^PL solicon-sulik)
          (N^D^PL saldõn-salitha)))
      (CODE <C>)
      (VB seggean-seggian)
      (MDDS^3^SG muosti-motan)))
  (. .-.))
(ID OSHeliandC.470.871-872))

```

### **Non-subject arguments**

There are three types of non-subject argument: -OB1, -OB2, and -PRD.

▣ The two **types of object** are used, broadly speaking, for accusative objects and for dative objects respectively. With certain verbs, genitive objects can also be either -OB1 or -OB2, depending on the case of the other object. Consult the **treatment of individual words** for details. With one verb, *lerian*, both objects can be accusative objects, with the people being taught as -OB2.

Single-object clauses:

```

( (IP-MAT (CODE <R_57>)
  (HVDI^3^PL habdun-hebbian)
  (PP (P fan-fan)
    (NP (NPR^D^SG Rumuburg-Rumuburg)))
  (CODE <MS_6b>)
  (CODE <C>))

```

```

(NP-OB1 (N^A^SG riki-riki))
(GE+VN giuunnan-winnan)
(CODE <R_58>)
(NP-SBJ (N^N^PL helmgitrosteon-helmgitrostio))
(. :-:))
(ID OSHeliandC.16.57-58))

```

```

( (IP-MAT-SPE (CODE <R_1612>)
(CONJ ac-ak)
(VBI^2^SG hilf-helpan)
(NP-OB2 (PRO^D^1^PL us-we))
(PP (P uuiðar-withar)
(NP (Q^D^PL allon-al)
(CODE <C>)
(ADJ^D^PL ubilon-uvil)
(N^D^PL dadeon-dad)))
(. .-.))
(" "-"))
(ID OSHeliandC.857.1612))

```

#### Double-object clauses:

```

( (IP-MAT (CODE <C>)
(ADVP-TMP (ADV Than-than))
(HVDI^3^SG habda-hebbian)
(ADVP-TMP (ADV thuo-tho))
(NP-SBJ (N^N^SG drohtin-drohtin)
(NPR^N^SG god-god))
(CODE <R_54>)
(NP-OB2 (NP-POS (NPR^G^PL Romano-Roman))
(N^D^PL liudeon-liudi))
(VN farliuuan-farlihan)
(CODE <C>)
(NP-OB1 (NP-POS (N^G^PL rikeo-riki))
(QS^A^SG mesta-mer))
(. :-:))
(ID OSHeliandC.14.53-54))

```

```

( (IP-MAT (CODE <R_5308>)
(IP-MAT-0 (NP-SBJ *con*)
(MDDI^3^SG uuelda-willian)
(NP-OB1 (NP-POS (N^G^PL manno-man))
(N^A^PL barn-barn))
(CODE <C>)
(NP-OB2 (N^G^SG morthies-morth))
(VB atuomian-atomian))
(, ,-,)
(CODE <R_5309>)
(CONJP (IPX-MAT=0 (VB nerian-nerian)
(PP (P af-af)
(NP (N^D^SG nodi-nod))))))
(. .-.))
(ID OSHeliandC.3143.5308-5309))

```

```

( (IP-MAT (CODE <R_2271>)
(NP-SBJ *con*)
(VBDI^3^SG lerda-lerian)
(NP-OB2 (PRO^A^3^PL sia-he))
(NP-OB1 (PRO$^A^3^SG iro-iru)
(N^A^SG giloðon-gilovo)))
(ID OSHeliandC.1201.2271))

```



**Predicates**, unlike other arguments, may be NPs or other phrases, most usually ADJP. They are usually nominative, but may also be genitive, or (in small clauses) accusative.)

```
( (IP-MAT (CODE <P_10>)
  (CODE <R_76>)
  (NP-PRD (NPR^N^SG Zacharias-Zacharias))
  (BEDI^3^SG uuas-wesan)
  (NP-SBJ (PRO^N^3^SG hie-he))
  (VN hetan-hetan)
  (. .-.))
(ID OSHeliandC.24.76))

( (IP-MAT (CODE <R_82>)
  (NP-SBJ *con*)
  (BEDI^3^PL uuarun-wesan)
  (ADJP-PRD (ADV so-so)
    (ADJ^N^PL gihoriga-gihorig))
  (CODE <C>)
  (NP-OB2 (N^D^SG hebancuninge-hevankuning))
  (. ,-,))
(ID OSHeliandC.31.82))

( (IP-MAT (CODE <R_347>)
  (VBDS^3^SG quami-kuman)
  (PP (P te-te)
    (NP (D^D^SG them-the)
      (N^D^SG cnuosla-knosal)
      (CP-REL *ICH*-1)))
  (NP-SBJ (GE+WPRO^N gihue-gihwe))
  (CODE <C>)
  (CP-REL-1 (WADVP-DIR-2 (ADV thanan-thanan)
    (C 0)
    (IP-SUB (ADVP-DIR *T*-2)
      (NP-SBJ (PRO^N^3^SG hie-he))
      (NP-PRD (N^G^SG cunneas-kunni))
      (BEDI^3^SG uuas-wesan)
      (, ,-,)
      (CODE <R_348>)
      (PTP-PRD-PRN (GE+VN giboran-beran)
        (PP (P fan-fan)
          (NP (D^D^PL them-the)
            (N^D^PL burgion-burg))))))
    (. .-.))
  (ID OSHeliandC.194.347-348))

( (IP-MAT-SPE-0 (CODE <C>)
  (HVI^2^PL Hebbeat-hebbian)
  (IP-SMC-SPE (NP-SBJ (PRO$^A^2^SG iuuuan-iuwa)
    (N^A^SG muod-mod))
    (PP (P uuider-withar)
      (NP (D^D^PL them-the)))
    (CODE <R_1877>)
    (ADJP-PRD (ADV so-so)
      (ADJ^A^SG glauuan-glau))
    (ADVP-DIR (ADV tegegnes-tegegnes)))
  (CODE <C>)
  ...
  (ID OSHeliandC.979.1876-1880))
```

Verbs which take predicates include particularly *wesan* (BE) and *weorthan* (RD), but also *hetan* (in the passive) and a number of more minor constructions. Predicates are usually nominative, but can also be genitive ("they were of his kin").

### Adjuncts

Adjuncts in the HeliPaD behave as they do in the YCOE. -TMP (temporal), -DIR (directional), -LOC (locative), -LFD (left-dislocation), and -VOC (vocative) should all be treated as subcases of adjuncts, which if not further specified are marked -ADT.

The division into arguments and adjuncts may not be watertight in all cases. Specialists in argument structure using the corpus are advised to check the annotation carefully and use their own judgement.

### Case attraction

The determiners in some relative clauses may agree in case with the antecedent rather than bearing the case of the gap. If possible, these are treated as instances of apposition.

```
(NP-OB1-PRN (D^A^SG thena-the)
  (CP-REL-SPE (WNP-SBJ-3 (D^N^SG the-the))
    (C 0)
    (IP-SUB-SPE (NP-SBJ *T*-3)
      (NP-OB1 (Q^A^SG all-al)
        (N^A^SG reht-reht))
      (VBPI^3^SG bikann-bikunnan))))
```

(OSHeliandC.1017.1957-1961)

☐ True case attraction is never annotated at phrase level (since the HeliPaD does not have phrase-level case labels), but there may be a mismatch between (gap) grammatical function and word-level case if a treatment as apposition would lead to a relative clause with both an empty complementizer and an empty operator.

```
( (IP-MAT (CODE <C>)
  (ADVP-TMP (ADV Thuo-tho))
  (NP-SBJ (D^N^SG thiu-the)
    (N^N^SG magat-magath))
  (HVDI^3^SG habda-hebbian)
  (CODE <R_2767>)
  (GE+VN githionot-thionon)
  (PP (P ti-te)
    (NP (N^D^SG thanke-thank)))
  (CODE <C>)
  (NP-OB2 (NP-OB2 (N^D^SG thiodcuninge-thiodkuning))
    (CODE <MS_77b>)
    (CODE <R_2768>)
    (CONJP (CONJ endi-endi)
      (NP-OB2 (Q^D^SG allon-al)
        (D^D^SG them-the)
        (N^D^SG erlscipie-erlskepi)
        (CODE <C>)
        (CP-REL (WNP-SBJ-1 (D^D^SG them-the))
          (C 0)
          (IP-SUB (NP-SBJ *T*-1)
            (PP (ADVP-LOC (ADV thar-
              (P inne-inne))
              (BEDI^3^SG uuas-wesan))))
          (CODE <R_2769>)
          (NP-POS (ADJ^G^PL godaro-god)
```

(N^G^PL gumono-gumo)))))

(. :-:))

(ID OSHeliandC.1494.2766-2769))

### *Adjective Phrases (ADJP)*

ADJP is an endocentric phrasal category, headed by an adjective or a first- or second-person possessive pronoun (which can be thought of as a subclass of adjective). It is normally only projected when an adjective is modified (or takes a complement - though this is rare in the HeliPaD) or when the adjective is predicative, scrambled or extraposed and therefore not contained within an NP.

The most common modifiers for adjectives are adverbs or ADVPs. As usual for modifiers, a single adverb does not project a phrase, and adverbs only do so if they are themselves modified.

```
( (IP-MAT (CODE <C>)
  (NP-SBJ (D^N^SG That-the))
  (BEDI^3^SG uuas-wesan)
  (NP-PRD (ADJP (ADV so-so)
    (ADJ^N^SG salig-salig))
    (N^N^SG man-man))
  (, , -, )
  ...
  (ID OSHeliandC.25.76-78))
```

```
( (IP-MAT (CODE <C>)
  (NEG ni-ne)
  (CODE <MS_7a>)
  (MDDI^3^SG muosta-motan)
  (NP-OB2 (PRO^D^3^PL im-he))
  (NP-SBJ (N^N^SG erbiuuard-erwiward))
  (CODE <R_80>)
  (PP (P an-an)
    (NP (PRO$^D^3^SG iro-iru)
      (N^D^SG iuguthedi-juguthhed)))
  (CODE <C>)
  (ADJP-PRD (ADJ^N^SG gibithig-gividig))
  (RD uuerthan-werthan)
  (. ---))
  (ID OSHeliandC.28.79-80))
```

```
( (IP-MAT (CODE <C>)
  (ADVP-TMP (ADV Thuo-tho))
  (RDDI^3^SG uuarth-werthan)
  (NP-SBJ (D^N^SG that-the)
    (CP-QUE-PRN *ICH*-1))
  (ADJP-PRD (ADVP (ADV so-so)
    (ADV uuido-wido))
    (ADJ^N^SG cuth-kuth))
  ...
  (. :-:))
  (ID OSHeliandC.1074.2071-2074))
```

Instrumental determiners and adjectives are taken as modifiers, as with **ADVPs**.

```
(CP-THT (C that-that)
  (IP-SUB (NP-SBJ (PRO^N^3^SG hie-he))
    (NEG ni-ne)
    (BEDS^3^SG uuari-wesan)
    (NP-OB2 (D^D^SG them-the)
      (NP-POS (N^G^PL manno-man))
```

```

(N^D^SG folke-folk))
(, , -, )
(CODE <R_5542>)
(NP-OB2-PRN (D^D^SG them-the)
(N^D^SG uuerode-werod))
(ADJP-PRD (D^I^SG thiu-the)
(ADJR^N^SG uurethra-wreth)))

```

(OSHeliandC.3278.5540-5542)

```

( (IP-MAT-SPE (CODE <R_1781>)
(ADVP-TMP (ADV Thann-than))
(VBPI^3^SG ligit-liggian)
(ADVP-TMP (ADV eft-eft))
(ADJP-1 (ADJ^N^SG oder-othar))
(CODE <C>)
(ADJP-2 (ADJR^N^SG engera-engi)
(ADJ^I^SG mikilu-mikil))
(CODE <R_1782>)
(NP-SBJ (ADJP *ICH*-1)
(ADJP *ICH*-2)
(N^N^SG uueg-weg))
(PP (P an-an)
(NP (D^D^SG thesaro-these)
(N^D^SG uueroldi-werold)))
(. , -, ))
(ID OSHeliandC.936.1781-1782))

```

Adjective phrases may also contain an NP-POS, especially when superlative.

```

( (IP-MAT-SPE (CODE <R_2595>)
(NP-SBJ (D^N^SG that-the))
(BEPI^3^SG is-wesan)
(ADJP-PRD (NP-POS (Q^G^PL allero-al)
(N^G^PL beuuo-beo))
(ADJS^N^SG bredost-bred))
(. ; -; ))
(ID OSHeliandC.1397.2595))

```

As in the YCOE, ordinal numbers are labelled as adjectives.

```

( (IP-MAT (CODE <C>)
(NP-SBJ (Q^N^SG sum-sum))
(ADVP-LOC (ADV thar-thar))
(ALSO oc-ok)
(ADVP-TMP (ADV sithor-sithor))
(VBDI^3^SG quam-kuman)
(CODE <R_3422>)
(PP (P an-an)
(NP (D^A^SG thia-the)
(ADJ^A^SG elliftun-ellifto)
(N^A^SG tid-tid)))
(. .-.))
(ID OSHeliandC.1946.3421-3422))

```

### Quantifier Phrases (QP)

Quantifier phrases are only found in modifying function in the HeliPaD, unlike ADJPs. Quantifiers and QPs can be either pre- or postnominal. The QP is only projected when the quantifier is modified or when it is floated, scrambled, or extraposed.

```
(NP-OB1 (QP (ADV so-so)
             (Q^A^SG manag-manag))
         (N^A^SG gibod-gibod)
         (NP-POS (NPR^G^SG godes-god)))
```

(OSHeliandC.4.12-15)

```
( (IP-MAT (CODE <C>)
         (ADVP-TMP (ADV Thuo-tho))
         (RDDI^3^SG uuarth-werthan)
         (NP-SBJ (NP-POS (D^G^SG thes-the)
                       (N^G^SG uuiβes-wif))
                (N^N^SG hugi-hugi)
                (QP *ICH*-1))
         (CODE <R_282>)
         (PP (P after-aftar)
            (NP (D^D^SG them-the)
                (N^D^SG arundie-arundi)))
         (CODE <C>)
         (QP-1 (Q^N^SG all-al))
         (GE+VN gihuorban-hwervan)
         (CODE <R_283>)
         (PP (P an-an)
            (NP (NP-POS (NPR^G^SG godes-god))
                (N^A^SG uuilleon-willio)))
         (. .-.))
  (ID OSHeliandC.155.281-283))
```

▣ Like other **modifiers** separated from their head, floated quantifiers are traced to the head, as in the above example.

Consult the **treatment of individual words** to find out which items are treated as quantifiers.

▣ *al* may be treated as an adverb (ADV) in some cases, particularly when introducing *al so*-clauses.

```
( (IP-MAT (CODE <R_2656>)
         (NP-SBJ *con*)
         (VBDI^3^SG auuohs-awahsan)
         (ADVP (ADV all-al))
         (PP (P under-undar)
            (NP (D^D^SG theson-these)
                (N^D^SG uueroda-werod)))
         (. :-:))
  (ID OSHeliandC.1428.2656))

( (IP-MAT (CODE <C>)
         (ADVP-TMP (ADV Thuo-tho))
         (HVDI^3^SG habda-hebbian)
         (NP-SBJ (PRO^N^3^SG hie-he))
         (NP-OB1 (NP-POS (PRO$^G^1^SG usas-usa)
                       (N^G^SG uualdandes-waldand))
                (CODE <R_191>))
```

```

                (N^A^SG geld-geld))
(GE+VN gilestid-lestian)
(, , -, )
(CODE <C>)
(CP-ADV (ADVP (ADV all-al))
        (C so-so)
        (IP-SUB (NP-SBJ (PRO$^N^3^SG is-is)
                        (N^N^SG gigengi-gigengi))
                (BEDI^3^SG uuas-wesan)
                (CODE <R_192>)
                (GE+VN gimarcod-markon)
                (PP (P mid-mid)
                    (NP (N^D^PL mannon-man))))))
(. .-.))
(ID OSHeliandC.89.190-192))

```

### *Number Phrases (NUMP)*

In most respects, number phrases behave like **quantifier phrases**. They are endocentric, though they may contain multiple "heads" (in which case they are treated like compounds). NUMP is only used when there are multiple words in the same number, when the numbers are modified in some way, or when the numbers have been separated from their head. Complex numbers are not given any internal structure, except when involving conjunction.

```

( (IP-MAT (CODE <C>)
      (ADVP-TMP (ADV Thuo-tho))
      (BEDI^3^SG uuas-wesan)
      (NP-SBJ (PRO^N^3^SG siu-siu))
      (NP-PRD (N^N^SG uuidua-widowa))
      (PP (P after-aftar)
          (NP (D^D^SG thiu-the))))
  (CODE <R_513>)
  (PP (P at-at)
      (NP (D^D^SG them-the)
          (N^D^SG friduuiihe-frithuwih)))
  (CODE <C>)
  (NP-TMP (NUMP (NUMP (NUM^A^PL fiuuar-fiuwar)
                     (CONJP (CONJ endi-endi)
                             (NUMP (NUM^A^PL ahtoda-ahto))))
          (CODE <R_514>)
          (NP-POS (N^G^PL uuintro-wintar)))
  ...
(ID OSHeliandC.272.512-516))

( (IP-MAT (CODE <C>)
      ...
      (CP-ADV (C antthat-untat)
              (IP-SUB (ADVP-LOC (ADV thar-thar))
                      (NP-SBJ (N^N^PL uueros-wer))
                      (ADVP-DIR (ADV ostan-ostan))
                      (CODE <R_542>)
                      (NP-SBJ-PRN (ADJP (ADV suitho-switho)
                                       (ADJ^N^PL glauua-glau))
                                  (N^N^PL gumon-gumo)
                                  (NUMP *ICH*-1))
                      (CODE <C>)
                      (IP-INF-NCO (VB gangan-gangan))
                      (VBDI^3^PL quamun-kuman)
                      (CODE <R_543>)
                      (NUMP-1 (NUM^N^PL threa-thrie))
                      (PP (P te-te)
                          ...

```

```

(NP (D^D^SG thero-the)
(N^D^SG thiodo-thiod))
...
(. :-:))
(ID OSHeliandC.284.540-544))

```

The number *en* ("one") is always tagged as a number, even when it means "alone".

```

(NP-SBJ (NPR^N^SG god-god)
(NUM^N^SG eno-en))

```

(OSHeliandC.1234.2322-2324)

## Other constituents

### *Adverb Phrases (ADVP)*

Adverb phrases are a straightforward endocentric phrasal category. They may take the extended labels -LOC (locative), -DIR (directional) or -TMP (temporal). They normally consist of an adverb on its own, potentially with another adverb modifying it (usually *switho*, *so*, *than*, or with temporal adverbs *lango*). Adverbs other than these are, as a rule, separated into different ADVPs even when adjacent and functionally equivalent.

```

( (IP-MAT (CODE <C>)
(NP-SBJ (D^N^SG that-the))
(BEPI^3^SG is-wesan)
(ADVP-TMP (ADV noh-noh))
(ADVP-TMP (ADV lango-lango))
(ADJP-PRD (ADJ^N^SG scin-skin))
(, ,-,)
(CODE <MS_21b>)
(CODE <R_648>)
(PTP-PRD-PRN (GE+VN gicuthid-kwethan)
(NP-SBJ (N^N^SG craft-kraft)
(NP-POS (NPR^G^SG godes-god))))
(. .-.))
(ID OSHeliandC.342.647-648))

```

Instrumental determiners and adjectives are taken as modifiers, as with **ADJPs**.

```

( (IP-MAT (CODE <C>)
(VBDI^3^PL truodun-truon)
(NP-SBJ (PRO^N^3^PL sia-he))
(ADVP-TMP (ADV sithor-sithor))
(CODE <R_2070>)
(ADVP (D^I^SG thiu-the)
(ADVR mer-mer))
(PP (P an-an)
(NP (PRO$^A^3^SG is-is)
(N^A^SG mundburd-mundburd)))
...
(. .-.))
(ID OSHeliandC.1073.2069-2071))

( (IP-MAT (CODE <C>)
(NP-SBJ (N^N^PL erlos-erl))
(VBDI^3^PL thrungun-thringan)

```

```

(CODE <R_182>)
(ADVP-DIR (ADVR nahor-nah)
           (ADJ^I^SG mikilu-mikil))
(. :-:))
(ID OSHeliandC.85.181-182))

```

▣ Adverbs are not treated as taking complements in HeliPaD. Their apparent complements (e.g. the complement of *nah* and *giliko* below) are parsed as clausal adjuncts or arguments.

```

( (IP-MAT (CODE <R_2381>)
          (CONJ ac-ak)
          (VBDI^3^SG geng-gangan)
          (NP-OB2-RFL (PRO^D^3^SG im-he))
          (ADVP-TMP (ADV thuo-tho))
          (NP-SBJ (NP-SBJ (D^N^SG thie-the)
                        (ADJ^N^SG guodo-god))
                 (CODE <C>)
                 (CONJP (CONJ endi-endi)
                        (NP-SBJ (PRO$^N^3^PL is-is)
                                (N^N^PL iungron-jungaro))))
          (PP (P mid-mid)
              (NP (PRO^D^3^SG im-he)))
          (, , -,))
  (CODE <R_2382>)
  (NP-SBJ-PRN (N^N^SG fridubarn-frithubarn)
              (NP-POS (NPR^G^SG godes-god)))
  (CODE <C>)
  (NP-LOC (D^D^SG them-the)
          (N^D^SG fluode-flod))
  (ADVP-DIR (ADVR nahor-nah))
  (CODE <R_2383>)
  (PP (P an-an)
      (NP (NUM^A^SG en-en)
          (N^A^SG scip-skip))
      (ADV innan-innan))
  (. , -,))
(ID OSHeliandC.1270.2381-2383))

( (IP-MAT-SPE-0 (CODE <R_2603>)
                (ADVP-LOC (ADV thar-thar))
                (MDPI^3^PL sculun-skulan)
                (NP-SBJ (PRO^N^3^PL sia-he))
                (PTP-ADT (GE+VNI^N^PL gibundana-bindan))
                (CODE <C>)
                (NP-OB1 (ADJ^A^SG bittra-bittar)
                       (N^A^SG logna-logna))
                (, , -,))
  (CODE <R_2604>)
  (NP-OB1-PRN (N^A^SG thrauerc-thrawerck))
  (VB tholon-tholon)
  (, , -,))
  (CODE <C>)
  (IPX-MAT-PRN-SPE=0 (CONJ endi-endi)
                    (CODE <MS_73a>)
                    (NP-SBJ (D^N^PL thia-the)
                            (ADJ^N^PL odra-othar))
                    (NP-OB1 (N^A^SG thioduelon-thiodwelo))
                    (CODE <R_2605>)
                    (PP (P an-an)
                        (NP (N^D^SG hebanriekie-hevanriki)))
                    (, , -,))

```



```

(CODE <C>)
(NP-OB2 (ADJ^D^SG huitero-hwit)
        (N^D^SG sunnun-sunno))
(CODE <P_180>)
(CODE <R_2606>)
(VB luhtian-liuhtian)
(ADVP (ADV gilico-giliko)))
(. .-.))
(ID OSHeliandC.1405.2603-2606))

```

### *Prepositional Phrases (PP)*

Prepositional phrases in the HeliPaD are an endocentric category that take NPs (unmarked for function) as well as some types of CPs and ADVP-LOC as their complement.

```

(PP (P ti-te)
    (NP (D^D^SG them-the)
        (N^D^SG giuuirkie-giwerki)))

```

(OSHeliandC.7.19-20)

```

(PP (P neuan-newan)
    (CP-THT (C that-that)
            (IP-SUB (NP-SBJ (PRO^N^3^PL sia-he)
                            (NUM^N^PL fiori-fiuwar))
                    (PP (P te-te)
                        (NP (D^I^SG thio-the)))
                    (CODE <R_17>)
                    (PP (P thuru-thurh)
                        (NP (N^A^SG craft-kraft)
                            (NP-POS (NPR^G^SG godas-god))))
                    (CODE <C>)
                    (GE+VNI^N^PL gecorana-kiosan)
                    (RDDI^3^PL uurdun-werthan))))

```

(OSHeliandC.5.15-17)

☐ In the HeliPaD, subordinating conjunctions are not treated as prepositions.

There are several **multi-word prepositions** in the HeliPad; these are RPs within PPs (preceding the preposition), or, more frequently, ADVs within PPs (usually locative or directional elements following the complement).

```

(PP (RP up-up)
    (P te-te)
    (NP (N^D^SG himile-himil)))

```

(OSHeliandC.797.1484-1491)

```

(PP (P an-an)
    (NP (N^A^SG felis-felis))
    (ADV uppan-uppan))

```

(OSHeliandC.948.1801-1810)

Where the latter class occurs without an NP complement, the ADVP-LOC is taken to be the complement itself.

```
(PP (P fan-fan)
    (ADVP-LOC (ADV ostan-ostan)))
```

(OSHeliandC.298.563-566)

Elements can occur before the P of the PP, in modifying function.

```
(PP (ADV suitho-switho)
    (P an-an)
    (NP (N^D^PL sorogon-sorga)))
```

(OSHeliandC.1514.2801-2802)

In several constructions, the NP or an ADVP-LOC complement can occur before the P. This is most common with *thar* or a personal pronoun.

```
( (IP-MAT (CODE <C>)
    (CONJ endi-endi)
    (NP-SBJ *con*)
    (NP-OB2 (PRO^D^3^SG im-he))
    (NP-OB1 (D^A^PL tha-the)
            (N^A^PL geþa-geva))
    (VBDI^3^PL druogun-dragan)
    (, , -,))
  (CODE <P_50>)
  (CODE <R_674>)
  (NP-OB1-PRN (NP-OB1-PRN (N^A^SG gold-gold))
              (CONJP (CONJ endi-endi)
                    (NP-OB1-PRN (N^A^SG uuihroc-wihrok))))
  (CONJP *ICH*-1))
  (CODE <C>)
  (PP (P bi-bi)
      (NP (NP-POS (NPR^G^SG godes-god))
          (N^D^PL teknon-tekan)))
  (CODE <R_675>)
  (CONJP-1 (CONJ endi-endi)
           (NP-OB1-PRN (N^A^SG mirra-myrra)))
  (PP (ADVP-LOC (ADV thar-thar))
      (P midi-mid))
  (. .-.))
(ID OSHeliandC.361.673-675))
```

```
( (IP-MAT (CODE <C>)
    (ADVP-TMP (ADV Thuo-tho))
    (VBDI^3^SG geng-gangan)
    (PP (NP (PRO^D^3^PL im-he))
        (P tuo-to))
    (NP-SBJ (D^N^SG thie-the)
            (NP-POS (N^G^SG landes-land))
            (N^N^SG uuard-ward))
    (. , -,))
  (ID OSHeliandC.1764.3155))
```

### *Interjection Phrases (INTJP)*

Interjections (INTJ) are dominated by an INTJP when modified or taking a complement. Otherwise they occur alone. Either way, INTJ(P) is itself at the highest phrasal level.

```
( (IP-MAT-SPE (CODE <R_1522>)
```

```

(NP-SBJ *con*)
(VBPS^3^SG quede-kwethan)
(INTJ ia-ja)
(CP-ADV-SPE (C ef-ef)
             (IP-SUB-SPE (NP-SBJ (PRO^N^3^SG it-it))
                          (BEPS^3^SG si-wesan)))
(. , -,))
(ID OSHeliandC.807.1522))

```

Interjections may take a question or *that*-clause as complement.

```

( (QTP (CODE <R_3024>)
    (' '-'))
  (INTJP (INTJ Uuola-wola)
        (CP-THT-SPE (C that-that)
                    (IP-SUB-SPE (NP-SBJ (PRO^N^2^SG thu-thu))
                                (NP-VOC (N^N^SG uuib-wif))
                                (HVPI^2^SG habis-hebbian)
                                (CODE <C>)
                                (NP-OBJ (N^A^SG uuilleon-willio)
                                        (ADJ^A^SG guodan-god))))))
(. !-!))
(ID OSHeliandC.1672.3024))

```

INTJP is usually an endocentric category, but can also dominate other elements - usually *wh*-*"adverbial"* *hwat* (which may well not really be an interjection here, and not merit the following punctuation, but is readily retrievable this way).

```

( (IP-MAT-SPE (CODE <C>)
  (INTJP (WADV Huat-hwe))
  (, , -,)
  (NP-SBJ (PRO^N^1^PL uui-we)
          (QP *ICH*-1))
  (NP-OBJ (D^A^SG thia-the))
  (ADVP-LOC (ADV hier-her))
  (VBPI^1^PL uuitun-witan)
  (QP-1 (Q^N^PL alla-al))
  (. , -,))
(ID OSHeliandC.1426.2654))

```

### *Participle Phrases (PTP)*

Phrases headed by a participle that are not part of the main sequence of verbs (e.g. perfect or future periphrasis or modal construction) are labelled PTP. The participle can be either inflected (e.g. VNI, VGI) or uninflected (e.g. VN, VG). PTP is a formal, not a functional, label. On the whole, PTPs behave like **ADJPs**, and may be marked as adjuncts or predicates (the latter only when modified or in apposition). PPs that are adjacent to an NP they modify are contained within that NP.

```

(CP-ADV (ADVP-LOC (ADV thar-thar))
  (C 0)
  (IP-SUB (NP-SBJ (PRO^N^3^SG hie-he))
        (PP (P an-an)
            (NP (PRO$^D^3^SG is-is)
                (N^D^SG benki-bank)))
        (VBDI^3^SG sat-sittian)
        (CODE <R_2747>)
        (PTP-ADT (NP-ADT (N^I^SG uuinu-win)
                        (GE+VN giuulenkid-wlenkian))))))

```

(OSHelianC.1483.2745-2747)

```

( (IP-MAT (CODE <C>)
  (NP-SBJ (PRO^N^3^SG Hie-he))
  (NEG ni-ne)
  (BEDI^3^SG uuas-wesan)
  (ADVP (ADV thoh-thoh))
  (PP (P mid-mid)
    (NP (N^D^PL sibbeon-sibbia))
    (NP-PRN *ICH*-1)
    (NP-PRN *ICH*-2))
  (ADJP-PRD (ADJ^N^SG bifang-bilang))
  (CODE <R_65>)
  (NP-PRN-1 (N^D^PL auaron-avaro)
    (NP-POS (NPR^G^SG Israheles-Israhel)))
  (CODE <C>)
  (NP-PRN-2 (N^D^SG ediligiburdi-ethiligiburd)
    (, ,-,))
  (CODE <R_66>)
  (PTP-PRD-PRN (VN cuman-kuman)
    (PP (P fon-fan)
      (NP (PRO$^D^3^SG iro-iru)
        (N^D^SG cnuosle-knosal))))
    ...
    (. .-.))
  (ID OSHeliandC.20.64-72))

```

PTPs have a sentential internal syntax, and may therefore take arguments, adjuncts, etc.

```

(PTP-ADT (NP-ADT (N^G^SG tionon-tiono))
  (VN atomid-atomian))

```

(OSHeliandC.555.1014-1018)

```

(PTP-PRD-PRN (GE+VN gicuthid-kwethan)
  (NP-SBJ (N^N^SG craft-kraft)
    (NP-POS (NPR^G^SG godes-god)))

```

(OSHeliandC.342.647-648)

### *Foreign phrases (LATIN)*

Foreign phrases, including isolated Latin titles, are labelled with the exocentric phrasal category LATIN and treated as a separate token. These phrases usually only include foreign words (FW).

```

( (LATIN (CODE <R_1600>)
  (FW Pater-pater)
  (FW noster-noster)
  (. :-:))
  (ID OSHeliandC.848.1600))

```

### *Wh-phrases (W\*P)*

There are in principle five types of *wh*-phrase, each corresponding to a non-*wh*-phrase: WNP, WQP, WPP, WADJP, WADVP. These phrases may be endocentric (i.e. headed by a *wh*-word), but need not be. See the [POS manual](#) and [Treatment of individual words](#) for details of how *wh*-words are tagged.

WQP would be used in instances of a quantifier modified by a *wh*-word (e.g. *hwo manag*), but there are no such examples in the *Heliand*.

In relative clauses, the head of a *wh*-phrase is a demonstrative or *thar*, not a *wh*-word. Free relatives allow both types. *Wh*-phrases may also be empty.

(WNP-SBJ-2 (D^N^PL thia-the))

(OSHeliandC.3.9-12)

(WADV-LOC-1 (ADV thar-thar))

(OSHeliandC.129.249-252)

(WNP-PRD-1 (WPRO^N huat-hwe))

(OSHeliandC.105.210-211)

(WNP-OB1-5 0)

(OSHeliandC.9.32-42)

☐ Morphological *wh*- elements need not project a *W\*P* if they are not used as part of an extraction structure. This differs from their treatment in the YCOE, where they are treated as (for instance) plain quantifiers in this situation.

(NP-OB1 (NP-POS (ADJ^G^SG godcundeas-godkund))  
(WPRO^A huat-hwe))

(OSHeliandC.88.187-190)

(NP-SBJ (WADJ^N^SG huilik-hwilik)  
(ADJ^N^SG helag-helag)  
(N^N^SG man-man))

(OSHeliandC.283.537-540)

## 6. HeliPaD and the Penn Historical Corpora

- Introduction
- General differences
- Morphological differences
  - Major differences
  - Minor differences
- Syntactic differences
  - Major, following Penn
  - Major, different policy
  - Minor differences

### Introduction

This page is for experienced users of other Penn historical corpora such as the YCOE, PPCME2, IcePaHC, ENHG Parsed Corpus, etc. If you're a novice user of such corpora, you'd do better to consult the full **POS annotation manual** and **syntactic annotation manual**. On the whole, the HeliPaD closely follows the conventions of the **York-Toronto-Helsinki Parsed Corpus of Old English Prose (YCOE)**.

☞ In the HeliPaD manual, where the HeliPaD follows other Penn historical corpora, the text will be marked like this.

☞ Where the HeliPaD does its own thing, on the other hand, the text will be marked like this.

This page is a one-stop shop for the differences between the HeliPaD and the YCOE.

### General differences

☞ The corpus is in UTF-8, and thus, as in the IcePaHC but unlike in the YCOE, special characters such as barred b and d are not given any special annotation: they are simply present in the text. Like the IcePaHC, the HeliPaD is also **lemmatized**, with the lemma given after the word form and separated by a hyphen.

☞ The corpus contains specific **textual and metrical annotation**, which, however, follows the general principles of the Penn corpora: these annotations are enclosed in angle brackets and receive the POS-tag CODE.

### Morphological differences

#### *Major differences*

☞ The most major difference between the HeliPaD and the YCOE is that the HeliPaD makes much more extensive use of attributes. In the YCOE, **case** is the only attribute that is regularly annotated, and is indicated by means of a caret delimiter (e.g. N^D for a dative noun). The HeliPaD extends the logic of this approach and adds **person** and **number**. Where the three attributes co-occur, case precedes person precedes number. All nominal elements are annotated for case and number; pronouns are also annotated for person, e.g. *ik* would be PRO^N^1^SG. All finite verbs are annotated for person and number. For full details see **Additional attributes**. The overall approach is "maximalist": any element that can receive attributes must receive them, even where there is formal ambiguity. Ambiguity mostly relates to case, and the hierarchy of preference N over A over D over G over I is followed. Where there is number ambiguity, SG is preferred over PL is preferred over DU.

☒ In connection with the attributes, the HeliPaD has an extra tag for each type of participle when it is inflected: VNI alongside VN, VGI alongside VG, etc. These forms take attributes exactly as adjectives do. If it is possible to treat an adjective as a participial form, I have done so. Formally uninflected elements are treated as VN, VG etc.; only participles with agreement endings are tagged as VNI, VGI etc.

☒ The Penn corpora treat many subordinators as prepositions, with corresponding levels of structure. There is no rationale for this in Old Saxon, where prepositions and subordinators are almost completely distinct classes. Those subordinators that are homophonous with adverbs are tagged as such, and form ADVPs in SpecCP. Other subordinators are treated as **C elements**. Three P elements - *butan*, *newa* and *newan* - can take a CP complement, though *newa* is the only one that usually introduces clauses without *that*, and is usually treated as C.

☒ Morphologically *wh*- elements such as *wh*-indefinites like *hwilik* are tagged using the W\* tags even when they are not part of an extraction structure. Since Old Saxon is particularly flexible in using *wh*-words as indefinites, this is quite important. Syntactic role is disambiguated at phrasal level.

### *Minor differences*

☒ Forms of *werthan* are tagged RD\*/RG\*/RN\*, as in the IcePaHC and ENHG Parsed Corpus.

☒ Proper nouns are tagged as NPR (not NR as in the YCOE).

☒ Adverbs do not bear extended tags ^T, ^L and ^D for temporal, locative and directional, as they do in the YCOE. This information is retrievable from the **phrasal extended label** and from the lemma.

☒ The word *ok* is tagged ALSO (the cognate is ADV in the YCOE). ALSO does not head a phrase, may modify adjectives, and often co-occurs with conjunctions within a CONJP.

☒ Inflected infinitives are not given special treatment, unlike in the YCOE. They can always be retrieved due to their co-occurrence with TO within an IP-INF.

☒ The ambiguity tags VBP and VBD etc., for formally ambiguous indicative/subjunctive/imperative verbs, are not used. (Verb form classification follows Köbler.)

☒ The tag AX\*, for auxiliary verbs, is not used.

☒ The FP (focus particle) and XX (problematic word) tags are not used in the HeliPaD, mainly since there is no call for them in the current material.

☒ The tag RP is closed class, and used for the particles *an*, *to*, *up* and *ut*. It does not occur prefixed to verbs as in the YCOE.

☒ The tag GE has a one-to-one mapping with the prefix *gi*-. It never occurs independently, but always prefixed/cliticized to a verbal form. Nominal *gi*- is not tagged in this way, except in the context of *gihwilik* and *gihwe*.

☒ Unlike in the YCOE, where a weak adjective is used nominally (i.e. without a noun head), it normally retains its adjectival tag.

☒ The adjectives *mikil* and *luttil* are tagged as adjectives, even when they are clearly quantifiers. Cognates in the YCOE and other Penn corpora are treated in the exact opposite way.

☒ Some apparently quantificational elements such as *wiht*, *eowiht* etc. are treated as nouns in the HeliPaD rather than as quantifiers as the corresponding items are in the YCOE.

☒ *al* may be treated as an adverb (ADV) in some cases, particularly when introducing *al* so-clauses.

## Syntactic differences

### *Major, following Penn*

☐ In general, nominal extended labels for arguments in the HeliPaD work like they do in the PPCME2 and IcePaHC and not as they do in the YCOE. The nominal extended labels for arguments are -SBJ, -OB1, -OB2, and -PRD. These replace the YCOE's phrasal case labels, though the two types of object are not used in exactly the same way (see [below](#)). See [Noun phrase extended labels](#) and the following subsections for detail.

☐ Within other phrases, excluding IPs, PTPs and CPs but including noun phrases, NPs are indicated as possessive (NP-POS) if this is their function (as it usually is within NPs themselves), and unmarked otherwise (as is always the case within PPs). This differs from the approach taken in the YCOE, which labels constituents for case. The default treatment of NPs is to attach them as high as possible in the structure: this means treating constituents as NP-ADT rather than arguments of a non-verbal element.

### *Major, different policy*

☐ In the YCOE, nominal appositive constituents are either contained within, or indexed to, the constituent to which they are in apposition. In the HeliPaD, they are instead treated as sisters to that constituent (or indexed to a sisterhood position). This enables a much less cluttered clausal representation, and the intended apposition relations are usually very clear semantically.

☐ Unlike in the YCOE, traces are marked with all the same extended labels as the moved element itself. The extended labels of these traces are identical in every way to those of overt constituents.

☐ The HeliPaD's approach to conjunction differs in three important ways from that of the YCOE and other Penn corpora. First, single-word conjuncts are treated in exactly the same way as other conjuncts. Secondly, any extended labels borne by the root node are inherited by the two conjuncts. Thirdly, shared pre- and post-head modifiers are simply included under the root node. For examples see the section on [Conjunction](#).

☐ The two types of object -OB1 and -OB2 are used, broadly speaking, for accusative objects and for dative objects respectively. With certain verbs, genitive objects can also be either -OB1 or -OB2, depending on the case of the other object. Consult the [treatment of individual words](#) for details. With one verb, *lerian*, both objects can be accusative objects, with the people being taught as -OB2.

☐ When modifiers (for instance, floated quantifiers) are separated from a head with which they agree, these are traced to the head. Unlike in the YCOE, this is the case regardless of whether they are case-marked.

### *Minor differences*

☐ Arbitrary PRO in ECM infinitives is indicated by \*arb\*, as it is in the PPCME2 and IcePaHC (but not the YCOE).

☐ In the HeliPaD, single-word modifiers do not project a phrase, even when they follow the head.

☐ If an NP immediately dominates only a modified modifier, the modifier is treated as the head, and the extra level omitted: for instance, an NP headed by a quantifier that is itself modified by an adverb.

☐ ADJPs can be headed only by adjectives, inflected participles, and possessive pronouns, and not by participle phrases or by quantifiers as in the YCOE.

☐ In the YCOE, the first independent clause following a verb of saying is included in the parse as the complement of the verb of saying, whereas later independent clauses are treated as separate tokens. In the HeliPaD, these clauses are always treated as independent tokens.



- ☒ In the HeliPaD, unlike in the YCOE, the label -LFD is used systematically with all clausal categories, including CP-ADV (e.g. in *if ... then* constructions).
  
- ☒ Raising to subject is not usually explicitly represented in the HeliPaD.
  
- ☒ For IPs, the extended label -SUB-CON is not used in the HeliPaD, as it is redundant and all instances of conjoined subordinate clauses can be retrieved in other ways. The label -ABS (for infinitival absolutes) is also not used.
  
- ☒ For CPs, the HeliPaD does not include -CAR (clause-adjoined relatives), -CLF (clefts), -EXL (exclamatives), and -EOP (gapped infinitival relative/purpose clauses), as these structures are essentially not found in the *Heliand*.
  
- ☒ In the HeliPaD, non-argumental *that*-clauses (e.g. with resultative meaning) are still labelled CP-THT and not CP-ADV. The HeliPaD is more liberal than the YCOE in its use of CP-THT, which is used broadly for any clause introduced by *that*, as well as some in the scope of negation introduced by *ne*, unless they are instances of **CP-DEG**. CP-THT is thus essentially a formal rather than functional label. CP-THT does not need to be the complement of a verb or adjective, and is usually unindexed at IP level.
  
- ☒ Purpose clauses headed by a genitive demonstrative are not treated as CP-ADV but as CP-FRL-ADT.
  
- ☒ Titles are never labelled as separate appositive phrases in the HeliPaD, whether modified or not.
  
- ☒ NP-COM is not used, and nouns are assumed not to have nominal complements. NP-POS or clause-level adjuncts take over much of the work of this label in the HeliPaD.
  
- ☒ True case attraction is never annotated at phrase level (since the HeliPaD does not have phrase-level case labels), but there may be a mismatch between (gap) grammatical function and word-level case if a treatment as apposition would lead to a relative clause with both an empty complementizer and an empty operator.
  
- ☒ Adverbs are not treated as taking complements in HeliPaD. Their apparent complements are parsed as clausal adjuncts or arguments.

## 7. Known issues

- General
- Morphology
- Syntax
- Specific tokens

On this page a record will be kept of issues that are raised with the corpus and that will be fixed in the next version. If you spot an error, please **contact me!**

### **General**

No issues at present.

### **Morphology**

No issues at present.

### **Syntax**

No issues at present.

### **Specific tokens**

No issues at present.