

## TOWARD THE MORPHO-SYNTACTIC ANNOTATION OF AN OLD ENGLISH CORPUS WITH UNIVERSAL DEPENDENCIES

Javier Martín Arista

Universidad de La Rioja, Spain

**Abstract:** *The aim of this article is to take the first steps toward the compilation of a treebank of Old English compatible with the framework of Universal Dependencies (UD). Such a treebank will comprise morphological and syntactic annotation of Old English texts adequate for cross-linguistic comparison, diachronic analysis and natural language processing. The article, therefore, engages in four tasks: (i) identifying the Old English exponents of UD lexical categories; (ii) selecting the Old English exponents of UD morphological features; (iii) finding the areas of Old English morphology that require token indexing in the UD format; and (iv) checking on the relevance of the universal set of dependency relations. The data have been extracted from ParCorOEv2, an open access annotated parallel corpus Old English-English. The main conclusions are that the annotation format calls for two additional fields (gloss and morphological relatedness) and that enhanced dependencies are required in order to account for some syntactic phenomena.*

**Keywords:** *Universal Dependencies, Treebanks, Syntactic Annotation, Old English.*

### 1. AIMS, SOURCES AND OVERVIEW

The aim of this article is to lay the foundations of the compilation of a treebank of Old English based on the Universal Dependencies (hereafter UD) descriptive framework and data storage format. The application of UD to Old English consists of two main steps, the first one focussed on the categorial level and the second on the relational level. At the categorial level, the steps taken in this article consist of three tasks, namely identifying the Old English exponents of UD lexical categories; selecting the Old English exponents of UD morphological features; and finding the areas of Old English morphology that require token indexing in the UD format. At the relational level, the relevance of the universal set of dependency relations is discussed. This involves the additional task of determining whether or not enhanced dependencies are necessary for the morpho-syntactic annotation of Old English.

With respect to the sources, the identification of the exponents of lexical categories and the selection of the relevant features from the universal set draw on Old English grammars (Campbell 1987; Hogg and Fulk 2011) and studies in Old English morphology such as Ringe and Taylor (2014). The data, including the lemma list and the POS (part-of-speech) tags when relevant, have been extracted from ParCorOEv2, an open access annotated parallel corpus Old English-English (Martín Arista et al., 2021). Additional information has been retrieved from the knowledge base that automatically annotates ParCorOEv2.

The relevance of the undertaking lies in the lack of a lemmatised corpus of Old English and, above all, in the lack of an analysis of dependencies. The corpora of reference of Old English are unlemmatised. This includes *The Helsinki Corpus of English Texts* (300,000 words; Rissanen et al. 1991), *The York-Helsinki Parsed Corpus of Old English Poetry* (50,000 words; Pintzuk and Plug 2001), *The York-Toronto-Helsinki Parsed Corpus of Old English Prose* (1,500,000 words; Taylor et al. 2003) and *The Dictionary of Old English web corpus* (3,000,000 words; Healey et al. 2004). The only lemmatised corpus of Old English is ParCorOEv2. An open access annotated parallel corpus Old English-English (Martín Arista et al., 2021), but its second version comprises 110,000 words only. *The York-Helsinki Parsed Corpus of Old English Poetry* (Pintzuk and Plug 2001) and *The York-Toronto-Helsinki Parsed Corpus of Old English Prose* (Taylor et al. 2003) have POS tagging and are parsed for the syntax, although they provide a constituency annotation based on theoretical insights from the 1980's.

**To cite this article:** Martín Arista, J. (2022). "Toward the morpho-syntactic annotation of an Old English corpus with Universal Dependencies". *Revista de Lingüística y Lenguas Aplicadas*, 17, 85-97. <https://doi.org/10.4995/rlyla.2022.16787>

**Correspondence author:** javierma76@gmail.com



With these premises, the article is structured as follows. Section 2 summarises the main aspects of the UD framework and briefly reviews its theoretical basis. Section 3 presents the method of this research, which is geared to the adaptation of the Old English data to the storage format of UD. Section 4 starts the application of UD to Old English by selecting the relevant lexical classes from the universal set of categories adopted by UD. Section 5 identifies the relevant features from the UD morphological inventory and adapts them to the data from this historical stage of the English language. Section 6 presents the morphological phenomena that require token indexing in Old English, namely verbal contractions and compounding. Section 7 discusses the relevance of dependency relations to Old English and offers an initial version of enhanced dependencies. Section 8 draws the conclusions of the article. The inventories that result from the discussion are listed in the Appendix.

## 2. BACKGROUND

Universal Dependencies (De Marneffe et al. 2021) is a model of morphological and syntactic annotation devised for the compilation of computerised data sets (or *treebanks*; Jurafsky and Martin *fc.*) that facilitate cross-linguistic comparison (De Marneffe et al. 2014) aimed to natural language processing (Nivre 2015) and to areas of applied linguistics like language acquisition and translation (MacDonald et al. 2013; Nivre 2016). The annotation includes UPOS (universal part-of-speech tags; Petrov et al. 2012), XPOS (language-specific part-of-speech tags), Feats (universal morphological features), lemmas, and dependency heads and labels (Nivre et al. 2016). The 2015 release of the Universal Dependencies (hereafter UD) dataset consisted of ten treebanks representing ten languages, whereas the 2021 release comprises 183 treebanks over 104 languages (Nivre et al. 2020).

While the descriptive procedure of UD resorts to graph theory (Nilsson et al. 2006), its theoretical basis draws on Dependency Grammar, on Linguistic Universals and cross-linguistic typology as well as on the models that resulted from the generative semantics controversy.

In Dependency Grammar (Tesnière 1959, 2015; Mel'čuk 1988; Mel'čuk et al. 2016), clause structure is explained by means of hierarchically arranged horizontal and vertical relations that involve governors (or *dependency heads*) and their dependents. In the classical view, the concept of *dependency* (or *determination*; Böhmová et al. 2003) can be defined as a relation ruled by valence restrictions that stipulate the number and type of dependents which can depend on a certain governor. In recent developments of the theory (Kübler and Zinsmeister 2015: 57), emphasis is made on the differences between dependency and constituency: dependency is concerned with the semantic-syntactic relations that link a given dependent to its governor, as well as a certain governor to a governor of a higher order (typically, a nominal predicate and a verbal predication); whereas constituency describes the gradual expansion and geometry of syntactic projections comprised of syntactic categories (phrases headed by certain lexical classes) that contain the minimal constituents representing the final step of tree diagram analysis. Constituent trees may represent null elements whereas dependency trees don't.

UD has a strong typological foundation, as it is directly linked to the Linguistic Universals research programme (Greenberg 1966) and linguistic typology as represented in the works by Foley and Van Valin (1984), Van Valin and LaPolla (1997), Foley (2007), Croft et al. (2017) and Haspelmath (2019). UD opts for a simplified syntactic architecture that maximises the points of convergence across languages rather than a fine-grained theoretical apparatus whose cross-linguistic applicability may not be so direct. UD, for instance, does not incorporate the verb phrase node to its representations, thus avoiding certain language-specific and theory-internal issues.

Generative Semantics postulated a semantic abstract level of sentence representation (Lakoff 1971) in contradistinction to the canonically syntactic deep structure of mainstream Generative Grammar (Chomsky 1970). A group of linguistic theories originated in the Generative Semantics controversy which partly inspire UD representations. This includes linguistic theories that focus on clausal relations and diathesis alternations, such as Lexical-Functional Grammar (Kaplan and Bresnan 1982), Relational Grammar (Perlmutter 1983), and Word Grammar (Hudson 1984); linguistic theories concerned with deep semantic roles like Case Grammar (Fillmore 1968), as well as with generalised semantic roles, such as Role and Reference Grammar (Van Valin & LaPolla 1997); and linguistic theories of a functional-typological persuasion, such as Functional Grammar (Dik 1997a, 1997b) that distinguish quantitative and qualitative valencies that focus on the semantic relations holding between the predicate, on the one hand, and the head of the noun phrase, and their arguments, adjuncts and modifiers, on the other.

Against this background, the website of UD (<https://universaldependencies.org/introduction.html>) summarises the crucial aspects and developments of the model as follows.

The dependency representation of UD evolves out of Stanford Dependencies (SD), which itself follows ideas of grammatical relations-focused description that can be found in many linguistic frameworks. That is, it is centrally organized around notions of subject, object, clausal complement, noun determiner, noun modifier, etc. The goal

of the new universal version was to add or refine relations to better accommodate the grammatical structures of typologically different languages and to clean up some of the quirkiest and more English-specific features of the original version.

In other words, the framework of UD relies on hard-core linguistic theory but avoids theory-centered debates and local phenomena not only to improve its typological adequacy but also to avoid biases that can be put down to Anglo-centrism or Euro-centrism. The following sections apply the UD framework as reviewed above to Old English. As has been said above, this application constitutes a significant update of Old English treebanks, which currently include *The York-Helsinki Parsed Corpus of Old English Poetry* (Pintzuk & Plug 2001) and *The York-Toronto-Helsinki Parsed Corpus of Old English Prose* (Taylor et al. 2003), given that they provide constituency annotation and are based on syntactic theory from the 1980's.

### 3. METHOD. FROM THE PARALLEL CORPUS TO THE CONLL-U FORMAT

This section unfolds the methodological procedure of this research, by presenting the sources of data and the target format used for structuring such data.

The 2021 release of *An open access annotated parallel corpus Old English-English* (ParCorOEv2; Martín Arista et al., 2021) comprises 109,985 records. ParCorOEv2 has been implemented in Filemaker database software and published with Filemaker WebDirect. Each token (database file) is provided with a concordance showing the prefield and the postfield of the concorded term, with the translation of the fragment into English, and with the sources of the Old English and the Present-Day English texts (<https://www.nerthusproject.com/corpus-description>). The current records correspond to the following texts (<https://uploads.strikinglycdn.com/files/e7113fe2-7c23-4af8-895e-e55fe14d1a2c/ParCorOEManual1-Sources.pdf?id=3387163>): BOET (23,972 records), HERB (10,863 records), LACN (8,901 records), LEEC (11,883 records), MARK (12,036 records), MART (24,355 records), OROS (13,638 records) and QUAD (4,337 records). Another 240,000 records are currently in preparation, with which ParCorOEv3 is expected to file 250,000 records by December 2022. As can be seen in Figure 1, each record displays one token (an inflectional form) and its tagging (file number, lemma, lexical category, inflectional category and gloss).

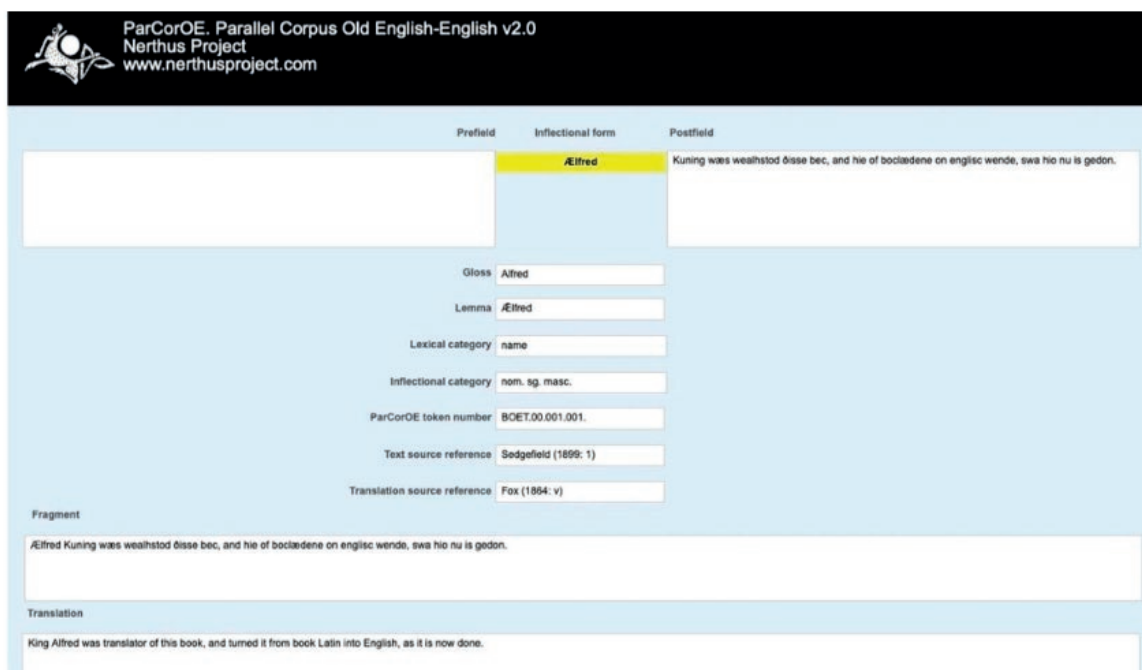


Figure 1. *An open access annotated parallel corpus Old English-English* (ParCorOEv2; Martín Arista et al., 2021).

It is a two-fold undertaking to apply the framework of UD to Old English data retrieved from a parallel corpus annotated with a knowledge base. In the first place, the data available from ParCorOE must conform to the storage format of UD, called *CoNLL-U Format* ([universaldependencies.org](http://universaldependencies.org)). This calls for the tokenization of the corpus at word and token level and the assignment of lemma and POS tag (including category and morphological analysis). In the second place, the periods or sentences made up from sets of tokens must be parsed for dependency relations.

Tokenisation and annotation in CoNLL-U are encoded in plain text files in UTF-8 format. The LF character is used as line break. Three types of lines are distinguished: word lines, which contain the annotation of words and token in fields separated by a tab character; blank lines, which mark sentence boundaries within a sense unit or period; and comment lines, which start with hash (#). Sentences comprise at least one-word line. Each word line contains ten fields that can be seen in Figure 2.

ID	Word index, integer starting at 1 for each new sentence; may be a range for tokens with multiple words.
FORM	Word form or punctuation symbol.
LEMMA	Lemma or stem of word form.
UPOSTAG	Universal part-of-speech tag drawn from the revised version of the Google universal POS tags.
XPOSTAG	Language-specific part-of-speech tag.
FEATS	List of morphological features from the universal feature inventory or from a defined language-specific extension.
HEAD	Head of the current token, which is either a value of ID or zero (0).
DEPREL	Universal Stanford dependency relation to the HEAD (root iff HEAD = 0).
DEPS	List of secondary dependencies.
MISC	Any other annotation.

Figure 2. Fields in CoNLL-U annotation (from <https://universaldependencies.org/docs/format.html>).

The fields in Figure 2 can neither be empty nor contain space characters. The underscore ( \_ ) is used to denote unspecified values in all fields except ID. Annotation segments can belong to either the word or the multi-word type. Multi-word segments consist of two or more tokens. Words are indexed with integers 1, 2, 3, etc., while tokens are indexed with integer ranges which are inserted before the first word in the range. Words and tokens have a FORM (the segment that appears in the sentence). This is illustrated in Figure 3, which presents the first three fields of the English sentence *She's at home*.

ID	FORM	LEMMA
1-2-	She's	_
1	she	she
2	is	be
3	at	at
4	home	home
5	.	.

Figure 3. ID, form and lemma.

The UPOS field specifies a POS tag from the universal set of POS tags (<https://universaldependencies.org/u/pos/index.html>). The FEATS field provides a list of morphological features separated by a bar and with underscore to represent an empty set. Features are represented as attribute-value pairs with an equal sign ( = ) separating attributes from values. Features are picked from the universal feature inventory and sorted alphabetically. As is shown in Figure 4, morphological annotation is only provided for words.

ID	FORM	LEMMA	UPOSTAG	FEATS
1-2	She's	–	–	–
1	she	she	PRON	Case=Nom   Gender=Fem   Number=Sing   Person=3
2	is	be	VERB	Number=Sing   Person=3   Tense=Pres
3	at	at	ADP	–
4	home	home	NOUN	Number=Sing
5	.	.	PUNCT	–

Figure 4. ID, form, lemma, category and morphological features.

Two extra specific fields are required in order to adequately annotate the morphology and syntax of Old English. The first is GLOSS. As we are annotating a historical language, a translation into Contemporary English facilitates the annotator's task. Inflectional forms, lemmas and glosses are automatically imported from ParCorOEv2, or the knowledge base, which relates canonical and non-canonical (alternative spellings) inflectional forms to their corresponding lemma. For example, the inflectional forms *leasung*, *leasunga*, *leasungæ*, *leasunge*, *leasungan* and *leasungum* are attributed to the noun lemma *lēasung* and glossed as 'leasing'.

The second specific field proposed in this paper is MORPHREL (morphological relatedness). In a language characterised by the existence of large derivational families with transparent morphological relations (Kastovsky 1992) and generalised inflectional inheritance of the prefix *ge-* (Martín Arista 2012), a paradigmatic field specifying short-distance and long-distance morphological relatedness constitutes a remarkable explanatory resource. For instance, the derived adjective *unābrecedlic* 'inextricable' is morphologically related (short-distance) to the adjective *\*ābrecedlic* as well as to the primitive strong verb BRECAN 'to break, tear, crush, shatter, burst, break up, destroy, demolish' (long-distance morphological relatedness). These facts are indicated in the MORPHREL field as *\*ābrecedlic* / BRECAN.

Once the relevant fields of the CoNLL-U format have been defined and illustrated, the next step in the research methodology is to establish the correspondence between the fields in the parallel corpus (or the knowledge base on which it is largely based) and the fields in the CoNLL-U format. Although the correspondence is not direct because some fields need editing (morphological features, for instance, should be adapted to the universal inventory), the data flow could be represented as can be seen in Figure 5.

PacCorOE	KB
Word indexing	
Inflectional form >>> FORM	Inflectional form >>> FORM
Lemma >>> LEMMA	Lemma >>> LEMMA
Lexical category >>> UPOSTAG	
Inflectional category >>> FEATS	
	Base of derivation, lexical prime >>> MORPHREL
Token indexing	
	Base/adjunct of compounding >>> LEMMA
	Inflectional form >>> FORM

Figure 5. Tagging from the parallel corpus and the knowledge base.

It must be noted that token indexing must be imported from the knowledge base, as the parallel corpus does not distinguish units below the syntactic word. The lemmas and inflections not found in the parallel corpus are also drawn from the knowledge base.

#### 4. UNIVERSAL PART-OF-SPEECH TAGS AND THEIR OLD ENGLISH EXPONENTS

This section turns to the categorial aspects of the application of UD to Old English. Figure 6 compares the traditional set of parts of speech and the universal parts-of-speech set of UD. It also presents the corresponding Old English exponents. Given that there is no full coincidence between the Old English exponents and the UPOS tags, the exponents will be shown in the XPOS tag column of the CoNNL-U format tabulation.

Standard POS	UPOS	OE exponent	OE realisations
Noun	noun	common noun	<i>hlāfweard</i> ‘steward’, <i>mūða</i> ‘mouth (of a river)’, <i>sæcol</i> ‘jet’
	propn	proper noun	<i>Egipste</i> ‘the Egyptians’, <i>Iringes weg</i> ‘Milky Way’, <i>Legaceaster</i> ‘Chester’
Verb	verb	main verb	<i>sleacian</i> ‘to slow’, <i>ðoterian</i> ‘to cry’, <i>twengan</i> ‘to pinch’
	aux	auxiliary verb	<i>bēon</i> ‘to be’, <i>habban</i> ‘to have’, <i>weorðan</i> ‘to become’
Adjective	adj	adjective	<i>gnēað</i> ‘frugal’, <i>inwit</i> ‘thick’, <i>meagol</i> ‘mighty’
–	det	demonstrative-article	<i>se</i> ‘the’
	num	numeral (cardinal and ordinal)	<i>ðrie</i> ‘three’, <i>fēowertiene</i> ‘fourteen’, <i>hundseofontigoða</i> ‘seventieth’
Adverb	adv	adverb	<i>grundlinga</i> ‘horribly’, <i>hedendlīce</i> ‘strictly’, <i>hindanweard</i> ‘from behind’
Pronoun	pron	pronoun	<i>sumhwilc</i> ‘some’, <i>ðu</i> ‘you’, <i>ðin</i> ‘your’
Preposition	adp	adposition	<i>betweox</i> ‘between’, <i>gēan</i> ‘against’, <i>tōweard</i> ‘towards’
Conjunction	cconj	coordinating conjunction	<i>and</i> ‘and’, <i>ge</i> ‘and also’, <i>oððe</i> ‘or’
	sconj	subordinating conjunction	<i>hwær</i> ‘where’, <i>ðēah</i> ‘although’, <i>ðy</i> ‘because’
Interjection	intj	interjection	<i>æ</i> ‘oh!’, <i>ēuwā</i> ‘wow!’, <i>nū</i> ‘lo!’
–	part*	–	–
	x	foreign word	<i>meretrix</i> ‘prostitute’, <i>silua</i> ‘forest’, <i>torre</i> ‘tower’
–	sym	–	–
–	punct	punctuation	.,,: ? !***

Figure 6. UPOS tags and Old English exponents.

With respect to Figure 6, the particle (part) is a single word marker in some languages, but is not present in Old English. A non-punctuation symbol is a hash (#) or an emoji, neither of which is relevant to Old English. As regards punctuation, it must be recalled that the punctuation of Old English is not consistent.

#### 5. UNIVERSAL MORPHOLOGICAL FEATURES AND THEIR OLD ENGLISH EXPONENTS

Figure 7 displays the universal features of UD, along with the corresponding exponent and value of Old English. It must be noted that the nominal features case, gender and number are also found in verbs because present and past participles can be inflected as adjectives in Old English. Another question worth mentioning is the reflexive. The reflexive with *self* has not been grammaticalised in Old English yet, so that the reflexive is expressed by means of the corresponding personal pronoun, inflected for the accusative or the dative.

Feature		Values
pronominal type	PronType	Dem-Art Ind Int Prs Rel
numeral type	NumType	Card Ord
possessive	Poss	Yes
foreign word	Foreign	Yes
abbreviation	Abbr	Yes
wrong spelling	Typo	Yes
gender	Gender	Fem Masc Neut
animacy	–	–
noun class	–	–
number	Number	Dual Plur Sing
case	Case	Acc Nom Dat Gen Ins
definiteness	–	–
reflexive	Reflex	Yes
comparison	Degree	Cmp Pos Sup
person	Person	1 2 3
verbal form	VerbForm	Fin Inf Part
mood	Mood	Imp Ind Sub
tense	Tense	Past Pres
aspect	–	–
voice	Voice	Act Mid Pass
polarity	Polarity	Neg
person	Person	1 2 3
politeness	–	–
clusivity	–	–
gender	Gender	Fem Masc Neut
number	Number	Dual Plur Sing
case	Case	Acc Nom Dat Gen In
invariable form	Uninflected	Yes

Figure 7. Universal morphological features, Old English exponents and values.

The following abbreviations are used in Figure 7: 1 (first person), 2 (second person), 3 (third person), Acc (accusative), Act (active), Art (article), Card (cardinal), Cmp (comparative), Dat (dative), Dem (demonstrative), Dual (dual), Fem (feminine), Fin (finite), Gen (genitive), Imp (imperative), Ind (indefinite), Ind (indicative), Inf (infinitive), Ins (instrumental), Int (interrogative), Masc (masculine), Mid (middle), Neut (neuter), Nom (nominative), Ord (ordinal), Part (participle), Pass (passive), Past (past), Plur (plural), Pos (positive), Pres (present), Prs (personal), Refl (reflexive), Rel (relative), Sing (singular), Sub (subjunctive), Sup (superlative).

## 6. TOKEN INDEXING

Word indexing requires the segmentation of fragments into periods, the separation of punctuation marks and the insertion of the blank lines that divide a given period from the previous and the following ones. This can be imported automatically from ParCorOEv2 but requires manual revision because the punctuation of Old English is not always consistent. Token indexing is in point when dealing with complex words that result from compounding two free lexical forms. Derivatives, which involve a free and a bound form are annotated as words only. This limits

the scope of token indexing to contractions and compounds. Contractions can be broken down into a negative grammatical word and a content word, as *nyllað* (*ne-willan*), *nys* (*ne-bēon*) and *nysstest* (*ne-wītan*). Compounds are imported from the knowledge base, as can be seen in Figure 8, by base and adjunct, including the categorial tag of the word and the tokens.

FORM	LEMMA	UPOSTAG	TOKEN 1	UPOSTAG	TOKEN 2	UPOSTAG
upga	ūpgān	verb	ūp	adverb	gān(ge)	verb
syueræte	sýferæte	adjective	sýfre	adjective	æ̆t	noun
eall	ealcyn	adjective	eall	adjective	cynn	noun
upcymð	ūpcuman	verb	ūp	adverb	cuman(ge)	verb
herinne	hērinne	adverb	hēr	adverb	inne	adverb

Figure 8. Compounds as tokens.

## 7. DEPENDENCY RELATIONS

To recapitulate, the tokenisation as well as the fields FORM, LEMMA and GLOSS are imported automatically from ParCorOEv2. Units smaller than a word indexed as tokens (abbreviations and compounds) can be imported from the knowledge base. The fields UPOSTAG and XPOS are adapted from the morphological tags of ParCorOEv2. At this stage, MORPHREL, HEAD and DEPREL are inserted manually. The rest of this section delves into the question of the set of dependency relations and enhanced dependencies that are relevant for Old English.

In the UD framework (De Marneffe et al., 2021), dependency relations are defined by means of dependency heads and labels. Dependency heads are content words, whereas function words do not usually have dependents. The basic distinction is drawn between core arguments (subjects, objects and clausal complements) and oblique modifiers (adjuncts and oblique arguments). Of these, the following universal dependency relations are found in Old English: *acl* (clausal modifier of noun), *acl:relcl* (relative clause modifier), *advcl* (adverbial clause modifier), *advmod* (adverbial modifier), *advmod:emph* (emphasizing word, intensifier), *advmod:lmod* (locative adverbial modifier), *amod* (adjectival modifier); *appos* (appositional modifier), *case* (case marking), *cc* (coordinating conjunction), *cc:preconj* (preconjunct), *ccomp* (clausal complement), *csubj:pass* (clausal passive subject), *conj* (conjunct), *cop* (copula), *csubj* (clausal subject), *det* (determiner), *det:poss* (possessive determiner), *discourse* (discourse element), *dislocated* (dislocated elements), *fixed* (fixed multiword expression), *flat* (flat multiword expression), *flat:foreign* (foreign words), *flat:name* (names), *goeswith* (goes with), *iobj* (indirect object), *list* (list), *mark* (marker), *nmod* (nominal modifier), *nmod:poss* (possessive nominal modifier), *nmod:tmod* (temporal modifier), *nsubj* (nominal subject), *nummod* (numeric modifier), *obj* (object), *obl* (oblique nominal), *obl:agent* (agent modifier), *obl:arg* (oblique argument), *obl:lmod* (locative modifier), *obl:tmod* (temporal modifier), *orphan* (orphan), *parataxis* (parataxis), *punct* (punctuation), *root* (root) and *vocative* (vocative). An illustration of some of the commonest dependencies can be seen in (1). Fragments and fragment numbers have been taken from *The Dictionary of Old English web corpus* (Healey et al. 2004).

(1)

a. Copula

[Bede 2 010100 (4.108.8)]

*Se wæs feorða biscop þære Romaniscan burge from Sancti Gregorii.*

‘He was the fourth bishop of Rome from Saint Gregory.’

b. Object

[Bede 2 007400 (3.104.12)]

*Da wæs æfter ðissum þætte Agustinus Breotone ærcebiscop gehalgade twegen biscopas.*

‘After that, Augustinus, archbishop of Britain, consecrated two bishops.’



## c. Indirect object

[ChronC (O'Brien O'Keeffe) 019000 (755.1)]

*Her Cynewulf benam Sigebrihte his rice & Wæstseaxna witan for unrihtum dædum buton Hamtunscyre.*

‘This year Cynewulf and the witan of the West-Saxons deprived Sigebriht of his kingdom except Hampshire for evil deeds.’

## d. Oblique argument

[ChronA (Bately) 019510 (675.2)]

*& by ilcan geara Wulfhere forþferde, & Eþelræd feng to rice.*

‘This year Wulfhere died and Æthelred inherited the kingdom.’

The dependency relations *aux* (auxiliary), *aux:pass* (passive auxiliary), *nsubj:pass* (passive nominal subject), *expl* (expletive) and *expl:impers* (impersonal expletive) are relevant for the syntactic annotation of Old English, but this calls for some comment. Although there are clear-cut instances of these dependencies in Old English, there are others on the grammaticalisation cline. This includes, at least, pre-auxiliaries, expletives and the passive (Denison 1993; Ringe & Taylor 2014; Petré 2014; Martín Arista et al., 2018). For instance, the passive has not been fully grammaticalised in Old English yet for three reasons. Firstly, there is fluctuation between *bēon* ‘to be’ and *weorðan* ‘to become’ as passive auxiliaries in Old English. Secondly, the past participle is frequently inflected as an adjective and agrees with the subject, as in *Ac heo wæran cumene wið Westseaxna þeode to gefeohte* ‘They accompanied the West-Saxon people into the battle’ (Bede 2 014100 (5.114.2)), in which the masculine plural nominative past participle *cumene* ‘come’ agrees in gender, number and case with the subject *heo* ‘they’. Thirdly, it is still possible for the agent to preserve the dative case in the corresponding passive, as in *And him wæs gedemed fram unrihtwisum demum* ‘And he was judged by folly judges’ (ÆAdmon 1 006000 (4.31)), where the patient is case-marked dative both in the active and the passive (*him*).

In practical terms, the dependency relations *aux* (auxiliary), *aux:pass* (passive auxiliary), *nsubj:pass* (passive nominal subject), *expl* (expletive) and *expl:impers* (impersonal expletive) are distinguished in the annotation of Old English for descriptive reasons, although they are not fully grammaticalised yet.

Nothing has been said so far of the DEPS field in the CoNLL-U format. This field is used for secondary dependencies. In general, null elements are not analysed in the UD framework. However, it is sometimes necessary to make explicit some implicit relations between words. Cross-linguistically, enhanced dependencies are used to account for empty elements, antecedents in relative constructions and arguments shared by the main clause and the subordinate clause in control and raising constructions. Let us consider example (2).

## (2) [Ad 001000 (5.1)]

*Saga me hwær is seo eorðe þe næfre sunne on ne scean, ne mona, ne næfre wind on ne bleow, nane tid dæges ne ær ne æfter.*

‘Tell me where is the earth that never the sun on shines, nor the moon, nor wind ever blows on, in no time of the day, either early or later.’

In Old English, the order of constituents is freer than in Contemporary English because inflection is more explicit. In the context of a relative construction, the relative word *þe* ‘that’ is invariable and cannot mark the agreement with the antecedent explicitly. If, moreover, the preposition is stranded, as *on* in *þe næfre sunne on ne scean*, the relation between the antecedent *eorðe* ‘earth’ and the relative word *þe* ‘that’ on the one hand; and between *on* ‘on’, *sunne* ‘sun’, *monna* ‘moon’ and *wind* ‘wind’, on the other, should be made explicit by means of a secondary dependence.

## 8. CONCLUSIONS AND FURTHER RESEARCH

This article has listed the lexical and morphological categories that should be distinguished in a UD treebank of Old English. It has also identified the dependency relations that are relevant to the syntax of Old English. With respect to the CoNLL-U format, this study has pointed out that two additional fields are required, namely,

GLOSS (a gloss Old English-Contemporary English) and MORPHREL (morphological relatedness). The article has also insisted on the need for enhanced dependencies in order to adequately annotate the morpho-syntax of Old English. Last but not least, the limits of the automatic importation from the parallel corpus used as source have been explored. It remains for future research to determine if additional enhanced dependencies are required and to automatise some data importation tasks.

## ACKNOWLEDGEMENTS

Grant IPID2020-119200GB-100, funded by MCIN/ AEI/10.13039/501100011033/.

## REFERENCES

- Böhmová, A., Hajič, J., Hajičová, E., & Hladká, B. (2003). “The Prague Dependency Treebank”, in A. Abeillé (ed.) *Treebanks: Building and Using Parsed Corpora*. Dordrecht: Kluwer Academic Publishers, 103-127.
- Campbell, A. (1987). *Old English Grammar*. Oxford: Oxford University Press.
- Chomsky, N. (1970). “Remarks on Nominalization”, in R. Jacobs & P. Rosenbaum (eds.) *Readings in English Transformational Grammar*. Waltham, MA.: Blaisdell, 184-221.
- Croft, W., Nordquist, D., Looney, K., & Regan, M. (2017). “Linguistic typology meets Universal Dependencies”, in M. Dickinson, J. Hajic, S. Kübler & A. Przepiórkowski (eds.) *Proceedings of the 15th International Workshop on Treebanks and Linguistic Theories (TLT15)*, 63-75.
- De Marneffe, M.C., Dozat, T., Silveira, N., Haverinen, K., Ginter, F., Nivre, J. & Manning, C. (2014). “Universal Stanford Dependencies: a cross-linguistic typology”, in *Proceedings of LREC 2014*, 4585–4592.
- De Marneffe, M.C., Manning, C., Nivre, J. & Zeman, D. (2021). “Universal Dependencies”. *Computational Linguistics* 47/2, 255–308.
- Denison, D. (1993). *English Historical Syntax*. London: Longman.
- Dik, S.C. (1997a). *The Theory of Functional Grammar. Part 1: The Structure of the Clause*. Berlin: Mouton de Gruyter. Edited by K. Hengeveld.
- Dik, S.C. (1997b). *The Theory of Functional Grammar. Part II: Complex and Derived Constructions*. Berlin: Mouton de Gruyter. Edited by K. Hengeveld.
- Fillmore, C.J. (1968). “The Case for Case”, in E. Bach and R. Harms (eds.) *Universals in Linguistic Theory*. New York: Holt, Rinehart & Winston, 1-88.
- Foley, W. (2007). “A typology of information packaging in the clause”, in T. Shopen (ed.) *Language Typology and Syntactic Description*. Volume I: Clause Structure. Cambridge: Cambridge University Press, 362–446.
- Foley, W. & R.D. Van Valin, Jr. (1984). *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.
- Greenberg, J.H. (1966). “Some universals of grammar with particular reference to the order of meaningful elements”, in J.H. Greenberg (ed.), *Universals of grammar*. Cambridge, Mass.: MIT Press, 73–113.
- Haspelmath, M. (2019). “Indexing and flagging, and head and dependent marking”. *Te Reo* 62/1, 93–115.
- Healey, A. (ed.), Wilkin, J. & Xiang, X. (2004). *The Dictionary of Old English web corpus*. Toronto: Dictionary of Old English Project, Centre for Medieval Studies, University of Toronto.
- Hogg, R.M. & Fulker, R.D. (2011). *A Grammar of Old English. Volume 2: Morphology*. Oxford: Blackwell.
- Hudson, R.A. (1984). *Word Grammar*. Oxford: Blackwell.
- Jurafsky, D. & J.H. Martin. *Speech and Language Processing* (3<sup>rd</sup>. edition). Forthcoming.
- Kaplan, R. & Bresnan, J. (1982). “Lexical-Functional Grammar: A formal system for grammatical representation”, in J. Bresnan (ed.) *The Mental Representation of Grammatical Relations*. Cambridge, Mass.: MIT Press, 173–281.
- Kastovsky, D. (1992). “Semantics and vocabulary”, in R. Hogg (ed.) *The Cambridge history of the English language I: The beginnings to 1066*. Cambridge: Cambridge University Press, 290–408.
- Kübler, S. & Zinsmeister, H. (2015). *Corpus Linguistics and Linguistically Annotated Corpora*. London: Bloomsbury.
- Lakoff, G. (1971). “On generative semantics”, in D.D. Steinberg & L.A. Jakobovits (eds.), *Semantics: An interdisciplinary reader in philosophy, linguistics and psychology*. Cambridge: Cambridge University Press, 232-296.
- Martín Arista, J. (2012). “The Old English prefix *ge-*: A panchronic reappraisal”. *Australian Journal of Linguistics* 32/4, 411–433.

- Martín Arista, J. & Ojanguren López, A.E. (2018). “Grammaticalization and deflexion in progress. The past participle in the Old English passive”. *Studia Neophilologica* 90/2, 155-175.
- Martín Arista, J., Domínguez Barragán, S., García Fernández, L., Ruíz Narbona, E., Torre Alonso, R. & Veá Escarza, R. (comp.). (2021). *ParCorOE2. An open access annotated parallel corpus Old English-English*. Nerthus Project, Universidad de La Rioja, [www.nerthusproject.com](http://www.nerthusproject.com).
- McDonald, R.T., Nivre, J., Quirnbach-Brundage, Y., Goldberg, Y., Das, D., Ganchev, K., Hall, K., Petrov, S., Zhang, H., Täckström, O., Bedini, C., Bertomeu, N. & Lee, J. (2013). “Universal Dependency Annotation for Multilingual Parsing”, in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, 92–97.
- Mel’čuk, I., Bresnan, J., Asudeh, A., Toivonen, I. & Wechsler, S. (2016). *Lexical-Functional Syntax*. Chichester: Wiley-Blackwell.
- Mel’čuk, I. (1988). *Dependency Syntax: Theory and Practice*. State University of New York Press.
- Nilsson, J., Nivre, J., & Hall, J. (2006). Graph transformations in data-driven dependency parsing, in *COLING 21 and ACL 44*, 257–264.
- Nivre, J. 2015. “Towards a universal grammar for natural language processing”, in A. Gelbukh (ed.) *Computational Linguistics and Intelligent Text Processing*. New York: Springer, 3–16.
- Nivre, J. 2016. “Universal Dependencies: A Cross-Linguistic Perspective on Grammar and Lexicon”, in *Proceedings of the Workshop on Grammar and Lexicon: Interactions and Interfaces*, 38-40.
- Nivre, J., de Marneffe, M.C., Ginter, F., Hajič, J., Manning, C., Pyysalo, S., Schuster, S., Tyers, F. & Zeman, D. (2020). “Universal Dependencies v2: An evergrowing multilingual treebank collection”, in *Proceedings of the Twelfth International Conference on Language Resources and Evaluation (LREC 2020)*, 4027–4036.
- Nivre, J., de Marneffe, M.C., Ginter, F., Golberg, Y., Hajič, J., Manning, B., C., Pyysalo, S., Schuster, S., Silveira, Tsarfaty, R. & Zeman, D. 2016. “Universal Dependencies v1: a multilingual treebank collection, in *Proceedings of the 10th International Conference on Language Resources and Evaluation*, 1659-1666.
- Perlmutter, D. (ed.). (1983). *Studies in Relational Grammar*. Chicago: The University of Chicago.
- Petré, P. (2014). *Constructions and environments: Copular, passive, and related constructions in Old and Middle English*. Oxford: Oxford University Press.
- Petrov, S., Das, D. & R. McDonald. 2012. “A Universal Part-of-Speech Tagset”, in *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC’12)*, 2089-2096.
- Pintzuk, S. & Plug, L. (eds.) (2001). *The York-Helsinki Parsed Corpus of Old English Poetry* [<http://www-users.york.ac.uk/~lang18/pcorpus.html>].
- Ringe, D. & Taylor, A. (2014). *The development of Old English. A linguistic history of English*, vol. 2. Oxford: Oxford University Press.
- Rissanen, M., Kytö, M., Kahlas-Tarkka, L., Kilpiö, M., Nevanlinna, S., Taavitsainen, I., Nevalainen, T. & Raumolin-Brunberg, H. (eds). (1991). *The Helsinki Corpus of English Texts*. Department of Modern Languages, University of Helsinki.
- Taylor, A., Warner, A., Pintzuk, S. & Beths, F. (2003). *The York-Toronto-Helsinki Parsed Corpus of Old English Prose* [<https://www-users.york.ac.uk/~lang22/YcoeHome1.htm>].
- Tesnière, L. 2015 (1959). *Elements of Structural Syntax*. Translation by T. Osborne and S. Kahane of Tesnière (1959). Amsterdam: John Benjamins.
- Van Valin, R.D., Jr. & R. LaPolla. (1997). *Syntax: structure, meaning and function*. Cambridge: Cambridge University Press.

## APPENDIX

### 1. Inventories of grammatical categories

- Demonstrative-article

se ‘the’; þes ‘this’.

- Relative pronoun

ðe ‘that, who, which’.

- Interrogative pronouns

hwilic ‘of what sort’; hwā ‘who; what; anyone, someone, anything, something; each’; hwæðer ‘which of two, whether.’; hwæðer...ðe each of two, both; one of two, either’; hwæðer...or each of two, both; one of two, either’;

hwæm ‘where’; hwæt ‘what’; hwæt ‘who; what; anyone, someone, anything, something; each’; hwilc ‘which, what; whosoever, whichever; any (one), some (one)’; swā hwā ‘whosoever’; swā hwā swā ‘whosoever’; swā hwæðer swā ‘whichever’; swā hwær swā ‘wherever’; swā hwæt swā ‘whatsoever’; swā hwilc swā ‘whosoever’; swā...swā ‘whether...or; either...or’; swæðer ‘whichever, whosoever’; tō hwæm ‘wherefore’.

- Indefinite pronouns

æghwā ‘each one, everyone, everything, who or whatever’; æghwæðer ‘everyone, either, both; each’; æghwæt ‘each one, everyone, everything, who or whatever’; æghwæt ‘anything’; ælc ‘each; any’; ænig ‘any, any one’; āhwa ‘any one’; āhwæðer ‘someone, something; any one; anything; either, each, one or other’; ānra gehwā ‘everyone’; edwihte ‘anything, something’; gehwā ‘each one, every one, any one, whoever’; gehwæðer ‘both, either, each’; gehwilc ‘each, any, every (one), all, some, many, whoever, whatever’; anra gehwilc ‘each one’; hwilcwega ‘some, any, someone; little, some, not much or great; alone, anything, something’; ilca ‘the same’; nānhtwæt ‘something unknown’; nāhwæðer ‘neither’; samhwylc ‘some’; sum ‘a certain one, someone, something, one’; sumhwilc ‘a certain’; swā ‘so, the same, such, that’; swilc ‘such a one, he, the same; such; (as a relative) which’; swilc...hwilc ‘such...as; so...as’; swilc...swilc ‘so much (many)...as; as much (many)...as’; ðullic ‘such, such a’; welhwā ‘every one, everything’; welhwilc ‘each, any, nearly every’.

- Personal pronouns

gē ‘you’; git ‘you two’; he ‘he, she, it; (pl.) they; (reflex. pron.) himself, herself, itself’; hēo ‘she, they’; hīe ‘they’; hit ‘it’; ic ‘I’; tū ‘thou’; ðæge ‘they, these’; ðu ‘thou’; uncer ‘of us two, our (of two persons)’; we ‘we’; wit ‘we two’.

- Coordinating conjunctions

ac ‘but; but also, moreover, nevertheless, however; because, for’; æghwæðer ge...ge ‘both...and, as well as’; æghwæðer...and ‘both...and, as well as’; and ‘and; but; or’; būtan ‘except, except that, but, only’; būtan ðæt ‘except; unless, save that; except, but, besides, if only, provided that’; eornostlice ‘therefore, but’; ge ‘and, also’; ge...ge ‘both... and; not only... but also; whether... or’; nāhwæðer ‘neither’; ne ‘neither, nor’; oððe ‘or; and’; oððe...oððe ‘either...or’; ofðe ‘or’; sam ‘whether, or’; sam ðe...sam ðe ‘whether...or’; sam ge...sam ge ‘whether...or’; sam...sam ‘whether...or’; samðe...samðe ‘as well...as’.

- Subordinating conjunctions

ær ‘before that’; for ‘for, because’; for ðy ðe ‘because’; forðæm ‘for (the reason) that, owing to (the fact) that, for, because, on that account, therefore, seeing that’; forðy ‘for that, because, therefore’; forūton ‘without, besides, except’; gif ‘if; whether, though’; hwæðer ðe...ðe ‘whether...or’; hwær ‘where, whither, somewhere, anywhere, everywhere’; hweðer ‘whether’; hwý ‘why’; mid ðy ‘while, when’; nemne ‘unless, except, save, only’; nū ‘now that, inasmuch as, because, since, when’; oð ‘until’; oððæt ‘until’; sið ‘after, afterwards’; siððan ‘as soon as, when, since, after that, inasmuch as’; siððan...siððan ‘when...then’; sōðhwæðere ‘however, yet, nevertheless’; swā ‘so as, consequently, just as, so far as, in such wise, in this or that way, thus, so that, provided that’; swā ðeah ‘nevertheless, yet.’; swilce ‘as if, as though’; tō ðy ðæt ‘for the purpose that, in order that’; ðā ‘then, at that time; after that time, thereupon; when, at the time that, whilst, during; there, where; seeing that, inasmuch as, if, when, since, as, because’; ðā ðā ‘when’; ðā hwile ðe ‘while, whilst, so long as’; ðā...ðā ‘then...when’; ðær ‘there, thither, yonder; where, whither; then; when; though, if, so far as, whilst, provided that; in that respect’; ðær ðær ‘where, wherever’; ðær wið ‘in regard to that’; ðær...of ‘therefrom’; ðærforan ‘before that’; ðæt ‘that, so that, in order that, after that, then, thence’; ðæt ðe ‘that’; ðætte ‘that, so that, in order that. tō ðon ðætte so that’; ðe ‘when; or; then; where. (with comparatives) than’; ðe...ðe ‘the...or, either...or’; ðeah ‘though, although, even if, that, however, nevertheless, yet, still; whether’; ðeah ðe ‘although’; ðeah... ðeah ‘although, still, yet’; ðenden ‘meanwhile, while, as long as, until’; ðonne ‘then; therefore, wherefore; yet; while, when; thereafter, henceforth; rather than; since; although; (with comparatives) than’; ðonne...ðonne ‘when...then’; ðonne...ðe ‘since’; ðonne...gýt ‘as yet, even’; ðonne...hwæðere ‘yet, nevertheless’; ðy ‘because, since, on that account; therefore; then; (with comparatives) the’; ðy læs ðe ‘lest’; ðy...ðy ‘the...the’; ðylæs ‘lest’; weald ‘in case’; weald ðeah ‘perhaps, possibly’; wið ‘until’.

2. Token indexing

- Abbreviations (based on Taylor et al. 2003)

nabbað (ne-habban), nabban (ne-habban), nabbap (ne-habban), nabbe (ne-habban), nabbeð (ne-habban), nabbon (ne-habban), nactes (nā-wiht), nað (ne-āgan), naðær (nā-hwæðer), naðer (nā-hwæðer), naðere (nā-hwæðer), naðor (nā-hwæðer), naðre (nā-hwæðer), naðres (nā-hwæðer), naðron (nā-hwæðer), naðrum (nā-hwæðer), næbbað (ne-habban), næbban (ne-habban), næbbe (ne-habban), næbben (ne-habban), næbbon (ne-habban), næbre (ne-æfre), næfd (ne-habban), næfð (ne-habban), næfdan (ne-habban), næfde (ne-habban),

næfde (ne-habban), næfden (ne-habban), næfdest (ne-habban), næfdon (ne-habban), næfedð (ne-habban), næffre (ne-æfre), næfræ (ne-æfre), næfre (ne-æfre), næfst (ne-habban), næfþ (ne-habban), nægnigne (ne-æinig), næne (ne-ān), nænegu (ne-æinig), nænegum (ne-æinig), nænes (ne-ān), nænge (ne-æinig), nænges (ne-æinig), nængum (ne-æinig), nænig (ne-æinig), nænig (ne-æinig), nænige (ne-æinig), næniges (ne-æinig), nænigne (ne-æinig), nænigo (ne-æinig), nænigra (ne-æinig), nænigre (ne-æinig), nænigu (ne-æinig), nænigum (ne-æinig), nænigpinga (ne-æinig+ðing), nænine (ne-æinig), nænne (ne-ān), nænum (ne-æinig), nænygum (ne-æinig), næran (ne-bēon), nære (ne-bēon), næren (ne-bēon), næron (ne-bēon), nærre (ne-bēon), nærun (ne-bēon), næs (ne-bēon), nafa (ne-habban), nafað (ne-habban), nafast (ne-habban), nafaþ (ne-habban), nafst (ne-habban), nagan (ne-āgan), nage (ne-āgan), nagon (ne-āgan), nah (ne-āgan), nah (ne-habban), naht (nā-wiht), nahte (nā-wiht), nahte (ne-āgan), nahtes (nā-wiht), nahtest (ne-āgan), nahton (ne-āgan), nahton (ne-habban), nahu (nā-hū), nahwæðer (nā-hwæðer), nahwær (nā-hwær), nahwanen (nā-hwanon), nahwer (nā-hwær), nahwider (nā-hwider), nahwonan (nā-hwanon), nahwyder (nā-hwider), nalæs (nā-læs), nalas (nā-læs), nalde (ne-willan), nales (nā-læs), nallæs (nā-læs), nallas (nā-læs), nalles (nā-læs), nals (nā-læs), nam (ne-ān), namare (na+mare), nan (ne-ān), nana (ne-ān), nanan (ne-ān), nanðinc (ne-ān+ðing), nanðing (ne-ān+ðing), nane (ne-ān), nanegum (ne-æinig), nanen (ne-ān), nanes (ne-ān), nang (ne-ān), naniges (ne-æinig), nannæ (ne-ān), nanne (ne-ān), nanno (ne-ān), nanon (ne-ān), nanra (ne-ān), nanre (ne-ān), nanuht (nān+wiht), nanum (ne-ān), nanwiht (nān+wiht), nanwit (nān+wiht), nanwith (nān+wiht), nanwuht (nān+wiht), nanwyht (nān+wiht), nanþing (ne-ān+ðing), nanþingc (ne-ān+ðing), nast (ne-āgan), nat (ne-āgan), nat (ne-wītan), nateshwan (nātes+hwōn), nateshwon (nātes+hwōn), nateþæshwon, natoðæshwon (nātes+hwōn), natoðeshwon (nātes+hwōn), natohwon (nātes+hwōn), natopæshwon (nātes+hwōn), natopeshwon (nātes+hwōn), natoposhwon (nātes+hwōn), natopyshwon (nātes+hwōn), nauðer (nā-hwæðer), nauðrum (nā-hwæðer), nauht (nā-wiht), nauhtas (nā-wiht), nauhte (nā-wiht), nauhtes (nā-wiht), nauþær (nā-hwæðer), nauþer (nā-hwæðer), nauþres (nā-hwæðer), nawaðer (nā-hwæðer), nawærn (nā-hwærn), nawar (nā-hwær), nawcte (nā-wiht), nawðær (nā-hwæðer), nawðer (nā-hwæðer), nawer (nā-hwær), nawern (nā-hwærn), nawht (nā-wiht), nawiht (nā-wiht), nawit (nā-wiht), naworn (nā-hwærn), nawt (nā-wiht), nawuht (nā-wiht), nauhtes (nā-wiht), nauuihte (nā-wiht), nawyht (nā-wiht), nawyt (nā-wiht), nauþer (nā-hwæðer), napær (nā-hwæðer), naper (nā-hwæðer), naping (nā-ðing), napor (nā-hwæðer), naprum (nā-hwæðer), nealles (nā-læs), nealt (ne-willan), neart (ne-bēon), nebbe (ne-habban), nefdan (ne-habban), nefdon (ne-habban), nefræ (ne-æfre), nefre (ne-æfre), nel (ne-willan), nele (ne-willan), nellað (ne-willan), nellan (ne-willan), nellaþ (ne-willan), nelle (ne-willan), nelleð (ne-willan), nellen (ne-willan), nellendes (ne-willan), nellendum (ne-willan), nellon (ne-willan), nelt (ne-willan), nene (ne-ān), nenegum (ne-æinig), nenge (ne-æinig), nenig (ne-æinig), neniggra (ne-æinig), nenigne (ne-æinig), nenne (ne-ān), neom (ne-bēon), neron (ne-bēon), nes (ne-bēon), nestan (ne-wītan), nestes (ne-wītan), neton (ne-wītan), nile (ne-willan), nis (ne-bēon), niste (ne-wītan), niston (ne-wītan), nitendum (ne-wītan), niton (ne-wītan), noðer (nā-hwæðer), noht (nā-wiht), nohte (nā-wiht), nohtes (nā-wiht), nohuæðer (nā-hwæðer), nohwæðer (nā-hwæðer), nohwær (nā-hwær), nohwæpre (nā-hwæðer), nohwider (nā-hwider), nolæs (nā-læs), noldan (ne-willan), nolde (ne-willan), nolden (ne-willan), noldes (ne-willan), noldest (ne-willan), noldon (ne-willan), non (ne-ān), nonne (ne-ān), nouðer (nā-hwæðer), nowðer (nā-hwæðer), nower (nā-hwær), nowiht (nā-wiht), nowuht (nā-wiht), nowyht (nā-wiht), nowþer (nā-hwæðer), noþer (nā-hwæðer), nuste (ne-wītan), nyle (ne-willan), nyllað (ne-willan), nyllan (ne-willan), nylle (ne-willan), nylleð (ne-willan), nyllen (ne-willan), nylt (ne-willan), nys (ne-bēon), nysðon (ne-wītan), nysse (ne-wītan), nysste (ne-wītan), nysstest (ne-wītan), nystan (ne-wītan), nyste (ne-wītan), nysten (ne-wītan), nystes (ne-wītan), nystest (ne-wītan), nyston (ne-wītan), nytað (ne-wītan), nytan (ne-wītan), nyte (ne-wītan), nyten (ne-wītan), nytende (ne-wītan), nytendum (ne-wītan), nyton (ne-wītan).