

Article

Toward a Universal Dependencies Treebank of Old English: Representing the Morphological Relatedness of Un-Derivatives

Javier Martín Arista 

Department of Modern Languages, Faculty of Humanities and Education, University of La Rioja, 26006 Logroño, Spain; javier.martin@unirioja.es

Abstract: This article deals with one of the aspects involved in the compilation of a treebank of Old English within the framework of Universal Dependencies. More specifically, this study addresses the question of how to account for the remarkable degree of Old English morphological relatedness in a type of treebank designed to stress syntactic similarities across languages. The solution proposed and assessed in this study is the addition of an extra field of annotation for morphological relatedness. The data of this analysis comprise 1106 derivatives attaching the prefix *un-*. Out of these, there are around 80 morphologically complex nouns, adjectives, verbs, and adverbs whose derivation cannot be described gradually, 33 of which are unique formations or *hapax legomena* according to the attestations provided by the *Dictionary of Old English Corpus*. The main conclusion is that the specification of short-distance and long-distance morphological relatedness provides the Old English treebank with a paradigmatic dimension that can be particularly relevant for languages with relatively generalised and transparent derivational morphology.

Keywords: Old English; corpus linguistics; computational linguistics; Universal Dependencies



Citation: Martín Arista, Javier. 2024. Toward a Universal Dependencies Treebank of Old English: Representing the Morphological Relatedness of Un-Derivatives. *Languages* 9: 76. <https://doi.org/10.3390/languages9030076>

Academic Editor: Julien Longhi

Received: 6 November 2023

Revised: 23 February 2024

Accepted: 23 February 2024

Published: 27 February 2024



Copyright: © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Aims and Scope

This article intends to contribute to the compilation of a treebank of Old English based on the annotation model of Universal Dependencies. Universal Dependencies, hereafter UD, (de Marneffe et al. 2021) is a model of morphological and syntactic annotation devised for the compilation of computerised datasets geared to cross-linguistic comparison, natural language processing, language acquisition, and translation (McDonald 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 standards of adequacy of the model include that UD should meet criteria for linguistic analysis, typology, human annotation, accessibility to non-linguists, habitable design favoring traditional grammar, accurate computer parsing, and support for various language-understanding tasks (<https://universaldependencies.org/introduction.html>, accessed on 15 October 2023). The 2015 UD dataset consisted of ten treebanks representing ten languages, whereas the 2021 release comprises 183 treebanks and over 100 languages (Nivre et al. 2020).

Old English is the historical stage of the English language spoken in England between approximately the 5th and the 11th centuries (CE). From the typological point of view, it belongs to the Anglo-Frisian branch of the West-Germanic group of the Indo-European family of languages. Written records, which can be traced back to the 7th century onwards, comprise approximately 3 million words in around 3000 texts. The main lexicographical sources of Old English include the dictionaries by Bosworth and Toller (1973), Sweet (1976), and Clark Hall (1986), as well as the *Dictionary of Old English* (Healey 2018). The main textual sources of Old English are *The Dictionary of Old English web corpus* (3,000,000 words; Healey 2012) and *The York Toronto Helsinki Parsed Corpus of Old English Prose* (Taylor et al. 2003).

Against this background, this article aims at adapting the annotation format of Universal Dependencies, called CoNLL-U, to the linguistic characteristics of Old English. Two

aspects of this diachronic stage of the English language are relevant for morpho-syntactic annotation: the historical character of the language, which requires a gloss or equivalent in Present-Day English in order to guarantee the searchability and the comparability of the treebank, and the remarkable role played by derivational morphology in the organisation of the Old English lexicon, which includes numerous derivational families with transparent morphological relations. A detailed annotation of the derivational morphology, comparable to the one that the model has already adopted for inflectional morphology, may reinforce the paradigmatic dimension of the annotation and allow for syntactic and semantic generalisations that would increase the explanatory power of the model. This is compatible with the general approach adopted by the UD framework, which aims for comparability and stresses similarity while allowing for language-specific descriptions. It must be remarked in this respect that the UD framework of morpho-syntactic annotation has not been primarily devised for historical languages, with which the evidence gathered from them should be considered in its own terms, at least in some respects. This can be seen as a step taken towards the convergence with the framework of Universal Derivations (<https://ufal.mff.cuni.cz/universal-derivations>, accessed on 10 October 2023), which provides an exhaustive account of morphological relations which the UD, as a general rule, lacks. In the wider context of the compilation of an Old English treebank, the enriched annotation model proposed in this study may contribute to giving visibility to historical languages in general and Old English in particular by gathering, digitising, and disseminating structured linguistic data of Old English, and to provide scholars with an intermediate expertise in Old English and/or theoretical linguistics with an open access set of fully comparable and searchable data. In this line, the access to the glosses of the tokens guarantees a contextualised translation of the lexical item while the paradigmatic analysis of derivational morphology, which underlines the lexical consistency of the language, provides a Present-Day English lexical correlate.

Adapting the annotation format of Universal Dependencies to Old English requires the justification of the fields that should be incorporated to the CoNLL-U format, defining their content, implementing them in a dataset, and assessing the results. The dataset selected for this study comprises prefixal derivatives with *un-*. There are 1106 derivatives attaching this prefix in Old English. Out of these, around 80 morphologically complex nouns, adjectives, verbs, and adverbs have been found whose derivation cannot be described gradually (a maximum of one process takes place at a time), 34 of which are unique formations or *hapax legomena* according to the attestations provided by the *Dictionary of Old English Corpus*. This analysis focuses on these unique formations because they represent a real challenge in derivational morphology and lexicographical practice. The line is taken in this respect that the morphological relatedness of a lexical item in general and a unique formation in particular should be described with respect to both productive (analysable and more transparent) processes and unproductive (non-analysable and opaque) processes, which can only be recovered on the diachronic axis. This approach subsumes the presentation of the derivatives of a given entry that some dictionaries offer, while increasing the power of generalisation and reinforcing the paradigmatics of the annotation model.

On that account, different aspects of historical linguistics, morphology, corpus linguistics, natural language processing, and lexicography converge in this study. The application of the framework of UD to Old English entails the morphological and syntactic annotation of a large dataset, which is currently unlemmatised. This study also deals with aspects of natural language processing such as the components of the CoNLL-U format, which can be adapted to the characteristics of the languages it tabulates. Finally, this study also addresses some theoretical questions of derivational morphology related to lexicographical practice, like the treatment of synchronic word-formation and etymology.

With these premises, this article is organised as follows. Section 2 sets the descriptive basis of the study. Section 3 reviews the lexicographical and textual sources. Section 4 underlines the basic aspects of the theoretical basis of this study, the framework of UD, with special emphasis on the CoNLL-U format. It is argued that the consistency and

paradigmatic organisation of the Old English lexicon advises the annotation of short-distance and long-distance morphological relations within the CoNLL-U format. A gloss field is also proposed in this section. Section 5 presents the data and method of analysis. Section 6 analyses the data by lexical category and the distance of the derivational relation. It also implements the results in the CoNLL-U Plus format and assesses the results. The main conclusions of the research are presented in Section 7.

2. Descriptive Basis

The relevance of this undertaking is related to the consistently Germanic lexicon of Old English, characterised by its associative status, which is defined as the existence of large morphological families with relatively transparent relations of form and meaning (Kastovsky 1992, p. 294). Although some authors, like Haselow (2011), point to a rise in analytic tendencies as a result of the loss of efficiency of the derivational resources of the language, and others, like Martín Arista (2011, 2012, 2019), describe areas of inefficiency and opaqueness in Old English word-formation, derivational morphology constitutes the main organising principle of the Old English lexicon. In Lass's (1994, p. 198) words, the older an IE language, the more transparent and complex its derivational morphology, and, indeed, the more derived forms there appear to be, and the more central derivation appears to be to overall lexical structure.

Strong verbs are, as a general rule, the starting point of derivational processes in the old Germanic languages (Bammesberger 1965; Hinderling 1967; Seebold 1970; Kastovsky 2006, among others). For instance, the compounds and derivatives presented in (1) are related to the strong verb *bacan* 'to bake'.

(1)

ābacan 'to bake', *asbacen* 'baked on ashes', *bacan* 'to bake', *bæcere* 'baker', *bæcering* 'gridiron', *bæcern* 'bakery', *bæcestre* 'baker', *ealdbacen* 'stale', *elebacen* 'cooked in oil', *gebæc* 'bakemeats', *heorðbacen* 'baked on the hearth', *nīwbacen* 'newly baked', *ofenbacen* 'baked in an oven'.

This is not to say, however, that other lexical categories do not turn out large families of derivatives. Seebold (1970, p. 501) provides Germanic *teldam* and *teldō* for, respectively, Old English *teld* 'tent' and *teldian* 'to provide with tents', which belong in the derivational family listed in (2).

(2)

beteldan 'to cover', *būrgeteld* 'pavilion', *ganggeteld* 'portable tent', *(ge)teld* 'tent', *(ge)teldan* 'to spread a covering', *geteldung* 'tabernacle', *geteldwurðung* 'feast of tabernacles', *oferteldan* 'to cover over', *teldgehlīwung* 'tabernacle', *teldian* 'to spread (net)', *teldsticca* 'tent-peg', *teldtrēow* 'tent-peg', *teldwyrhta* 'tent-maker', *tyldsyle* 'tent'.

As regards the adjective, Heidermanns (1993, p. 576) opts for the Germanic adjective **sweiga* 'still' as the etymon of the derivational family displayed in (3).

(3)

ānswēge 'harmonious', *āswēgan* 'to thunder, intone', *bencswēg* 'bench-rejoicing', *(ge)swēge* 'harmonious', *geswēgsumlice* 'unanimously', *geswōgung* 'swooning', *hāsswēge* 'sounding hoarsely', *hearpwēg* 'sound of the harp', *hereswēg* 'martial sound', *hlūdswēge* 'loudly', *onāswēgan* 'to sound forth', *samodswēgende* 'consonantal', *samswēge* 'sounding in unison', *selfswēgend* 'vowel', *swēg* 'sound, melody, voice, musical instrument', *swēgan* 'to make a noise', *swēgcræft* 'music', *swēgdynn* 'noise, crash', *swēgendlic* 'vocal', *swēghlēodor* 'sound', *swēging* 'sound', *swēglic* 'sonorous', *swētswēge* 'agreeable (of sound)', *swīðswēge* 'strong-sounding', *swōgan* 'to sound', *ungeswēge* 'inharmonious', *welswēgende* 'melodious'.

Examples (1), (2), and (3) also show that, while meaning and form similarities justify the gathering of derivatives, two types of morphological relatedness arise, namely, long-distance and short-distance. Long-distance morphological relatedness holds between two items whose relation is recoverable on the diachronic axis (Stark 1982, p. 62), as is the case with the Germanic **sweiga* 'still' and the Old English *swēgan* 'to make a noise' in (3). Short-distance morphological relatedness holds between two items whose relation can be

described productively on the synchronic axis, as in the suffixation of *-ere* to *bæcere* 'baker' in (1). Short-distance morphological relatedness shows different degrees of productivity and formal and semantic transparency. The suffixation of *-lic* to *swēgendlic* in (3) is fully productive and transparent, the prefixation of *ā* to *āswēgan* has been lexicalised (Hiltunen 1983; Brinton and Traugott 2005), and the alternation *swōgan-swēg* (Kastovsky 1968, p. 109) is fully unproductive and opaque.

Examples (1), (2), and (3) follow Kastovsky (1992) in gathering flat derivational families. A hierarchical model is needed, though, that distinguishes the combination of free forms (compounding) from the attachment of bound forms to free forms (affixation), and non-recursive derivatives and compounds from recursive ones. Such a hierarchical model is called a *derivational paradigm* in this study (Nichols 2014). A derivational paradigm is headed by the stem that is common to all the compounds and derivatives and classifies its members by process and by degree of recursivity (Martín Arista 2013). In (1), for instance, *bæcestre* 'baker' is a suffixed derivative whereas *ealdbacen* 'stale' is a compound. In (3), *onāswēgan* 'to sound forth' is a recursive derivative that results from the successive prefixation of *ā-* and *on-*, while *swēging* 'sound' is a non-recursive suffixal derivative with *-ing*.

3. Review of the Sources

Beginning with lexicographical sources, the entries to dictionaries of natural languages tend to include information on the origin of words. For instance, the main components of an entry to *The Oxford English Dictionary* (<https://public.oed.com/the-oed-today/rewriting-the-oed/editing-of-entries/>, accessed on 12 October 2023) are spelling, pronunciation, illustrative quotations, definition, and etymology. The scope of the information on the origin of words may be restricted to word-formation that is transparent in synchronic analysis, as in the entry for *truthful* in *The Cambridge Dictionary* (<https://dictionary.cambridge.org/es/diccionario/ingles/truthful?q=truthfulness>, accessed on 12 October 2023), which relates both the adverb *truthfully* and the noun *truthfulness* to the adjective *truthful*, or may widen to accommodate formations that can only be recovered by the diachronic axis of analysis, as can be said of *lordless* in *The Merriam–Webster Dictionary* (<https://www.merriam-webster.com/dictionary/lordless>, accessed on 12 October 2023): *Middle English* lordles, *alteration of* loverdles, *from Old English* hlāfordlēas, *from* hlāford *lord* + *-lēas* *-less*. This said, dictionaries usually vary as to the explicitness with which etymology is distinguished from word formation. *The Oxford English Dictionary*, for example, draws a distinction between compounding (involving existing English words) and derivation (based on other English words and comprising regular processes of word formation), on the one hand, and etymology, *dealing with the origin and derivation of the word* (<https://public.oed.com/how-to-use-the-oed/glossary/>, accessed on 12 October 2023), on the other.

In the specific field of the lexicography of Old English, the existing dictionaries significantly differ as to the questions of etymology and word formation. Bosworth and Toller (1973) is the only dictionary that provides the correlates in the Germanic languages and other Indo-European languages. It also lists derivatives and marks primitive verbs with capitals, as in the entry to BRECAN. In this entry, the user gets a list of cognates including Old Saxon *brekan*, Old Frisian *breka*, Old High German *brechan*, and Gothic *brikan*. There is, as a general rule, no reconstructed Proto-Germanic form, but Latin, Greek, and Sanskrit cognates are often available and the user can find Middle English evolutions and related words in the modern Germanic languages on a consistent basis. In the entry to BRECAN, as in the entries to other primitive verbs, no distinction is made between zero derivations (*bræc*, *brec*, *broc*), prefixal derivatives (*a-brecan*, *be-*, *for-*, *ge-*, etc.), suffixal derivatives (*brecendlic*), and compounds (*burh-*, *cyric-*, *eodor-*; *-cōðu*, *-seóc*, *-seócnes*, etc.). Nevertheless, Bosworth and Toller (1973) is the most reliable Old English dictionary when it comes to finding etymological and derivational information: it ranks entries according to their role

in lexical derivation, provides cognates in the Germanic dialects and other branches of Indo-European, and lists derivatives and compounds.

While [Clark Hall \(1986\)](#) does not engage in etymology or word formation at all, [Sweet \(1976\)](#) explicitly gathers word-families, in the line of [Ettmüller \(1968\)](#). As [Metola Rodríguez \(2017, p. 182\)](#) notices, the entry to *scīr* ‘transparent, bright; clear’ is followed by its derivatives and compounds, indentation marking the discontinuity in alphabetical sorting: *scīrbaso* ‘bright purple’, *scīre* ‘brightly’, *scīrecg* ‘bright-edged’, *scīrham* ‘in bright armour’, *scīrmæled* ‘with bright ornaments (sword)’, *scīrwered* ‘bright (light)’, and *scīran* ‘declare, tell, speak’. For this reason, [Sweet \(1976\)](#) is praiseworthy for its presentation of derivatives and compounds, although it provides a significantly lower number of entries than the other dictionaries reviewed in this section: around 26,000 entries as opposed to 36,000 in Bosworth and Toller and Clark Hall and 16,000 in the A–I segment of the *Dictionary of Old English*.

The *Dictionary of Old English* (henceforth DOE; [Healey 2018](#)) does not deal with word formation explicitly. It sometimes provides correlates in the Germanic dialects, but it does not offer Proto-Germanic etymons. Word formation is treated in this dictionary in terms of cross-references, displayed as links to morphologically related words, and correlates in Middle English and in the modern Germanic languages, including those available from the *Oxford English Dictionary*. It is certainly an asset of the DOE that the Latin correspondences of glosses are listed. It is also noteworthy that cross-references to less transparent derivatives are given, such as *drāf*, *drīf*, *drāfan* in the case of *drīfan* ‘to drive’. Regarding primitive lexical items like *brecan* ‘to break’, the DOE does not acknowledge this status, in such a way that no principled and explicit hierarchical distinction is drawn between items that give rise to many lexical derivations, such as *brecan*, and others that merely constitute the final stage of derivations and do not have derivatives of their own, like *breahtmian* ‘to make a noise’. More importantly, the rendering of word formation is less consistent if primitives of lexical derivation are not defined. For instance, while *æbrucol* ‘sacrilegious’ is listed in the DOE under *brecan* ‘to break’, *æbrecð* ‘sacrilege’ is not. Its entry is cross-referenced to *æ*, *brecð* ‘breach’ and *æbrucol* ‘sacrilegious’. *Bræclian* ‘to crackle’ is not listed under *brecan*, although the entry to *bræclian* is cross-referenced to *brecan*. *Brocenlic* ‘fragile’ is listed under *brecan*, but *broccian* ‘to tremble’ is not and its entry is not cross-referenced to *brecan*. *Broclīc* ‘full of hardship’ is not listed under *brecan*, to which it is cross-referenced indirectly through *broc*¹ ‘affliction’. The same can be said of *brocung* ‘sickness’, which is also cross-referenced to *brecan* through *broc*¹. These inconsistencies could be avoided, or at least reduced, by relating all the derivatives to the lexical prime, while also acknowledging the relatedness among non-primitive bases and their derivatives.

Regarding textual sources, the main lesson that can be learned from a review of the corpora of Old English is that some of the most important are neither annotated nor lemmatised. This includes *The Helsinki Corpus of English Texts* (400,000 words; [Rissanen et al. 1991](#)) and *The Dictionary of Old English* web corpus (3,000,000 words; [Healey 2012](#)). Even those that are annotated or lemmatised have some shortcomings. *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 (part-of-speech) tagging and are parsed for the syntax, although they provide a constituency annotation based on theoretical insights from the 1980s and are not lemmatised. The only lemmatised corpus of Old English is ParCorOEv2, *An open access annotated parallel corpus Old English-English* ([Martín Arista et al. 2021](#)). The 2021 release comprises 110,000 records, tagged with file number, lemma, lexical category, inflectional category, and gloss. Against this backdrop, the field is in need of large lemmatised and annotated datasets that guarantee full searchability of a language whose spelling is not standardised.

4. UD and the CoNLL-U Format

As has been remarked above, this article intends to reinforce the paradigmatic dimension of the framework of UD as applied to Old English. To this aim, this section presents the standard annotation format of UD and discusses its extension.

The design of a UD treebank of Old English comprises two main steps, namely, the segmentation of texts and the annotation of the resulting fragments. Three types of units result from segmentation: sentences, words, and tokens. Sentences are syntactic periods with full meaning. Words are defined as syntactic words, written without hyphens or spaces between their parts. A key aspect of the segmentation of Old English is token indexing (the identification and segmentation of complex words), which requires a principled inventory of the free forms that partake in contractions. Since Old English, as a general rule, does not separate or hyphenate compounds, they are considered as a product of the morphology and annotated at token level. Consequently, the dependency relation compounded is not marked in the DEPREL or DEPS columns.

The UD annotation format is known as the CoNLL-U format. CoNLL (Conference on Natural Language Learning) is the name of tab-separated formats in natural language processing. The CoNLL-X format (Buchholz and Marsi 2006) established the foundations of the annotation based on tab-separated text files, which can be summarised as follows. An annotation is the value of a particular word on a column. Every column presents one annotation. All words are annotated on the same columns. Every word takes up one line. One empty line is inserted before and after each sentence. Fragmentation is carried out at word level and at token level (analysable components of syntactic words, such as clitics and contractions). The annotation includes part-of-speech tagging and, in some models, dependency relations.

The revised version of the CoNLL-X format is known as the CoNLL-U format. Tokenisation and annotation in CoNLL-U are encoded in plain text files in the UTF-8 format. The LF character is used as a 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 a hash (#). Sentences comprise at least one word line. Each word line contains the ten fields described in Figure 1.

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.
DEPS	List of secondary dependencies.
MISC	Any other annotation.

Figure 1. The columns in the CoNLL-U annotation format (based on <https://universaldependencies.org/docs/format.html>, accessed on 7 April 2023).

Although the model of UD is concerned with patterns of convergence across languages, it also favours the annotation of language-specific phenomena. Extensions of the CoNLL-U format constitute instances of the CoNLL-U Plus format. An extension is necessary because the MISC column can be used for these purposes, but, in the annotation of Old English, it has been taken up by the indication that there is no space after a token.

Two extra specific fields are required to annotate adequately the morphology and syntax of Old English. The first is GLOSS (gloss). As we are annotating a historical language, a translation into Present-Day English facilitates the annotator's task and, more importantly,

allows for diachronic and cross-linguistic comparison and improves the accessibility and searchability of the treebank. Of the two types of gloss available in the UD framework, namely morphological tag and modern form, the latter is preferred because the morphological tag is redundant with the columns that tabulate part-of-speech annotation (LEMMA, XPOSTAG, and FEATS) and because the modern form increases the comparability of the treebank. For example, glossing *besibb* as ‘related’ refers the user to ‘relation’, ‘relationship’, ‘sibling’, German *Geschwister*, Latin *frater*, etc.

The second specific field proposed in this article is MORPHREL (morphological relatedness). MORPHREL annotates the morphological relatedness of the lemmas from the major lexical classes: nouns, adjectives, and verbs. This means that tokens and lemmas from minor lexical and grammatical classes are not specified as to morphological relatedness. Nevertheless, in a language characterised by an associative lexicon (Kastovsky 1992), 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*. Notice that the asterisk indicates diachronic reconstruction.

Figure 2 illustrates the implementation of the CoNLL-U Plus format in a dataset from Old English. Technically speaking, this extension constitutes the implementation of a CoNLL-U Plus file, in which a comment line must be inserted that starts with the # character and lists the names of the columns used in this file.

```
# global.columns ID FORM LEMMA UPOSTAG XPOSTAG FEATS HEAD DEPREL DEPS MISC GLOSS MORPHREL
```

ID	FORM	LEMMA	UPOSTAG	XPOSTAG	FEATS	HEAD	DEPREL	DEPS	MISC	GLOSS	MORPHREL
1	ac	ac	CCONJ	coord. conj.	Uninflected=Yes	6	cc	-	-	but	-
2	him	hē	PRO	pronoun	Case=Dat Gender=Masc Number=Sing Person=3 PronType=Prs	6	obl	-	-	him	-
3	næs	-	-	-	-	-	-	-	-	not be	ø/BĒON
3.1	ne	ne	ADV	adverb	Polarity=Neg Uninflected=Yes	3	advmod	-	SpaceAfter=No	-	-
3.2	wæs	bēon	AUX	aux. verb	Mood=Ind Number=Plur Tense=Pres VerbForm=Fin	6	aux:pass	-	-	-	-
4	þa	þā	ADV	adverb	Uninflected=Yes	6	obl	-	-	-	-
5	nan	-	-	-	-	-	-	-	-	nothing	-
5.1	ne	ne	ADV	adverb	Polarity=Neg Uninflected=Yes	5	nmod	-	SpaceAfter=No	-	-
5.2	an	ān	PRO	pronoun	Case=Nom Gender=Masc Number=Sing Person=3 PronType=Ind	6	nsubj	-	-	-	-
6	geseald	gesellan	VERB	main verb	Case=Nom Number=Sing Tense=Past VerbForm=Part	0	root	-	-	given	sellan/SELLAN
7	.	-	PUNCT	punctuation	Uninflected=Yes	6	punct	-	-	-	-

DEPREL abbreviations: cc (coordinating conjunct), obl (oblique), advmod (adverbial modifier), aux:pass (passive auxiliary), nmod (nominal modifier), nsubj (nominal subject), root (sentence node), punct (punctuation).

Figure 2. The CoNLL-U Plus format of Old English.

As shown in Figure 2, columns in the CoNLL-U format cannot be empty. They have the underscore if they are not used. Notice that macrons, which represent vowel quantity, are inserted in the LEMMA column, but not in the FORM column. The contractions in Figure 2 are followed by two extra lines with decimal ID. The contractions, such as *næs*, are marked with the underscore on all the columns except ID, FORM, and MISC and the extra columns GLOSS and MORPHREL. The column MISC has the value *SpaceAfter=No* in the line corresponding to the first element of the contraction to indicate that the contraction is written as one word. The elements of the contraction, like *ne* and *wæs*, are fully specified because they belong to two different categories, display various morphological features, and have different heads and relations of dependency. The principle of lexical integrity does not apply to inflectionally complex words, given that the syntax can be seen in the elements of the word. In contradistinction, derivationally complex words (affixal derivatives and compounds) are not broken down into their bases and adjuncts because the complex word belongs to one category and partakes in one dependency relation. For example, the prefix *ge-* is not separated from the base of derivation *sellan*.

Notwithstanding the appeal of the UD model and its remarkable growth despite its relative novelty, this framework is focused on the syntagmatic dimension. On the syntagmatic axis, syntactic compounding is exclusive to compounds written as two different

words. Since the CoNLL-U format specifies the lemma of each form (syntactic word), it is possible to collect all the forms that share a lemma tag, thus gathering the inflections of the form in question and, thanks to the MORPHREL column, the compounds in which it can be found. On the other hand, forms related to one another by means of affixation are not linked through the CoNLL-U format, leaving alone unique formations, which cannot always be related either by their base of derivation or by the affix attached to it. The successive processes involved in recursive derivation cannot be accounted for either. For this reason, the CoNLL-U Plus format is implemented in a set of hapaxes in Section 5.

5. Data and Method

The CoNLL-U Plus format proposed for the annotation of a treebank of Old English is implemented and assessed with respect to the affixal derivatives to which the prefix *un-* is attached. This prefix can convey a pejorative meaning, as in the nouns *ungifu* ‘evil gift’ (<*giefu* ‘gift’) and *unlæce* ‘bad physician’ (<*læce* ‘physician’); an oppositive meaning, as in the nouns *unānrædnes* ‘inconstancy’ (<*ānrædnes* ‘constancy’) and *ungȳmen* ‘carelessness’ (<*gȳemen* ‘care’) and the adjectives *unforboden* ‘lawful’ (<*forbēodan* ‘to forbid’) and *ungedēfe* ‘improper’ (<*gedēfe* ‘suitable’); and a counterfactual meaning, as in the verbs *unwrēon* ‘to uncover’ (<*wrēon* ‘to cover’) and *unbindan* ‘to unbind’ (<*bindan* ‘to bind’).

The dictionaries of Old English by [Bosworth and Toller \(1973\)](#), [Sweet \(1976\)](#), and [Clark Hall \(1986\)](#) have been searched for *un-* prefixal derivatives. The prefix *un-* representing an alternative spelling of *an-* or *on-* has been discarded on the basis of the meaning conveyed by the prefix. *The Dictionary of Old English* ([Healey 2018](#)) has not been searched because it has not published the letter U yet. The spelling has been normalised to the dictionary by Clark Hall–Meritt. The search has turned out a total of 1106 derivatives attaching the prefix *un-*. By lexical category, there are 799 adjectives, 194 nouns, 42 verbs, and 70 adverbs. As regards the categories to which *un-* is attached, 360 derivatives are based on adjectives, 196 on nouns, 350 on verbs (which can be broken down into 115 based on strong verbs and 235 based on weak verbs), and 54 *un-*derivatives are based on adverbs. Turning to the output categories, *un-* derives nouns from nouns, as in *unaga* ‘one who owns nothing’ (<*aga* ‘owner’), *ungemet* ‘excess’ (<*gemet* ‘proper’), and *unāblinn* ‘irrepressible state’ (~*āblinnan* ‘to cease’); adjectives from nouns, thus *untæle* ‘blameless’ (<*tæl* ‘blame’), *undæslīc* ‘inappropriate’ (<*ðæslīc* ‘suitable’), and *onspornend* ‘not stumbling’ (~*spurnan* ‘to stumble’); verbs from nouns, as is the case with *unmihtan* ‘to deprive of strength’ (~*miht* ‘power’), *unrōtian* ‘to be or make sad’ (~*rōt* ‘glad’), and *unwindan* ‘to unwind’ (<*windan* ‘to wind’); and, finally, the prefix *un-* derives adverbs from nouns, adjectives, verbs, and other adverbs, as is the case, respectively, with *ungewyrhtum* ‘without a cause’ (~*gewyrht* ‘cause’), *ungescēad* ‘exceedingly’ (<*gescēad* ‘reasonable’), *unbedōhte* ‘unthinkingly’ (~*ðencan* ‘to think’), and *unbeorhte* ‘not brightly’ (<*beorhte* ‘brightly’). The prefix *un-* is very frequently attached at the final stage of the derivation, or, put differently, it qualifies as a closing prefix in recursive derivations like *untōdælednes* ‘undividedness’ (<*tōdælednes* ‘separation’), *unāscyrigendlic* ‘inseparable’ (<*āscyrigendlic* ‘disjunctive’), and *unðurhscēotendlic* ‘impenetrable’ (~*ðurhscēotan* ‘to shoot through’).

Assessing the morphological relatedness of some *un-* derivatives is not a straightforward task. There are around 80 morphologically complex nouns, adjectives, verbs, and adverbs whose derivation cannot be described gradually. Of these, 33 are unique formations or hapax legomena according to the attestations provided by the *Dictionary of Old English Corpus* ([Healey 2012](#)). The rest of this section deals with this set for two reasons: firstly, because of the role played by hapax legomena in lexical creation, on the basis of which they are usually taken as an indicator of productivity ([Baayen 2008](#), p. 309; [Baayen 2009](#), p. 15; [Trips 2009](#), p. 38), and secondly, because hapax legomena pose an additional challenge to historical dictionaries and datasets when it comes to stating their morphological relatedness either on synchronic or on diachronic grounds.

The method pursued in this article puts forward a gradual model of lexical derivation that links a primitive of a description to the headword in question through successive

steps of affixation and compounding, as well as inflection when relevant. Under certain constraints, derivations can display hypothetical forms. Otherwise, the information on morphological relatedness is restricted to the lexical prime. In either case, the lexical prime, which is the form closest to the Proto-Germanic etymon, constitutes the main target of the description, so that the distance is reduced between the account of availability on the synchronic axis and of recoverability in diachrony. For example, the derived adjective *unbecēas* ‘indisputable’ is morphologically related to the primitive noun *CĒAS*, which can be traced back to Gmc, *KEUS-A- (Seebold 1970, p. 293; Orel 2003, p. 213), and is also related to the first preterite of the Old English Class II strong verb *cēosan* ‘to choose’ (*cēas*). The application of this method results in the compilation of a set of derivational paradigms. Derivational paradigms are hierarchical sets of morphologically related lexical items headed by a common stem. The definition of a paradigm requires that certain meaning components are shared, in such a way that all the derivatives that share a given affix do not constitute a derivational paradigm. Derivational paradigms are classified on the basis of the type of derivational process and the degree of recursivity.

With these premises, two types of morphological relatedness are distinguished: morphological relatedness involving a derivation for which an immediate base is available (this is marked with the symbol <), and morphological relatedness for which derivatives that belong to the same paradigm can be identified or, at least, a primitive noun, adjective, or verb can be found (this is indicated with the symbol ~). Derivations are organised by the lexical category of the derivative.

This said, most Old English *un-* derivatives have immediate bases of derivation, that is to say, to obtain them, it suffices to add this prefix to the immediate base of derivation, as in *unāberendlic* ‘intolerable’ (<*āberendlic* ‘tolerable’). Others cannot be directly derived from the nearest form that is available on the stepwise chain, as is the case with *unāreccendlic* ‘unexplainable’. The items morphologically related to this derived adjective include *reccend* ‘ruler’, *reccenddōm* ‘ruling’, *reccere* ‘teacher’, *reccing* ‘narration’, *recceliest* ‘carelessness’, *reccelic* ‘firm’, (*ge*)*reccan* ‘to offer, present; to get, seize, obtain, attain, reach; to tend to, hold out; to extend, stretch; to explain, interpret, instruct; to tell, narrate, relate, record; to quote; to translate; to correct, reprove, reproach; to direct, control, rule, set in order; to decide, give judgement; to wield (authority); to count, reckon; to subdue; to extend’, and *reccan* ‘to take care of’. However, neither **unāreccend* nor the more likely derivative **āreccendlic* are attested. In the absence of an immediate base, the starting point of the derivation can be selected, in order not only to state the morphological relatedness of the derivative but also to stress less transparent morphological bonds that link the derivative in question and other derivatives to the lexical primitive. Regarding *unāreccendlic*, the feminine noun *racu* ‘explanation, exposition, narrative, observation; reason, argument; rhetoric, the art of exposition; comedy; direction, guidance, correction; account, reckoning’ relates this adjective to other derivatives from the same derivational paradigm, such as *āreccende* ‘explaining’, *bereccan* ‘to explain’, *oferreccan* ‘to convince’, etc.

While some derivations, like the one of *unāreccendlic*, miss a derivational step, others lack an inflectional element. For example, *unācenned* ‘unbegotten’ can be related to *ācennan* ‘to beget’ if the past participle is inserted into the derivation, thus *ācennan*, *ācenned* > *unācenned*. Inflectional forms are not considered hypothetical when other forms from the inflectional paradigm of the reference form are attested, as in *ācenned* with respect to *ācennan*.

These facts seem to suggest that it is unclear whether we are dealing with issues of textual transmission or with lexical gaps. Given the data on the prefix *un-*, such lexical gaps could be attributed to the facts that negation distributes rather freely across all the major lexical categories and, moreover, that the result of the attachment of the negative prefix is semantically predictable and formally transparent. Consequently, parasynthetic affixation cannot be completely ruled out, involving, for example the prefix *un-* and the suffix *-lic*. As a matter of fact, 96 *un--lic* adjectives can be found whose base of derivation is not immediately

available, as happens in *unāblinnendlic* ‘unceasing’ (~*āblinnan* ‘to cease’), *unābrecedlic* ‘inextricable’ (~*breccan* ‘to break’), and *unymbwendedlic* ‘unalterable’ (~*ymbwendan* ‘to turn’).

On the grounds of this kind of evidence, the following steps are taken in order to fill in the gaps in lexical derivations and to provide derivational paradigms with heads. Derivations are defined gradually or in a stepwise manner, so that each step of the derivation involves the attachment of a maximum of one derivational affix. Hypothetical forms are proposed if there is evidence for similar formations and if they comply with two requirements, namely, that they are attached neither at the first nor at the last step of the derivation and that a maximum of one hypothetical form is inserted into a lexical derivation.

Regarding the restriction on the maximum of hypothetical forms by derivation, consider the derived adjective *unācnyccendlic* ‘that cannot be untied or loosened; indissoluble’. Its morphological relatedness needs to be sought in the set of derivatives *gecnycc* ‘bond’, (*ge*)*cnyccan* ‘to bind, tie’, *tōcnyccan* ‘to join’, and *ācnyht* ‘joined’. Two derivations might be proposed, depending on whether the negative prefix *un-* is attached earlier or later in the derivation. If it is attached to the hypothetical adjective **ācnyccendlic*, the derivation can be stated thus: *gecnycc* > (*ge*)*cnyccan* > **ācnyccan* > **ācnyccend* > **ācnyccendlic* > *unācnyccendlic*. If *un-* is prefixed to the hypothetical verb **unācnyccan*, the derivation can be defined as *gecnycc* > (*ge*)*cnyccan* > **ācnyccan* > **unācnyccan* > **unācnyccend* > *unācnyccendlic*. These derivations clearly show that if the number of hypothetical forms is not restricted, the question of bracketing paradoxes cannot be addressed in a principled way. On the other hand, if a maximum of one hypothetical form is allowed into the gradual derivation, bracketing paradoxes can be solved according to type frequency. For example, *ungewyldendlic* ‘impatient’ may be the result of two derivations: *gewealdan* > *gewieldan*, *gewieldend* > **gewyldendlic* > *ungewyldendlic*; or *gewealdan* > *gewieldan*, *gewieldend* > **ungewyldend* > *ungewyldendlic*. If type frequency is taken into account, it turns out that there are 29 derivatives of the type **gewyldendlic* > *ungewyldendlic* and none of the type **ungewyldend* > *ungewyldendlic*. The bracketing paradox, therefore, is solved in favour of attaching the prefix *un-* at the last step of the derivation.

6. The Morphological Relatedness of *un-* Derivatives

Section 6 focuses on 34 *un-* hapaxes whose derivation cannot be described gradually. The analysis links a primitive of description to a headword through successive steps of affixation and compounding, as well as inflection when relevant. The main target of the description, therefore, is the lexical prime, which represents the closest form to the Proto-Germanic etymon. Bracketing paradoxes are solved on the basis of type frequency. The discussion is divided into short-distance and long-distance relations. Short-distance relations are organised by category: derivatives related to a primitive noun (Section 6.1), derivatives related to a primitive adjective (Section 6.2), and derivatives related to a primitive verb (Section 6.3). Long-distance relations are considered in Section 6.4. Section 6.5 presents the results of the analysis.

6.1. Derivatives Related to a Primitive Noun

The derived adjective *unbecēas* ‘indisputable, incontestable’ is morphologically related to the primitive noun *CĒAS* ‘strife, contention, reproof, quarrelling, scandal; sedition’ [Gmc. *KEUS-A- (Seibold 1970, p. 293; Orel 2003, p. 213), Old English *cēosan*, 1st. preterite *cēas*], through the derived adjective **becēas*, which is formed by means of the attachment of the prefix *be-* to the primitive noun. Thus, the proposed derivation is *cēas* > **becēas* > *unbecēas*. Regarding the motivation of the hypothetical adjective **becēas*, another three adjectives prefixed with *be-* and based on a noun have been found: *bebyrd* ‘set with nails’ (<*byrd* ‘embroidering’), *besorg* ‘anxious’ (<*sorg* ‘sorrow’), and *besibb* ‘related’ (<*sibb* ‘relationship’).

The derived adjective *unlȳfendlic* ‘illicit, unlawful’ is morphologically related to the primitive noun *LĒAF* ‘permission, license, leave, privilege’ [Gmc. *LEIB-A- (Seibold 1970, p. 326; Orel 2003, p. 232); *LAIBŌ (Orel 2003, p. 232)], on which the weak verb *liefan* ‘to allow, permit’ is based. The substantivised participle of *liefan*, *lȳfend*, can then function as

the base of derivation of the hypothetical adjective **lȳfendlic* ‘licit, lawful’, to which the prefix *un-* is attached. As an argument in favour of the hypothetical form, it may be pointed out that a total of 151 adjectives derived from verbs and attaching the suffixal sequence *-end-lic* can be found, including *hierwendlic* ‘contemptible’, *lādiendlic* ‘excusable’, *ðencendlic* ‘thoughtful’, etc. The resulting stepwise derivation is: *lēaf* > *liefan*, *liefend* > **lȳfendlic* > *unlȳfendlic*.

The derivation of the adjective *unscamig* ‘unashamed, unabashed’ from the primitive noun SCAMU ‘shame, disgrace, dishonour; confusion; insult; modesty, bashfulness; private parts’ [Gmc. **SKAMŌ* (Orel 2003, p. 333)], takes an *-ig* adjective based on this noun, **scamig*. The existence of 192 adjectives attaching the suffix *-ig* to nouns, such as *hlīsīg* ‘famous’, *hrīmīg* ‘frosty’, and *mōdīg* ‘brave’, counts as evidence for the hypothetical form, as do the thirty *un-* prefixed adjectives based on adjectives attaching the suffix *-ig*, such as *unblōdig* ‘bloodless’, *ungeðyldīg* ‘impatient’, *unsælig* ‘unfortunate’, etc. That said, the full gradual derivation can be stated as follows: *scamu* > **scamig* > *unscamig*.

The derived adjective *unsefful* ‘senseless, irrational’ is based on the primitive noun SEFA ‘mind, understanding; spirit, heart’ [Gmc. **SAF-JA-* (Seebold 1970, p. 383; Orel 2003, p. 311); **SAFJŌN* (Orel 2003, p. 311)]. As in another eighty-two instances, a *-ful* adjective like **sefful* can be formed on a noun, thus *hohful* ‘careful’, *spræcful* ‘talkative’, and *tælful* ‘blameful’. There is also evidence of the subsequent attachment of the prefix *un-* to these adjectives, of which fifteen instances can be found, such as *ungetingful* ‘ineloquent’, *unmæðful* ‘immoderate’, and *unsideful* ‘immodest’. The gradual derivation containing the hypothetical form is *sefa* > **sefful* > *unsefful*.

The derivation of the adjective *untellendlic* ‘innumerable’ involves the weak verb *tellan* ‘to tell, state, count, consider’, which is ultimately related to the primitive noun TALU ‘tale, series, account, list; calculation; statement, deposition, relation, communication; talk, discussion, dispute; story, narrative, fable, tale; case, action at law, accusation, charge, claim, excuse, defence’ [Gmc. **TALŌ* (Orel 2003, p. 401)]. On the substantivised form of the participle of this verb, *tellend*, the hypothetical adjective **tellendlic* may be formed. As has been said above, there are 151 adjectives attaching the suffix *-lic* to *-end*. The gradual derivation can be stated as follows: *talū* > *tellan*, *tellend* > **tellendlic* > *untellendlic*.

6.2. Derivatives Related to a Primitive Adjective

The noun *unandcȳðignes* ‘ignorance’ is morphologically related to the primitive adjective CŪÐ ‘known, usual, well-known, famous, noted; familiar, friendly, related, intimate; known, plain; certain; manifest; excellent’ [Gmc. **KUNÞAZ* (Orel 2003, p. 224)], and its derivatives *cȳðan* ‘to tell, make know’, *cȳðīg* ‘known’, and *oncȳðīg* (= *uncȳðīg*) ‘not acquainted with’, in such a way that the following derivation can be put forward: *cūð* > *cȳðan* > *cȳðīg* > *oncȳðīg* > **oncȳðignes* > *unoncȳðignes*. The hypothetical form *oncȳðignes* is motivated by the well-attested formation of *-nes* nouns from *-ig* adjectives, which produces around fifty-nine nouns (*cystignes* ‘liberality’, *ēadignes* ‘happiness’, *nēadignes* ‘obligation’, etc.), as well as the derivation of *un-* prefixed nouns from *nes-* suffixed ones, of which twenty-seven instances turn up, including *ungecnyrdnes* ‘indifference’, *ungewīsnes* ‘uncertainty’, and *unsmēðnes* ‘roughness’.

The point of departure of the formation of the noun *ungedæftnes* ‘importunity, untimely intervention or interruption, unseasonableness’ can be traced back to the primitive adjective GEDAFEN ‘suitable, fit, proper, becoming’ [**DHABH-* (Pokorny 1959–1969, vol. 1, p. 234)]. A derivation can be proposed for *ungedæftnes* that includes the inflection for the past participle of the verb *gedæftan* ‘to prepare, make ready, put in order, arrange’ (*gedafte*), on which the *-nes* suffixed noun might have been formed, eventually turning out the *un-* derivative *ungedæftnes*: *gedafen* > *gedæftan*, *gedafte* > **gedæftnes* > *ungedæftnes*. As has been said above, there are twenty-seven *un-* prefixed nouns based on *-nes* suffixed derivatives, similar to the formation **gedæftnes* > *ungedæftnes*.

The primitive adjective WEORÐ ‘worth, of value; worthy, honoured, noble, honourable, excellent, of high rank; valued, dear, precious; becoming, fit, meet; capable,

properly qualified for, entitled to, possessed of' [Gmc. *WERÐAN (Heidermanns 1993, p. 675; Orel 2003, p. 457)] is found in the hypothetical compound **leahtorwyrðe* 'culpable', to which the prefix *un-* is attached to form the derived adjective *unleahtorwyrðe* 'unblameable; not culpable': **leahtor-wyrðe* > *unleahtorwyrðe*. The compound *leahtorcwide* 'injurious, insulting or opprobrious speech, blasphemy' is similar to the hypothetical formation proposed as input to this derivation.

The primitive adjective ORNE 'not mean' can function as the base of **orn(e)lic*, thus constituting an instance of the derivation of *-lic* adjectives from other adjectives, which amounts to 375 types, such as *geornlic* 'desirable', *gesīenelic* 'visible', and *gramlic* 'fierce'. Additional evidence for this derivation is provided by the formation of *un-* prefixed derived adjectives based on adjectives attaching the suffix *-lic*, of which there are ninety-three instances, including *unābȳgendlic* 'inflexible', *ungehælendlic* 'incurable', and *unmærlīc* 'inglorious'. The resulting stepwise derivation of the derived adjective *unornlic* 'mean' is *orne* > **ornlic* > *unornlic*.

The derived adverb *untrāglīce* 'well, honestly, frankly' can be related to the adjectival prime TRĀG 'bad, mean, evil' [Gmc. *TREG-A- (Seebold 1970, p. 506; Orel 2003, p. 409); *TREGÖN (Orel 2003, p. 409)]. The gradual derivation of *untrāglīce* needs a hypothetical adverb that is derived from the primary adjective, **trāglīce*. Two types of evidence support this proposal: there is a total of 155 adverbs derived with the suffix *-līce* that have an adjectival base, like *orenlīce* 'excessively', *rōtlīce* 'gladly', and *werodlīce* 'sweetly'; while forty-eight adverbial derivatives with *-līce* are further derived by means of the attachment of the prefix *un-*, as is the case with *unfrēondlīce* 'unkindly', *ungecoplīce* 'unsuitably', and *unhīersumlīce* 'disobediently'. The resulting stepwise derivation is *trāg* > **trāglīce* > *untrāglīce*.

6.3. Derivatives Related to a Primitive Verb

The derived adjective *unābrecondlic* 'inextricable' is morphologically related to the primitive strong verb BRECAN 'to break, tear, crush, shatter, burst, break up, destroy, demolish; to bruise; to curtail, oppress; to injure, violate; to tame, subdue; to press, urge, force; to interrupt, break into, storm, capture (city); to break or crash through, burst forth, spring out; to retch; to make a noise or crash; to sail; to intersect' [Gmc. *BREK-A- (Seebold 1970, p. 132; Orel 2003, p. 55)]. Its derivation requires a hypothetical *-lic* adjective based on the substantivised participle of the prefixed strong verb *ābrecan* 'to break, break up, break to pieces, break down, break off, separate forcibly, destroy; to assault, vanquish, to take by storm', namely *ābrecond*. The existence of another 150 adjectives formed by means of the attachment of the suffix *-lic* to an *-end* form, and of ninety-three *un-* prefixed derived adjectives based on adjectives attaching the suffix *-lic*, can be adduced as evidence for the following derivation: (*ge*)*brecan* > *ābrecan*, *ābrecond* > **ābrecondlic* > *unābrecondlic*.

The derived adjective *unhrædspræce* 'slow of speech' is morphologically related to the strong verb SPRECAN 'to speak, say; to utter, make a speech; to converse; to declare, tell off; to agree; to settle' [Gmc. *SPREC-A- (Seebold 1970, p. 457; Orel 2003, p. 366)], as well as its nominal derivative *spræce* 'talk, discourse'. A stepwise description of the formation of *unhrædspræce* calls for a hypothetical adjectival compound **hrædspræce* 'quick of speech', which can be justified on the grounds of similar instances of compounding. Indeed, four adjectival compounds have been found that present the adjective *hræd* in the leftmost position: *hrædmōd* 'hasty, quick-tempered', *hrædrīpe* 'ripe, premature, *hrædtæfle* 'quick at throwing dice', and *hrædwyrde* 'quick or hasty of speech'. Of these, *hrædmōd* and *hrædtæfle* take a noun as base, as *unhrædspræce* does. Consequently, the following derivation is proposed for *unhrædspræce*: *sprecan* > *spræce* > **hrædspræce* > *unhrædspræce*.

The derived adjective *unstydful* 'unstable, inconstant, apostate' is morphologically related to the primitive strong verb STANDAN 'to stand, occupy a place; to remain, continue, stand firm, stand up, be upheld; to reside, abide; to be valid, stand good; to be, exist, take place, last; to oppose; to oppress, attack, assail; to resist attack; to reprove; to stop, cease to move, stand still; to appear, arise, come; to be present to, come upon (of fear); to be fixed as a law or regulation; to urge; to seize' [Gmc. *STAND-A- (Seebold 1970,

p. 460; Orel 2003, p. 371)]. A hypothetical form is needed in order to propose a full stepwise derivation: **stydful* ‘stable’. The evidence that can be furnished to support this view comprises eighty-three adjectives based on nouns and attaching the suffix *-ful* like *glengful* ‘adorned’, *hearmful* ‘harmful’, *hepful* ‘severe’, etc., as well as fifteen *un-* prefixed derivatives based on adjectives suffixed with *-ful*, such as *unfremful* ‘unprofitable’, *ungeornful* ‘careless’, *unsideful* ‘immodest’, etc. The derivation can be stated as follows: *standan* > *stýde* > **stydful* > *unstydful*.

The derived adjective *ungewyldendlic* ‘impatient’ belongs in the paradigm of the primitive strong verb WEALDAN ‘to control, rule, direct, command, regulate, determine, ordain, have power or dominion over, bear sway, wield power; to govern; to wield (a weapon); to possess, have at command, be master of; to cause, bring about, author; to have power to do, be able’ [Gmc. **WALD-A-* (Seebold 1970, p. 536; Orel 2003, p. 443)]. Along with the prime, the weak verb *gewieldan* ‘to control, rule, have power or dominion over, compel, restrain; to subdue, make submissive, tame; to seize, conquer, take into one’s power; to temper; to secure; to force’; and the substantivised participle *wieldend* ‘subjugator; tamer’ can be found. On the basis of this evidence, and the abovementioned existence of 151 *-end-lic-* derived adjectives and ninety-three *un-* prefixed adjectives based on adjectives with the suffix *-lic*, the following gradual derivation is proposed: *(ge)wealdan* > *gewieldan*, *gewieldend* > **gewyldendlic* > *ungewyldendlic*.

The derived adverb *untōlāetendlice* ‘incessantly, unremittingly’ is related to the strong verb prime LĀETAN ‘to let, allow, permit; to remain, leave behind; to depart from, let alone; to think, consider, estimate, regard as; to leave undone; to cause to do; to behave, treat; to set free, let out; to assert, profess; to appear, pretend; to assign, allot; to let go, forsake, give up, dismiss; to desist’ [Gmc. **LĀET-A-* (Seebold 1970, p. 334; Orel 2003, p. 237)]. Considering that there are nineteen *-end-lic* adverbs related to verbs of the type *āgendlice* ‘properly’ *strūdgendlice* ‘greedily’, and *swīgiendlice* ‘silently’, and, as has been pointed out above, forty-eight adverbial derivatives with *-lice* attached to the prefix *un-*, the hypothetical derivative **tōlāetendlice*, based on the attested verb *tōlāetan* ‘to disperse; to let go, release, cause to go; to relax’, is proposed, so that the gradual derivation of the adverb is *lāetan* > *tōlāetan*, *tōlāetend* > **tōlāetendlice* > *untōlāetendlice*.

6.4. Long-Distance Morphological Relations and Undefined Primitives

The previous sections have presented the stepwise derivation of *un-* formations to which an immediate base can be assigned, on the conditions that a maximum of one hypothetical form is proposed and that such a hypothetical form is not the starting point of the derivation. This section lists the complex adjectives and adverbs whose gradual derivation entails more than one hypothetical form and, consequently, the lexical primitive, rather than the full gradual derivation, is provided. This section also renders the inventory of derivatives for which no derivation can be defined. They are presented along with the other members of the derivational paradigms to which they belong.

The full inventory of derivatives that require more than one hypothetical form includes the adjectives *unācnycendlic* ‘that cannot be untied or loosened; indissoluble’ (~CNYCC ‘bond’ [Gmc. **KNUTTŌN* (Orel 2003, p. 237)]); *unābindendlic* ‘indissoluble’ (~BINDAN ‘to bind’ [Gmc. **BEND-A-* (Seebold 1970, p. 102; Orel 2003, p. 41)]); *unāfeohtendlic* ‘inevitable, not to be overcome or contented’ (~FEOHTAN ‘to fight’ [Gmc. **FEHT-A-* (Seebold 1970, p. 190; Orel 2003, p. 96)]); *unāhefendlic* ‘unbearable, insupportable’ (~HEBBAN ‘to heave’ [Gmc. **HAFJ-A-* (Seebold 1970, p. 245; Orel 2003, p. 149)]); *unāscended* ‘unharmful, unharmed, unhurt’ (~SCAND ‘shame, disgrace’ [Gmc. **SKANDŌ* (Orel 2003, p. 334)]); *ungefērenlic* (~FARAN ‘to fare’ [Gmc. **FAR-A-* (Seebold 1970, p. 186; Orel 2003, p. 93)]); *ungelæccendlic* ‘unreprovable, irreprehensible’ (~LĀCAN ‘to fight, contend’ [Gmc. **LAIK-A-* (Seebold 1970, p. 321; Orel 2003, p. 231)]); *ungrāpigende* ‘not grasping’ (~GRĪPAN ‘to take, seize, grasp’ [Gmc. **GREIP-A-* (Seebold 1970, p. 237; Orel 2003, p. 143)]); *unsamwōræde* ‘contrary, opposed; incongruous’ (~WRĀED ‘union’ [Gmc. **WREIT-* (Pokorny 1959–1969, vol. 1, p. 1159)]); *unandwendlic* ‘unchangeable, immovable; unceasing’ (~WINDAN ‘to fly, leap,

spring; to start; to wheel; to waver, swing; to twist, roll; to wave, brandish' [Gmc. *WEND-A- (Seebold 1970, p. 554; Orel 2003, p. 454)]; *untamcol* 'untameable; invincible' (~TAM 'tame, tractable, gentle, mild' [Gmc. *TAMAZ (Heidermanns 1993, p. 589; Orel 2003, p. 401)]); *unđurhscēotendlic* 'impenetrable (~SCĒOTAN 'to shoot; to hit, strike, push, thrust, press forward; to move quickly' [Gmc. *SKEUT-A- (Seebold 1970, p. 417; Orel 2003, p. 339)]); *unāsēđendlic* 'insatiable' (~SÖĐ 'sooth, true, very; just, righteous; real, genuine' [Gmc. *SANĀPAZ (Orel 2003, p. 319)]); as well as the adverbs *unāgæledlice* 'unremittingly' (~GĀL 'pleasant; licentious; wicked; proud' [Gmc. *GAILAZ (Heidermanns 1993, p. 226; Orel 2003, p. 122)]), *unoflinnedlice* 'unceasingly, without desisting or leaving off (~LINNAN 'to cease from, leave off, desist; to yield up; to part from; to lose' [Gmc. *LENN-A- (Seebold 1970, p. 331; Orel 2003, p. 246)]).

No primitive can be found for the gradual derivation of *untēorig* 'untiring; unceasing'. It belongs in a derivational paradigm that comprises *ātēorian* 'to fail, cease, leave off, come to an end; to become exhausted or weary, faint; to be defective'; *ātēorigendlic* 'failing, transitory, perishable, fleeting; defective'; *ātēorodnes* 'cessation; exhaustion'; *ātēorung* 'failing, fainting; exhaustion'; *(ge)tēorian* 'to tire, weary, exhaust; to be tired or exhausted, become weary; to faint, fail, cease, perish, come to an end'; *(ge)tēorigendlic* 'exhausted, failing'; *(ge)tēorung* 'weariness; fainting, failing'; *tēorodnes* 'weariness; fainting, exhaustion'; *unātēoriende* 'unwearying, indefatigable'; *unātēorigendlic* 'permanent; unwearied, indefatigable'; *unātēorigendlice* 'indefatigably; unceasingly; without failing'; *unātēorod* 'unwearied, unceasingly; without failing'; *ungetēorigendlic* 'inexhaustible, unceasingly; without failing'; *ungetēorigendlice* 'indefatigably; unceasingly; without failing'; *ungetēorod* 'unwearied, unceasingly, unceasingly; without failing'; and *untēorig* 'untiring; unceasing'. The best candidate for lexical prime is the verb *(ge)tēorian*, but weak verbs are derived from nouns, adjectives, and other verbs and can hardly be considered primitives of lexical description. The derived adjective *unđoligendlic* 'intolerable' raises the same issue. No lexical prime is available because all the linked items display affixes, although the morphologically related weak verb *đolian* 'to suffer' belongs in the derivational paradigm consisting of *āđolian* 'to suffer, endure; to hold out, sustain'; *āđylgian* 'to bear up, sustain; to be patient, wait patiently'; *đolung* 'passion'; *forđolian* 'to be deprived of, lack, want; to go without'; and *(ge)đolian* 'to undergo, bear, suffer, endure, sustain; to bear with, tolerate; to allow; to persevere, hold out, survive; to continue; to lose, lack, forfeit, dispense with, be deprived of; to remain, wait; to stop; to stick, cleave'. Finally, no etymon has been found that corresponds to the nominal lexical prime MOLSAN 'decay, corruption', as represented by *unformolsniendlic* 'undecaying, incorruptible'.¹

6.5. Results

This section discusses the results both on the descriptive and on the theoretical side.

Out of the 33 hapaxes, only 3 are completely unrelated in the context of the whole lexicon: *untēorig*, *unđoligendlic* and *unformolsniendlic*. Some suggestions have been made but further research is needed in these derivatives. Of the remaining 30, 15 have been annotated for both short-distance and long-distance morphological relatedness while another 15 have been annotated for long-distance morphological relatedness only. It is worth pointing out in this respect that all long-distance derivatives are adjectives because the prefix *un-* also forms nouns, verbs, and adverbs. Type frequency has restricted the generation of hypothetical steps of derivation, of which a maximum of one has been allowed per derivation. The lexical prime has linked the two levels of morphological relatedness to the etymon provided by authoritative etymological dictionaries of Germanic.

Figure 3 summarises the results of the analysis. The columns LEMMA, GLOSS, and MORPHREL of the CoNNL-U Plus format are presented. On the MORPHREL column, the short-distance morphological relatedness precedes the long-distance morphological relatedness. Lexical primes are rendered in capitals. The asterisk marks hypothetical steps of gradual derivation. The \emptyset symbol indicates that no short-distance relatedness has been found. In compounds, relatedness is acknowledged by base of compounding.

LEMMA	GLOSS	MORPHREL
unbecēas	indisputable	*becēas/CĒAS
unlȳfendlic	unlawful	*lȳfendlic/LĒAF
unscamig	unashamed	*scamig/SCAMU
unsefful	senseless	*sefful/SEFA
untellendlic	innumerable	*tellendlic/TALU
unandcȳðignes	ignorance	*oncȳðignes/CŪD
ungedæftnes	importunity	*gedæftnes/GEDAFEN
unleahtorwyrðe	unblameable	*leahtorwyrðe/WEORD
unornlic	mean	*ornlic/ORNE
untrāglice	honestly	*trāglice/TRĀG
unābrecendlic	inextricable	*ābrecendlic/BRECAN
unhrædspræce	slow of speech	*hræd-spræce/SPRECAN
unstydful	unstable	*stydful STANDAN
ungewyldendlic	impatient	*gewyldendlic/WEALDAN
untōlætendlice	incessantly	*tōlætendlice LĒTAN
unācnycendlic	indissoluble	∅/CNYCC
unābindendlic	indissoluble	∅/BINDAN
unāfeohtendlic	inevitable	∅/FEOHTAN
unāhefendlic	unbearable	∅/HEBBAN
unāscended	unharmed	∅/SCAND
ungefērenlic	unbearable	∅/FARAN
ungelæccendlic	irreprehensible	∅/LĀCAN
ungrāpigende	not grasping	∅/GRĪPAN
unsamwrāde	contrary	∅/WRĒD
unandwendlic	unchangeable	∅/WINDAN
untamcol	untameable	∅/TAM
unðurhscēotendlic	impenetrable	∅/SCĒOTAN
unāsēðendlic	insatiable	∅/SŌD
unāgāledlice	unremittingly	∅/GĀL
unoflinnedlice	unceasingly	∅/LINNAN

Figure 3. LEMMA, GLOSS and MORPHREL fields in the CoNLL-U Plus format.

On the theoretical side, the main results of the analysis presented in this section have to do with the status of the primitives of description and the stepwise character of derivations.

Primitive nouns, adjectives, and verbs constitute a set of anchor terms that can be used for lemmatisation, normalization, and glossing. Considering entries by derivational paradigm can also ease the tasks of editing the dataset. More importantly, lexical primitives link synchronic availability to diachronic recoverability by associating the base of the lexical derivation to the etymon via the lexical prime. Short-distance and long-distance morphological relations are, therefore, considered side by side. For instance, the stepwise derivation of the adjective *unābrecendlic* ‘inextricable’ (*brecan* > *ābrecan*, *ābrecend* > **ābrecendlic* > *unābrecendlic*) relates this derivative, on the synchronic axis of analysis, to the strong verb base and, on the diachronic axis, to Germanic BREK-A (Seibold 1970, p. 132; Orel 2003, p. 55).

Explicit stepwise lexical derivations can be defined on the basis of lexical paradigms. Derivations with gaps could of course be presented, but the inclusion of hypothetical predicates maximises paradigmatic generalisation, as the evidence furnished in favour of the hypothetical derivatives included in the derivations shows. For example, the derived adjective *unsefful* ‘senseless, irrational’, when considered in the derivational paradigm of *sefa* ‘mind, understanding; spirit, heart’, is related to the compounds *brēostsefa* ‘heart, mind’, *ferhðsefa* ‘mind, thought, intellect’, *mōdsefa* ‘soul, spirit, mind, thought, imagination, heart; purpose; character’, and *wīssefa* ‘wise-minded or souled person’. However, when a gradual derivation containing a hypothetical derivative is defined, *unsefful* is related to eighty-three adjectives resulting from the attachment of the suffix *-ful* to a noun like *glengful* ‘adorned’, *hefful* ‘severe; grievous’, and *mæðful* ‘humane, courteous’, and eighteen adjectives prefixed

with *un-* and based on adjectives attaching the suffix *-ful*, such as *unbealaful* ‘innocent’, *unfremful* ‘unprofitable’, and *ungeornful* ‘careless’.

7. Conclusions

This article has addressed the question of how to account for the generalised derivational relatedness of Old English in a type of treebank originally designed for stressing syntactic similarities across languages. The main thrust of the article is that the annotation of bases of derivation and derivatives within a hierarchical model of morphology reinforces the paradigmatics of UD and, above all, increases the power of generalisation of the framework. The annotation for derivational morphology and gloss facilitates lexical semantics studies that can focus on the syntax of certain meanings or sets of morphologically related words. It also allows us to carry out diachronic studies that involve Old English and Present-Day English, and comparative studies of Old English with other historical languages or, through the gloss annotation, with natural languages available from UD treebanks.

From the descriptive point of view, the solution of morphologically unrelated derivatives has been addressed. In the CoNLL-U Plus format, the MORPHREL column displays three types of information: availability of base of derivation, lack of base of derivation, and lexical prime. The lexical prime links the two levels of morphological relatedness to the etymon, given that the lexical paradigm, gathered around the prime, also contains the related lexical items. The generation of hypothetical steps of derivation, of which a maximum of 1 has been allowed in each derivation, has been restricted on the basis of type frequency. A total of 30 out of 33 hapaxes have been fully annotated with this method and format.

On the theoretical side, this discussion has delved into the question of derivational relatedness (made explicit by means of derivational paradigms based on synchronic productivity and diachronic recoverability). It has been shown that a rich description of word formation that includes short-distance as well as long-distance relatedness allows the annotator to gather derivational paradigms, stresses some points of contact of word formation and etymology, and takes steps towards bridging the gap between synchronic and diachronic description. Furthermore, a paradigmatic model based on lexical primes and gradual derivation has two main advantages. Firstly, explicit stepwise lexical derivations can be defined on the basis of lexical paradigms. Secondly, primitive nouns, adjectives, and verbs constitute a set of anchor terms that can be used for lemmatisation, normalisation, and glossing.

Funding: The research reported in this article has been funded through grant IPID2020-119200GB-100, funded by MCIN/ AEI/10.13039/501100011033/, which is gratefully acknowledged.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Research data will be available at <https://investigacion.unirioja.es/investigadores/112/publicaciones>. See also www.nerthusproject.com/search-nerthus, accessed on 12 October 2023.

Acknowledgments: This research has been funded by the National Agency of Research (Ministry of Science and Innovation, Government of Spain), which is gratefully, which is gratefully acknowledged.

Conflicts of Interest: The authors declare no conflict of interest.

Note

¹ Cf. **malta* ‘soft, gone bad (?)’ (Kroonen 2013, p. 351).

References

- Baayen, R. Harald. 2008. *Analyzing Linguistic Data. A Practical Introduction to Statistics*. Cambridge: Cambridge University Press.
- Baayen, R. Harald. 2009. Corpus linguistics in morphology: Morphological productivity. In *Corpus Linguistics. An International Handbook*. Edited by Anke Lüdeling and Merja Kytö. Berlin: Mouton De Gruyter, pp. 900–19.

- Bammesberger, Alfred. 1965. *Deverbative jan-Verba des Altenglischen, Vergleichend mit den Übrigen Altgermanischen Sprachen Dargestellt*. München: Ludwig-Maximilians Universität.
- Bosworth, Joseph, and Thomas N. Toller. 1973. *An Anglo-Saxon Dictionary*. Oxford: Oxford University Press.
- Brinton, Laurel, and Elizabeth Closs Traugott. 2005. *Lexicalization and Language Change*. Cambridge: Cambridge University Press.
- Buchholz, Sabine, and Erwin Marsi. 2006. CoNLL-X shared task on Multilingual Dependency Parsing. Paper presented at the 10th Conference on Computational Natural Language Learning (CoNLL-X), New York City, NY, USA, June 8–9; pp. 149–64.
- Clark Hall, John R. 1986. *A Concise Anglo-Saxon Dictionary*. With a supplement by Herbert D. Merritt. Toronto: University of Toronto Press.
- de Marneffe, Marie C., Christopher Manning, Joakim Nivre, and Daniel Zeman. 2021. Universal Dependencies. *Computational Linguistics* 47: 255–308. [CrossRef]
- Ettmüller, Ludwig. 1968. *Lexicon Anglosaxonicum*. Basse: Quedlinburg & Leipzig.
- Haselow, Alexander. 2011. *Typological Changes in the Lexicon: Analytic Tendencies in English Noun Formation*. Berlin: De Gruyter Mouton.
- Healey, Antonette. 2012. *The Dictionary of Old English Web Corpus*. With John Price Wilkin and Xin Xiang. Toronto: Dictionary of Old English Project, Centre for Medieval Studies, University of Toronto. Available online: <http://www.doe.utoronto.ca/pub/webcorpus.htm> (accessed on 7 January 2023).
- Healey, Antonette. 2018. *The Dictionary of Old English: A to I*. Toronto: Dictionary of Old English Project, Centre for Medieval Studies, University of Toronto.
- Heidermanns, Frank. 1993. *Etymologisches Wörterbuch der Germanischen Primäradjektive*. Berlin: Walter de Gruyter.
- Hiltunen, Risto. 1983. *The Decline of the Prefixes and the Beginnings of the English Phrasal Verb*. Turku: Turun Yliopisto.
- Hinderling, Robert. 1967. *Studien zu den Starken Verbalabstrakten des Germanischen*. Berlin: Walter de Gruyter.
- Kastovsky, Dieter. 1968. *Old English Deverbal Substantives Derived by Means of a Zero Morpheme*. Esslingen: Bruno Langer Verlag.
- Kastovsky, Dieter. 1992. Semantics and vocabulary. In *The Cambridge History of the English Language I: The Beginnings to 1066*. Richard Hogg. Cambridge: Cambridge University Press, pp. 290–408.
- Kastovsky, Dieter. 2006. Typological Changes in Derivational Morphology. In *The Handbook of The History of English*. Edited by Ans van Kemenade and Bettelou Los. Oxford: Blackwell, pp. 151–77.
- Kroonen, Guus. 2013. *Etymological Dictionary of Proto-Germanic*. Leiden: Brill.
- Lass, Roger. 1994. *Old English. A Historical Linguistic Companion*. Cambridge: Cambridge University Press.
- Martín Arista, Javier. 2011. Projections and Constructions in Functional Morphology. The Case of Old English HRĒOW. *Language and Linguistics* 12: 393–425.
- Martín Arista, Javier. 2012. Lexical database, derivational map and 3D representation. *RESLA-Revista Española de Lingüística Aplicada* 1: 119–44.
- Martín Arista, Javier. 2013. Recursivity, derivational depth and the search for Old English lexical primes. *Studia Neophilologica* 85: 1–21. [CrossRef]
- Martín Arista, Javier. 2019. Another Look at Old English Zero Derivation and Alternations. *ATLANTIS* 41: 163–82. [CrossRef]
- Martín Arista, Javier, Sara Domínguez Barragán, Laura García Fernández, Esaúl Ruíz Narbona, Roberto Torre Alonso, and RaquelVea Escarza. 2021. *ParCorOE2. An open access annotated parallel corpus Old English-English*. Logroño: Nerthus Project, Universidad de La Rioja. Available online: www.nerthusproject.com (accessed on 10 October 2023).
- McDonald, Ryan T., Joakim Nivre, Yvonne Quirnbach-Brundage, Yoav Goldberg, Dipanjan Das, Kuzman Ganchev, Keith Hall, Slav Petrov, Hao Zhang, Oscar Täckström, and et al. 2013. Universal Dependency Annotation for Multilingual Parsing. Paper presented at the 51st Annual Meeting of the Association for Computational Linguistics, Sofia, Bulgaria, August 4–9; pp. 92–97.
- Metola Rodríguez, Darío. 2017. On the applicability of the dictionaries of Old English to linguistic research. *Journal of English Studies* 15: 173–91. [CrossRef]
- Nichols, Johanna. 2014. Derivational paradigms in diachrony and comparison. In *Paradigm Change: In the Transeurasian Languages and Beyond*. Edited by Martine Robbeets and Walter Bisang. Amsterdam: John Benjamins, pp. 61–88.
- Nivre, Joakim. 2016. Universal Dependencies: A Cross-Linguistic Perspective on Grammar and Lexicon. Paper presented at the Workshop on Grammar and Lexicon: Interactions and Interfaces, Osaka, Japan, December 11; pp. 38–40.
- Nivre, Joakim, Marie C. de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajič, Christopher D. Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, and et al. 2016. Universal Dependencies v1: A multilingual treebank collection. Paper presented at the 10th International Conference on Language Resources and Evaluation, Portorož, Slovenia, May 23–28; pp. 1659–66.
- Nivre, Joakim, Marie C. de Marneffe, Filip Ginter, Jan Hajič, Christopher D. Manning, Sampo Pyysalo, Sebastian Schuster, Francis Tyers, and Daniel Zeman. 2020. Universal Dependencies v2: An evergrowing multilingual treebank collection. Paper presented at the Twelfth International Conference on Language Resources and Evaluation (LREC 2020), Marseille, France, May 11–16; pp. 4027–36.
- Orel, Vladimir. 2003. *A Handbook of Germanic Etymology*. Leiden: Brill.
- Petrov, Slav, Dipanjan Das, and Ryan McDonald. 2012. A Universal Part-of-Speech Tagset. Paper presented at the Eighth International Conference on Language Resources and Evaluation (LREC'12), Istanbul, Turkey, May 21–27; pp. 2089–96.
- Pintzuk, Susan, and Lendert Plug. 2001. *The York-Helsinki Parsed Corpus of Old English Poetry*. York: Department of Language and Linguistic Science, University of York.
- Pokorny, Julius. 1959–1969. *Indogermanisches Etymologisches Wörterbuch*. 2 vols. Bern: Francke.

- Rissanen, Matti, Merja Kytö, Leena Kahlas-Tarkka, Matti Kilpiö, Saara Nevanlinna, Irma Taavitsainen, Terttu Nevalainen, and Helena Raumolin-Brunberg. 1991. *The Helsinki Corpus of English Texts*. Helsinki: Department of English, University of Helsinki.
- Seebold, Elmar. 1970. *Vergleichendes und Etymologisches Wörterbuch der Germanischen Starken Verben*. The Hague: Mouton.
- Stark, Detlef. 1982. *The Old English Weak Verbs. A Diachronic and Synchronic Analysis*. Tübingen: Max Niemeyer Verlag.
- Sweet, Henry. 1976. *The Student's Dictionary of Anglo-Saxon*. Oxford: The Clarendon Press.
- Taylor, Ann, Anthony Warner, Susan Pintzuk, and Frank Beths. 2003. *The York-Toronto-Helsinki Parsed Corpus of Old English Prose*. York: Department of Language and Linguistic Science, University of York.
- Trips, Carola. 2009. *Lexical Semantics and Diachronic Morphology. The Development of -Hood, -Dom and -Ship in the History of English*. Tübingen: Max Niemeyer Verlag.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.