Why Jesus and Job spoke bad Welsh : the origin and distribution of V2 orders in Middle Welsh

Meelen, M.

CHAPTER 8

Conclusions

In this thesis I aimed to address the question of the puzzling observations in Middle Welsh word order. First of all, the most-frequently found patterns involve verb-second order. This is 'abnormal' from a Modern Welsh preferred VSO point of view. A further puzzling fact is the large number of possible word orders in Middle Welsh. The verb-second orders alone can take various forms with the sentence-initial constituent and the agreement pattern as the main variables. Finally, it is unclear where these verb-second orders come from, because the limited amount of data available for older stages of the language suggests that sentences with verb-initial orders were more commonly used. In this study, I therefore tried to answer two crucial questions:

1. How can we explain the distribution of the various word order patterns in Middle Welsh?

2. Where do the various verb-second orders (including those with and without subject-verb agreement) come from?

To a certain extent, these questions have been "vexed" and are "by now tormented" by various Welsh scholars in the past decades (see Chapters 1 and 4 in particular). Much progress was made over the years, but there we still find variation in Middle Welsh word order that "frustratingly defies easy explanation" (Poppe, 2014:73). I argue that there are two ways to solve this problem and that we need both if we want to make significant progress in elucidating obscure patterns in word order variation found in any (historical stage of a) language. We first of all need a (large)

digitised corpus that is morpho-syntactically annotated. In second place, we need a consistent methodology to analyse information-structural (or any other) notions that can influence the order of the words in a sentence. Apart from answering the above questions for Middle Welsh, this thesis furthermore presents a sound methodology on how to approach word order phenomena in historical corpora.

In Chapter 2 I formulated my arguments for the use of annotated corpora in more detail. When conducting historical linguistic research, in particular syntactic research, we can only rely on the distribution of the different forms and constructions that we can find. The extent to which our observations reflect the language at the time is likely to increase when we use larger corpora. If a particular pattern occurs often in one text, we cannot jump to the conclusion that this is the case in all textual evidence. Exactly because the amount of extant data is extremely limited, we must try and retrieve the most information we possibly can. This can be achieved by providing detailed part-of-speech tags. This elaborate morpho-syntactic annotation helps to automatically extract the necessary linguistic information from the corpus. Ideally, we create an annotated corpus containing all extant texts, but building such a corpus is a tremendous task. For the present study, I took the first steps on the way to create a fully annotated Treebank of Middle Welsh by selecting, preprocessing, tagging, correcting and parsing 15 texts from the early to the late Middle Welsh period.

I trained a memory-based part-of-speech (PoS) tagger to automatically assign morpho-syntactic tags to the Middle Welsh texts. The choice of PoS-tagger was mainly based on the good results achieved with minimal preprocessing of the difficult data. The difficulty for any automated task lies mainly in the highly irregular orthography found in the Middle Welsh manuscripts and furthermore, the concept of initial consonant mutation found in all Insular-Celtic languages. I furthermore extended the conventional UPenn tagset tremendously to include highly detailed morpho-syntactic information that can facilitate much more future research. With a Global Accuracy of over 90%, the memory-based tagger performed reasonably well considering the difficult data and large tagset (consisting of $>$200 tags). The amount of time needed for subsequent manual correction was thus fairly limited. I then designed a rule-based grammar for Middle Welsh and used the NLTK regular expression parser to add phrase structure to the corpus based on the corrected PoS-tags. With an extremely detailed grammar and a double loop, the parser assigned hierarchical structures to the corpus. These automatic parses were again manually corrected and subsequently converted to bracketed formats to enable searches via CorpusSearch of XQuery facilitating any queries concerning word order patterns. The main result is a reasonably large corpus (15 texts) from which over 9,000 well-annotated positive declarative main clauses could be extracted. In the future, this corpus can be extended to include more texts from different genres, manuscripts and stages of the Welsh language.

In Chapter 3 I outlined a consistent methodology for the investigation of information structure in historical corpora. I discussed three core information-structural notions in detail: Givenness, Topic (vs. Comment) and Focus (vs. Background). I outlined their main characteristics in a systematic way so that they can be used to annotate a corpus consistently. I annotated the referential status of subjects and objects (i.e. their 'Givenness') in the Middle Welsh corpus according to the Pentaset developed by Komen (2013). In Chapter 5 I showed how this type of annotation can help identify effects in word order distributions in combination with annotated syntactic features. Concerning the second core information-structural notion of Topic, I identified three different kinds of topics in the Middle Welsh corpus: Aboutness, Contrastive and Familiar topics. In the next part of Chapter 3 I presented a detailed overview of different kinds of Focus structures including systematic 'algorithms' to find the right focus articulation (Presentational/Thetic, Predicate or Constituent Focus) and the numerous subtypes of Constituent Focus. I furthermore discussed two further notions that are relevant to information structure: Point of departure and Information Flow. The Principle of Natural information flow stipulates that old information usually precedes new information. In sentences with the reverse order, the 'flow' of information, or in particular the referential status of the core arguments, is 'marked'. This helps to give an accurate description of object-initial word orders in Middle Welsh, as I discussed in Chapter 5. Finally, the 'Points of Departure' of a sentence appear mainly in the form of temporal or circumstantial clauses. In effect, they function as frame setters delimiting the context of the rest of the sentence. The clear definitions and guidelines to find the right labels presented in this chapter facilitate annotation of large corpora. A consistent analysis in turn is indispensable for the type of research historical syntacticians are interested in.

Chapter 4 and 5 presented the data and core observations concerning Middle Welsh word order variation. In the compiled corpus, I found a large number of different word order patterns in positive declarative main clauses. I categorised them based on purely formal reasons into nine different main types:

I Verb-initial (VSO)
  (a) VSO (verb absolute clause-initial)
  (b) particle VSO

II Periphrastic constructions with initial auxiliary (AuxSVO)
  (a) with auxiliary *bod*
  (b) with auxiliary *gwneud*
  (c) with auxiliary *ddaru*

III Verb-second after adjuncts ('Abnormal Sentence')
  (a) AdjP *y* VSO
  (b) PredP *y* VSO
  (c) AspP *y* VSO

(d) AdvP *y* VSO
(e) PP *y* VSO

IV Verb-second after arguments and VNs ('Abnormal Sentence')
(a) S *a* V_{agree} O
(b) O *a* V S
(c) patient *a* $V_{impersonal}$
(d) VN *a* DO_{infl} (*gwneuthur*-periphrasis)

V Verb-second after focussed items ('Mixed Sentence')
(a) *(ys)* focussed noun/pronoun *a* V_{3sg}
(b) *(ys)* focussed adjunct *y* V_{3sg}

VI Bare verbal nouns
(a) VN + agent
(b) VN + *o* + agent
(c) *a(c)* VN (continuing previous finite clause)

VII Copular clauses
(a) SCP
(b) PCS
(c) CPS
(d) C S *yn* P
(e) C S *(ys)sydd* P

VIII Identificational Focus construction
(a) Sef + DP (+ relative)
(b) Sef + *yw/oed*
(c) Sef + *a/y*

IX Non-verbal clauses
(a) *dyma/dyna/llyma/llyna* + S (truncated copular clause)
(b) S *(yn)* P
(c) PS
(d) Absolutive: Ac S P(P)

Sentences with verb-initial word order are rare in Middle Welsh, although variants with sentence-initial conjunctions or declarative particles like *neu(r)* directly followed by the verb are found somewhat more frequently. The second type is a periphrastic construction with the auxiliary form of the verb *bod* 'to be', rendering AuxSVO order. This type is also rarely found. Its frequency increases towards the end of the Middle Welsh period. Word order Types I and II (VSO and AuxSVO) are the predominant patterns found in Modern Welsh. The verb-second pattern (the 'Abnormal Sentence') in one of its various forms (Types III, IV or even the focussed type V, the 'Mixed Sentence') is the most commonly found pattern in

Middle Welsh. The adjunct-initial order can appear in many forms and multiple adjuncts are possible too, as long as the 'topicalised' constituent functions as an adjunct. The other type of 'Abnormal Sentence', Type IV, on the other hand places a core argument (Subject or Direct Object) in sentence-initial position. A variant of this type consists of sentences with verbal nouns in initial position followed by the pre-verbal particle *a* and the auxiliary *gwneuthur* 'to do'. This type most commonly appears in contexts of narrative continuity. In subject-initial sentences, the verb usually agrees with the pre-verbal subject. This is what formally distinguishes the 'Abnormal Sentence' from the 'Mixed Sentence' in which the verb shows default third-person singular inflection (Type V). Sentences with verbal nouns *instead of* finite verbs (Type VI) were mainly possible in (Early) Middle Welsh. In early Middle Welsh texts such as *Culhwch*, the verbal noun could appear in non-finite main clauses on their own followed by the subject. These 'verbal noun + agent' almost disappear in independent main clauses. Only sentence-initial verbal nouns in co-ordinated sentences depending on preceding finite clauses continued to exist much longer. Types VII and VIII only describe sentences with copular verbs. The copula could also be left out in Middle Welsh. These non-verbal sentences were finally labelled as Type IX. If we leave out the copular clauses, we can see a clear trend in the distributional of the various word order patterns presented in rough chronological order in Figure 8.1.

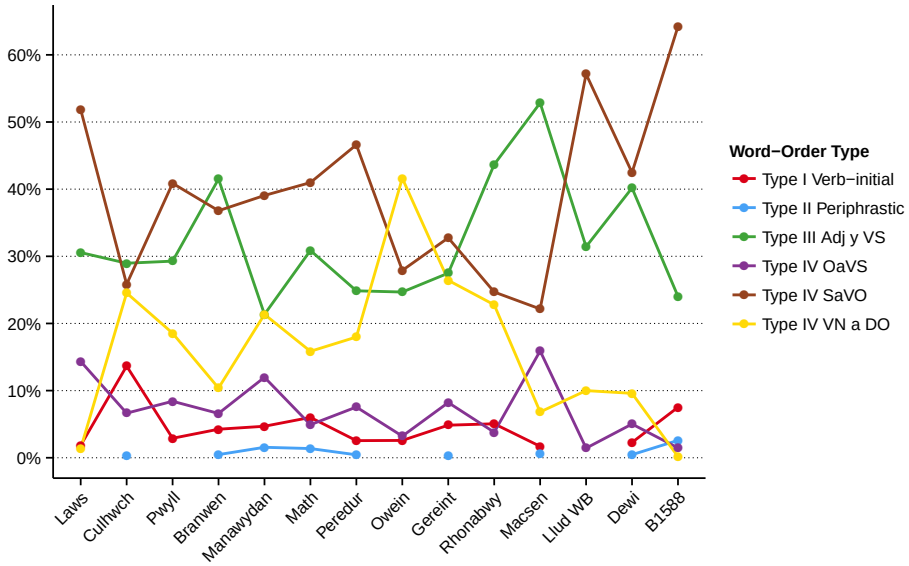

Figure 8.1: Main word order types per text from Early to Late Middle Welsh

It is clear from the above graph that language is already changing at the end of the Middle Welsh period. The preferred word order is still the verb-second 'Ab-

normal' order, but an overwhelming number of sentences are now subject-initial. Verb-initial orders (Type I) and in particularly auxiliary-initial periphrastic orders (Type II) are on the rise. The 1588 Bible translation is particularly interesting, because it was very influential. Most prose texts in Early Modern Welsh are of a religious nature written by people who were very familiar with this translation. As pointed out in the introduction, for the 19th-century Oxford reformers, it was "embarrassing" to hear Jesus and Job speak 'bad Welsh'. The prevalent V2 order in the 1588 translation is indeed different from the Modern Welsh V1-language they spoke. Interesting, however, from this study its becomes clear that the syntax and word order preferences in the 1588 Bible translation also differ from the general patterns in Middle Welsh. The clear preference for subject-initial sentences in the 1588 translation is not found earlier.

In Chapter 5 I systematically presented all possible factors that could influence the word order of the Middle Welsh sentence. Starting with possible grammatical factors, verb-second sentences with verbal nouns in initial position (Type IVc VNaDO) almost exclusively occur with verbs in the preterite tense. The significance of (preterite) tense as a factor is likely to be related to the fact that these verbal-noun patterns are the basic word order in indirect speech passages of narrative tales. In direct speech, on the other hand, subject-initial orders are most frequently attested. The corpus study furthermore shows that impersonal verbs are most frequently found in verb-second sentences with initial adjuncts (Type III). Finally, there seems to be a limited role for Animacy of the core constituents. For subjects, there are no significant results, but inanimate objects tend to appear in object-initial orders more frequently than expected.

Only once all language-internal and language-external factors are systematically tested in this way (to the extent this is possible with the information we have), we can determine whether other factors, such as information-structural notions play a role. The first information-structural notion under investigation was Givenness. Direct objects in initial position almost exclusively convey New information. This indicates that the 'Natural information flow' of the sentence (going from old to new) is reversed and these object-initial sentences are thus marked in this way. The only exceptions to this generalisation are so-called Familiar topics. These are topics that appear in sentence-initial position mainly in the form of demonstrative pronouns. They refer back to the last-mentioned item/person/concept in the immediately preceding context. The corpus study revealed two further observations in terms of textual cohesion. First of all 'points of departure' or frame-setters occur most often in verb-second sentences with adjunct-initial order (Type III) in which they function as the topic. A second observation in this context concerns textual continuity. In order to achieve close cohesion, verbal nouns can be placed in sentence-initial position. They are either relying on an inflected verb in the previous sentence (Type VI) or are continued with an inflected form of the auxiliary 'to do' (Type IVc). Again this is likely to be part of the narrative style in this genre. Finally, focus can first of all be observed in the dedicated (reduced) cleft order called the 'Mixed Sentence'

(Type V). Focus of the identificatory predicate can furthermore be found in the special *sef*-construction (Type VIII), but not all sentences with *sef* are focussed.

Chapter 6 and 7 focussed on the synchronic and diachronic syntactic analysis of the different word order patterns. In Chapter 6 I presented four different case studies related to the most important information-structural features in Middle Welsh. The aim of this chapter was to provide a syntactic analysis for those information-structural phenomena and to see how notions like topic, focus and givenness are implemented in the syntax of the language. Middle Welsh only allowed one topic position, but V3 and even V4 structures are attested. In the discussion I mainly focussed on the puzzling variation and agreement observations in the verb-second 'Abnormal Sentence'. Two different types of analyses were presented and discussed in detail: a movement and a base-generated approach. I argued that agreement with sentence-initial plural DP topics can be explained by adopting a base-generated approach, but not by a movement approach. The topic is base-generated in the left periphery of the clause, but it is co-indexed with a minimal pronoun lower down in the structure. The φ-features on the verb can be checked in the C-domain via λ-predication in the same way this is possible in relative clauses (cf. Kratzer (2009)). The lack of agreement is due to an operator that moves to SpecCP as a remnant of the reduced cleft in an earlier stage of the language. A movement approach (but not a base-generated approach) can account for sentences without subject-verb agreement and I argued that in some particularly difficult coordinate structures exhibiting both plural and singular agreement, a mixed analysis is the best solution. In general, however, movement approaches create problems for sentence-initial subjects, because Middle Welsh seems to adhere to the Complementarity Principle. According to this principle (that is also found in Breton), any form of agreement with plural full DPs is unexpected. The same holds for the lack of agreement with focalised pronouns. Both of these observed structures thus present problems for a movement analysis. Under a base-generated approach, however, these different agreement patterns can be explained. There are, however, also examples that present a greater challenge for a base-generated analysis, such as sentence-initial constituents that must be (locally?) bound by a quantifier and argumental PPs. The Middle Welsh corpus most likely reflects two possible patterns: movement and base-generation.

In the final chapter I discussed various approaches to the study of diachronic syntax, including socio-linguistic variationist, construction grammar and generative approaches. I argued that adopting an generative acquisitional framework has various benefits in the study of Middle Welsh diachronic syntax, because it allows us to use insights from synchronic studies on variation in syntax. The tools and mechanisms tested within the Minimalist Program can furthermore help us define the exact conditions and context in which innovations can and cannot occur and how they can trigger any subsequent changes.

I presented two case studies of syntactic change in the history of Welsh to demonstrate this. The first of these case studies is concerned with a very specific type of focus strategy: identificatory predicate focus. I showed how this construction arose from the cleft construction still found in Old Welsh and led to the emergence of the focus marker *sef*. When the focussed interpretation was lost, *sef* was reinterpreted as an expletive and, finally, as a linker in reformulative appositional structures ("i.e.").

In the second case study I addressed the second research question of the present study. I showed how the the verb-second orders came into existence in Middle Welsh by careful comparison with other Celtic languages. I focussed on the reconstruction of the functional particles in the C-domain that can still be found in the Brythonic languages. I then described the further developments of reanalysis of hanging topics and relative clauses (the 'Mixed Sentence') and the extension of information-structural functions leading to the postulation of a generalised Edge Feature on the C-head. On the basis of further possible sentence types like the periphrastic construction with the auxiliary *bod* 'to do' in Middle Welsh, I further argued that this Edge Feature must be on a lower C-head, $C_{\texttt{Fin}}$. The phonological erosion of the C-particles in the Early Modern Welsh period eventually resulted again in the loss of V2. Finally, I put these diachronic developments in a wider cross-linguistic context and sketched a tentative feature hierarchy for word order patterns including V2:

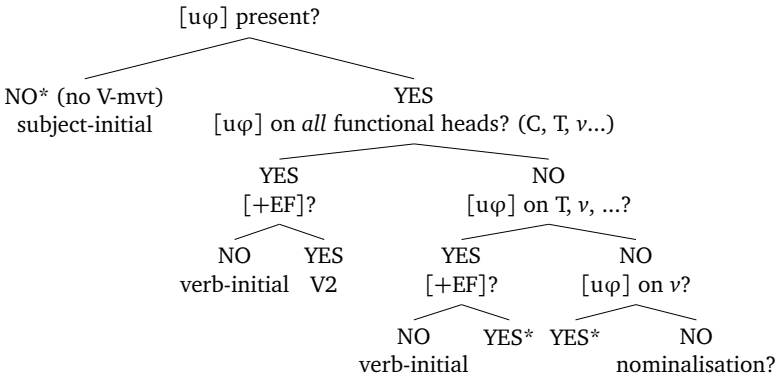

Figure 8.2: Hierarchy for verb-movement via [uφ], including [+EF] yielding V2

This present study finally aimed to investigate the interaction between syntax and information structure and their respective (or combined) effects on word order. From a synchronic point of view, the distribution of word order patterns in Middle Welsh is the result of a combination of both grammatical and information-structural factors. Focus was expressed with a reduced cleft construction, the so-called 'Mixed Order'. In identificatory copular clauses, however, focus was expressed by means of the focus marker *sef* (< *ys* + *ef* 'it is that'). Givenness and textual cohesion

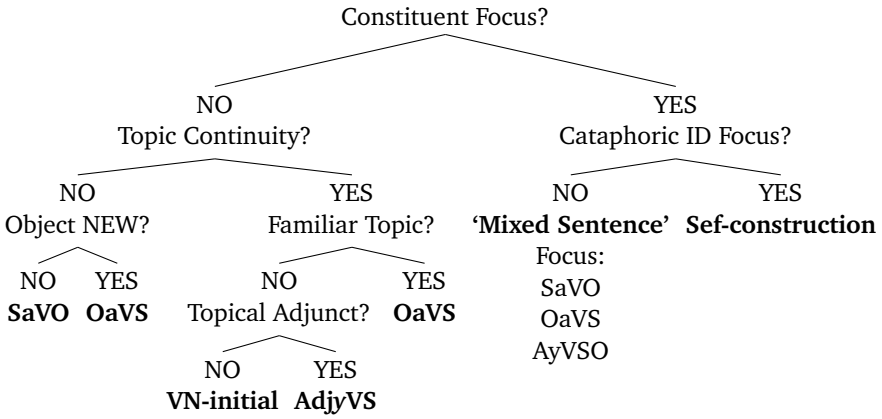

Figure 8.3: Decision algorithm 'predicting' the word order pattern in Middle Welsh

furthermore play a role as described above. Based on the present corpus study,[1] we can establish an algorithm to 'predict' or 'choose' the right option from the wide range of possible word orders; in this way the exact distribution of the various patterns can be explained. With the intended message ready in the Numeration, the syntax can build the sentence that will ultimately yield one of the word order types. In transitive statements in narrative contexts, the basic decision-making algorithm in figure 8.3 can 'predict' the word order (leaving additional adjuncts and extra-clausal constituents like hanging topics aside):

From a diachronic point of view, I showed that information-structural features play a role in syntactic innovations and reanalyses. The extension of information-structural functions of the sentence-initial constituent in verb-second sentences in Middle Welsh (from Contrastive Focus $>$ Contrastive Topics and New Information Focus $>$ Familiar and Aboutness topics) is a good example of this. The ultimate triggers for syntactic changes sometimes remain hard to define, but a detailed and consistent description of the synchronic variation systematically checking different variables as presented in this study is indispensable in diachronic syntactic research.

[1] Needless to say if the corpus is extended with more Middle Welsh texts a similar study needs to be conduct to see if we still arrive at the same result with the extended dataset. I leave this for future research