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

Citation
Meelen, M. (2016, June 21). *Why Jesus and Job spoke bad Welsh : the origin and distribution of V2 orders in Middle Welsh*. *LOT dissertation series*. LOT, Utrecht. Retrieved from https://hdl.handle.net/1887/40632

Cover Page

Author: Meelen, M.
Title: Why Jesus and Job spoke bad Welsh : the origin and distribution of V2 orders in Middle Welsh
Issue Date: 2016-06-21

Appendix - Annotation Manual

1   Introduction

This brief manual describes the guidelines used to add part-of-speech and phrasal annotation to the corpus of Middle Welsh prose. This corpus was initially built for the present investigation in Middle Welsh word order. The focus therefore lies on facilitating queries concerning word order. These query codes are presented at the end of this Appendix.

1.1   Philosophy and goals

The main aim of this project was not to give a correct syntactic analysis or provide a detailed parsed structure. The part-of-speech tags contain highly detailed morphological information, but the phrasal annotation is only a slightly more elaborate shallow parse. In this way, the annotated corpus could remain theory-neutral. At the same time, queries for linear order and hierarchical phrase structure are still possible. And finally, future enrichment of the chunk-parsed corpus is not excluded, because of its flexible XML-format.

Any controversial decisions are avoided as much as possible. The same goes for constructions that are changing over time. A good example is the *sef*-construction in Welsh. The information-structural status of this sentence changes from initial identificatory focus to plain predicate focus in the course of the Middle Welsh period. Since most texts are difficult to date exactly, throughout the corpus I used the specific tag SEF for any occurrence of this type of sentence. In this way, all these sentences can be easily found and investigated by future researchers as well.

1.2   File formats

All mark-up is stripped from the texts, which are then saved as plain text files (.txt). Further preprocessing involved the insertion of sentence-final punctuation (if that was not present in the manuscript already) and the deletion of sentence-internal full

stops (in order to make it readable to the PoS tagger). Finally, utterance boundaries were inserted semi-automatically (automatically after a full stop and manually if the full stop did not exist in the manuscript). The PoS-tagged files created by the Memory-Based Tagger include tags to words in the following fashion: `word/TAG`. These tagged files are saved as text files as well.

The NLTK regular expression chunkparser requires a list of words and tags. Therefore, the PoS-tagged text files were converted to the right format using the script in Figure 4. Chunk-parsed files contain bracketed structures representing phrasal and morphological annotation. The plain text files in this format are thus parsed (.psd). These types of parsed files are searchable with CorpusSearch and other querying tools. The Cesax Software package designed by Erwin Komen (cf. Komen (2013)) converts text and psd files into xml-format (.psdx). In this way, corpus searches are also possible via XQuery.

```
import sys
import os
import re

def make_nltk_readable(file_name):
"""
function takes one argument (file_name), and returns a list
containing (for every sentence) a list of word-pos pairs
"""

all_text = open(file_name)
corpus = []

for line in all_text:
    sentence = []
    pairstrings = re.split("\s", line)
        #split line in word-pos-pair-strings WPPS
        #delete final pairstring
    for p in pairstrings:
            sentence.append(tuple(re.split("\/{1,2}", p)))
        #for each WPPS, split word from PoS and add to sentence
        #print sentence
    sentence = sentence[:-2]

    corpus.append(sentence)
      # add sentence to corpus
    return corpus
```

Figure 4: Script to make output files of the automatic PoS-tagger ready for Chunkparsing

1.3 Text markup

For the markup, I chose the TEI P5 header that is suitable for philological data, translations and annotation in XML format. Any information about the philological background of the text can be stored in this header and easily retrieved for future online usage. In the textual markup, any changes to the annotation, can be indicated as well to trace the history of the annotated text and corpus as a whole. Finally, it would ultimately be possible to combine different versions of the texts (i.e. diplomatic and critical editions) into one xml file to make sure invaluable philological information is not lost.

2 Splitting and joining words

As became clear from the initial pilot, the huge amount of orthographical variation complicates the PoS-tagging task tremendously. The Memory-Based Tagger could filter those out on the basis of the context most of the time. In this way, there was no real need for time-consuming preprocessing of the text in terms of splitting merged tokens. Some tokens, however, were particularly challenging for the automated tagger, since very few generalisations could be made from the small training set (cf. Meelen and Beekhuizen (2013)). Below is a list of items that were split or combined to facilitate automatic tagging.

2.1 Items that are split

— combined words with nasalising prepositions, e.g. *ymwyt > y* + mwyt* 'in food'
— conjunctions combined with definite articles: *ar > a* + r* 'and the'
— particle combined with pronouns, e.g. *ae > a* + e* 'PRT 3MS'

2.2 Combined conjunctions and prepositions

Welsh employs combined prepositions: a combination of a preposition plus a grammaticalised noun. Pronominal objects of these type of prepositions appear in between the two prepositions as a possessive pronoun, e.g. *yn eu herbyn* 'against/towards them' (PKM 65.6-7) from *yn* 'in' + *eu* 'their' + *erbyn* 'opposition'. In this particular case of combined prepositions, a more conservative annotation scheme, acknowledging the nominal origin of the construction yielding the tag sequence 'P 3P N' (preposition third-person plural possessive noun) was preferred to facilitate rule-based chunk-parsing. The most commonly combined prepositions annotated in this way are:

— *ach/ger law* 'beside' (Lit. 'by hand')
— *am law* 'in addition to' (Lit. 'at hand')
— *am/ar/uch ben* 'on top of' (Lit. 'on head')
— *amgylch* 'about' (Lit. 'on circle')

- *ar ffuryf* 'like, as' (Lit. 'in form')
- *ar drws* 'in front of' (Lit. 'at door')
- *ar gefyn* 'on' (Lit. 'on back')
- *ar ol* 'after' (Lit. 'on track')

- *ger/rac bronn* 'by, before' (Lit. 'by breast')
- *heb law* 'past' (Lit. 'without hand')
- *y maes o* 'outside' (Lit. 'in field of')
- *is gil* 'behind' (Lit. 'below back')
- *o achaws* 'because of' (Lit. 'from cause')
- *yn lle* 'instead of' (Lit. 'in place')
- *ym penn* 'after' (Lit. 'in head')

Prepositions in Welsh could also be combined with other prepositions, e.g. *y dan* 'under, below' from *y* 'to' + *tan* 'under'. These complex prepositions were tagged PSUB + PSUB, so they could be recognised as separate, but also as combined prepositions. A further advantage of this is that the automatic tagger looking at the tags preceding and following the focus word, will not encounter the odd sequence of two prepositions. For combined conjunctions, a similar extension was used: *o + herwydd* CONJSUB + CONJSUB meaning 'because'. The most commonly combined prepositions and conjunctions are:

- *hyt ar* 'as far as, up to'
- *hyt at* 'as far as, to'
- *hyt yn* 'until'
- *y am* 'about, towards'
- *y ar* 'on, upon'
- *y gan* 'by, because'
- *y dan* 'under, below'

- *y wrth* 'from
- *y vewn* 'into'
- *o vywn* 'within'
- *y dros* 'for, instead of'
- *y tu ac* 'towards'
- *yr mwyn* 'for the sake of'
- *yn erbyn* 'against'

2.3   Fused forms

Middle Welsh manuscripts exhibit some fused forms as well. The combination found most commonly is the preposition *y* 'to' and the infixed third-person pronoun 'him, her, them' that is often written as *y* as well. These fused forms are annotated with hyphenated tags 'P-PRO'.

3   List of PoS tags

Adjectives and adverbs (ADJ, ADV)

Adjectives appear in various forms:
- positive adjectives, e.g. *coch* 'red' ⇒ ADJ
- comparative adjectives, e.g. *clotuorach* 'more famous' ⇒ ADJR
- superlative adjectives, e.g. *dewraf* 'bravest' ⇒ ADJS
- plural adjectives, e.g. *ieueinc* 'young' ⇒ ADJPL
- ordinal number, e.g. *eil* 'second' ⇒ ADJNUM

Adverbs can appear on their own as true adverbial lexical items, but they can also be adjectives following the predicative particle *yn*, e.g. *yn gyflym* 'quickly'. In these cases, the adjectives are tagged ADJ, but the phrase - a combination of predicative *yn* + ADJ - is labeled as an adverbial phrase ADVP.

Particles (PCL)

There are many different kinds of particles in Middle Welsh:
— preverbal particles, e.g. *a/y* ⇒ PCL
— question particles, e.g. *a* ⇒ PCL-QU
— negative particles, e.g. *ny* ⇒ PCL-NEG
— negative focus particles, e.g. *na* ⇒ PCL-NEG-FOC
— negative question particles, e.g. *oni* ⇒ PCL-QU-NEG
— negative focus question particles, e.g. *onid* ⇒ PCL-FOC-QU-NEG
— focus particles, e.g. *panyw* ⇒ PCL-FOC

Cardinal numbers (NUM)

Cardinal numbers are tagged NUM, regardless of whether they are used as substantives or as adjectives:
— substantives ⇒ *y pedwar hynny* 'those four', *pym mil o wyr* '5,000 men' (lit. '5 thousand of men'), *tri o wyr* 'three men'
— adjective ⇒ *deu wr* 'two men', *teir llong ar dec* 'thirteen ships' (lit. '3 ship on ten'), *pedwar meib ar hugeint* '24 sons' (lit. '4 sons on twenty')

Inflected verbs (VB) and Verbal nouns (VN)

Verbs appear with and without inflection. The uninflected forms can function as nouns or infinitival verbs. To avoid any linguistic interpretation, they are consistently tagged VN. The inflection of the verb is reflected in the tag following VB-. Tense, aspect, mood, person and number are all indicated separately:
— present indicative, e.g. *caraf* 'I love' ⇒ VBPI-1SG
— present subjunctive, e.g. *carhych* 'you would love' ⇒ VBPS-2SG
— preterite verb, e.g. *carawd* 'he loved' ⇒ VBD-3SG
— imperfect indicative, e.g. *carem* 'we loved' ⇒ VBAI-1PL
— imperfect subjunctive, e.g. *carhit* 'was loved' ⇒ VBAS-4
— pluperfect, e.g. *carassewch* 'you (pl) had loved' ⇒ VBG-2PL
— imperative, e.g. *car* 'love!' ⇒ VBI-2SG

Some present and imperfect forms are ambiguous between indicative and subjunctive mood, e.g. *carem* 'we loved'. Whenever they are ambiguous, they are tagged without mood indication: VBA as 'imperfect verb'. Verbs that function as auxiliaries as well have specific tags, e.g. *cael* 'to get' HV-, *bod* 'to be' BE- (unless it is the verbal noun or complementiser, both tagged as BOD), *gwneuthur* 'to do' DO-.

Nominals (N, NPR, PRO)

Singular nouns are N, plural nouns NPL and proper nouns are NPR. There are various types of pronouns in Middle Welsh:

— regular pronouns, e.g. *mi* 'I' ⇒ PRO
— conjunctive pronouns, e.g. *enteu* 'he (too)' ⇒ PROC
— reduplicated pronouns, e.g. *tydi* 'YOU (not him)' ⇒ PROR
— accusative pronouns (infixed clitics), e.g. *e* 'her' ⇒ PRO-A
— genitive pronouns (infixed clitics), e.g. *fy* 'my' ⇒ PRO-G
— indefinite pronoun, *un* 'one' ⇒ ONE

Prepositions (P)

Some prepositions can be inflected in Welsh. The inflection is tagged like verbal endings, e.g. *iddo* 'to him' P-3SGM, *amdanaf* 'about me' P-1SG.

Wh-words

There are various wh-words in Middle Welsh:

— wh-adverbs, e.g. *pryd* 'when?', *sut* 'how?' ⇒ WADV
— wh-determiners, e.g. *pa* 'which, what X' ⇒ WD
— wh-pronouns, e.g. *pwy* 'who?' ⇒ WPRO
— wh-quantifiers, e.g. *sawl* 'how many?' ⇒ WQ
— unidentified wh-item, e.g. *beth* 'what?' ⇒ W

Other tags

Finally, there are some remaining tags:

— Demonstratives, e.g. *hwnnw* 'that' ⇒ DEM
— Determiners, e.g. *yr* 'the' ⇒ D
— Conjunctions, e.g. *a* 'and', *pan* 'when' ⇒ CONJ
— Complementisers, e.g. *y* 'that' ⇒ C
— Quantifiers, e.g. *rai* 'some' ⇒ Q
— Foreign words, e.g. *lama* 'why?' (Aramaic) ⇒ FW
— Predicative markers, e.g. *yn* ⇒ PRED
— Progressive markers, e.g. *yn* ⇒ PROGR
— Reflexives, e.g. *hun* '-self' ⇒ REFL
— Interjections, e.g. *o* 'oh' ⇒ INTJ
— Punctuation ⇒ PUNC

Generating a Middle Welsh PoS-tagger

The tagger is first of all created with the standard parameter settings. Each of these settings can be adjusted, according to what works best for the corpus used. The optimal settings for a certain corpus could be retrieved automatically by running

a script trying all possible options and evaluating the results with a 10-fold cross-validation (see results below).

There are many possible parameter settings (see the MBT reference manual Daelemans et al. (2010)). You can first of all choose which features you would like to take into account when assigning tags to known or unknown words. The letter sequences following -p (known words) and -P (unknown words) indicate the specific context and characters at the beginning and or the end of the word that the tagger should take into account. For the Middle and Modern Welsh taggers, the following features gave the best results:

```
-p dfa -P sssdFawchn
```

The letter 'F' is the focus word that can be examined with the following features. The letter 's', for instance, indicates that the final character should be taken into account. The triple repetition of the letter 's' means that it will take the last three characters into account. Not surprisingly for a language that relies on inflectional suffixes, the last three final characters were important to guess the correct tag for unknown words. 'd' and 'a' refer to the tag of the left and the right context words respectively; 'w' is used for the left or right context words themselves. The letters 'c', 'h' and 'n' stand for capital letters, hyphens or numbers. Features like these help the tagger assign the correct tag for a word it has not 'seen' in the training set and is thus labeled as 'unknown'.

On the basis of this MBTg (the tagger generator) first creates an ambitag lexicon. This is a list of words associated with the different tags it can have according to the training corpus. When a word-tag combination occurs less than 5% (by default, this too is an adjustable setting), it is not included.

Then it creates a frequency list of the 100 (by default, but 200 gave better results for Welsh) most frequent words in the corpus. All words not in the most-frequent-words list are transformed into special symbols: HAPAX-<code> (<code> is either 0, or a combination of H (hyphen), C (capital letter), and N (number)). Instances are created using the specified information sources for known words (as indicated with -p in the parameter settings), then the case base is generated from that (see Daelemans et al. (2010) for further technical details on this process).

On the basis of this, the case base for known words is generated by TiMBL. By default, a lazy-learning algorithm like IGTREE is used, but for this particular corpus, I got better results with the alternative IB1 algorithm for both known and unknown words.

For unknown words, the tagger uses a k-nearest-neighbour algorithm (based on Aha, Kibler, and Albert (1991) but with added *Information Gain* weighting). In addition to that, the selected feature metric is set to –mM 'MVDM' (Modified Value difference metric), which allows for partial feature matches (cf. Stanfill and Waltz (1986), Cost and Salzberg (1993) and Daelemans and Van den Bosch (2005)). Finally, weighting of features can be done in an inverse linear fashion with the parameter setting –dIL. This means that the neighbour with the smaller distance

is weighted more heavily than the one with a greater distance. From all this, a settings file is created that can be used to annotate unseen texts in the rest of the corpus.

Since there is no need to understand or adjust any of the above-mentioned algorithms or parameter settings to generate a tagger, the MBTg offers a simple and quick way to generate a tagger for any new language or corpus. Thousands of words can be tagged per second and there is no need for any additional smoothing for sparse data since this is already part of the similarity-based model (Zavrel and Daelemans (1997)). Spelling, morphology, context and the words themselves are all sources of information integrated in the weighted similarity metric.

4   List of phrasal tags

The following phrasal tags were used for chunkparsing the corpus:

- verb phrase, combining the preverbal particle and the verb (including direct object) ⇒ VP
- noun phrase, projection of any noun ⇒ NP
- determiner phrase, any determiner/adjective/demonstrative + noun (no internal hierarchy) ⇒ DP
- prepositional phrase, any preposition with a following NP or DP ⇒ PP
- inflected prepositional phrase, projection of inflected prepositions ⇒ PPROP
- adjectival phrase, projection of any adjective ⇒ ADJP
- adverbial phrase, projection of any adverb or adjective + predicative marker ⇒ ADVP
- aspectual phrase, combination of aspectual marker + verbal noun ⇒ ASPP
- numeral phrase, projection of any numeral ⇒ NUMP
- numeral determiner phrase, NUMP + determiners/demonstratives ⇒ NUMDP
- complementiser phrase, main or subordinate clause ⇒ CP or CP-SUB
- quantifier phrase, projection of any quantifier ⇒ QP

Chunking Middle Welsh

The NLTK modules are based on Python; their rule-based regular expression parser works best under version 2.7. In order to chunkparse the PoS-tagged texts, the (manually corrected) output of the Memory-Based Tagger needs to be converted to a format that is readable to the parser. The text files were automatically converted with a Python-based text-preparation script ('chunkprep.py')[2]:

---

[2]Many thanks to Barend Beekhuizen for helping me develop the Python scripts presented here.

```
import sys
import os
import re

def make_nltk_readable(file_name):
"""
function takes one argument (file_name), and returns a list
containing (for every sentence) a list of word-pos pairs
"""

all_text = open(file_name)
corpus = []

for line in all_text:
   sentence = []
   pairstrings = re.split("\s", line)
        #split line in word-pos-pair-strings WPPS
        #delete final pairstring
   for p in pairstrings:
            sentence.append(tuple(re.split("\/{1,2}", p)))
        #for each WPPS, split word from PoS and add to sentence
        #print sentence
   sentence = sentence[:-2]

   corpus.append(sentence)
     # add sentence to corpus
   return corpus
```

Figure 5: Script to make output files of the automatic PoS-tagger ready for Chunkparsing

The chunkparser was originally not meant to perform parses with such extensive hierarchical structures as required for the present study, but by adjusting the option to loop through the grammar multiple times, these structures can be created nonetheless.

The python module 'pprint' can finally be used to ensure the newly parsed text is printed in the right .psd format to enable search queries via, for example, CorpusSearch. Figure 6 shows the step-by-step commands in Python to chunkparse text X. 'Xgold' refers to the gold standard, the version of the PoS-tagged text that has been manually corrected.

```
>>>import nltk, re, pprint, chunkprep
>>>grammar = r"""
...
VP:{<PCL-PRO-G|PCL-PRO-A|PCL|PCL-NEG|PCL-NEG-PRO-A>?|VBPS-2PL|...>}
    PROP: {<PRO|PROC|PROR|PROX>}
    VNP: {<PRO-G>?<VN|DON|HVN><PROP>?}
    ASPP: {<PROGR|PERF><PRO-G>?<VNP|HVN|DON|BOD>}
    DEMP: {<D><DEM>}
    NUMP: {<NUM>?<NUM|ONE><P><NUM>}
    WDP: {<WD><N|NPL|ONE|QP>}
    NP: {<N|NPL|NPR>}
    NUMDP: {<NUM>?<NUM|ONE><NP><P><D>?<NUM|NP>}
    NUMP: {<NUM|ONE><NP><ADJP>?}
    DP: {<D><NUMP>}
    DP: {<PRO-G><ADJP>?<NP><PROP>}
    DP: {<D><NUM><NUM>?<DEM>?}
    DP: {<NP|D><ADJP>}
    REFLP: {<PRO-G><REFL>}
    P: {<PSUB><PSUB>}
    P: {<P><P>}
    PP: {<P><PP><VNP>}
    PWP: {<P><WPRO|WDP>}
    ONEP: {<ONE><PP|OTHER>}
    PP: {<P><ONEP>}
    DEMP: {<DEM>}
    ADJQP: {<ADJQ><ADJQ>}
    PREDP: {<PRED><ADV|PRO-G>?<NP|ADJP|ADJQP|QP|DP>}
    ADJP: {<ADJP><PP>}
..."""
>>>cp = nltk.RegexpParser(grammar, loop=2)
>>>text = chunkprep.make_nltk_readable('Xgold.txt')
>>>results = []
>>>for t in text:
... result = cp.parse(t)
... results.append(result)
>>>f = open('Xchunked.psd', 'w')
>>>for r in results:
... f.write(r.pprint())
... f.write('\n')
>>>f.close()
```

Figure 6: Adopting & implementing the Python NLTK Chunkparser

5 Known annotation issues

In the current stage, the annotation of the corpus was done in such a way to optimise the search queries specific to the present thesis. The focus lies on the part-of-speech annotation. The highly detailed tag set facilitates future research in any linguistic framework. The same goes for the relatively flat structure of the chunk-parsed files. This can be extended to a full parse on the basis of the manually corrected .psd(x) files or on the basis of the PoS-tagged .txt files.

From a syntactic point of view, the difference between subject and object constituents was initially not indicated. Since Middle Welsh allows subject- and object-initial orders as well as pro-drop it was impossible to do this automatically. The DP-initial orders that could be ambiguous were manually disambiguated at a later stage, dividing them into SVO and OVS orders.

The most important elements that are not included in the current annotated corpus are empty categories. The main reason for not including these at this stage was because they were not necessary for the present investigation in word-order. Furthermore, the aim was to keep the annotated corpus as theory-neutral as possible and empty categories are very theory-specific. The flexible xml-based nature (compatible with the psd file structure) means that those can be added at a later stage as well. This can be done by developing a context-free grammar and/or manual insertion.

Finally, at various stages in the process of creating the corpus, manual correction was necessary. Since there was only one annotator available to build the present corpus, checking cross-annotator agreement was no option. Although an effort has been made to double-check all the corrected versions, some errors no doubt remain. In future, when making the annotated files accessible for everyone online, a final check will be done to filter out any possible mistakes and/or inconsistencies.

6 Coding queries

Figure 7 shows a sample of algorithms in Xquery code used to retrieve values for features like Negation, Focus or Tense, Aspect and Mood for different kinds of verbs (DO- 'to do', BE- 'to be', HV- 'to get', VB any other verb) from the PoS-tagged and Chunkparsed database (converted to XML format). The queries employ standard XQuery code plus additional functions built into the software package CorpusStudio (cf. Komen (2009b)), like `ru:matches` to match labels of PoS-tags indicated in the query with those in the database.[3]

---

[3]This is just a sample excerpt of the entire query that works with an accompanying definition file in which specific variables like `$vp` and `$sentence` are defined.

```
(: Find Focus particles :)
  let $Foc := $sentence/descendant::eTree[ru:matches
    (@Label, '*FOC*')][1]
  let $strFoc := ru:NodeText($Foc)
  let $feat_Foc := if ($strFoc = '') then '-' else $strFoc

(: Find Negation :)
  let $Neg := $sentence/descendant::eTree[ru:matches
    (@Label, '*NEG*')][1]
  let $strNeg := ru:NodeText($Neg)
  let $feat_Neg := if ($strNeg = '') then '-' else $strNeg

(: Find Mood :)
  let $feat_Mood :=
    if (exists($vp/descendant::eTree[ru:matches
(@Label, 'VBI*|DOI*|BEI*|HVI*')]))
then 'Imperative'
    else if(exists($vp/descendant::eTree[ru:matches
(@Label, 'VBPS*|VBAS*|BEPS*|BEAS*|DOPS*|DOAS*|HVPS*|HVAS*')]))
then 'Subjunctive' else 'Indicative'

(: Find Tense and Aspect :)
  let $feat_TenseAspect :=
    if (exists($vp/descendant::eTree[ru:matches
(@Label, 'VBP-*')]))
then 'Perfect'
else if(exists($vp/descendant::eTree[ru:matches
(@Label, 'VBAI*|VBAS*|DOAI*|DOAS*|BEAI*|BEAS*|HVAI*|HVAS*')]))
then 'Imperfect'
else if(exists($vp/descendant::eTree[ru:matches
(@Label, VBG*|DOG*|BEG*|HVG*')]))
then 'Pluperfect'
    else if(exists($vp/descendant::eTree[ru:matches
(@Label, 'VBD*|DOD*|BED*|HVD*')]))
then 'Preterite' else 'Present'
```

Figure 7: XQuery code to retrieve Focus, Negation and Tense/Aspect feature values

Figure 8 shows some excerpts of the complex query to find the various word order types.[4]

```
(: Look through each text for ['S'] :)
 for $search in //eTree[ru:matches(@Label, 'S')]

(: Determine what the first constituent is, excl. CONJ and C :)
  let $initialCns :=
  $search/child::eTree[not(ru:matches
  (@Label, 'CONJ*|INTJ|C|PCL-QU'))][1]

(: Determine the main type of this sentence :)
  let $mainType :=
  tb:WelshMainCat($initialCns, $search)

 return ru:back($search, '', $cat)

(...)

(: Get the VP :)
   let $vp := tb:WelshVP($sentence)

(: Determine the main category :)
   let $mainCat :=  if ($initialCns/@Label = 'SEF')
then 'Type VI Sef'
else if (ru:matches($initialCns/@Label,'W*|PCL-QU')
and not(exists($sentence/child::eTree[ru:matches
(@Label, 'QP')])) )
then 'Type X Question'
else if (ru:matches($initialCns/@Label, '*FOC*'))
then 'Type XI Focus'
(...)
else if ($initialCns/@Label = 'VNP' and
(some $ch in $initialCns/following-sibling::eTree satisfies
        ($ch is $vp and exists($vp/child::eTree[ru:matches
        (@Label, 'DO*')])) ) )
then 'Type IIIc VNaDO'
```

Figure 8: Sample XQuery definition & query algorithm to find the main word order type

---

[4]Many thanks to Erwin Komen for teaching me how to use XQuery.

References

Aarts, J. (1991). Intuition-based and Observation-based Grammar. In K. Aijmer & B. Altenberg (Eds.), *English corpus linguistics.* London: Longman.

Aboh, E. O. (2010). Information structuring begins with the numeration. *Iberia: An International Journal of Theoretical Linguistics, 2*(1), 12-42.

Abraham, W., & de Meij, S. (Eds.). (1986). *Topic, Focus and Configurationality: Papers from the 6th Groningen Grammar Talks, Groningen, 1984*. Amsterdam, Philadelphia: John Benjamins Publishing Company.

Adams, J. N. (2007). *The regional diversification of Latin 200 BC-AD 600*. Cambridge: Cambridge University Press.

Adger, D. (2011). Clefted situations: A note on expletives in Scottish Gaelic clefts. In A. Carnie (Ed.), *Formal Approaches to Celtic Linguistics* (p. 3-15). Cambridge Scholars Publishing.

Adger, D. (2013a). Constructions and grammatical explanation: comments on Goldberg. *Mind & Language, 28*(4), 466-478.

Adger, D. (2013b). *Constructions are not explanations.* Available online on http://ling.auf.net/lingbuzz/001675.

Adger, D., & Ramchand, G. (2003). Predication and equation. *Linguistic inquiry, 34*(3), 325-359.

Adger, D., & Ramchand, G. (2005). Merge and move: *Wh*-dependencies revisited. *Linguistic Inquiry, 36*(2), 161–193.

Aelbrecht, L., Haegeman, L., & Nye, R. (2012). *Main clause phenomena: New horizons* (Vol. 190). Amsterdam, Philadelphia: John Benjamins Publishing Company.

Aha, D. W., Kibler, D., & Albert, M. K. (1991). Instance-based learning algorithms. *Machine learning, 6*(1), 37-66.

Alderson, J. C. (2007). Judging the frequency of English words. *Applied Linguistics, 28*(3), 383-409.

Allen, C. (1995). *Case marking and reanalysis: grammatical relations from Old to Early Modern English*. Oxford: Oxford University Press.

Alter, S. G. (1999). *Darwinism and the linguistic image: language, race, and natural*

theology in the nineteenth century. Johns Hopkins University Press.

Ambridge, B., Pine, J. M., & Lieven, E. V. (2014). Child language acquisition: Why universal grammar doesn't help. *Language, 90*(3), e53-e90.

Andersen, H. (1973). Abductive and deductive change. *Language, 49*, 765-793.

Anderson, S. (1982). Where's morphology? *Linguistic Inquiry, 13*(4), 571–612.

Andor, J. (2004). The master and his performance: An interview with Noam Chomsky. *Intercultural Pragmatics, 1*(1), 93-111.

Anwyl, E. (1899). *A Welsh Grammar for Schools, part II - Syntax* (Vol. 2). London: Swan Sonnenschein & Co. Ltd.

Ariel, M. (1999). *Accessing Noun-Phrase Antecedents*. London/New York: Routledge.

Badan, L., & Del Gobbo, F. (2011). On the syntax of topic and focus in Chinese. In P. Beninca & N. Munaro (Eds.), *Mapping the left periphery* (p. 63-91). New York & Oxford: Oxford University Press.

Bailey, N. A. (2009). *Thetic constructions in koine greek*. Unpublished doctoral dissertation, Amsterdam: Vrije Universiteit.

Baker, M. (2008). The macroparameter in a microparametric world. In T. Biberauer (Ed.), *The limits of syntactic variation* (Vol. 132, p. 351-373). Amsterdam, Philadelphia: John Benjamins Publishing Company.

Baker, M. C. (2009). Is head movement still needed for noun incorporation? *Lingua, 119*(2), 148-165.

Baker, P. (2006). *Using corpora in discourse analysis*. Bloomsbury Publishing.

Barðdal, J. (2011). The rise of dative substitution in the history of Icelandic: A diachronic construction grammar account. *Lingua, 121*(1), 60-79.

Barðdal, J., & Eythórsson, T. (2012). Reconstructing syntax: Construction grammar and the comparative method. In H. C. Boas & I. Sag (Eds.), *Sign-based construction grammar* (pp. 257–308). Stanford, California: CSLI Publications.

Barðdal, J., Smirnova, E., Sommerer, L., & Gildea, S. (2015). *Diachronic Construction Grammar* (Vol. 18). Amsterdam, Philadelphia: John Benjamins Publishing Company.

Baxendale, T. (2009). *Y rhyfel oeraf* (T. D. Jones, Trans.). London: Rily Publications.

Beaver, D. I. (2004). The optimization of discourse anaphora. *Linguistics and Philosophy, 27*(1), 3–56.

Beekes, R. S. (1995). *Comparative Indo-European linguistics: an introduction*. Amsterdam, Philadelphia: John Benjamins Publishing Company.

Beekhuizen, B. (2015). *Constructions Emerging*. Utrecht: LOT Dissertation Series.

Beekhuizen, B., Bod, R., & Verhagen, A. (2014). The linking problem is a special case of a general problem none of us has solved: Commentary on Ambridge, Pine, and Lieven. *Language, 90*(3), e91-e96.

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied linguistics, 25*(3), 371-405.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

Biberauer, T. (2015). *The limits of syntactic variation: en emergentist comparative perspective. (Invited talk given at the Workshop on Language Variation and Change and Cultural Evolution)*. Centre for Linguistics History and Diversity, York University, 14 February 2015.

Biberauer, T., Holmberg, A., & Roberts, I. (2014). A syntactic universal and its consequences. *Linguistic Inquiry*, *45*(2), 169-225.

Biberauer, T., Holmberg, A., Roberts, I., & Sheehan, M. (2014). Complexity in comparative syntax: the view from modern parametric theory. In F. J. Newmeyer & L. B. Preston (Eds.), *Measuring grammatical complexity* (p. 103-127). Oxford: Oxford University Press.

Biberauer, T., & Richards, M. (2006). True optionality: When the grammar doesn't mind. *Minimalist essays*, *91*, 35-67.

Biberauer, T., & Roberts, I. (2008). Cascading parameter changes: internally driven change in Middle and Early Modern English. In T. Eythórsson (Ed.), *Grammatical change and linguistic theory: the Rosendal papers* (p. 79-113).

Biberauer, T., & Roberts, I. (2015). *The significance of what hasn't happened (Invited talk given at the Workshop on Language Variation and Change and Cultural Evolution)*. Centre for Linguistics History and Diversity, York University, 14 February 2015.

Biberauer, T., Sheehan, M., & Newton, G. (2010). Impossible changes and impossible borrowings. In A. Breitbarth, C. Lucas, S. Watts, & D. Willis (Eds.), *Continuity and change in grammar* (p. 35-60). Amsterdam: John Benjamins.

Biberauer, T., & Walkden, G. (2015). *Syntax Over Time: Lexical, Morphological, and Information-structural Interactions* (Vol. 15). Oxford: Oxford University Press.

Birch, S., & Clifton, C. (1995). Focus, accent, and argument structure: Effects on language comprehension. *Language and speech*, *38*(4), 365-391.

Birner, B. (2006). Inferential relations and noncanonical word order. In *Drawing the boundaries of meaning: Neo-Gricean studies in pragmatics and semantics in honor of Laurence R. Horn.* (p. 31-51). Amsterdam: John Benjamins.

Bloom, P. (1990). Subjectless sentences in child language. *Linguistic Inquiry*, *21*(4), 491–504.

Bod, R., Hay, J., & Jannedy, S. (2003). *Probabilistic linguistics*. Cambridge, MA: MIT Press.

Bonelli, E. T. (2010). Theoretical overview of the evolution of corpus linguistics. In A. O'ÂÃÂŹKeeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (p. 14-28). London: Routledge.

Borer, H. (1984). *Parametric syntax*. Dordrecht, Holland: Foris Publications.

Bornkessel, I., Schlesewsky, M., & Friederici, A. (2003). Contextual information modulates initial processes of syntactic integration: The role of inter-versus intrasentential predictions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(5), 871-882.

Borsley, R. (1989). A note on ellipsis and case. *Linguistic Inquiry*, *20*(1), 125–130.

Borsley, R., & Stephens, J. (1989a). Agreement and the position of subjects in

Breton. *Natural Language and Linguistic Theory*, *7*(3), 407–427.

Borsley, R., & Stephens, J. (1989b). Agreement and the position of subjects in Breton. *Natural Language & Linguistic Theory*, *7*(3), 407-427.

Borsley, R., Tallerman, M., & Willis, D. (2007). *The syntax of welsh*. Cambridge University Press.

Bresnan, J. (2001). *Lexical-functional syntax*. Oxford: Blackwell.

Bromiley, G. (1997). *The International Standard Bible Encyclopedia: Vol I: A-D*. Grand Rapids, MI: William B. Eerdmans.

Brown, J. S. (1988). Patch use as an indicator of habitat preference, predation risk, and competition. *Behavioral Ecology and Sociobiology*, *22*(1), 37–47.

Brugmann, K. (1876). Zur Geschichte der stammabstufenden Deklinationen, Erste Abhandlung: Die Nomina auf *-ar-* und *-tar-*. *Curtius' Studien*, *9*, 361-406.

Bucholtz, M. (2008). Theories of discourse as theories of gender: Discourse analysis in language and gender studies. In J. Holmes & M. Meyerhoff (Eds.), *The handbook of language and gender* (p. 43-68). Oxford: Blackwell.

Büring, D. (1997). *The meaning of topic and focus: the 59th Street Bridge accent*. London: Routledge.

Büring, D. (2003). On D-trees, beans, and B-accents. *Linguistics and philosophy*, *26*(5), 511-545.

Bury, D. (2002). A reinterpretation of the loss of verb-second in Welsh. In D. Lightfoot (Ed.), *Syntactic effects of morphological change* (p. 215-231). Oxford: Oxford University Press.

Busa, R. (1992). Half a Century of Literary Computing: Towards a 'New' Philology. *Literary and Linguistic Computing*, *7*, 69-73.

Campbell, L. (2000). What's wrong with grammaticalization? *Language sciences*, *23*(2), 113-161.

Campbell, L., & Janda, R. (2000). Introduction: conceptions of grammaticalization and their problems. *Language sciences*, *23*(2), 93-112.

Cappelle, B. (2009). Can we factor out free choice? *Describing and modeling variation in grammar*, *204*, 183.

Cardinaletti, A., Cinque, G., & Giusti, G. (1988). *Constituent structure*. Dordrecht: Foris Publications.

Chafe, W. L. (1976). Givenness, contrastiveness, definiteness, subjects, topics and point of view. In L. C. N. (Ed.), *Subject and topic* (p. 27-55). New York: Academic Press.

Chambers, E. (1728). *Cyclopaedia; or, an universal dictionary of arts and sciences*. London.

Charles-Edwards, T. (2001). The Textual Tradition of Medieval Welsh Prose Tales and the Problem of Dating. In B. Maier, S. Zimmer, & C. Bakte (Eds.), *150 Jahre "Mabinogion"-Deutsch-Walisische Kulturbeziehungen* (p. 23-39). Tübingen: De Gruyter.

Charles-Edwards, T. M. (2013). *Wales and the Britons, 350-1064*. Oxford: Oxford University Press.

Cheng, L. L.-S., & Downing, L. J. (2012). Against FocusP: arguments from Zulu.

In *Information structure. contrasts and positions* (p. 247-267). Cambridge: Cambridge University Press.

Choi, H.-W. (1999). *Optimizing structure in context: Scrambling and information structure*. Stanford: CSLI Publications.

Chomsky, N. (1964). *Current issues in linguistic theory*. The Hague: Mouton, Den Haag.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, Massachusetts: M.I.T. Press.

Chomsky, N. (1969). Quine's Empirical Assumptions. In D. Davidson & J. Hintikka (Eds.), *Words and Objections* (Vol. 21, p. 53-68). Springer Netherlands.

Chomsky, N. (1986). *Barriers*. Cambridge, Massachusetts: MIT Press.

Chomsky, N. (1995). *The minimalist program*. Cambridge, Massachusetts: MIT Press.

Chomsky, N. (2000). Minimalist inquiries: The framework. In R. Martin, D. Michaels, & J. Uriagereka (Eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik* (pp. 89–156). MIT Press.

Chomsky, N. (2001). Derivation by phase. In M. Kenstowicz (Ed.), *Ken hale: A life in language* (p. 1-52). Cambridge, MA: MIT Press.

Chomsky, N. (2005). Three factors in language design. *Linguistic inquiry*, *36*(1), 1-22.

Chomsky, N. (2013). Problems of projection. *Lingua*, *130*, 33-49.

Church, K. W., & Mercer, R. L. (1993). Introduction to the special issue on computational linguistics using large corpora. *Computational linguistics*, *19*(1), 1-24.

Cinque, G. (1977). The movement nature of left dislocation. *Linguistic inquiry*, 397-412.

Cinque, G. (1999). *Adverbs and functional heads: a cross-linguistic perspective*. New York: Oxford University Press.

Clahsen, H. (1991). *Child language and developmental dysphasia: Linguistic studies of the acquisition of German* (Vol. 2). Amsterdam, Philadelphia: John Benjamins Publishing.

Clark, R., & Roberts, I. (1993). A computational model of language learnability and language change. *Linguistic Inquiry*, *24*, 299-345.

Comrie, B. (1989). *Language universals and linguistic typology: Syntax and morphology*. University of Chicago press.

Conrad, S. (2010). What can a corpus tell us about grammar. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (p. 227-240). London: Routledge.

Cost, S., & Salzberg, S. (1993). A weighted nearest neighbor algorithm for learning with symbolic features. *Machine learning*, *10*(1), 57-78.

Coupland, N. (2007). *Style: Language variation and identity*. Cambridge: Cambridge University Press.

Cowles, H. W. (2012). The psychology of information structure. In M. Krifka & R. Musan (Eds.), *The expression of information structure* (Vol. 5, p. 287-317).

Berlin: Walter de Gruyter.

Cowles, H. W., Walenski, M., & Kluender, R. (2007). Linguistic and cognitive prominence in anaphor resolution: topic, contrastive focus and pronouns. *Topoi*, *26*(1), 3-18.

Craik, K. (1943). *The nature of explanation*. Cambridge: Cambridge University Press.

Crisma, P., & Longobardi, G. (2009). Introduction: change, relatedness and inertia in historical syntax. In P. Crisma & G. Longobardi (Eds.), *Historical syntax and linguistic theory* (p. 1-13). Oxford: Oxford University Press.

Culy, C. (1996). Formal properties of natural language and linguistic theories. *Linguistics and Philosophy*, *19*(6), 599–617.

Curme, G. (1978). *A Grammar of the English Language*. Verbatim Books.

Currie, O. (2000). Word order stability and change from a sociolinguistic perspective. *Stability, Variation, and Change of Word-order Patterns Over Time*, *213*, 203-230.

Currie, O. (2013). The history of gradual change and continual variation. In A. G. Ramat, C. Mauri, & P. Molinelli (Eds.), *Synchrony and diachrony: A dynamic interface* (Vol. 133, p. 43-78). Amsterdam, Philadelphia: John Benjamins Publishing Company.

Curzan, S. (2008). Historical corpus linguistics and evidence of language change. In A. Lüdeling & M. Kytö (Eds.), *Corpus Linguistics. An International handbook* (Vol. 2, p. 1097-1109). Berlin: Walter de Gruyter.

Cutler, A., & Fodor, J. A. (1979). Semantic focus and sentence comprehension. *Cognition*, *7*(1), 49-59.

Daelemans, W., & Van den Bosch, A. (2005). *Memory-based language processing*. Cambridge: Cambridge University Press.

Daelemans, W., Zavrel, J., Van den Bosch, A., & Van der Sloot, K. (2010). Mbt: memory-based tagger - version 3.2 Reference Guide. *ILK Technical Report - ILK 10-04*.

D'Alessandro, R., & Van Oostendorp, M. (2016). Prosody, phi-features and deixis in Southern Italian: what vocatives can tell us on the architecture of language. *Available from* `http://ling.auf.net/lingbuzz/002721`.

Daneš, F. (1970). One instance of Prague school methodology: functional analysis of utterance and text. *Method and theory in linguistics*, 132-146.

Daneš, F. (1974). Functional sentence perspective and the organization of the text. In F. Daneš (Ed.), *Papers on functional sentence perspective* (p. 106-128).

Davies, J. (1621[1809]). *Antiquae linguae britannicae, nunc communiter dictae cambro-britannicae, a suis cymraecae vel cambricae, ab aliis wallicae rudimenta: Juxta genuinam naturalemque ipsius linguae proprietatem, qua fieri potuit accurata methodo et brevitate conscripta*. London: Oxonii.

Davies, P., & Deuchar, M. (2010). Using the Matrix Language Frame model to measure the extent of word-order convergence in Welsh-English bilingual speech. In A. Breitbarth, C. Lucas, S. Watts, & D. Willis (Eds.), *Continuity and change in grammar* (p. 77-96).

Davies, S. (1995). *Crefft y Cyfarwydd: Astudiaeth o dechnegau naratif yn Y Mabinogion*. Caerdydd: Gwasg Prifysgol Cymru.

Davies, S. (1998). Written text as performance: the implications for Middle Welsh prose narratives. In *Literacy in medieval celtic societies* (p. 133-48). Cambridge.

Davies, W. D., & Dubinsky, S. (2004). *The grammar of raising and control: A course in syntactic argumentation*. John Wiley & Sons.

Davis, H., Gillon, C., & Matthewson, L. (2014). How to investigate linguistic diversity: Lessons from the Pacific Northwest. *Language, 90*(4), e180-e226.

Delbrück, B. (1900 [1982]). *Einleitung in das Sprachstudium: ein Beitrag zur Geschichte und Methodik der vergleichenden Sprachforschung*. Cambridge: Cambridge University Press.

Déprez, V., & Pierce, A. (1993). Negation and functional projections in early grammar. *Linguistic Inquiry, 24*(1), 25–68.

De Saussure, F., Bailly, C., & Séchehaye, A. (1916). *Cours de linguistique générale*. Paris: Grande Bibliothèque Payot.

Destruel, E., & Velleman, L. (2014). Refining Contrast: Empirical Evidence from the English it-Cleft. In *Empirical issues in syntax and semantics* (pp. 197–214). Colloque de syntaxe et sémantique à Paris.

De Swart, H., & De Hoop, H. (1995). Topic and focus. *Glot international, 1*(7), 3-7.

Deutscher, G. (2002). On the misuse of the notion of 'abduction' in linguistics. *Journal of Linguistics, 38*(03), 469-485.

Di Eugenio, B. (2003). Discourse processing. In L. Nadel (Ed.), *Encyclopedia of cognitive science* (Vol. 1, p. 976-983). London: Nature Publishing Group.

Dik, S. C. (1997). *The theory of functional grammar: the structure of the clause*. Berlin/New York: Mouton de Gruyter.

Dreschler, G. (2015). *Passives and the loss of verb second: A study of syntactic and information-structural factors*. Utrecht: LOT dissertation series.

Dresher, B. E. (1999). Charting the learning path: Cues to parameter setting. *Linguistic Inquiry, 30*(1), 27–67.

Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational linguistics, 19*(1), 61-74.

Eckhoff, H., & Haug, D. (2011). Personal pronouns with articles: a quantitative approach. In *Information structure and corpus annotation: theoretical and practical perspectives*. Oslo, Lysebu: University of Oslo.

É.Kiss, K. (1995). *Discourse configurational languages*. Oxford: Oxford University Press.

É.Kiss, K. (1998). On generic and existential bare plurals and the classification of predicates. In S. Rothstein (Ed.), *Events and grammar* (pp. 145–162). Dordrecht: Kluwer Academic Publishers.

É.Kiss, K. (2001). Parasitic chains revisited. In P. Culicover & P. Postal (Eds.), *Parasitic gaps* (pp. 99–124). Cambridge, Massachusetts: MIT Press.

Ellis, N., O'Dochartaigh, C., Hicks, W., Morgan, M., & Laporte, N.

(2001). *Cronfa Electroneg o Gymraeg (CEG): A 1 million word lexical database and frequency count for Welsh.* Available online from `http://www.bangor.ac.uk/canolfanbedwyr/ceg.php.en`. Last accessed d.d. 20 September 2013.

Engdahl, E., & Vallduví, E. (1996). Information packaging in HPSG. *Edinburgh working papers in cognitive science*, *12*, 1-32.

Erteschik-Shir, N. (2007). *Information structure: The syntax-discourse interface* (Vol. 3). Oxford: Oxford University Press.

Evans, D. S. (1968). The sentence in early Modern Welsh. *Bulletin of the Board of Celtic Studies*, *22*, 311-337.

Evans, D. S. (1971). Concord in Middle Welsh. *Studia Celtica*, *6*, 42-56.

Evans, D. S. (1990). Insular Celtic and the emergence of the Welsh language. In *Britain 400-600: Language and history*. Heidelberg: Carl Winter.

Evans, D. S. (2003 [1964]). *A grammar of Middle Welsh*. DIAS, Dublin.

Evans, E. (1958). Cystrawennau 'sef' mewn Cymraeg Canol. *Bulletin of the Board of Celtic Studies*, *18*, 38-54.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and brain sciences*, *32*(05), 429-448.

Evers, A., & Van Kampen, J. (2008). Parameter setting and input reduction. *The Limits of Syntactic Variation*, *132*, 483-515.

Faarlund, J. T. (1990). *Syntactic change: toward a theory of historical syntax* (Vol. 50). Berlin & New York: Mouton de Gruyter.

Fairclough, N. (1992). Discourse and text: Linguistic and intertextual analysis within discourse analysis. *Discourse & Society*, *3*(2), 193-217.

Falileyev, A. (2003). Languages of Old Wales: A Case for Co-existence. *Dialectologia et Geolinguistica*, *2003*(11), 18-38.

Fanselow, G., & Lenertová, D. (2011). Left peripheral focus: mismatches between syntax and information structure. *Natural Language & Linguistic Theory*, *29*(1), 169-209.

Ferstl, E. C., & von Cramon, D. Y. (2001). The role of coherence and cohesion in text comprehension: an event-related fMRI study. *Cognitive Brain Research*, *11*(3), 325-340.

Féry, C., & Ishihara, S. (2016). *The Oxford Handbook of Information Structure*. Oxford: Oxford University Press.

Féry, C., & Krifka, M. (2008). Information structure. Notional distinctions, ways of expression. In *Unity and diversity of languages* (p. 123-136). Amsterdam, Philadelphia: John Benjamins Publishing Company.

Fife, J. (1986). The semantics of gwneud inversions. *BBCS*, *33*, 133-144.

Fife, J. (1988). *Functional syntax: a case study in Middle Welsh*. Lublin: Redakcja Wydawnictw Katolickiego Uniwersytetu Lubelskiego.

Fife, J. (1991). Some constituent-order frequencies in Classical Welsh Prose. In J. Fife & E. Poppe (Eds.), *Studies in Brythonic word order* (Vol. 83, p. 251-274). Amsterdam, Philadelphia: John Benjamins Publishing Company.

Fife, J. (2010). Typological aspects of the Celtic languages. In *The Celtic Languages* (p. 1-21). London: Routledge.

Fife, J., & King, G. (1991). Focus and the Welsh 'Abnormal Sentence': A cross-linguistic perspective. In J. Fife & E. Poppe (Eds.), *Studies in Brythonic word order* (Vol. 83, p. 81-153). Amsterdam, Philadelphia: John Benjamins Publishing Company.

Fillmore, C. J. (1992). "Corpus linguistics" or "computer-aided armchair linguistics". In *Directions in corpus linguistics. Proceedings of Nobel Symposium 82, Stockholm, 4-8 August 1991* (Vol. 82, p. 35-60).

Fillmore, C. J., Kay, P., & O'Connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions: The case of let alone. *Language*, 501-538.

Firbas, J. (1964). On defining the theme in functional sentence analysis. *Travaux linguistiques de Prague*, *1*, 267-280.

Fischer, O. (1992). Syntax. In N. F. Blake, R. Lass, & S. Romaine (Eds.), *The Cambridge history of the English language: Volume II, 1066-1476* (p. 207-408). Cambridge: Cambridge University Press.

Fischer, O., Van Kemenade, A., Koopman, W., & Van der Wurff, W. (2000). *The syntax of early English*. Cambridge: Cambridge University Press.

Fodor, J. A. (1966). How to learn to talk: Some simple ways. In *The genesis of language* (p. 105-122). Cambridge Massachusetts: MIT Press.

Fodor, J. D. (1998). Unambiguous triggers. *Linguistic Inquiry*, *29*(1), 1–36.

Foley, W. A. (1994). Information Structure. In R. Asher (Ed.), *The Encyclopedia of Language and Linguistics* (p. 1678-1685). New York: Pergamon Press.

Fortson, B. (2010). *Indo-European language and culture: an introduction (Second Edition)*. Cambridge: Cambridge University Press.

Fox, D. (2002). Antecedent-contained deletion and the copy theory of movement. *Linguistic Inquiry*, *33*(1), 63–96.

Francis, W., & Kučera, H. (1964). *A Standard Corpus of Present-Day Edited American English, for use with Digital Computers*. Department of Linguistics, Brown University, Providence, Rhode Island, USA.

Frascarelli, M. (2000). *The syntax-phonology interface in focus and topic constructions in italian*. Dordrecht: Kluwer.

Frascarelli, M. (2007). Subjects, topics and the interpretation of referential pro. *Natural Language & Linguistic Theory*, *25*(4), 691-734.

Frascarelli, M., & Hinterhölzl, R. (2007). Types of topics in German and Italian. In *On information structure, meaning, and form* (p. 87-116). Amsterdam: John Benjamins.

Fried, M. (2009). Construction grammar as a tool for diachronic analysis. *Constructions and Frames*, *1*(2), 262-291.

Galves, C., Cyrino, S., Lopes, R., Sandalo, F., & Avelar, J. (2012). *Parameter Theory and Linguistic Change* (Vol. 2). Oxford: Oxford University Press.

Garrett, A. (2012). The historical syntax problem: Reanalysis and directionality. In D. Jonas, J. Whitman, & A. Garrett (Eds.), *Grammatical change: Origins, nature, outcomes* (p. 52-72). Oxford: Oxford University Press.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Erlbaum.

Gianollo, C., Guardiano, C., & Longobardi, G. (2008). Three fundamental issues in parametric linguistics. In T. Biberauer (Ed.), *The limits of syntactic variation* (pp. 109–142). Amsterdam, Philadelphia: John Benjamins Publishing Company.

Gibson, E., & Wexler, K. (1994). Triggers. *Linguistic Inquiry*, *25*, 407-454.

Givón, T. (1984). *Syntax: A functional-typological Introduction* (Vol. 1). Amsterdam, Philadelphia: John Benjamins Publishing Company.

Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory and language*, *26*(1), 69-83.

Goldberg, A. (1995). *Constructions. A Construction Grammar Approach to Argument Structure.* Chicago, IL: Chicago University Press.

Goldberg, A. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Götze, M., Weskott, T., Endriss, C., Fiedler, I., Hinterwimmer, S., Petrova, S., ... Stoel, R. (2007). Information structure. In S. Dipper, M. Götze, & S. Skopeteas (Eds.), *Working Papers of the SFB632, Interdisciplinary Studies on Information Structure (ISIS)* (p. 147-187). Potsdam: Universitätsverlag Potsdam.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological review*, *101*(3), 371-395.

Granger, S. (2003). The International Corpus of Learner English: a new resource for foreign language learning and teaching and second language acquisition research. *Tesol Quarterly*, *37*(3), 538-546.

Greaves, C., & Warren, M. (2010). What can a corpus tell us about multi-word units? In A. O'ÂĂÂŹKeeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (p. 212-226). London: Routledge.

Green, J. (2005). Reading poetic texts in isaiah. *Leaven*, *13*(2), 60-62.

Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. Greenberg (Ed.), *Universals of language.* Cambridge, Massachusetts: MIT Press.

Gregory, M. L., & Michaelis, L. A. (2001). Topicalization and left-dislocation: A functional opposition revisited. *Journal of pragmatics*, *33*(11), 1665-1706.

Grice, H. P. (1989). *Study in the way of words*. : Cambridge, MA: Harvard University Press.

Gries, S. T., & Stefanowitsch, A. (2007). *Corpora in cognitive linguistics: corpus-based approaches to syntax and lexis*. Berlin: Walter de Gruyter.

Grosz, B. J., Weinstein, S., & Joshi, A. K. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational linguistics*, *21*(2), 203-225.

Guest, L. C. (1849). *The Mabinogion. From the Llyfr Coch o Hergest, and other ancient Welsh manuscripts, with an English translation and notes.* London: Longmans.

Gundel, J. K. (1974). *The role of topic and comment in linguistic theory*. Unpublished doctoral dissertation, Austin: University of Texas.

Gundel, J. K. (1988). Universals of topic-comment structure. *Studies in syntactic typology*, *17*, 209-239.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 274-307.

Gwenogvryn Evans, J. (1898-1910). *Report on Manuscripts in the Welsh Language*. London: Royal Commision on Historical Manuscripts.

Haeberli, E., & Ihsane, T. (2015). Revisiting the loss of verb movement in the history of English. *Natural Language and Linguistic Theory*, 1-46.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304*(5669), 438-441.

Hale, K. (1983). Warlpiri and the grammar of non-configurational languages. *Natural Language and Linguistic Theory*, *1*(1), 5–48.

Hale, M. (1998). Diachronic syntax. *Syntax*, *1*(1), 1–18.

Halliday, M. A. (1967). Notes on transitivity and theme in english: Part 2. *Journal of linguistics*, *3*(02), 199-244.

Handford, M. (2010). What can a corpus tell us about specialist genres? In A. O'ÂÃÂŹKeeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (p. 255-269). London: Routledge.

Hannay, M., & Mackenzie, J. L. (2002). *Effective writing in English: a sourcebook*. Bussum: Coutinho.

Harbour, D. (2011). Mythomania? methods and morals from 'the myth of language universals'. *Lingua*, *121*(12), 1820-1830.

Harlos, A., Poppe, E., & Widmer, P. (2014). Decoding Middle Welsh clauses or "Avoid Ambiguity". *Indogermanische Forschungen*, *119*(1), 125-148.

Harris, A. C., & Campbell, L. (1995). *Historical syntax in cross-linguistic perspective* (Vol. 74). Cambridge: Cambridge University Press.

Hartmann, K., & Zimmermann, M. (2007). Focus strategies in Chadic–the case of Tangale revisited. *Studia Linguistica*, *61*(2), 95-129.

Haug, D. (2009). Info-structural annotation in the PROIEL corpus. In *Annotating and analysing information structure in historical corpus texts*.

Hawkins, J. A. (1990). A parsing theory of word order universals. *Linguistic Inquiry*, *21*(2), 223–262.

Hedberg, N., & Sosa, J. M. (2007). The prosody of topic and focus in spontaneous english dialogue. In C. Lee, M. Gordon, & D. Büing (Eds.), *Topic and focus* (p. 101-120). Springer.

Heine, B., & Kuteva, T. (2002). *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.

Heller, D. (1999). *The syntax and semantics of specificational pseudoclefts in hebrew*. Unpublished doctoral dissertation, Tel-Aviv University.

Hémon, R. (1975). *A historical morphology and syntax of Breton* (Vol. 3). Dublin institute for advanced studies.

Hinterhölzl, R. (2009). The role of information structure in word order variation and word order change. In R. Hinterhölzl & S. Petrova (Eds.), *Information structure and language change: New approaches to word order variation in germanic* (p. 45-66).

Hirt, B. (2012). *Measuring the Award's impact*. London: The Duke of Edinburgh's International Award Foundation.

Hirt, H. (1913). Fragen des Vokalismus und der Stammbildung im Indogermanischen. *Indogermanische Forschungen*, *32*, 236-247.

Hirt, H. (1921). *Indogermanische Grammatik II: Der indogermanische Vocalismus*. Heidelberg.

Hoey, M. (2005). *Lexical priming: A new theory of words and language*. Psychology Press.

Holmberg, A. (2005). Is there a litte pro? evidence from Finnish. *Linguistic Inquiry*, *36*(4), 533–564.

Holmberg, A. (2013). Verb second. In T. Kiss & A. Alexiadou (Eds.), *Syntax: an International Handbook of Contemporary Syntactic Research*. Berlin: Walter de Gruyter Verlag.

Holmberg, A., & Roberts, I. (2005). On the role of parameters in Universal Grammar: A reply to Newmeyer. In H. Broekhuis, N. Corver, R. Huybregts, U. Kleinhenz, & J. Koster (Eds.), *Organizing grammar: Linguistic studies in honor of henk van riemsdijk* (p. 538-553). Berlin: Mouton de Gruyter.

Hopper, P. J. (1975). *The syntax of the simple sentence in Proto-Germanic* (Vol. 143). Den Haag: De Gruyter Mouton.

Hopper, P. J., & Traugott, E. C. (2003). *Grammaticalization*. Cambridge: Cambridge University Press.

Horvath, J. (1981). *Aspects of Hungarian syntax and the theory of grammar*. Unpublished doctoral dissertation, University of California, Los Angeles.

Horvath, J. (2010). "discourse features", syntactic displacement and the status of contrast. *Lingua*, *120*(6), 1346-1369.

Hruska, C., Alter, K., Steinhauer, K., & Steube, A. (2000). Can wrong prosodic information be mistaken by the brain. *Journal of Cognitive Neuroscience*, Supplement 122: E82.

Hunston, S. (2010). How can a corpus be used to explore patterns? In A. O'ÂĂÂŹKeeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (p. 152-166). London: Routledge.

Huws, D. (1991). Llyfr Gwyn Rhydderch. *CMCS*, *21*, 1-37.

Jackendoff, R. (1972). *Semantic interpretation in generative grammar*. Cambridge, Massachusetts: MIT Press.

Jamison, S. W. (1983). *Function and Form in the -áys-formations of the Rig Veda and Atharva Veda*. Göttingen: Vandenhoeck & Ruprecht.

Johansson, S., Leech, G., & Goodluck, H. (1978). *Manual of Information to Accompany the Lancaster-Olso/Bergen Corpus of British English, for Use with Digital Computers*. Oslo: Department of English, Oslo University.

Johnson, K. (2003). In search of the English middle field.

Johnson, S. (1755). *A dictionary of the English Language*. London, Consortium.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness* (No. 6). Cambridge, MA: Harvard University Press.

Johnson-Laird, P. N. (1989). Mental models. In *Foundations of cognitive science* (p. 469-499). Cambridge, MA: MIT Press.

Johnson-Laird, P. N. (2013). Mental models and cognitive change. *Journal of Cognitive Psychology*, *25*(2), 131-138.

Johnson-Laird, P. N., Byrne, R. M., & Schaeken, W. (1992). Propositional reasoning by model. *Psychological review*, *99*(3), 418-439.

Johnston, T. (2010). From archive to corpus: transcription and annotation in the creation of signed language corpora. *International journal of corpus linguistics*, *15*(1), 106-131.

Kaiser, E., & Trueswell, J. C. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition*, *94*(2), 113-147.

Karttunen, L. (1974). Presupposition and linguistic context. *Theoretical linguistics*, *1*(1-3), 181-194.

Kay, P., & Fillmore, C. J. (1999). Grammatical constructions and linguistic generalizations: the What's X doing Y? construction. *Language, 75*, 1-33.

Kayne, R. S. (2000). *Parameters and universal grammar*. Oxford: Oxford University Press.

Keenan, E. L. (2002). Explaining the creation of reflexive pronouns in English. In D. Minkova & R. Stockwell (Eds.), *Studies in the history of the english language: a millennial perspective.* (p. 325-355). Berlin: Mouton De Gruyter.

Kelly, E. P., & Sikora, M. (2011). *Reading the Fadden More Psalter: an introduction.* Dublin: National Museum of Ireland.

Kennedy, G. (1998). *An introduction to corpus linguistics*. London, New York: Longman.

King, G. (1993). *A comprehensive grammar of Modern Welsh*. Oxford: Oxford University Press.

Kintsch, W. (1989). The representation of knowledge and the use of knowledge in discourse comprehension. *Language processing in social context*, 185-209.

Kintsch, W., & Rawson, K. A. (2005). Comprehension. In M. Snowling & C. Hulme (Eds.), *The Science of Reading: A Handbook* (p. 209-226). Oxford: Blackwell.

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological review*, *85*(5), 363-394.

Kiparsky, P. (1995). Indo-european origins of Germanic syntax. In A. Battye & I. Roberts (Eds.), *Clause structure and language change* (pp. 140–172). Oxford: Oxford University Press.

Kirk, A. (2012). *Word order and information structure in New Testament Greek*. Utrecht: LOT dissertation series.

Koch, J. (1991). On the prehistory of Brittonic Syntax. In J. Fife & E. Poppe (Eds.), *Studies in Brythonic word order* (Vol. 83, p. 1-43). Amsterdam, Philadelphia: John Benjamins Publishing Company.

Koch, J. (1992). "Gallo-brittonic" Vs. "insular Celtic": The Inter-relationships of the Celtic Languages Reconsidered. In *Bretagne et Pays Celtiques: Langues, Histoire, Civilisation.* Rennes: Presses Universitaires Rennes, Saint-Brieuc:Skol.

Komen, E. (2009a). CESAC: Coreference Editor for Syntactically Annotated Corpora. In *7th York-Newcastle-Holland Symposium on the History of English Syntax (SHES7)* (Vol. 8). Nijmegen: CLS Department English Language and Culture.

Komen, E. (2009b). *Corpus Studio manual*. Nijmegen: Radboud University Nijmegen.

Komen, E. (2013). *Finding Focus*. Utrecht: LOT Dissertation Series.

Komen, E., & Los, B. (2012). *The pentaset: Annotating information state primitives.*

Koopman, H. (1984). *The syntax of verbs*. Dordrecht, The Netherlands: Foris Publications.

Koster, J. (2000). Extraposition as parallel construal. *Ms. University of Groningen*.

Krámsky, J. (1972). A Contribution to the Investigation of the Frequency of Occurrence of Nominal and Verbal Elements in English. *Prague Studies in Mathematical Linguistics*, *4*, 35-45.

Kratzer, A. (2009). Making a Pronoun: Fake Indexicals as Windows into the Properties of Pronouns. *Linguistic Inquiry*, *40*(2), 187–237.

Krifka, M. (1999). Additive particles under stress. *Semantics and Linguistic Theory (SALT)*, *IX*, 111-128.

Krifka, M. (2008). Basic notions of information structure. *Acta Linguistica Hungarica*, *55*(3), 243-276.

Krifka, M., & Musan, R. (2012). Information structure: Overview and linguistic issues. In M. Krifka & R. Musan (Eds.), *The expression of information structure.* Berlin: De Gruyter Mouton.

Kroch, A. (1989). Reflexes of grammar in patterns of language change. *Language variation and change*, *1*(03), 199-244.

Kroch, A. (2000). Verb-object order in early Middle English. In S. Pintzuk, S. Tsoulas, & A. Warner (Eds.), *Diachronic syntax: Models and mechanisms* (p. 132-163). Oxford University Press.

Kroch, A., Santorini, B., & Delfs, L. (2004). *The Penn-Helsinki Parsed Corpus of Early Modern English (PPCEME)*. Department of Linguistics, University of Pennsylvania. CD-ROM, first edition, `http://www.ling.upenn.edu/hist-corpora`.

Kroch, A., Santorini, B., & Diertani, A. (2010). *The Penn-Helsinki parsed corpus of Modern British English (PPCMBE)*. Department of Linguistics, University of Pennsylvania. CD-ROM, first edition, `http://www.ling.upenn.edu/hist-corpora`.

Kroch, A., & Taylor, A. (2000). *The Penn-Helsinki Parsed Corpus of Middle English (PPCME2)*. Department of Linguistics, University of Pennsylvania. CD-ROM, second edition, `http://www.ling.upenn.edu/hist-corpora`.

Kruijff, G.-J., & Duchier, D. (2003). Information structure in topological dependency grammar. In *Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics-Volume 1* (p. 219-226).

Kučerová, I., & Neeleman, A. (2012). *Contrasts and positions in information structure*. Cambridge: Cambridge University Press.

Kuno, S. (1987). *Functional Syntax: Anaphora, Discourse, and Empathy*. Chicago: University of Chicago Press.

Kuryłowicz, J. (1927). Les effets du ə en indoiranien. *Prace Filologiczne, 11*, 201-243.

Kuryłowicz, J. (1949). La nature des procès dits «analogiques». *Acta linguistica, 5*(1), 15-37.

Kutas, M., Van Petten, C., & Kluender, R. (2006). Psycholinguistics electrified ii (1994-2005). In M. Gernsbacher & M. Traxler (Eds.), *Handbook of psycholinguistics* (p. 659-724). New York: Elsevier.

Kytö, M. (1991). *Manual to the Diachronic Part of "The Helsinki Corpus of English Texts" Coding Conventions and Lists of Source Texts*. University of Helsinki, Department of English.

Kytö, M., & Rissanen, M. (1992). A language in transition: The Helsinki Corpus of English Texts. *ICAME Journal, 16*, 7-27.

Labov, W. (2001). *Principles of linguistic change, ii: social factors*. Oxford: Blackwell.

Lambrecht, K. (1994). *Information structure and sentence form: A theory of topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press.

Langacker, R. W. (1988). A usage-based model. *Topics in cognitive linguistics, 50*, 127-163.

Lash, E. J. F. (2011). *A Synchronic and Diachronic Analysis of Old Irish Copular Clauses*. Unpublished doctoral dissertation, Cambridge University.

Lass, R. (1980). *On explaining language change*. Cambridge: Cambridge University Press.

Lass, R. (1997). *Historical linguistics and language change* (Vol. 81). Cambridge:Cambridge University Press.

Ledgeway, A. (2016). Introduction. In A. Ledgeway & I. Roberts (Eds.), *Cambridge Handbook of Historical Syntax*. Cambridge: Cambridge University Press.

Lee, D. Y. (2010). What corpora are available? In A. O'ÂÃÂŹKeeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (p. 107-121). London: Routledge.

Leech, G. (1992). Corpora and theories of linguistic performance. *Directions in corpus linguistics: Proceedings of the Nobel Symposium 82, Stockholm, 4-8 August 1991*, 105-122.

Leech, G. (2004a). *Meaning and the english verb*. London: Pearson Education.

Legate, J. A. (2002). *Warlpiri: theoretical implications*. Unpublished doctoral dissertation, Massachusetts Institute of Technology.

Lehmann, W. P. (1972). Proto-Germanic syntax. *Toward a grammar of Proto-Germanic*, 239-268.

Lehmann, W. P. (1973). A structural principle of language and its implications. *Language, 49*, 42–66.

Levinsohn, S. (2009). *Self-instruction materials on narrative discourse analysis.*

SIL-International, available at `www.sil.org/ levinsohns/narr.pdf`.

Levinson, S. C. (1983). *Pragmatics*. Cambridge: Cambridge University Press.

Levinson, S. C., & Evans, N. (2010). Time for a sea-change in linguistics: Response to comments on the myth of language universals. *Lingua*, *120*(12), 2733-2758.

Li, C. N., & Thompson, S. A. (1976). Subject and Topic: A New Typology of Language in Subject and Topic. In C. N. Li & S. A. Thompson (Eds.), *Subject and topic* (p. 458-489). New York: Academic Press.

Lidz, J., & Williams, A. (2009). Constructions on holiday. *Cognitive linguistics*, *20*(1), 177-189.

Lightfoot, D. (1979). *Principles of diachronic syntax*. Cambridge: Cambridge University Press.

Lightfoot, D. (1991). *How to set parameters: Arguments from language change*. Cambridge: Cambridge University Press.

Lightfoot, D. (1999). *The development of language: Acquisition, change, and evolution*. Oxford: Blackwell.

Lightfoot, D. (2002). Myths and the prehistory of grammars. *Journal of Linguistics*, *38*(01), 113-136.

Lightfoot, D. (2006). *How new languages emerge*. Cambridge: Cambridge University Press.

Lightfoot, D., & Westergaard, M. (2007). Language Acquisition and Language Change: Inter-relationships. *Language and Linguistics Compass*, *1*(5), 396-416.

Lions, J. (1996). *Ariane 5: Flight 501 Failure*. Available from `http://sunnyday.mit.edu/accidents/Ariane5accidentreport.html`. Last accessed d.d. 3 September 2014.

Lipták, A. (2011). The structure of the topic field in Hungarian. In P. Beninca & N. Munaro (Eds.), *Mapping the left periphery* (p. 163-200).

Longobardi, G. (2001). How comparative is semantics? a unified parametric theory of bare nouns and proper names. *Natural Language Semantics*, *9*(4), 335–369.

Longobardi, G. (2003). Methods in parametric linguistics and cognitive history. In P. Pica (Ed.), *Linguistic variation yearbook* (Vol. 3, p. 101-138). Amsterdam, Philadelphia: John Benjamins Publishing Company.

Longobardi, G., & Roberts, I. (2010). Universals, diversity and change in the science of language: Reaction to "The Myth of Language Universals and Cognitive Science". *Lingua*, *120*(12), 2699-2703.

López, L. (2009). *A derivational syntax for information structure* (Vol. 23). Oxford: Oxford University Press.

Lozano, C. (2006). Focus and split-intransitivity: the acquisition of word order alternations in non-native Spanish. *Second Language Research*, *22*(2), 145–187.

Lubotsky, A. (1990). La loi de Brugmann et *$*H_3e$*. La reconstruction des laryngales*, *CCLIII*, 129–136.

Lubotsky, A. (1997). Review of: Marianne Volkart, Zu Brugmanns Gesetz im Altindischen. *Kratylos*, *42*, 55-59.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.

Lüdeling, A., & Kytö, M. (2008). Introduction. In A. Lüdeling & M. Kytö (Eds.), *Corpus Linguistics. An International handbook. Volume 1* (p. i-xii). Berlin: Walter de Gruyter.

MacCana, P. (1973). On Celtic Word and the Welsh 'Abnormal' Sentence. *Ériu*, 90-120.

MacCana, P. (1976). Latin influence on British: the pluperfect. In J. J. O'Meara & B. Naumann (Eds.), *Latin Script and Letters, AD 400-900: Festschrift presented to Ludwig Bieler on the Occasion of his 70th Birthday* (p. 194-203). Leiden: Brill.

MacCana, P. (1991). Further notes on constituent order in Welsh. In J. Fife & E. Poppe (Eds.), *Studies in Brythonic word order* (Vol. 83, p. 45-80). Amsterdam, Philadelphia: John Benjamins Publishing Company.

MacWhinney, B. (2000). The CHILDES project: Tools for analyzing talk: Volume i: Transcription format and programs, volume ii: The database. *Computational Linguistics*, *26*(4), 657-657.

Manning, H. P. (1995). Fluid intransitivity in Middle Welsh: Gradience, typology and 'unaccusativity'. *Lingua*, *97*(2), 171-194.

Manning, H. P. (1997). The geology of railway embankments: Oxford Welsh and the 'abnormal sentence'. In *Papers from the Panels on Linguistic Ideologies in Contact* (Vol. 33, p. 59-74). Chicago Linguistic Society.

Manning, H. P. (2004). The geology of railway embankments: Celticity, Liberalism, the Oxford Welsh reforms, and the word order (s) of Welsh. *Language & Communication*, *24*(2), 135-163.

Marriott, K., Meyer, B., & Wittenburg, K. B. (1998). A survey of visual language specification and recognition. In *Visual language theory* (p. 5-85). Springer.

Marty, A. (1884). Über subjektlose sätze und das verhältnis der grammatik zur logik und philosophie, iii: Von gewissen unterschieden der sprachlichen ausdrücke und speziell der aussagen, die nicht den durch sie bezeichneten gedanken betreffen ('innere sprachform'und deren wirkungen).". *Vierteljahrsschrift für wissenschaftliche Philosophie*, *8*, 292-340.

Mathesius, V. (1929 [1983]). Functional linguistics. In J. Vachek (Ed.), *Praguiana* (p. 121-42). Amsterdam, Philadelphia: John Benjamins Publishing Company.

Maynard, C., & Leicher, S. (2007). Pragmatic annotation of an academic spoken corpus for pedagogical purposes. In *Corpus Linguistics beyond the Word. Corpus Research from Phrase to Discourse.* (p. 107-115). Amsterdam: Rodopi.

McCloskey, J. (1992). Adjunction, selection and embedded verb second. *Ms University of California at Santa Cruz*.

McEnery, T., & Hardie, A. (2012a). *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.

McEnery, T., & Hardie, A. (2012b). *Corpus linguistics: Method, theory and practice*.

Cambridge: Cambridge University Press.

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. London: Routledge.

McEnery, T., & Xiao, Z. (2004a). The Lancaster Corpus of Mandarin Chinese: A corpus for monolingual and contrastive language study. In M. Lino, M. Xavier, F. Ferreire, R. Costa, & R. Silva (Eds.), (p. 1175-8).

McFadden, T. (2014). *Corpus research methodology*. Handout from seminar on Corpus research for historical syntax, Utrecht, April 2014.

Meelen, M., & Beekhuizen, B. (2013). PoS-tagging and chunking historical Welsh. In *Proceedings of the scottish celtic colloquium 2012.*

Meillet, A. (1958 [1912]). L'évolution des formes grammaticales. In *Linguistique historique et linguistique générale* (p. 130-58). Paris: Champion.

Meisel, J., Elsig, M., & Rinke, E. (2013). *Language Acquisition and Change: A Morphosyntactic Perspective*. Oxford: Oxford University Press.

Mereu, L. (2009). Introduction. In L. Mereu (Ed.), *Information Structure and its Interfaces* (Vol. 19, p. 1-11). Berlin: Walter de Gruyter.

Meurman-Solin, A., López-Couso, M. J., & Los, B. (2012). On the interplay of syntax and information structure: synchronic and diachronic considerations. In A. Meurman-Solin, M. J. López-Couso, & B. Los (Eds.), *Information structure and syntactic change in the history of English* (p. 3-18). Oxford: Oxford University Press.

Meyer, C. F. (2002). *English corpus linguistics: An introduction*. Cambridge: Cambridge University Press.

Meyer, C. F. (2008). Pre-electronic corpora. In A. Lüdeling & M. Kytö (Eds.), *Corpus Linguistics. An International handbook. Volume 1* (p. 1-13). Berlin: Walter de Gruyter.

Meyer, C. F., & Tao, H. (2005). Response to Newmeyer's 'grammar is grammar and usage is usage'. *Language*, *81*(1), 226–228.

Milroy, J. (1992). *Linguistic variation and change: on the historical sociolinguistics of English*. Oxford: Blackwell.

Mithun, M. (1987). Is basic word order universal? In R. S. Tomlin (Ed.), *Coherence and grounding in discourse* (p. 281-328). Amsterdam, Philadelphia: John Benjamins Publishing Company.

Morris Jones, J. (1913). *A Welsh grammar: historical and comparative*. Oxford: Clarendon Press.

Morris-Jones, J. (1931). *Welsh syntax: an unfinished draft*. Cardiff: the University of Wales Press.

Motut, A. (2010). Merge over Move and the empirical force of economy in Minimalism. *Toronto Working Papers in Linguistics*, *33*, 1-54.

Neeleman, A., & Van de Koot, H. (2008). Dutch scrambling and the nature of discourse templates. *Journal of Comparative Germanic Linguistics*, *11*(2), 137-189.

Nekula, M. (1999). Vilém mathesius. In J.-O. Verschueren, J. Östman, J. Blommaert, & C. Bulcaen (Eds.), *Handbook of Pragmatics* (p. 1-14). Amster-

dam/Philadelphia: John Benjamins Publishing Company.

Nelson, M. (2010). Building a written corpus: what are the basics? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (p. 53-65). London: Routledge.

Nevalainen, T., & Raumolin-Brunberg, H. (2003). *Historical Sociolinguistics: Language Change in Tudor and Stuart England*. London: Pearson Education.

Newmeyer, F. J. (2004). Against a parameter-setting approach to typological variation. In P. Pica, J. Rooryck, & J. Van Craenenbroek (Eds.), *Linguistic variation yearbook* (Vol. 4, p. 181-234). Amsterdam, Philadelphia: John Benjamins Publishing Company.

Newmeyer, F. J. (2005). A reply to the critiques of 'grammar is grammar and usage is usage'. *Language, 81*(1), 229–236.

Newton, G. (2006). *The development and loss of the Irish double system of inflection*. Unpublished doctoral dissertation, Cambridge University.

Nivre, J. (2008). Treebanks. In A. Lüdeling & M. Kytö (Eds.), *Corpus Linguistics. An International handbook. Volume 1* (Vol. 1, p. 225-241). Berlin: Walter de Gruyter.

Niyogi, P., & Berwick, R. C. (2009). The proper treatment of language acquisition and change in a population setting. *Proceedings of the National Academy of Sciences, 106*(25), 10124-10129.

Nolda, A. (2004). Topics detached to the left. On 'left dislocation', 'hanging topic' and related constructions in German. *ZAS Papers in Linguistics, 35*, 423-448.

Norde, M. (2009). *Degrammaticalization*. Oxford: Oxford University Press.

Nurmio, S. (2015). *Studies in grammatical number in Old and Middle Welsh*. Unpublished doctoral dissertation, St John's College, University of Cambridge.

Nurmio, S., & Willis, D. (2016). The rise and fall of a minor number: The case of the Welsh numerative. *Unpublished Ms.*.

O'Brien, E. J., Rizzella, M. L., Albrecht, J. E., & Halleran, J. G. (1998). Updating a situation model: a memory-based text processing view. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(5), 1200-1210.

O'Keeffe, A., & McCarthy, M. (2010). Historical perspective: what are corpora an how have they evolved? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (p. 3-13). London: Routledge.

O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge: Cambridge University Press.

Onea, E., & Beaver, D. (2011). Hungarian focus is not exhausted. *Semantics and Linguistic Theory, 19*, 342-359.

Osgood, C. E., & Sebeok, T. A. (1954). *Psycholinguistics: a survey of theory and research problems* (Vol. 49). American Psychological Association.

Osthoff, H., & Brugmann, K. (1878). *Morphologische Untersuchungen auf dem Gebiete der indogermanischen Sprachen* (Vol. 1). Leipzig: S. Hirzel.

Östman, J.-O., & Virtanen, T. (1999). Theme, comment, and newness as figures in information structuring. *Amsterdam studies in the theory and history of linguistic science series, 4*, 91-110.

Ouhalla, J. (1994). Verb movement and word order in Arabic. In D. Lightfoot & N. Hornstein (Eds.), *Verb movement* (pp. 41–72). Cambridge, Great Britain: Cambridge University Press.

OUP. (2014). *History of the OED*. Available online from `http://public.oed.com/history-of-the-oed` last accessed d.d. 3 September 2014.

Parker, W. (2007). *The Four Branches of the Mabinogi*. Dublin: Bardic Press.

Paul, H. (1920 [1880]). *Prinzipien der Sprachgeschichte*. Tübingen: Niemeyer.

Payne, D. (1987). Information structuring in Papago narrative discourse. *Language*, *63*, 783-804.

Pearl, L. (2014). Evaluating learning-strategy components: Being fair (Commentary on Ambridge, Pine, and Lieven). *Language*, *90*(3), e107-e114.

Pedersen, H. (1913). *Vergleichende Grammatik der keltischen Sprachen* (Vol. 2). Göttingen: Vandenhoeck und Ruprecht.

Pinker, S. (1984). Visual cognition: An introduction. *Cognition*, *18*(1), 1-63.

Pintzuk, S. (1991). *Phrase structures in competition: variation and change in Old English word order*. Unpublished doctoral dissertation, University of Pennsylvania.

Pintzuk, S. (2002). Verb-object order in Old English: variation as grammatical competition. In D. Lightfoot (Ed.), *Syntactic effects of morphological change* (p. 276-299). Oxford: Oxford University Press.

Pintzuk, S., & Plug, L. (2002). *The York-Helsinki parsed corpus of Old English poetry*. Department of Linguistics, University of York. Oxford Text Archive, first edition, `http://www-users.york.ac.uk`.

Plackett, R. L. (1983). Karl pearson and the chi-squared test. *International Statistical Review/Revue Internationale de Statistique*, 59-72.

Plein, K., & Poppe, E. (2014). Patterns of verbal agreement in "Historia Gruffud vab Kenan": norm and variation. *Études celtiques*(40), 145-164.

Pollard, C., & Sag, I. (1987). *Information-based Syntax and Semantics* (Vol. 1). Stanford: CSLI publications.

Pollard, C., & Sag, I. A. (1994). *Head-driven phrase structure grammar*. University of Chicago Press.

Poppe, E. (1989). Constituent ordering in 'Breudwyt Maxen Wledic'. *BBCS*, *36*, 43-63.

Poppe, E. (1990). Word-order patterns in Breudwyt Ronabwy. In M. Ball, J. Fife, E. Poppe, & J. Rowland (Eds.), *Celtic Linguistics. Ieithyddiaeth Geltaidd. Readings in the Brythonic Languages. Festschrift for TA Watkins* (p. 445-460). Amsterdam: John Benjamins.

Poppe, E. (1991). Word order in Cyfranc Lludd a Llefelys: note on the pragmatics of constituent-ordering in MW narrative prose. In J. Fife & E. Poppe (Eds.), *Studies in Brythonic word order* (Vol. 83, p. 155-205). Amsterdam, Philadelphia: John Benjamins Publishing Company.

Poppe, E. (1993). Word order in Middle Welsh: the case of Kedymdeithyas Amlyn ac Amic. *Bwletin y Bwrdd Gwybodau Celtaidd*, *40*, 95-117.

Poppe, E. (2009). The pragmatics of Middle Welsh word order: Some conceptual and descriptive problems. In *Pragmatische Kategorien. Form, Funktion und Diachronie.* (Vol. 24, p. 247-264).

Poppe, E. (2014). How to Achieve an Optimal Textual Fit in Middle Welsh Clauses. *Cambrian Medieval Celtic Studies*, *68*, 69-100.

Popper, K. (1935). *Logik der Forschung: zur Erkenntnistheorie der modernen Naturwissenschaft*. Vienna: Springer.

Popper, K. (1968). *The Logic of Scientific Discovery*. New York: Harper Torch Books.

Preminger, O. (2011). *Agreement as a fallible operation*. Unpublished doctoral dissertation, Massachusetts Institute of Technology.

Prince, E. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical pragmatics* (p. 223-255). Academic Press, New York.

Ramble, C. (2013). Both Fish and Fowl? Preliminary Reflections on Some Representations of a Tibetan Mirror-World. In F.-K. Ehrhard & P. Maurer (Eds.), *Nepalica-Tibetica: Festgabe for Christoph Cüppers* (Vol. 2, p. 75-89). International Institute for Tibetan and Buddhist Studies GmbH.

Randall, B., Taylor, A., & Kroch, A. (2005). *Corpussearch 2*. Available at: http://corpussearch.sourceforge.net/credits.html.

Reinhart, T. (1981). Definite NP anaphora and C-command domains. *Linguistic Inquiry*, *12*(4), 605–635.

Reinhart, T. (1982). Pragmatics and Linguistics: An Analysis of Sentence Topics. *Philosophica*, *27*(1), 53-94.

Repp, S. (2010). Defining 'contrast' as an information-structural notion in grammar. *Lingua*, *120*(6), 1333-1345.

Rhys, J., & Jones, B. (1902). *The Welsh People: Chapters on Their Origin, History, Laws, Language, Literature, and Characteristics*. T. Fisher Unwin.

Richards, M. (1938). *Cystrawen y frawddeg Gymraeg*. Caerdydd: Gwasg Prifysgol Cymru.

Riester, A., Lorenz, D., & Seemann, N. (2010). A Recursive Annotation Scheme for Referential Information Status. In *Proceedings of the seventh international conference of language resources and evaluation (LREC), Valletta, Malta* (p. 717-722).

Rissanen, M. (1989). Three problems connected with the use of diachronic corpora. *ICAME journal*, *13*, 16-19.

Rissanen, M. (1998). Towards an integrated view of the development of English: Notes on causal linking. *Trends in linguistics studies and monographs*, *112*, 389-406.

Rissanen, M. (2008). Corpus linguistics and historical linguistics. In *Corpus Linguistics. An International Handbook.* Berlin: Mouton de Gruyter.

Ritz, J., Dipper, S., & Götze, M. (2008). Annotation of Information Structure: an Evaluation across different Types of Texts. In *Proceedings of the 6th LREC-2008 Conference.* Marrakech, Morocco.

Rizzi, L. (1997). The fine structure of the left periphery. In L. Haegeman (Ed.), *Elements of grammar* (pp. 281–337). Dordrecht: Kluwer Academic Publishers.

Rizzi, L. (2004). On the cartography of Syntactic Structures. In L. Rizzi (Ed.), *The structure of CP and IP* (pp. 3–15). Oxford: Oxford University Press.

Roberts, I. (1997). Restructuring, head movement, and locality. *Linguistic Inquiry*, *28*(1), 423–460.

Roberts, I. (2004). The C-system in Brythonic Celtic languages, V2, and the EPP. In L. Rizzi (Ed.)*, The structure of CP and IP: the Cartography of syntactic Structures* (Vol. 2, pp. 297–328).

Roberts, I. (2005). *Principles and parameters in a VSO language: A case study in Welsh*. Oxford: Oxford University Press.

Roberts, I. (2007). *Diachronic syntax*. Oxford: Oxford University Press.

Roberts, I. (2009). A deletion analysis of null subjects. In T. Biberauer, A. Holmberg, I. Roberts, & M. Sheehan (Eds.), *Parametric variation: Null subjects in minimalist theory* (p. 58-87). Cambridge: Cambridge University Press.

Roberts, I. (2010). *Agreement and head movement: Clitics, incorporation, and defective goals*. MIT Press.

Roberts, I. (2012). Macroparameters and minimalism. In C. Galves, S. Cyrino, R. Lopes, F. Sandalo, & J. Avelar (Eds.), *Parameter Theory and Linguistic Change* (Vol. 2, p. 320-335). Oxford: Oxford University Press.

Roberts, I., & Holmberg, A. (2005). On the role of parameters in Universal Grammar: A reply to Newmeyer. In H. Broekhuis, N. Corver, M. Everaert, & J. Koster (Eds.), *Organizing grammar: Linguistic studies in honor of Henk van Riemsdijk* (p. 538-553). Berlin: Mouton de Gruyter.

Roberts, I., & Roussou, A. (2003). *Syntactic change: A minimalist approach to grammaticalization* (Vol. 100). Cambridge: Cambridge University Press.

Rodway, S. (2004). The Red Book Text of "Culhwch ac Olwen": A Modernising Scribe at Work. *Studi Celtici*, 93-161.

Rodway, S. (2013). *Dating Medieval Welsh Literature: Evidence from the verbal system*. CMCS.

Rooth, M. E. (1985). *Association with focus*. Unpublished doctoral dissertation, University of Massachusetts, Amherst.

Rosenkvist, H. (2010). A case of degrammaticalization in northern Swedish. In A. Breitbarth, C. Lucas, S. Watts, & D. Willis (Eds.), *Continuity and change in grammar* (Vol. 159, p. 303-320). Amsterdam, Philadelphia: John Benjamins Publishing Company.

Ross, J. R. (1986). *Infinite syntax*. New Jersey: Ablex Norwood.

Rouveret, A. (1994). *Syntaxe du gallois: principes généraux et typologie*. CNRS.

Rowland, T. (1876). *A grammar of the Welsh language*. Wrexham: Hughes & son.

Rühlemann, C. (2010). What can a corpus tell us about pragmatics? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (p. 288-301). London: Routledge.

Russell, P. (1999). What did medieval Welsh scribes do? The scribe of the 'Dingestow Court Manuscript'. *Cambrian Medieval Celtic Studies*(37), 79-96.

Russell, P. (2003). Texts in Contexts: Recent Work on the Medieval Welsh Prose Tales. *CMCS*, 59-72.

Russell, P. (2012). An habes linguam Latinam? Non tam bene sapio: Views of Multilingualism from the Early Medieval West. In A. Mullen & P. James (Eds.), *Multilingualism in the graeco-roman worlds* (p. 193-224). Cambridge: Cambridge University Press.

Sampson, G. R. (2007). Grammar without grammaticality. *Corpus linguistics and linguistic theory*, *3*(1), 1-32.

Sasse, H.-J. (1987). The thetic/categorical distinction revisited. *Linguistics*, *25*(3), 511–580.

Schmalhofer, F., Friese, U., Pietruska, K., Raabe, M., & Rutschmann, R. (2005). Brain processes of relating a statement to a previously read text: Memory resonance and situational constructions. In *Proceedings of the XVII Conference of The Cognitive Science Society* (p. 1949-1954).

Schmidt, K. H. (1990). Gallo-Brittonic or Insular Celtic. In *Studia indogermanica et palaeohispanica in honorem A. Tovar et L. Michelena* (p. 255-267). Universidad del País Vasco/Universidad de Salamanca.

Schrijver, P. (1995). *Studies in British Celtic historical phonology* (Vol. 5). Amsterdam: Rodopi.

Schrijver, P. (1997). *Studies in the History of Celtic Pronouns and Particles* (Vol. 2). Department of Old Irish, National University of Ireland.

Schrijver, P. (2002). The rise and fall of British Latin: Evidence from English and Brittonic. In M. Filppula, J. Klemola, & H. Pitkänen (Eds.), *The Celtic Roots of English* (p. 87-110). University of Joensuu.

Schrijver, P. (2007). What Britons spoke around 400 AD. In N. Higham (Ed.), *Britons in Anglo-Saxon England* (p. 165-71). Woodbridge.

Schrijver, P. (2014). *Language contact and the origins of the Germanic languages* (Vol. 13). London: Routledge.

Schumacher, S. (2011). Mittel- und Frühneukymrisch. In E. Ternes (Ed.), *Brythonic Celtic - Britannisches Keltisch: from Medieval British to Modern Breton* (Vol. 11, p. 85-236). Munich Studies in Historical Linguistics.

Schütze, C. T., Sprouse, J., & Caponigro, I. (2015). Challenges for a theory of islands: A broader perspective on Ambridge, Pine, and Lieven. *Language*, *91*(2), e31-e39.

Schwarzschild, R. (1999). Givenness, avoid F and other constraints on the placement of accent. *Natural Language Semantics*, *7*(2), 141–177.

Scott, M. (2010). What can corpus software do? In A. O'ÂĂÂŹKeeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (p. 136-151). London: Routledge.

Sheehan, M. L. (2007). *The EPP and null subjects in Romance*. Unpublished doctoral dissertation, University of Newcastle upon Tyne.

Sheehan, M. L. (2015). On the lack of consistency in (Romance) consistent null subject languages. *Paper presented at Cambridge Comparative Syntax (CamCos) 4, May 2015*.

Silk, J. (2014). Keeping Up With the Joneses: From William Jones to John James Jones. *Annual Report of The International Research Institute for Advanced*

Buddhology at Soka University(17), 427-441.

Sims-Williams, P. (2003). *The Celtic Inscriptions of Britain: Phonology and Chronology, c. 400-1200*. Publications of the Philological Society, Blackwell Publishing.

Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*. London: Routledge.

Speyer, A. (2008). Doppelte Vorfeldbesetzung im heutigen Deutsch und im Frühneuhochdeutschen. *Linguistische Berichte*, *2008*(216), 455-485.

Sproat, R. (1983). Vso languages and welsh configurationality. In *Proceedings of the harvard celtic colloquium* (Vol. 3, pp. 39–68).

Sproat, R. (1985). Welsh syntax and VSO structure. *Natural Language and Linguistic Theory*, *3*(2), 173–216.

Sprouse, J., & Lau, E. F. (2013). Syntax and the brain. In *The Cambridge Handbook of Generative Syntax* (p. 971-1005). Cambridge: Cambridge University Press.

Stalnaker, R. (1974). Pragmatic presuppositions. In P. K. Unger & M. K. Munitz (Eds.), *Semantics and philosophy* (p. 197-214). New York: New York University Press.

Stalnaker, R. (2002). Common ground. *Linguistics and philosophy*, *25*(5), 701-721.

Stanfill, C., & Waltz, D. (1986). Toward memory-based reasoning. *Communications of the ACM*, *29*(12), 1213-1228.

Stifter, D. (2011). The textual arrangement of Alise-Sainte-Reine [L-13]. *Zeitschrift fur Celtische Philologie*, *58*, 165-181.

Strachan, J. (1909). *An introduction to early Welsh*. Manchester: Manchester University Press.

Stump, G. T. (1989). Further remarks on Breton agreement. *Natural Language and Linguistic Theory*, *7*(3), 429–471.

Sturzer, N. (2001). How Middle Welsh expresses the unexpected. *CMCS*(41), 37-53.

Swales, J. M. (2002). Integrated and fragmented worlds: EAP materials and corpus linguistics. In J. Flowerdew (Ed.), *Academic discourse* (p. 150-164). Longman Harlow, Essex.

Szendrői, K. (2001). *Focus and the syntax-phonology interface*. Unpublished doctoral dissertation, University College London.

Takahashi, N., Miner, L. L., Sora, I., Ujike, H., Revay, R. S., Kostic, V., . . . Uhl, G. R. (1997). VMAT2 knockout mice: heterozygotes display reduced amphetamine-conditioned reward, enhanced amphetamine locomotion, and enhanced MPTP toxicity. *Proceedings of the National Academy of Sciences*, *94*(18), 9938–9943.

Tallerman, M. (1996). Fronting constructions in Welsh. In R. Borsley (Ed.), *The syntax of the Celtic languages: a comparative perspective* (p. 97). Cambridge: Cambridge University Press.

Tallerman, M. (1998). Celtic word order: Some theoretical issues. In *Constituent order in the languages of Europe* (Vol. 20, p. 599-648). Berlin, New York: Mouton de Gruyter.

Tallerman, M. (2011). *Understanding syntax: third edition*. London: Hodder

Education.

Tallerman, M., & Wallenberg, J. (2012). *The Middle Welsh historic infinitive.* University of Wales, July 2012.

Taylor, A., & Pintzuk, S. (2014). Testing the theory. In *Information Structure and Syntactic Change in Germanic and Romance Languages* (Vol. 213, p. 53-77). Amsterdam, Philadelphia: John Benjamins Publishing Company.

Taylor, A., & Pintzuk, S. (2015). Verb order, object position, and information status in Old English. In (Vol. 15, p. 318-335). Oxford: Oxford University Press.

Taylor, A., Warner, A., Pintzuk, S., & Beths, F. (2003). *The York-Toronto-Helsinki parsed corpus of Old English prose (YCOE).* Department of Linguistics, University of York. Oxford Text Archive, first edition, `http://www-users.york.ac.uk`.

Teleman, U. (1974). *Manual för grammatisk beskrivning av skriven och talad svenska.* Lund: Studentlitteratur.

Thomason, S. G., & Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics.* University of California Press.

Thorndike, E. L., & Lorge, I. (1944). *The teacher's wordbook of 30,000 words.* New York: Columbia University Press.

Thorne, D. (1993). *A comprehensive Welsh grammar.* Oxford: Blackwell.

Thurneysen, R. (2003 [1946]). *A grammar of Old Irish.* Dublin Institute for Advanced Studies.

Traugott, E. C., & König, E. (1991). The semantics-pragmatics of grammaticalization revisited. *Approaches to grammaticalization*, *1*, 189-218.

Traugott, E. C., & Pintzuk, S. (2008). Coding the York-Toronto-Helsinki Parsed Corpus of Old English Prose to investigate the syntax-pragmatics interface. In S. M. Fitzmaurice & D. Minkova (Eds.), *Studies in the History of the English Language IV: Empirical and Analytical Advances in the study of English language change* (p. 61-80).

Travis, L. (1984). *Parameters and effects of word order variation.* Unpublished doctoral dissertation, Massachusetts Institute of Technology.

Vallduví, E. (1992). *The information component.* New York: Garland.

Vallduví, E., & Vilkuna, M. (1998). On rheme and kontrast. *Syntax and semantics*, 79-108.

Van Berkum, J. J., Koornneef, A. W., Otten, M., & Nieuwland, M. S. (2007). Establishing reference in language comprehension: An electrophysiological perspective. *Brain Research*, *1146*, 158-171.

Van der Wal, J. (2009). *Word order and information structure in Makhuwa-Enahara.* Utrecht: LOT dissertation series.

Van der Wal, J. (2015). What you see is (not) what you get: information structure on the interface between syntax and discourse. *Lecture at MFil, 6-7 November 2015, University of Manchester*.

Van der Wurff, W., & Foster, T. (1997). Object-verb word order in 16th century English: A study of its frequency and status. In *Language history and language modelling: a festschrift for Jasek Fisiak on his 60th birthday* (Vol. 101, p. 439-

453). Berlin: Mouton De Gruyter.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Van Gelderen, E. (2004). *Grammaticalization as economy* (Vol. 71). Amsterdam, Philadelphia: John Benjamins Publishing Company.

Van Gelderen, E. (2009). Feature economy in the linguistic cycle. In P. Crisma & P. Longobardi (Eds.), *Historical syntax and linguistic theory* (p. 93-109).

Van Gelderen, E. (2011). *The linguistic cycle: Language change and the language faculty*. Oxford: Oxford University Press.

Van Kemenade, A. (2007). Formal syntax and language change Developments and outlook. *Diachronica*, *24*(1), 155-169.

van Koppen, M. (2007). Agreement with (the internal structure of) copies of movement. *The copy theory of movement*, *107*, 327.

Van Valin, R. (1993b). A synopsis of Role and Reference Grammar. In R. D. Van Valin (Ed.), *Advances in Role and Reference Grammar* (p. 1-164). Amsterdam, Philadelphia: John Benjamins Publishing Company.

Vendryes, J. (1912). La place du verbe en celtique. *Mémoires de la Société de Linguistique de Paris*, *17*, 337-51.

Vennemann, T. (1974). *Topics, Subjects, and Word Order: from SXV to SVX via TVX* (Vol. 2). Amsterdam: North-Holland Publishers.

Vincent, N. (1988). Latin. In M. Harris & N. Vincent (Eds.), *The romance languages* (p. 26-78). Londen: Routledge.

Vitt, A. (2011). *Peredur vab Efrawc: Edited Texts and Translations of the MSS Peniarth 7 and 14 Versions*. Unpublished doctoral dissertation, Aberystwyth University.

Volkart, M. (1994). *Zu Brugmanns Gesetz im Altindischen*. Universität Bern. Institut für Sprachwissenschaft. Arbeitspapier 33.

Von Fintel, K. (2000). What Is Presupposition Accommodation? *Unpublishe Manuscript, MIT*.

Vriezen, T. C., & Van der Woude, A. (2000). *Oudisraëlitische en vroegjoodse literatuur*. Kampen: Uitgeverij Kok.

Wade-Evans, A. (1909). *Welsh Medieval Law: Being a Text of the Laws of Howel the Good, Namely the British Museum Harleian Ms. 4353 of the 13th Century, with Translation, Introduction, Appendix, Glossary, Index, and a Map*. Oxford: Clarendon Press.

Wagner, H. (1959). *Das Verbum in den Sprachen der Britischen Inseln*. Tübingen: Niemeyer.

Walkden, G. (2009). *The comparative method in syntactic reconstruction.* MPhil dissertation: University of Cambridge.

Walkden, G. (2011). Abduction or Inertia? The logic of syntactic change. In C. Cummins, C.-H. Elder, T. Godard, M. Macleod, E. Schmidt, & G. Walkden (Eds.), *Proceedings of the Sixth Cambridge Postgraduate Conference in Language Research* (p. 230-239). Cambridge Institute of Language Research.

Walkden, G. (2012). Against inertia. *Lingua*, *122*(8), 891-901.

Walkden, G. (2014). *Syntactic Reconstruction and Proto-Germanic*. Oxford: Oxford

University Press.

Wallenberg, J., Ingason, A. K., Sigurdsson, E. F., & Rögnvaldsson, E. (2011). *Icelandic Parsed Historical Corpus (IcePaHC)*. `www.linguist.is/icelandic_treebank`.

Wallis, S. (2008). Searching treebanks and other structured corpora. In A. Lüdeling & M. Kytö (Eds.), *Corpus Linguistics. An International handbook* (Vol. 1, p. 738-758). Berlin: Walter de Gruyter.

Wasow, T. (2002). *Postverbal behaviour*. Stanford, California: CSLI Publications.

Watkins, C. (1976). Towards Proto-Indo-European syntax: problems and pseudo-problems. In S. Steever, C. Walker, & S. Mufwene (Eds.), *Papers from the parasession on diachronic syntax* (p. 306-326). Chicago Linguistic Society.

Watkins, C. (1999). Two Celtic notes. In P. Anreiter & E. Jerem (Eds.), *Studia celtica et indogermanica: Festschrift für wolfgang meid zum 70. geburtstag* (p. 539-543). Archaeolingua Alapítvány.

Watkins, T. A. (1977). Trefn yn y frawddeg Gymraeg. *Studia Celtica*, *12/13*, 367-395.

Watkins, T. A. (1987). Constituent order in the Old Welsh verbal sentence. *Bulletin of the board of Celtic Studies*, *34*, 51-60.

Watkins, T. A. (1988). *Constituent order in the positive declarative sentence in the medieval Welsh tale 'Kulhwch ac Olwen'* (Vol. 41). Institut für Sprachwissenschaft der Universität Innsbruck.

Watkins, T. A. (1993). Constituent order in main/simple verb clauses of Pwyll Pendeuic Dyuet. *Language Sciences*, *15*(2), 115-139.

Watkins, T. A. (1997). The sef[...] Realization of the Welsh Identificatory Copular Sentence. In A. Ahlqvist, C. R. Ó. Cléirigh, & V. Čapková (Eds.), *Dán do oide: essays in memory of Conn R. Ó Cléirigh.* Linguistics Institute of Ireland.

Watson, W. G. (1973). *The language and poetry of the book of Isaiah in the light of recent research in Northwest Semitic*. Unpublished doctoral dissertation, University of Aberdeen.

Weber-Wulff, D. (1992). *Rounding error changes Parliament makeup* (Vol. 13). Available from `http://catless.ncl.ac.uk/Risks`. Last accessed d.d. 3 September 2014.

Weinreich, U., Labov, W., & Herzog, M. (1968). *Empirical foundations for a theory of language change*. Austin: University of Texas Press.

West, M. (1953). *A general service list of english words*. London: Longman.

Westergaard, M. (2009). *The acquisition of word order: micro-cues, information structure, and economy* (Vol. 145). Amsterdam, Philadelphia: John Benjamins Publishing Company.

Widdowson, H. G. (2000). On the limitations of linguistics applied. *Applied linguistics*, *21*(1), 3-25.

Williams, S. (1980). *A Welsh grammar*. Cardiff: the University of Wales Press.

Williams ab Ithel, J. (1856). *Dosparth edeyrn davod aur, or The ancient Welsh grammar*. London.

Willis, D. (1998). *Syntactic Change in Welsh: A Study of the Loss of the Verb-second*.

Oxford: Oxford University Press.

Willis, D. (2006). Negation in Middle Welsh. *Studia Celtica*, *40*(1), 63-88.

Willis, D. (2007a). Specifier-to-head reanalyses in the complementizer domain: evidence from Welsh. *Transactions of the Philological Society*, *105*(3), 432-480.

Willis, D. (2007b). Syntactic lexicalization as a new type of degrammaticalization. *Linguistics*, *45*(2), 271-310.

Willis, D. (2011a). Reconstructing last week's weather: Syntactic reconstruction and Brythonic free relatives. *Journal of Linguistics*, *47*(02), 407-446.

Willis, D. (2011b). The limits of resumption in Welsh wh-dependencies. In A. Rouveret (Ed.), *Resumptive pronouns at the interfaces* (pp. 189–222). Amsterdam: John Benjamins.

Willis, D. (2014). Investigating geospatial models of the diffusion of morphosyntactic innovations: The Welsh strong second-person singular pronoun *chdi*. *Unpublished ms. University of Cambridge*.

Willis, D. (2015). Two predicate-phrase heads in Welsh copular clauses. *Paper presented at the Workshop on Copulas across Languages, Greenwich*.

Willis, D. (2016). Endogenous and exogenous theories of syntactic change. In A. Ledgeway & I. Roberts (Eds.), *Cambridge Handbook of Historical Syntax*. Cambridge: Cambridge University Press.

Wiltschko, M. (2014). *The universal structure of categories: Towards a formal typology* (Vol. 142). Cambridge University Press.

Winford, D. (2005). Contact-induced changes: Classification and processes. *Diachronica*, *22*(2), 373–427.

Wolfe, S. (2015). *Syntactic Microvariation in the Medieval Romance Languages and the Nature of V2 Reconsidered*. SyntaxLab Talk, Tuesday 9th June 2015, University of Cambridge.

Wright, S. (1993). In Search of History: English Language In the Eighteenth Century. In *English Language Corpora: Design, analysis and exploitation* (p. 25-39). Amsterdam: Rodopi.

Yang, C. D. (2000). Internal and external forces in language change. *Language variation and change*, *12*(03), 231-250.

Yang, C. D. (2002). *Knowledge and learning in natural language*. Oxford: Oxford University Press.

Yang, C. L., Perfetti, C. A., & Schmalhofer, F. (2007). Event-related potential indicators of text integration across sentence boundaries. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(1), 55-89.

Yates, F. (1934). Contingency tables involving small numbers and the χ^2 test. *Supplement to the Journal of the Royal Statistical Society*, 217-235.

Young, R. W., & Morgan, W. (1980). *The Navajo Language: A Grammar and Colloquial Dictionary*. Albuquerque: University of New Mexico Press.

Yule, G. (1981). New, current and displaced entity reference. *Lingua*, *55*(1), 41-52.

Zaring, L. (1996). "Two BE or not two BE": Identity, Predication and the Welsh Copula. *Linguistics and Philosophy*, *19*(2), 103-142.

Zavrel, J., & Daelemans, W. (1997). Memory-based learning: Using similarity for smoothing. *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, 436-443.

Zimmermann, M., & Féry, C. (2010). *Information structure: theoretical, typological, and experimental perspectives*. Oxford: Oxford University Press.

Zinsmeister, H., Hinrichs, E., Kübler, S., & Witt, A. (2008). Linguistically annotated corpora: Quality assurance, reusability and sustainability. In A. Lüdeling & M. Kytö (Eds.), *Corpus Linguistics. An International handbook. Volume 1* (p. 759-776). Berlin: Walter de Gruyter.

Zipf, G. K. (1935). *The psycho-biology of language*. Boston, Houghton-Mifflin.

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Oxford: Addison-Wesley Press.

Zubizarreta, M. L. (1998). *Prosody, focus, and word order*. Cambridge, Massachusetts: MIT Press.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological bulletin*, *123*(2), 162-185.