



Universiteit  
Leiden  
The Netherlands

## 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>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/40632>

**Note:** To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/40632> holds various files of this Leiden University dissertation.

**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

---

## Samenvatting in het Nederlands

---

Dit proefschrift gaat over de woordvolgorde in de Middelwelshe zin. De meest voorkomende woordvolgorde is de zogenaamde 'V2' volgorde, waarbij het werkwoord net als in het Nederlands op de tweede plaats in de zin staat. Deze volgorde wijkt af van de normale woordvolgorde van de zin in de moderne Welshe taal waarin het werkwoord voorop staat. Een tweede opvallende observatie in de Middelwelshe syntaxis betreft de grote variatie in woordvolgorde. Het werkwoord op de tweede plek kan diverse constituenten volgen en de congruentie tussen onderwerp en persoonsvorm neemt vaak ongebruikelijke vormen aan. Tot slot is het onduidelijk waar deze zinnen met V2 volgorde vandaan komen. De beperkte data uit eerdere stadia van de Welshe taal lijkt te wijzen op een voorkeur voor zinnen met het werkwoord voorop (VSO). Ik probeer in deze studie daarom de volgende vragen te beantwoorden:

1. Hoe kan de distributie van de verschillende patronen in woordvolgorde in Middelwelsch verklaard worden?
2. Waar komen de verscheidene V2 volgordes vandaan (inclusief de patronen met en zonder congruentie)?

Om deze vragen te kunnen beantwoorden is het om te beginnen noodzakelijk om een digitale database te creëren zodat alle teksten doorzoekbaar zijn en daarnaast voorzien van morfo-syntactische annotatie. Daarnaast is het belangrijk om een consistente methodologie te hanteren voor de analyse van de informatiestructuur en andere factoren die de volgorde van de constituenten in de zin kunnen bepalen.

In hoofdstuk 2 presenteer ik de argumenten voor een geannoteerd corpus. In historisch taalkundig onderzoek, met name onderzoek naar syntaxis, kunnen we enkel kijken naar de distributie van verschillende constructies die we vinden in manuscripten. Hoe meer manuscripten en teksten we kunnen digitaliseren en systematisch onderzoeken, hoe meer informatie we krijgen over de taalsituatie in die tijd. Als een bepaalde constructie in slechts een enkele tekst voorkomt, kun-

nen we niet direct concluderen dat deze constructie ook daadwerkelijk onderdeel was van de gangbare gesproken of geschreven taal. Omdat de hoeveelheid data die is overgeleverd vaak erg beperkt is, is het belangrijk om optimaal gebruik te maken van de data die ter beschikking is. Dit kan onder andere door de toevoeging van grammaticale informatie over de woordsoort en de daarbijbehorende vervoeging en verbuiging in zogenaamde 'Part-of-Speech' (PoS) tags. Deze gedetailleerde morfosyntactische informatie faciliteert de automatische extractie van de nodige taalkundige informatie. Idealiter voegen we alle beschikbare teksten en manuscripten toe aan een dergelijk geannoteerd corpus, maar dit kost heel veel tijd en moeite. Voor het onderzoek in dit proefschrift heb ik daarom de eerste stappen gezet om een dergelijke geannoteerde database te creëren voor het Middelwels: 15 teksten werden hiervoor zorgvuldig geselecteerd, klaargemaakt voor automatische annotatie, getagd en gecorrigeerd en uiteindelijk ook voorzien van de meest basale syntactische informatie.

Ik heb een PoS-tagger getraind die automatisch morfosyntactische labels toekent aan woorden in Middelwelshe teksten. Met een Global Accuracy van meer dan 90% presteerde de tagger redelijk goed gezien de complexe data en de zeer gedetailleerde tagset (bestaande uit meer dan 200 tags). De tijd die nodig was voor de handmatige correctie na de automatische toekenning van de tags werd hierdoor aanzienlijk beperkt. Naast een PoS-tagger heb ik ook een basale syntactische grammatica ontworpen voor het Middle Welsh met behulp van de NLTK parser. Na de handmatige correctie werden de bestanden geconverteerd naar diverse vormen die systematisch doorzocht kunnen worden met CorpusSearch of XQuery. Het belangrijkste resultaat in hoofdstuk 2 is een geannoteerde database van 15 teksten waarmee diverse taalkundige bijzonderheden systematisch onderzocht kunnen worden.

In hoofdstuk 3 werk ik een methodologie uit voor onderzoek naar informatiestructuur in historische corpora. Hierbij stonden drie onderwerpen centraal: Referentialiteit (Oude of Nieuwe Informatie), Topic (vs. Comment) en Focus (vs. Achtergrondinformatie). Ik heb de kenmerken van elk van deze onderwerpen systematisch omschreven, zodat ze gebruikt kunnen worden om historische corpora te annoteren. Er zijn daarnaast nog twee andere factoren die een rol spelen in de informatiestructuur van de zin: zogenaamde 'Points of Departure', het uitgangspunt van de zin, en de Information Flow, de manier waarop oude en nieuwe informatie elkaar volgen. De duidelijk uitgewerkte en omschreven definities en algoritmes faciliteren de annotatie van grote corpora. Een dergelijke consequente analyse is onmisbaar in elk historische syntactisch onderzoek.

In hoofdstuk 4 presenteer ik de data en de belangrijkste observaties betreffende de variatie in woordvolgorde in Middelwels. In de geannoteerde database heb ik diverse patronen gevonden in positieve hoofdzinnen. Ik heb deze onderverdeeld in negen hoofdtypen op basis van hun formele structuur. De meest voorkomende woordvolgorde aan het begin van de Vroegmiddelwelshe periode is nog steeds de

'Abnormal order' met het werkwoord op de tweede plaats van de zin, maar het overgrote deel van de zinnen in de bijbelvertaling van 1588 begint nu met het onderwerp. Zinnen met het werkwoord op de eerste plaats en met name zinnen met een hulpwerkwoord nemen in frequentie toe. De syntactische structuur van de bijbelvertaling wijkt dus af van zowel Middel- als Modernwelsh.

In Hoofdstuk 5 onderzoek ik systematisch all factoren die mogelijk invloed kunnen hebben op de woordvolgorde van de Middelwelshe zin. Beginnende met de grammaticale factoren zien we dat de 'Abnormal order' met een perifrastische constructie en een verbaalnomen vrijwel alleen maar voorkomt in de verleden tijd. Dit is waarschijnlijk gerelateerd aan het feit dit type woordvolgorde met name voorkomt in narratieve context. In direct taalgebruik, zoals in dialogen, staat meestal het onderwerp vooraan. Een grondige studie van het corpus laat verder zien dat werkwoorden met onpersoonlijke/passieve vervoeging vooral voorkomen in V2-zinnen met initiële adjuncten. Tot slot is er een beperkte rol voor 'Animacy' van het onderwerp en lijdend voorwerp. Als het lijdend voorwerp 'inanimate' is, staat het vaker in het begin van de zin dan verwacht.

Alleen als alle interne en externe factoren systematisch zijn onderzocht, kunnen we bepalen of andere factoren, zoals informatiestructuur werkelijk een rol spelen. De eerste factor betreffende de informatiestructuur die ik heb bekeken is 'referentialiteit' - de informatiestatus van het onderwerp en het lijdend voorwerp. Uit mijn onderzoek blijkt dat het lijdend voorwerp vrijwel alleen vooraan kan staan (OVS volgorde) als het Nieuwe Informatie bevat. Dit betekent dat de natuurlijke informatiestroom van de zin (normaal van Oude naar Nieuwe informatie) omgedraaid is. Deze lijdende voorwerpen worden dus gemarkeerd door de natuurlijke informatiestroom om te draaien. De enige uitzonderingen op deze generalisatie zijn de zogenaamde Familiar Topics. Dit zijn met name aanwijzende voornaamwoorden die de eerste plaats van de zin innemen. Ze verwijzen naar het laatstgenoemde concept of persoon in de voorafgaande context.

De corpusstudie leverde nog twee andere resultaten op die verband houden met de informatiestructuur, met name wat betreft de samenhang van de tekst ('tekstcohesie'). Framesetters of 'points of departure' komen om te beginnen meestal voor in V2-zinnen met adjuncten op de eerste plaats waar ze als topic fungeren. Een tweede observatie in deze categorie betreft de continuïteit. Om een sterke link tussen twee zinnen te bewerkstelligen, konden verbale nomina op de eerste plek van de zin geplaatst worden. Deze kunnen ofwel afhankelijk zijn van een vervoegd hulpwerkwoord in de voorafgaande zin of ze worden ondersteund door een vervoegde vorm van het hulpwerkwoord *gwneuthur* 'doen'. Dit maakt opnieuw deel uit van het narratieve karakter van de teksten in dit genre. Focus kan ten slotte gevonden worden in de speciale focusconstructie: de (gereduceerde) cleftzinnen ('Mixed Sentence'). Focus van een 'identity' predicaat komt meestal tot uiting in de specifieke *sef*-constructie, maar niet alle zinnen met *sef* bevatten focus, zeker niet aan het eind van de Middelwelshe periode.

In hoofdstuk 6 en 7 ligt de nadruk op de synchrone en diachrone syntactische analyse van de verschillende types woordvolgorde. Hoofdstuk 6 presenteert vier casussen die allemaal met informatiestructuur te maken hebben. Ik heb onderzocht hoe topics, focus en de referentiële status van constituenten samenhangen met de syntactische structuur. Het Middelwelsh kende maar een topicpositie, maar zinnen met V3- en V4-structuren komen ook voor. Ik heb twee verschillende analyses gepresenteerd voor V2-zinnen: een gebaseerd op verplaatsing van een constituent en een waar de constituent pas later aan de zin wordt toegevoegd. De constructie waarbij de constituent zelf verplaatst naar de eerste plek van de zin ontstond pas later in het Middelwelsh.

In het laatste hoofdstuk beschrijf ik verschillende manieren om historische syntaxis te analyseren: socio-linguïstiek, constructiegrammatica en generatief syntactische benaderingen. Ik laat zien dat een generatief framework diverse voordelen heeft in de studie van Middelwelshe historische syntaxis, omdat het gebruik kan maken van inzichten van studies naar synchrone variatie in andere talen. Deze technieken die grondig getest zijn in het Minimalisme (Chomsky's 'Minimalist Program') helpen ons de exacte condities en context te definiëren waarin syntactische vernieuwingen wel en niet kunnen plaatsvinden en hoe ze naar bepaalde innovaties kunnen leiden.

Ik presenteer opnieuw verschillende casussen, dit keer op het gebied van syntactische innovaties in de geschiedenis van het Welsh. De eerste gaat over een speciale focusconstructie: identificatiefocus van het predikaat. Ik laat zien hoe deze constructie is voortgekomen uit de cleftconstructie in het Oudwelsh en hoe de focusmarker *sef* is ontstaan. Toen de focus verdween werd *sef* opnieuw geïnterpreteerd als een expletief element en, uiteindelijk als een linker in reformulatieve appositieconstructies ("i.e."). In de tweede casus ga ik in op een van de belangrijkste onderzoeksvragen van dit proefschrift: het ontstaan van de V2-constructies. Door zorgvuldige vergelijking met andere Keltische talen en de reconstructie van de functionele partikels in het C-domein laat ik zien hoe de V2-zinnen in het Welsh zijn ontstaan. Na deze reconstructie volgde de herinterpretatie van zogenaamde 'hanging topics' en de uitbreiding van de functionele informatiestructuur van de constituenten op de eerste plaats van de zin. De fonologische erosie van de partikels in Vroegmodernwelsh leidde uiteindelijk tot het verdwijnen van de V2-constructie. Ten slotte vergelijk ik de resultaten van deze syntactische studie in het Middelwelsh met andere Middeleeuwse taalgroepen met het V2-fenomeen zoals het Romaans en Germaans.

In dit proefschrift probeer ik in het algemeen de interactie tussen syntaxis en informatiestructuur te analyseren en de invloed die beide uitoefenen op de woordvolgorde. De distributie van verschillende types woordvolgorde in het Middelwelsh het resultaat van zowel grammaticale factoren als informatiestructuur. Focus werd uitgedrukt door een gereduceerde cleftconstructie, de zogenaamde 'Mixed Order'. In identificerende copulazinnen daarentegen werd focus uitgedrukt door de focusmarker *sef* (< *ys + ef* 'dit is het'). Referentiële status en textcohesie speelden ook

een rol. Op basis van de huidige corpusstudie kunnen we een basaal algoritme ontwerpen om de juiste woordvolgorde te bepalen in transitieve zinnen in het Middelnederlands.

Vanuit historisch oogpunt heb ik laten zien dat informatiestructuur een rol kan spelen in syntactische innovaties, maar niet noodzakelijk de directe aanleiding vormt. De uitbreiding van informatiestructurele functies van constituenten op de eerste plaats van de zin is een goed voorbeeld hiervan. De uiteindelijke trigger voor syntactische veranderingen blijven soms moeilijk te achterhalen, maar een gedetailleerde en consequente beschrijving van de synchrone variatie waarbij alle variabelen systematisch worden gecontroleerd is onmisbaar in historisch syntactisch onderzoek.

