

Taalvariatie in het semantische domein NOT.UNTIL

De uitdrukking van NOT.UNTIL in het Frans, Duits, Nederlands, Engels en Zweeds

Lara Mutsaers (5901855)



Universiteit Utrecht

Bachelor Eindwerkstuk Kunstmatige Intelligentie (7.5 ECTS)

Utrecht University

Begeleider: Henriëtte de Swart

Tweede Lezer: Ruud Hortensius

April 2021

Inhoudsopgave

1. Introductie.....	3
2. Relevantie voor KI	6
3. Theoretische achtergrond.....	7
3.1. Drie formele semantische analyses ‘not...until’	7
3.1.1. De scope-analyse	8
3.1.2. De ambiguïteitshypothese.....	8
3.1.3. Benadering van lexicale compositie	9
3.1.4. Equivalentie van formele analyses.....	9
3.1.5. NOT.UNTIL in verschillende talen.....	9
3.2. Taalvariatie in connectieven in het temporele domein: <i>LATE-Clauses</i>	10
3.3. Motivatie onderzoek	14
4. Hypothesen	15
5. Opzet van het onderzoek bij <i>Time in Translation</i>	16
6. Methodologie	19
6.1. Het tekstcorpus Europarl.....	19
6.2. De dataset.....	19
6.3. Het annotatieproces	20
6.4. Data Analyseren.....	23
7. Resultaten	25
7.1. Tupelfrequentie	25
7.2. Voorkomens per taal	27
7.2.1. Het Zweeds.....	27
7.2.2. Het Engels.....	28
7.2.3. Het Nederlands	29
7.2.4. Het Duits	30
7.2.5. Het Frans	31
8. Discussie I: Toetsen van hypothesen	32
8.1. H0.1: Zweeds – ‘inte...förrän’	33
8.2. H0.2: Engels – ‘not...until’	33
8.3. H0.3: Nederlands – ‘pas...wanneer’ en ‘niet...zolang...niet’	33
8.4. H0.4: Duits – ‘erst...wenn’	34
8.5. H0.5: Frans – ‘ne...pas...tant que...ne...pas’	35
9. Discussie II: Algemeenheden in de data.....	35

9.1. Taalvoorkeur	35
9.2. Een overlap met het semantische domein van de conditionele connectieven	38
10. Conclusie.....	39
10.1. Beantwoorden hoofdvraag	39
10.2. Relevantie voor KI en vervolgonderzoek.....	40
11. Referenties	42
Bijlage 1: een overzicht van afkortingen gebruikt bij annotatie	43

ABSTRACT

Dit onderzoek gaat over de manier waarop de talen Frans, Duits, Engels, Zweeds en Nederlands het semantische domein NOT.UNTIL uitdrukken. Uit dit corpusonderzoek, dat wordt uitgevoerd aan de hand van het parallelle multilinguïstische tekstcorpus Europarl, zal blijken dat er in elke taal veel verschillende constructies worden gebruikt om dit domein uit te drukken. Bovendien worden er in elke taal andersoortige constructies gebruikt. Er is dus veel variatie aan te wijzen tussen en binnen talen. Vertalingen tussen deze talen zijn daarom niet stabiel. Wel lijkt elke taal een voorkeur te hebben voor het gebruik van een bepaalde constructie om NOT.UNTIL uit te drukken. De compositionele semantiek van een zin speelt daarin een rol. De polariteit van de hoofd- en bijzin is bepalend voor de betekenis. Er zal immers blijken dat elke voorkeursconstructie die NOT.UNTIL uitdrukt, steeds in dezelfde vorm voorkomt, met steeds eenzelfde polariteit.

1. Introductie

Temporele connectieven zijn voegwoorden die een temporele bijzin of een bijwoordelijke bepaling introduceren. De bijzin geeft een tijdsspecificatie mee aan de gebeurtenis die wordt beschreven in de hoofdzin (Wälchli, 2018). In dit onderzoek ligt de focus op temporele connectieven die een volledige bijzin introduceren, zoals de voorbeelden in (1). In dergelijke zinnen wordt zowel in de hoofd- als bijzin een gebeurtenis beschreven. In (1) wordt steeds in de hoofdzin de gebeurtenis ‘het naar de winkel gaan’ beschreven en in de bijzin steeds de gebeurtenis ‘het thuiskomen van mijn moeder’. Een temporeel connectief heeft het vermogen om de gebeurtenissen in de hoofd- en bijzin op een temporele manier aan elkaar te verbinden. Een temporeel connectief kan enerzijds een ordening van gebeurtenissen in de tijd aangeven. Dit kan bijvoorbeeld met een connectief zoals ‘voordat’, welke gebruikt wordt in voorbeeld (1a) (hier onderstreept). Anderzijds kan een temporeel connectief een overlap van tijd beschrijven. Dit kan bijvoorbeeld met een connectief zoals ‘terwijl’ zoals in voorbeeld (1b) (hier onderstreept).

- (1) a. Ik ga naar de winkel voordat mijn moeder thuiskomt.
b. Ik ga naar de winkel terwijl mijn moeder thuiskomt.

Bij het gebruik van sommige temporele connectieven is de temporele relatie wat complexer en worden er bovendien restricties opgelegd aan het aspectuele karakter van de hoofd- of bijzin. Dit is het geval bij het Engelse temporele connectief ‘until’. ‘Until’ combineert namelijk alleen met duratieve zinnen die een activiteit of gebeurtenis beschrijven. Beschouw bijvoorbeeld een zin als in (2a). Hier combineert het connectief ‘until’ met de duratieve activiteit ‘schrijven’. Zin (2b), waar ‘until’ wordt gebruikt in een niet-duratieve context, is grammaticaal incorrect (De Swart, 1996).

- (2) a. Susan wrote until midnight.
b. *Susan wrote a letter until midnight.

Wat opvallend is aan zinnen met 'until', is dat 'until' in een negatief polaire context wel te combineren valt met niet-duratieve activiteiten. Vergelijk zin (2b) en (3) (De Swart, 1996).

(3) Susan did not write a letter until midnight.

In de zinnen (2b) en (3) wordt precies dezelfde activiteit beschreven, maar de negatief polaire context van het voorbeeld in (3) zorgt ervoor dat de zin, in tegenstelling tot de zin in voorbeeld (2b), grammaticaal correct is. Dit is precies wat negatief polaire 'until'-constructies zo interessant maakt.

In de literatuur wordt er gediscussieerd over deze 'not...until'-constructie. De Swart (1996) werkt drie verschillende formele analyses van dergelijke constructies uit, die in de literatuur worden benoemd, om deze constructies beter te kunnen begrijpen. Dit zijn de *scope*-analyse, de ambigüiteitshypothese en de benadering van lexicale compositie. De Swart (1996) vraagt zich af welk van deze analyses het meest toereikend is en komt tot de conclusie dat deze analyses semantisch en pragmatisch equivalent aan elkaar zijn. Volgens De Swart zouden de verschillende besproken analyses wellicht kunnen voortkomen uit taalvariatie tussen talen onderling. Verschillende talen zouden verschillend omgaan in de uitdrukking van het semantische domein NOT.UNTIL.¹ Het onderzoek van De Swart (1996) wordt verder uitgewerkt in hoofdstuk 3.1. Deze variatie in de manier waarop NOT.UNTIL wordt uitgedrukt in verschillende talen, die De Swart (1996) al kort noemt, vormt het uitgangspunt van dit onderzoek. Het gaat hier dus om de vraag hoe in verschillende talen een zin zoals in (3) wordt uitgedrukt.

Met deze vraag houdt onder andere Bernhard Wälchli (2018) zich bezig. Hij doet onderzoek naar de distributie van temporele connectieven in het semantische domein in verschillende talen en vraagt zich af welke onderlinge taalvariatie er bestaat tussen die talen. Wälchli (2018) heeft met een corpus van 72 verschillende talen onderzocht hoe de meest voorkomende temporele connectieven, welke corresponderen met het Engelse 'until', 'before' en 'as long as', zijn verdeeld over de semantische ruimte. De vraag die hier wordt gesteld is of deze connectieven onder de connectieven vallen die een ordening van gebeurtenissen in de tijd beschrijven, zoals in zin (1a), een overlap beschrijven van gebeurtenissen in de tijd, zoals in zin (1b), of een heel andere relatie. Uit Wälchli's onderzoek bleek dat er in het semantische domein van de onderzochte connectieven, een globale clustering aan te wijzen viel van drie clusters die respectievelijk overeenkomen met de Engelse temporele connectieven 'until', 'before' en 'as long as'. Maar, met het toepassen van een meer granulaire analyse, bleek dat er nog een vierde cluster aan te wijzen viel, welke tussen de clusters lag die respectievelijk beschreven worden met de Engelse connectieven 'before' en 'until'. Dit cluster heeft een duidelijke categorisatie: een

¹ Een constructie geschreven tussen enkele aanhalingstekens, zoals 'until', refereert naar de vorm die wordt gebruikt om een semantische relatie uit te drukken. Een constructie geschreven in hoofdletters refereert naar de relatie in het semantische domein, dus naar de betekenis.

negatie in de hoofdzin. Dit is dan ook precies het cluster waar de constructie 'not...until' veelvuldig in voorkwam. In veel talen is deze zone niet gelexicaliseerd, waardoor er taalvariatie kan ontstaan. Wälchli stelt dat in talen waar de zone niet gelexicaliseerd is, het verschilt of deze zone op dezelfde manier wordt gecodeerd als UNTIL, of juist als BEFORE. Het gehele onderzoek van Bernhard Wälchli wordt uitgewerkt in Hoofdstuk 3.2.

Er lijkt dus inderdaad variatie te zitten in de manier waarop talen NOT.UNTIL uitdrukken. Het is daarom interessant om het corpusonderzoek van Wälchli (2018) verder uit te bouwen en in te zoomen op het vierde cluster waarin de constructie 'not...until' werd geclassificeerd in dat onderzoek. Mijn onderzoek zal een reproductie zijn van het onderzoek van Wälchli (2018), maar dan met een moderner tekstcorpus en een focus op Europese talen. Voor dit onderzoek worden de talen Duits, Nederlands, Engels, Frans en Zweeds beschouwd. Bovendien zal er in dit onderzoek de gehele constructie van de zin worden beschouwd en niet alleen het connectief zelf, zoals bij Wälchli (2018) het geval was. Er zit namelijk een duidelijk verschil in de classificatie van 'not...until' en 'until' in de semantische ruimte.

Vanuit de onderzoeken van De Swart (1996) en Wälchli (2018) krijgen we een verwachting over hoe NOT.UNTIL wordt uitgedrukt in het Duits, Nederlands, Zweeds en Engels. In eerder onderzoek van het onderzoeksteam van *Time in Translation*, waar dit onderzoek deel van uitmaakt, zijn er aanwijzingen gevonden voor hoe dit in het Frans wordt uitgedrukt. Het eerdere onderzoek van het onderzoeksteam van *Time in Translation* is kort uiteengezet in hoofdstuk 4.

Deze bovengenoemde assumpties kunnen vertaald worden naar hypothesen die in dit onderzoek aan de hand van een tekstcorpus worden getoetst. Deze hypothesen beschrijven allemaal een verwachting over hoe NOT.UNTIL wordt uitgedrukt in een van de bovengenoemde talen. Deze hypothesen zijn uitgewerkt in hoofdstuk 4. In dit onderzoek staat de vraag centraal in hoeverre deze hypothesen, geformuleerd aan de hand van de onderzoeken van De Swart (1996), Wälchli (2018) en *Time in Translation*, kloppen, om zo een beter beeld te krijgen van de semantische ruimte van NOT.UNTIL in de talen Frans, Duits, Engels, Zweeds en Nederlands. Het doel van dit onderzoek is dus om te onderzoeken of er inderdaad sprake is van taalvariatie zoals De Swart (1996) doet vermoeden en Wälchli (2018) al bewijst op het niveau van puur het connectief.

Bovenstaande onderzoeksvraag wordt, zoals eerder kort benoemd, onderzocht aan de hand van een parallel multilinguïstisch tekstcorpus. Dit corpus bestaat uit tekstdata waarbij elk datapunt parallel vertalingen omvat in verschillende vertalingen van dezelfde zin. Het belang van het gebruik van een dergelijk corpus is dat de betekenis tussen de vertalingen van eenzelfde zin altijd stabiel is. We kunnen aan de hand van een analyse van een parallel multilinguïstisch tekstcorpus onderzoeken hoe in verschillende talen NOT.UNTIL wordt uitgedrukt. Het corpus dat in dit onderzoek gebruikt wordt is het corpus Europarl. Vanuit dit corpus wordt er een dataset gegenereerd, waar elk datapunt bestond uit een tupel van de vijf verschillende talen die worden beschouwd in dit onderzoek. Door annotatie van deze

datapunten, met verschillende talen als brontaal, wordt de data inzichtelijk gemaakt. Het is belangrijk om vanuit verschillende talen te annoteren, om zo variatie in talen zichtbaar te maken. Op het overzicht van de annotaties wordt er een descriptieve analyse toegepast. Hierdoor kunnen we een uitspraak doen over welke constructies er in welke taal voorkwamen. De precieze onderzoeksmethode is verder uitgewerkt in hoofdstuk 6.

Uit dit onderzoek zal blijken dat er meer variatie in de semantische ruimte NOT.UNTIL bestaat dan er in eerste instantie in de literatuur gesuggereerd werd. De Swart (1996) en Wälchli (2018) gaan beide uit van het principe dat talen stabiel zijn in hun vertalingen. Het onderzoek van Wälchli bevestigde dit idee in eerste instantie, maar reproductie van zijn onderzoek op een ander tekstcorpus wijst iets anders uit. De resultaten van dit onderzoek zijn verder uitgewerkt in hoofdstuk 7 en worden besproken aan de hand van de geformuleerde hypothesen in hoofdstuk 8. In hoofdstuk 9 wordt er kort besproken wat voor regelmaat of tendensen we in de dataset kunnen aanwijzen.

Dit onderzoek is geschreven binnen het onderzoeksveld van kunstmatige intelligentie en zal uiteindelijk bijdragen aan het verbeteren van *Natural Language Processing*, een van de belangrijkste onderzoeksgebieden binnen dit veld. Praktisch zal dit onderzoek bijdragen aan het automatisch vertalen. In hoofdstuk 2 wordt er verder ingegaan op de relevantie van dit onderzoek voor dit vakgebied.

2. Relevantie voor KI

Dit onderzoek is uitgevoerd binnen het onderzoeksveld van kunstmatige intelligentie. Om deze reden wordt er in dit hoofdstuk aandacht besteed aan de relevantie van dit onderzoek binnen dit onderzoeksveld en wat dit onderzoek bijdraagt aan de kennis binnen dit vakgebied.

Het veld van kunstmatige intelligentie houdt zich bezig met het begrijpen van intelligentie met als doel om zelf entiteiten te bouwen die als intelligent kunnen worden beschouwd. Binnen het onderzoeksveld bestaan er verschillende ideeën over waaraan een kunstmatig intelligente entiteit moet voldoen om als zodanig beschouwd te worden. Eén daarvan is dat een entiteit moet handelen zoals mensen dat doen. Om dit criterium te toetsen is er in 1950 een Turing test ontwikkeld door Alan Turing (Russell en Norvig, 2010). Deze test wijst uit of een gemodelleerde entiteit (een computer of machine) daadwerkelijk over (kunstmatige) intelligentie beschikt. Tijdens de test is er sprake van een interactie tussen mens en machine, waarbij de mens simultaan interacteert met zowel een mens als een machine. Als deze persoon geen verschil kan benoemen tussen mens en machine, dan wordt de machine beschouwd als intelligent. Als een machine als intelligent wordt beschouwd, dan zou dat betekenen dat de machine over een aantal kwaliteiten of eigenschappen beschikt, waaronder het processen van natuurlijke taal (*natural language processing*, ofwel NLP) (Russell en Norvig, 2010).

Dit onderzoek draagt bij aan NLP. Dit is namelijk nog steeds een struikelblok voor computers of machines. Waar dit struikelblok sterk tot uitdrukking komt is in het

veld van het automatisch vertalen. Er zijn steeds meer programma's die automatisch vertalingen genereren van verschillende talen. Ondanks dat dit steeds beter wordt, voelen deze vertalingen nog niet altijd natuurlijk aan. Eén van de factoren waar dat aan zou kunnen liggen is dat er nog niet duidelijk is hoe sommige constructies die temporele relaties uitdrukken precies worden vertaald in verschillende talen. Er zijn wel verschillende ideeën over in de literatuur, maar er is nog te weinig empirisch onderzoek gedaan om specifieke algemeenheden aan te wijzen. Dit onderzoek bouwt voort op de ideeën in de literatuur en geeft door middel van een corpusonderzoek inzicht in het semantische domein van de temporele connectieven en specifiek de connectieven die het NOT.UNTIL domein uitdrukken. Hierdoor zullen we dus een beter idee krijgen over welke vertalingen gebruikt moeten worden voor een bepaalde constructie en zullen de vertalingen natuurlijker aanvoelen. In de resultaten van dit onderzoek kunnen er bepaalde tendensen geformuleerd worden, welke als input gegeven kunnen worden aan automatische vertalers.

3. Theoretische achtergrond

In dit hoofdstuk wordt de literatuur uiteengezet met betrekking tot de semantiek van de constructie 'not...until'. In het artikel van De Swart (1996) worden de formele analyses van 'not...until' onderzocht en in het artikel van Wälchli (2018) wordt de semantische ruimte van *LATE-clauses* geconstrueerd.

3.1. Drie formele semantische analyses 'not...until'

De Swart (1996) zet in haar artikel de discussie in de literatuur uiteen over het temporeel connectief 'until' in een negatief polaire context. 'Until' kan in een negatief polaire context wel gebruikt worden in combinatie met niet-duratieve activiteiten of gebeurtenissen, in tegenstelling tot het gebruik van 'until' in een positief polaire context. Dit verklaart het verschil tussen (2b) en (3). Dit wordt volgens De Swart (1996) als argument gebruikt dat negatie hier een aspectuele operator zou zijn. Een andere kijk op deze kwestie is de gedachte dat er twee verschillende varianten van UNTIL zouden bestaan: de duratieve en de punctuele. Laatstgenoemde zou zich alleen manifesteren als negatief polaire uitdrukking (NPI). Een NPI is een uitdrukking die uitsluitend voorkomt in negatieve contexten. Een voorbeeld van een NPI is het Nederlandse werkwoord 'hoeven'. Vergelijk bijvoorbeeld de zinnen in (4). Hier kan 'hoeven' alleen worden gebruikt in een negatief polaire context.

- (4) a. Jij hoeft geen huiswerk te maken.
b. *Jij hoeft huiswerk te maken.

Dit zou ook toepasbaar kunnen zijn op 'until', wat het verschil tussen (2b) en (3) zou kunnen verklaren.

In de literatuur wordt er ook gediscussieerd over de ambiguïteit die een constructie met 'not...until' teweegbrengt. Beschouw bijvoorbeeld een zin zoals in (5) welke op twee manieren gelezen kan worden. In de ene lezing heeft de prinses niet

de gehele tijd geslapen tot 9 uur (ze werd eerder wakker). In de andere lezing was er geen enkele situatie waarin de prinses sliep tot 9 uur (ze was al die tijd wakker).

(5) The princess did not sleep until nine o' clock.

De Swart (1996) beschrijft verschillende benaderingen die in de literatuur worden gebruikt om deze ambiguïteit te verklaren. Elke benadering doet dat op een andere manier en geeft zo een eigen manier van een formele analyse van 'not...until'. Hieronder volgt een overzicht van deze analyses zoals De Swart (1996) deze in haar artikel beschrijft.

3.1.1. De scope-analyse

De eerste formele semantische analyse die De Swart (1996) noemt is de *scope*-analyse. Smith (1974) en Mittwoch (1977) zijn hier de voornaamste aanhangers van. Volgens deze benadering is de ambiguïteit van een zin als in (5) te wijten aan een verschil in *scope* die de negatie in de zin kan aannemen. Deze zou zowel een brede als een smalle *scope* kunnen aannemen. De ambiguïteit van zin (5) kan dan worden weergegeven zoals in (6):

(6) a. \neg (until nine o'clock) (the princess slept))

Het is niet het geval dat de prinses de gehele tijd heeft geslapen tot 9 uur, ze werd eerder wakker.

b. (until nine o'clock (\neg (the princess slept)))

Tot minstens 9 uur was de prinses niet aan het slapen, ze was al die tijd wakker.

3.1.2. De ambiguïteitshypothese

Waar de *scope*-analyse uitgaat van één 'until', gaat de tweede analyse die De Swart (1996) noemt, de ambiguïteitshypothese van Karttunen (1974), uit van het bestaan van twee verschillende UNTILS. In deze analyse wordt uitgegaan van het idee dat negatie altijd een wijde *scope* aanneemt over zinnen met 'until'-constructies. (5a) is dus hier volgens Karttunen (1974) de juiste interpretatie. De ambiguïteit van deze zin valt volgens hem te wijten aan de ambiguïteit van het werkwoord 'sleep', wat zowel BE.ASLEEP als FALL.ASLEEP kan betekenen. Bovendien stelt Karttunen (1974) dat het negatief polaire gebruik van 'until' niet duratief is maar punctueel. 'Until' wordt gebruikt om gebeurtenissen te duiden in de tijd. Dit wordt ondersteund door het gegeven dat 'until' combineert met gebeurtenispredikaten in negatieve zinnen zoals in (3), maar niet in affirmatieve zinnen zoals in (2b). Karttunen stelt dat 'until' logisch equivalent is aan 'before'. Ook stelt Karttunen dat de focus van de zin ligt op het feit dat een gebeurtenis alleen plaatsvindt na een bepaalde tijdsperiode. Dit impliceert dat de gebeurtenis echt plaatsvindt, maar later dan verwacht.

3.1.3. Benadering van lexicale compositie

Een derde lijn van analyse behandelt negatie als iets dat altijd samengesteld is met het temporeel connectief. Decklerck (1995) behandelt 'not...until' in het Engels als één geheel wat ONLY betekent. Een argument hiervoor is dat de constructie 'not...until' in andere talen gelexicaliseerd is als één woord. In het Duits wordt volgens Decklerck immers 'erst' gebruikt en in het Nederlands 'pas'. Volgens Decklercks interpretatie vindt de beschreven gebeurtenis daadwerkelijk plaats, maar later dan verwacht omdat de alternatieven zijn geordend op een schaal.

3.1.4. Equivalentie van formele analyses

Deze drie verschillende formele semantische analyses van 'not...until' hebben volgens De Swart (1996) elk hun aantrekkelijke aspecten. De vraag is nu welke benadering het beste de semantiek en de pragmatiek van de constructie beschrijft. De Swart (1996) werkt deze analyses uit in een *event-based semantics* en laat zien dat de drie analyses semantisch en pragmatisch equivalent aan elkaar zijn. De analyses behandelen alle drie de constructie 'not...until' als een expressie die een exclusie beschrijft op een temporele schaal. Deze equivalentie betekent dat Engelse zinnen met de constructie 'not...until' semantisch en pragmatisch equivalent zijn aan zinnen in andere talen welke geen negatief polair gebruik kennen van 'until'. De formele analyse van deze equivalentie wordt hier niet uitgewerkt, omdat dit buiten de scope van dit onderzoek ligt. Slechts de notie dat deze drie formele analyses pragmatisch en semantisch equivalent aan elkaar zijn, is hier belangrijk.

3.1.5. NOT.UNTIL in verschillende talen

De equivalentie van de drie verschillende formele analyses van 'not...until' is volgens De Swart (1996) wellicht een indicatie dat er in verschillende talen op een andere manier uitdrukking gegeven wordt aan NOT.UNTIL. De Swart (1996) gebruikt hier een argument van Paul Kiparsky (p.c.) over het Fins als voorbeeld. Kiparsky (p.c.) stelt dat in Finse zinnen waar 'ennen kuin' ('before') inwisselbaar is met 'kunnen' ('until'), dit in het Engels vertaald kan worden met 'until'. Het Fins zou dus een taal zijn waar 'until' en 'before' in sommige contexten inwisselbaar zijn voor elkaar. Dit is volgens De Swart (1996) niet de enige taal waarin dit het geval is. Ook in het Engels zijn in sommige contexten deze twee temporele connectieven inwisselbaar voor elkaar. Beschouw de zinnen in (7):

- (7) a. But it's going to be a long time before it is over its problems.
- b. Thus it will take a long time until the pressure will show results.

De semantische analyse van Karttunen (1974) van de constructie 'not...until' zou hier wellicht uit kunnen voortkomen. Hij beschouwde 'until' logisch equivalent aan 'before'. In sommige talen zou dit wellicht het geval kunnen zijn en laat 'not...until' zich vertalen met een temporeel connectief uit het BEFORE-domein.

Wat talen volgens De Swart (1996) wel altijd gemeen hebben in de uitdrukking van NOT.UNTIL, is het feit dat constructies die NOT.UNTIL uitdrukken altijd een

exclusie doen op een schaal, namelijk een temporele. Dit noem je een scalaire constructie. De Swart (1996) illustreert dat aan de hand van een voorbeeld in het Duits (8):

- (8) a. Ich fahre nur am Donnerstag nach München.
I only go to Munich on Thursday.
b. Ich fahre erst am Donnerstag nach München.
I do not go to Munich until Thursday.

'Nur' in (7a) doet een exclusie van alle optionele dagen voor of na de genoemde dag. 'Erst' in (7b) doet een exclusie op een schaal en sluit alle dagen uit die vòòr de genoemde dag op de schaal vallen. In het Nederlands wordt om een dergelijke relatie uit te drukken volgens de Swart 'pas...wanneer'. 'Pas' is namelijk ook scalair. De Swart (1996) stelt dus dat de Duitse constructie die wordt gebruikt om NOT.UNTIL uit te drukken 'erst...wenn' is. In het Nederlands is dat 'pas...wanneer' en in het Engels is dat 'not...until'. Er wordt hier uitgegaan van het principe dat de vertalingen tussen deze talen stabiel zijn. Als we zouden kijken naar de manier waarop verschillende talen NOT.UNTIL uitdrukken, zouden we volgens De Swart voornamelijk afgebakende clusters zien.

3.2. Taalvariatie in connectieven in het temporele domein: *LATE-Clauses*

Iemand die verder bouwt op de notie van de Swart (1996) dat er tussen talen variatie zou bestaan in het uitdrukken van NOT.UNTIL, is Bernhard Wälchli (2018). Wälchli heeft onderzoek gedaan naar hoe temporele connectieven in *LATE-clauses* zijn verdeeld over de semantische ruimte. Hieronder worden in het Engels de connectieven 'before', 'until' en 'as long as' verstaan. Hij onderzoekt de relaties in het LATE-domein, dus ook de connectieven in andere talen die dergelijke relaties uitdrukken. Onder een relatie verstaat Wälchli de betekenis die het connectief uitdrukt. Er zouden volgens hem drie overkoepelende relaties in het LATE-domein bestaan: BEFORE, UNTIL, en AS.LONG.AS, maar hier bestaat in de literatuur discussie over. Het zou namelijk genuanceerder kunnen liggen. Wälchli (2018) noemt als voorbeeld een van de analyses die De Swart (1996) ook al in haar paper noemt, de argumentatie van Karttunen (1974), waarin wordt gesteld dat er twee verschillende soorten UNTIL zouden bestaan.

Wälchli (2018) construeert voor zijn onderzoek de opbouw van de semantische ruimte van het LATE-domein. Hij maakt gebruik van een hiervoor een cross-linguïstische benadering en gebruikt hiervoor een parallel multilinguïstisch tekstcorpus van 72 talen en dialecten. Een cross-linguïstische aanpak is essentieel om variatie in taal zichtbaar te maken. Verschillende vormen kunnen immers dezelfde betekenis uitdrukken.

Voordat Wälchli (2018) aan zijn corpusonderzoek begint formuleert hij een aantal verwachtingen over de daadwerkelijke opbouw van de semantische ruimte van de LATE-connectieven. Hij stelt dat de semantiek van LATE-connectieven op

een globaal niveau in te delen is in een schaal met als drie voornaamste clusters AS.LONG.AS, UNTIL en BEFORE zoals Kortmann (1997) formuleerde. Maar, zo stelt Wälchli (2018), als we de verschillen en nuances tussen talen willen aantonen, is het voordeliger om een groter aantal clusters te kiezen dan drie. Op een meer granulair niveau zal namelijk blijken dat de semantische ruimte veel complexer geordend is en dat we tussengelegen clusters zullen zien. Zo verwacht Wälchli dat er een cluster te zien zal zijn welke tussen BEFORE en UNTIL zal liggen en een cluster met constructies die een imperatief in de hoofdzin uitdrukken, welke tussen UNTIL en AS.LONG.AS zal liggen.

Om deze verwachtingen te toetsen genereert Wälchli (2018) semantische kaarten per taal door middel van *multidimensional scaling* (MDS). Op deze manier maakt hij de semantische ruimte van de 72 betrokken talen zichtbaar. MDS is erg nuttig als visualisatietool omdat het semantisch overeenkomstige datapunten dicht bij elkaar rangschikt. Hoe dichterbij twee punten naast elkaar liggen, hoe groter de kans dat in de bijbehorende passages hetzelfde connectief gebruikt is. Bovendien lenen de semantische kaarten zich erg goed voor het vergelijken van specifieke voorbeelden, omdat alle semantische kaarten van talen op dezelfde manier zijn opgebouwd.

Wälchli (2018) gebruikt als input voor MDS een database die hij heeft gegenereerd aan de hand van 133 datapunten uit een Bijbelcorpus, het Nieuwe Testament. Elk datapunt omvat parallel steeds hetzelfde stukje tekst in 72 verschillende talen of dialecten. De focus van het onderzoek van Wälchli (2018) ligt met name op de Baltische talen, waardoor veel van de talen uit dit tekstcorpus van deze categorie zijn. Ook worden er verschillende historische perioden van sommige talen meegenomen, omdat de typologie van *LATE-clauses* niet diachroon stabiel is. Doordat er in het corpus parallel verschillende vertalingen van dezelfde tekst zijn opgenomen, kan onderzocht worden of er semantische overeenkomsten bestaan tussen connectieven in verschillende talen.

Van elk paar punten van de 133 datapunten uit het Bijbelcorpus wordt eerst *Hamming distance* berekend als maat van dissimilariteit. Dit is de formule zoals in (a) waar s het aantal van dezelfde connectieven is in het paar datapunten en t het totaal aantal paren in de twee datapunten. t komt grotendeels overeen met het aantal talen dat is opgenomen in een datapunt.

$$(a) 1 - s/t$$

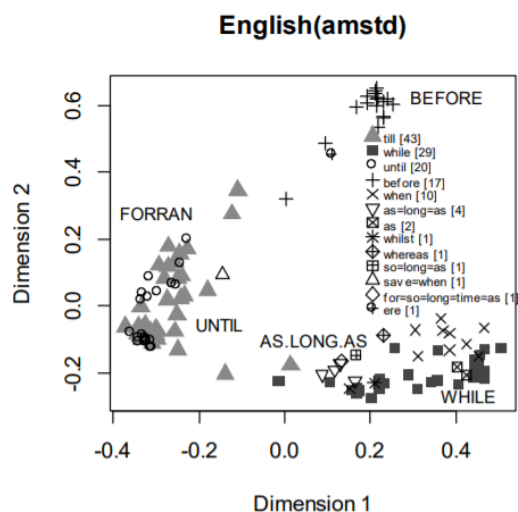
In **tabel 1** is een voorbeeld te zien van twee datapunten waarover de dissimilariteit kan worden berekend. Per rij is een datapunt uitgesplitst over de kolommen. In elke kolom is het connectief weergegeven dat in die taal voorkomt. In de tabel worden datapunt 003 en datapunt 004 weergegeven. In de bovenste rij wordt steeds de afkorting van de taal of het dialect weergegeven. De overeenkomstige paren in de twee datapunten zijn onderstreept. De dissimilariteit tussen deze datapunten is hier dus: $1 - 5/7$.

	aln	cat	hrv	dan	dut	eng- amstd	eng-leb
003	<u>deri sa</u>	<u>fins que</u>	sve dok	<u>indtil</u>	<u>totdat</u>	till	<u>until</u>
004	<u>deri sa</u>	<u>fins que</u>	dok	<u>indtil</u>	<u>totdat</u>	until	<u>until</u>

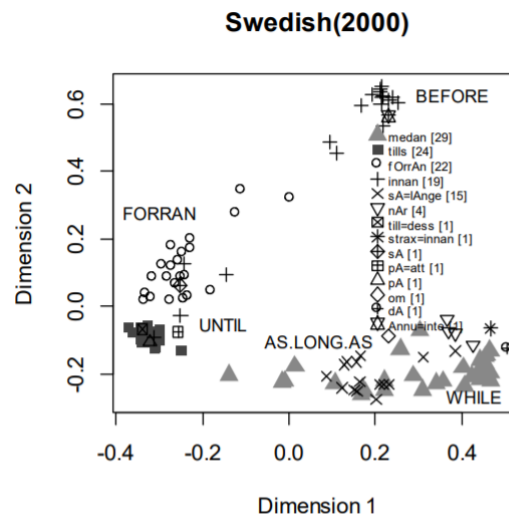
Tabel 1: datapunt 003 en datapunt 004 uitgelijnd over verschillende talen (Wälchli, 2018).

Deze berekeningen worden uiteengezet in een afstandsmatrix waarin de waardes variëren tussen 0.0 (alle talen gebruiken hetzelfde connectief voor dat specifieke paar passages) en 1.0 (alle talen gebruiken verschillende connectieven voor dat specifieke paar passages). Deze matrix wordt als input gebruikt voor MDS. Het resultaat van MDS is een probabilistische semantische ruimte van LATE-connectieven met een maximale dimensie van 132 ($n - 1$). MDS zorgt ervoor dat de twee dimensies met het laagste getal zo veel mogelijk informatie bevat. Elk punt in de semantische ruimte reflecteert een datapunt met parallel de verschillende talen vanuit het parallelle corpus. De afstand tussen elk paar punten reflecteert de kans dat beide worden uitgedrukt met dezelfde connectief in elke taal.

De semantische ruimte kan gevisualiseerd worden door het genereren van semantische kaarten per taal. In **afbeelding 1** is een voorbeeld te zien van een kaart van het Engels waarop de eerste twee dimensies te zien zijn. In elke semantische kaart wordt steeds dezelfde configuratie van symbolen gebruikt om connectieven uit te drukken. Er wordt dus hetzelfde symbool gebruikt voor vertaalequivalenten in verschillende talen. Voor het Engelse 'until' wordt bijvoorbeeld hetzelfde symbool gebruikt als het modern Zweedse 'förrän' (vergelijk **afbeelding 1** en **afbeelding 2**). Een overzicht van deze symbolen is rechts in de semantische kaart opgenomen in een legenda, welke gesorteerd is op de frequentie van connectieven (tussen haakjes weergegeven) van hoog naar laag. De eerste dimensie (x-as) spelt een UNTIL-cluster (negatief polair) uit tegen een WHILE/WHEN/AS.LONG.AS-cluster (positief polair). Voor het gemak zijn deze labels voor semantische clusters weergegeven in de figuur. De tweede dimensie (y-as) spelt een BEFORE-cluster (positief polair) uit. In de figuur zijn dus de clusters BEFORE en UNTIL, UNTIL en AS.LONG.AS met elkaar verbonden, maar BEFORE en AS.LONG.AS niet.



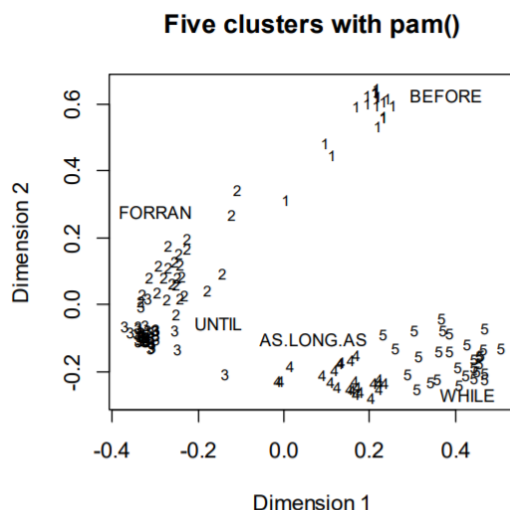
Afbeelding 1: De semantische kaart van connectieven in *LATE-clauses* in het Engels(amstd) (Wälchli, 2018).



Afbeelding 2: De semantische kaart van connectieven in *LATE-clauses* in het Zweeds(2000) (Wälchli, 2018).

Na een analyse van deze semantische kaarten concludeert Wälchli (2018) dat de relaties van de LATE-connectieven op de schaal liggen die Kortmann (1997) geïntroduceerd heeft en dat deze drie relaties beschouwd kunnen worden als drie aparte semantische clusters (AS.LONG.AS, BEFORE, UNTIL).

Deze notie wordt bevestigd door de tweede methode die Wälchli (2018) gebruikt in zijn onderzoek: *Partitioning Around Medoids* (PAM). Met deze methode worden er in de data clusters aangewezen, waarbij het optimale aantal clusters bepaald wordt. Elk cluster heeft hier een optimale dichtheid en komt overeen met een relatie. Ook hier blijkt dat drie het optimale aantal clusters is en dat deze overeenkomen met het Engelse 'before', 'until' en 'as long as'. Maar, zoals Wälchli (2018) al verwachtte aan het begin van zijn onderzoek, ziet hij duidelijke aanwijzingen voor een vierde semantische zone dat tussen BEFORE en UNTIL ligt, welke hij het FÖRRÄN-cluster noemt, naar het Zweedse connectief 'förrän'. Het Zweeds is een van de weinige talen waarin deze zone gelexicaliseerd is. Dit FÖRRÄN-cluster heeft een duidelijke categorisatie: een negatie in de hoofdzin. Zinnen met een 'not...until'-constructie worden in dit cluster geclassificeerd. Wälchli stelt dat in talen waarin deze zone niet gelexicaliseerd is, het verschilt of FÖRRÄN op dezelfde manier wordt gecodeerd als UNTIL, of als BEFORE. West-Europese talen hebben volgens Wälchli een UNTIL/FÖRRÄN overlap en Noord-Europese talen zouden een BEFORE/FÖRRÄN overlap hebben. Er is minder duidelijk bewijs voor een zone tussen AS.LONG.AS en UNTIL, maar het bewijs dat er is duidt erop dat de hoofdzin altijd een verzoek, een direct imperatief of een indirecte opdracht uitdrukt. In **afbeelding 3** is een voorbeeld van een clusterverdeling te zien die is gegenereerd door PAM.



Afbeelding 3: Vijf clusters in de semantische kaart van het *LATE-domein* (Wälchli, 2018).

3.3. Motivatie onderzoek

Uit het onderzoek van Wälchli (2018) blijkt dus dat er duidelijke aanwijzingen te vinden zijn voor een vierde cluster, het FÖRRÄN-cluster. In dit cluster worden zinnen geclassificeerd met een temporeel connectief in een negatief polaire context, zoals zinnen met de Engelse constructie ‘inte...förrän’. Deze zone is in weinig talen gelexicaliseerd. Uit het onderzoek van Wälchli (2018) blijkt dat het verschil of talen waarin er geen FÖRRÄN lexicalisatie is, er een FÖRRÄN/UNTIL of FÖRRÄN/BEFORE overlap is. Dit is een duidelijke aanwijzing voor taalvariatie in de uitdrukking van NOT.UNTIL, wat De Swart (1996) ook in haar onderzoek noemt. In dit onderzoek wordt deze taalvariatie verder onderzocht en wordt er ingezoomd op het FÖRRÄN-cluster, geïntroduceerd door Wälchli (2018). Dit wordt gedaan door een reproductie van het onderzoek van Wälchli (2018), maar met een andere focus. Het onderzoek van Wälchli (2018) beschouwde puur het temporele connectief in de zin, maar andere informatie in een zin, zoals de polariteit, lijkt ook bepalend te zijn voor de semantiek van een zin. Het is dus zinvol om de gehele temporele constructie van een zin mee te beschouwen in het onderzoek. Bovendien wordt er een moderner tekstcorpus gebruikt dan in het onderzoek van Wälchli (2018).

In de literatuur worden er al verschillende constructies genoemd die in verschillende talen gebruikt worden om NOT.UNTIL uit te drukken. De Swart (1996) heeft ideeën over welke constructies er in het Duits, Engels en Nederlands gebruikt worden. Wälchli (2018) heeft ons een mooi inzicht gegeven over het Zweeds. Deze inzichten worden beide meegenomen voor het formuleren van de hypothesen in het volgende hoofdstuk. Om deze reden worden er in dit onderzoek het Duits, Engels, Nederlands en Zweeds beschouwd. De Baltische talen die centraal stonden in het onderzoek van Wälchli (2018) worden omwille van de taalbarrière buiten beschouwing gelaten. Wel wordt er de Franse taal aan dit onderzoek toegevoegd, puur omdat er nog niks over het Frans op dit gebied bekend is. In het volgende

hoofdstuk worden er hypothesen geformuleerd over welke constructies we verwachten te vinden in elke taal.

4. Hypothesen

In de literatuur worden er verwachtingen geformuleerd over hoe NOT.UNTIL wordt uitgedrukt in verschillende talen. De Swart (1996) stelt dat in het Duits 'erst...wenn' en in het Nederlands 'pas...wanneer' zal worden gebruikt. Dit zijn volgens haar vertaalequivalenten van 'not...until'. In het Engels zal NOT.UNTIL worden uitgedrukt met 'not...until' (De Swart, 1996 en Wälchli, 2018). Uit het onderzoek van Wälchli (2018) kunnen we verwachten dat in het Zweeds de constructie 'inte...förrän' zal worden gebruikt. De motivatie voor de hypothese over het Frans, en het tweede deel van de hypothese over het Nederlands, volgt uit eerder onderzoek van het project *Time in Translation*, waarvan dit onderzoek een deelonderzoek is. In het volgende hoofdstuk wordt kort dit onderzoek uiteengezet en wordt de motivatie voor deze subhypothesen verhelderd.

De verwachting is dus dat vertalingen van de constructie 'not...until' stabiel zijn tussen talen. Dit betekent dat er in elke taal steeds één uitdrukking wordt gebruikt voor NOT.UNTIL. Het Nederlands is een uitzondering, daar verwachten we twee verschillende uitdrukkingen van NOT.UNTIL. Deze verwachtingen vertalen zich naar de volgende hypothesen:

H0: De vertaling van 'not...until' is stabiel tussen talen en daarom worden er in de talen Frans, Nederlands, Engels, Duits en Zweeds steeds met voornamelijk één (en in het geval van het Nederlands twee) constructies gebruikt om NOT.UNTIL uit te drukken:

H0.1: In het Zweeds vinden we voornamelijk de constructie 'inte...förrän' terug.

H0.2: In het Engels vinden we voornamelijk de constructie 'not...until' terug.

H0.3: In het Nederlands vinden we voornamelijk de constructies 'pas...wanneer' en 'niet...zolang...niet' terug.

H0.4: In het Duits vinden we voornamelijk de constructie 'erst...wenn' terug.

H0.5: In het Frans vinden we voornamelijk de constructie 'ne...pas...tant que...ne...pas' terug.

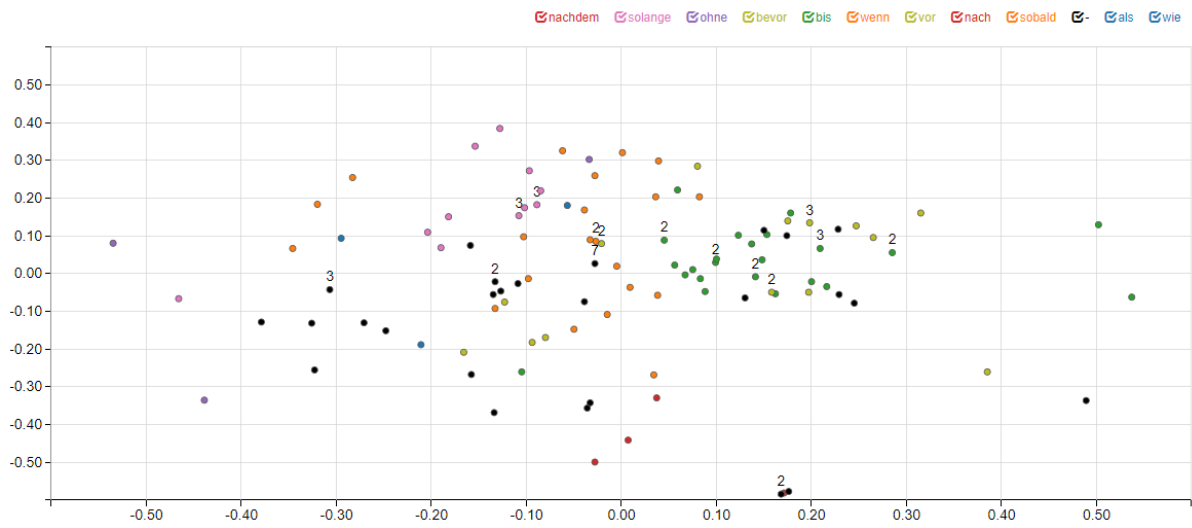
5. Opzet van het onderzoek bij *Time in Translation*

Dit onderzoek maakt deel uit van het onderzoeksproject *Time in Translation*. Het onderzoeksteam van dit project heeft naar aanleiding van de studies van Wälchli (2018) en De Swart (1996) een onderzoek opgezet over de compositionele semantiek van constructies als ‘not...until’ in verschillende talen. In dit hoofdstuk wordt ingegaan op welke inzichten er al zijn voortgekomen vanuit het onderzoek bij *Time in Translation*.

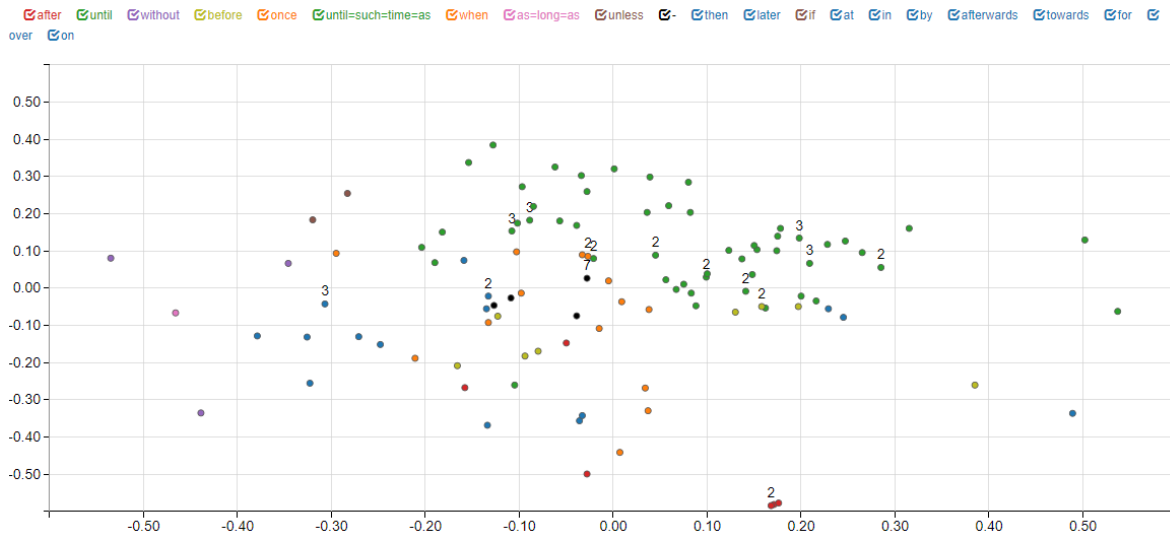
Vòòr de start van dit deelonderzoek is al heel wat werk verricht door het onderzoeksteam van *Time in Translation*. Bernhard Wälchli, Henriëtte de Swart en Jos Tellings zijn begonnen met het annoteren van 130 voorbeelden vanuit het Zweeds welke allemaal de constructie ‘inte...förrän’ bevatte. De verschillende doeltalen waren het Frans, Duits, Nederlands, Engels en Fins. De motivatie voor het annoteren vanuit het Zweeds komt uit het onderzoek van Wälchli (2018) waaruit bleek dat er nog een vierde cluster FÖRRÄN aan te wijzen valt dat ligt tussen de clusters UNTIL en BEFORE. Het Zweeds is een van de weinige clusters waarin deze zone gelexicaliseerd is. De insteek was dus om het FÖRRÄN-cluster inzichtelijker te maken. De motivatie voor het Engels als doeltaal komt voort uit hetzelfde onderzoek van Wälchli (2018). Hieruit bleek immers dat Engelse zinnen waar ‘until’ in een negatief polaire context in voorkwam in het FÖRRÄN-cluster werden geclassificeerd. De motivatie voor de andere doeltalen kwam uit het onderzoek van De Swart (1996). De Swart (1996) stelt namelijk dat het Engelse ‘not...until’ equivalent is aan het Nederlandse ‘pas...wanneer’, het Duitse ‘erst...wenn’ en het Finse ‘ennen...kuin.’ Voor mijn onderzoek is het Fins niet relevant en zal dan ook verder buiten beschouwing worden gelaten.

De Zweedse data werden, zoals ook het geval was met de data in het onderzoek van Wälchli (2018), gevisualiseerd in de semantische ruimte met behulp van MDS. In **afbeelding 4-6** zijn de semantische kaarten weergegeven van respectievelijk het Duits, Engels en Frans. De datapunten zijn steeds gelabeld met een kleur om zo connectief weer te geven wat in die taal gebruikt werd. Rechtsboven is in elke afbeelding een legenda weergegeven. In elke taal werd steeds dezelfde kleur gebruikt voor vertaalequivalenten. Zoals te zien is in de kaarten, bleek er veel variatie te zitten in het gebruik van connectieven. Veel meer dan vanuit de literatuur werd verwacht. Zoals op alle kaarten te zien is liggen de verschillende connectieven allemaal erg verspreid over de ruimte en is het dus ook niet mogelijk om afgebakende clusters aan te wijzen zoals Wälchli (2018) wel kon doen in zijn onderzoek. Om deze reden zijn deze kaarten voor mijn onderzoek niet informatief en is het interessanter om een descriptieve analyse uit te voeren van de geannoteerde data.

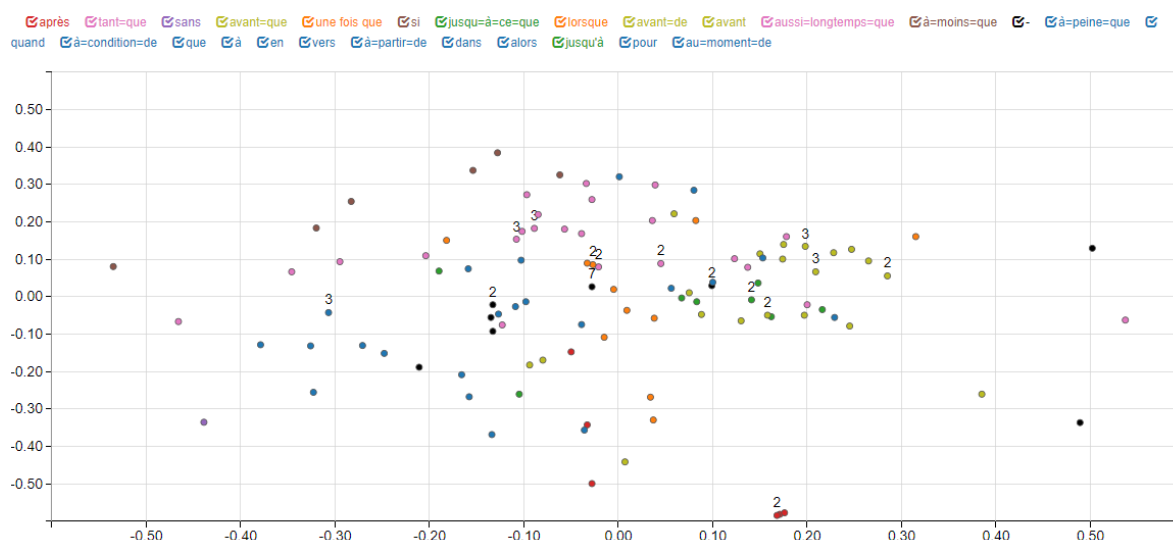
Wat er verder uit de analyse van de annotaties vanuit het Zweeds kwam was dat er in het Frans veelal de constructie ‘ne...pas...tant que...ne...pas’ voorkwam. Dit is de reden waarom deze constructie wordt genoemd in de hypothese over het Frans in hoofdstuk 4.



Afbeelding 4: visualisatie van de semantische ruimte van NOT.UNTIL in het Duits door *multidimensional scaling*.



Afbeelding 5: visualisatie van de semantische ruimte van NOT.UNTIL in het Engels door *multidimensional scaling*.



Afbeelding 6: visualisatie van de semantische ruimte van NOT.UNTIL in het Engels door *multidimensional scaling*.

Doordat er meer variatie bleek te zitten in het gebruik van connectieven in de verschillende doeltalen, is er besloten om ook andere talen als brontaal te annoteren. Er zijn toen 30 voorbeelden vanuit het Engels geannoteerd. De motivatie hiervoor was het resultaat uit het onderzoek van Wälchli. De Engelse constructie 'not...until' werd daar duidelijk geassocieerd met het FÖRRÄN-cluster.

Zelf heb ik voor dit onderzoek nog 60 voorbeelden vanuit het Nederlands aan toegevoegd. Ook over het Nederlands bestond er immers een verwachting vanuit de literatuur. Bovendien kunnen we op deze manier de variatie in het Engels en Zweeds zichtbaar maken. De Nederlandse data bestonden uit 30 voorbeelden van zinnen waar 'pas' in voorkwam en 30 voorbeelden van zinnen waar 'zolang' in voorkwam. De motivatie voor een constructie met 'pas' blijkt uit het onderzoek van De Swart (1996) die de constructie 'pas...wanneer' als vertaalequivalent noemt van het Engelse 'not until'. De motivatie voor de toevoeging van een constructie met 'zolang' blijkt niet uit de literatuur, maar uit eerdere annotaties door het onderzoeksteam van *Time in Translation*. Het vermoeden dat een constructie met 'zolang' in het FÖRRÄN-cluster zou worden geassocieerd blijkt uit het feit dat er in het Frans veelal de constructie 'ne...pas...tant que...ne...pas' voorkwam. Dit is in het Nederlands equivalent aan de constructie 'niet...zolang...niet'. Hierom is deze Nederlandse constructie ook meegenomen in het onderzoek. Er werden bij de annotaties vanuit de verschillende brontalen steeds dezelfde doeltalen gebruikt: het Frans, Nederlands, Engels, Zweeds en Duits.

6. Methodologie

In dit hoofdstuk wordt de methodologie beschreven die gebruikt wordt om de hoofdvraag van dit onderzoek te beantwoorden. In dit hoofdstuk wordt eerst het gebruikte tekstcorpus beschreven, daarna het genereren van de dataset en het annotatieproces en tot slot de methode die gebruikt wordt om uiteindelijk de annotaties te analyseren.

6.1. Het tekstcorpus Europarl

Het tekstcorpus dat in dit onderzoek gebruikt is, is het parallelle tekstcorpus Europarl. Dit corpus omvat parallel de notulen van het Europees parlement in de officiële talen van de Europese Unie. Een datapunt uit dit corpus bestaat steeds uit parallelle vertalingen van dezelfde zin uit een van de notulen. De vertalingen zijn gedaan door professionele vertalers. Voor dit onderzoek is een parallel tekstcorpus essentieel, omdat de betekenis van een zin in verschillende vertalingen stabiel is. We weten dus zeker dat een zin uit een parallel tekstcorpus in de parallelle talen dezelfde betekenis heeft. Aan de hand van een dergelijk tekstcorpus kunnen we dus onderzoeken hoe een bepaalde betekenis zoals NOT.UNTIL wordt uitgedrukt in verschillende talen.

6.2. De dataset

De dataset die gegenereerd is voor dit onderzoek vanuit het tekstcorpus Europarl bestaat grotendeels uit de reeds beschreven dataset die al door het onderzoeksteam van *Time in Translation* (Jos Tellings, Henriëtte de Swart en Bernhard Wälchli), is gegenereerd. Dit zijn de 130 voorbeelden vanuit het Zweeds en 30 voorbeelden vanuit het Engels. Zelf heb ik daar, zoals al eerder benoemd, nog 60 voorbeelden vanuit het Nederlands aan toegevoegd. Jos Tellings heeft in het Europarl tekstcorpus een zoekquery gedaan op respectievelijk de woorden 'pas' en 'zolang'. Hier heb ik handmatig van elke constructie 30 voorbeelden geselecteerd. Dit leverde een dataset op van 30 voorbeelden met het temporeel connectief 'zolang' in de zin en 30 voorbeelden met de constructies 'pas...zolang' en 'pas...wanneer'. De hoofdzin en de bijzin in die voorbeelden waren zowel positief als negatief polair.

De zinnen (9)-(13) geven een voorbeeld weer van een datapunt in deze dataset. Zin (9) geeft de zin weer in de brontaal, het Nederlands. Zinnen (10)-(13) geven de parallelle vertalingen van zin (8) weer in de doeltalen Duits, Engels, Frans en Zweeds. In elk voorbeeld is steeds de temporele constructie onderstreept.

(9) De heer Poetin, de Russische premier, heeft al aangekondigd dat Rusland het verbod niet op zal heffen zolang de Europese Unie geen details over de bron van de E.coli infectie heeft verschaft.

(10) Herr Putin, der russische Ministerpräsident, hat bereits angekündigt, dass Russland das Verbot erst aufheben wird, wenn die Europäische Union Einzelheiten über die Quelle der EHEC-Infektion bekannt gibt.

- (11) Mr Putin, the Russian Prime Minister, has already announced that Russia will not lift the ban until the European Union has provided details on the source of the E coli infection.
- (12) M. Poutine, le Premier ministre russe, a déjà annoncé que la Russie ne lèvera pas cette interdiction tant que l' Union européenne n' aura pas apporté de précisions sur l' origine de l' infection E. coli.
- (13) Rysslands premiärminister Vladimir Putin har redan meddelat att Ryssland inte kommer att upphäva förbudet förrän EU har tillhandahållit uppgifter om källan till E. coli-smittan.

Elke taal in bovenstaand voorbeeld drukt hier op zijn eigen manier NOT.UNTIL uit. In het Nederlands (9) zien we de constructie 'niet...zolang...niet' terug, in het Duits wordt 'erst...wenn' gebruikt, in het Engels zien we 'not...until' terug, in het Frans 'ne...pas...tant que...ne...pas' en in het Zweeds zien we de constructie 'inte...förrän'. Door het annoteren van dergelijke datapunten kunnen we op een rijtje krijgen welke constructie elke taal gebruikt om het semantische domein NOT.UNTIL uit te drukken.

6.3. Het annotatieproces

Door het onderzoeksteam van *Time in Translation* zijn reeds alle zinnen vanuit het Zweeds en Engels geannoteerd. Zelf heb ik daar de annotaties van 60 voorbeelden vanuit het Nederlands aan toegevoegd, samen met Henriëtte de Swart en Jos Tellings. De vertalingen van het Nederlands naar en Frans en Zweeds zijn geannoteerd door Henriëtte de Swart. Zelf heb ik de vertalingen vanuit het Nederlands naar het Engels en Duits geannoteerd. Dit is handmatig gedaan met behulp van de tools *PreSelect* en *TimeAlign*, ontwikkeld door het project *Time in Translation*.

De brontaal, in dit geval het Nederlands, is eerst geannoteerd met *PreSelect*. In **afbeelding 7** is een voorbeeld te zien van een scherm in *PreSelect*, waarin ook een voorbeeld van een annotatie te zien is. Eerst werd de constructie in de zin gemarkeerd die van belang was voor het onderzoek. Dit was in het geval van het Nederlands steeds '(niet...) zolang (...niet)', '(niet...) pas...wanneer (...niet)' of '(niet...) pas...zolang (...niet)'. Op deze manier werd gelijk inzichtelijk gemaakt welke constructie er in de zin voorkwam. Ook kon er worden aangegeven als de zin niet een dergelijke constructie bevatte. De belangrijkste annotatiestappen zijn het annoteren van (temporele) connectief van de zin, het temporele adverbium en de polariteit van de hoofd- en bijzin. Deze stappen breken de constructie op waar dit onderzoek in geïnteresseerd is en destilleren zo de relevante informatie uit de zin. Ook werd er geannoteerd op het zinstype. Verder werd er geannoteerd op de werkwoordstijd van de hoofd- en bijzin. De motivatie hiervoor is dat de relatie tussen temporele connectieven en werkwoordstijd nog niet onderzocht is (Wälchli, 2018). Dit idee valt echter buiten de scope van dit onderzoek, dus de annotaties op werkwoordstijd worden verder buiten beschouwing gelaten. Tot slot kon de volgorde

van de hoofd- en bijzin worden aangegeven. Dit was echter niet relevant voor dit onderzoek en komt daarom verder niet meer ter sprake.

Dutch

ep-10-01-20-003.xml - 442

Wij zijn het **niet**¹ eens met de geavanceerde status voor Marokko , **zolang**¹ het Sahariaanse volk het zelfbeschikkingsrecht **niet**¹ kan uitoefenen en zolang het permanent wordt vervolgd , ik herhaal , permanent wordt vervolgd door de Marokkaanse autoriteiten .

Selection tool:

This fragment does not contain a target

All targets have already been selected in this fragment

Connector

zolang

Polarity main clause

neg

Polarity dependent clause

neg

Clause type

tc

Temporal adverb

net zolang

Tense main clause

prs

Tense dependent clause

prs

Clause order

Comments

Comments

Afbeelding 7: een voorbeeldscherm in *PreSelect*, waarin een Nederlandse zin geannoteerd wordt.

Na het annoteren van de brontaal in *PreSelect*, werden de vertalingen van de brontaal geannoteerd in de tool *TimeAlign*. Dit waren, in het geval van het Nederlands als brontaal, het Frans, Duits, Engels en Zweeds. In **afbeelding 8** is een voorbeeldscherm te zien van een vertaling in *TimeAlign*. Hierin zijn twee parallelle fragmenten te zien (de brontaal en de vertaling). Links zien we hier het Nederlands, wat al geannoteerd is in *PreSelect* en rechts zien we de vertaling in het Duits. De vertaling kan hier op dezelfde manier geannoteerd worden als in de tool *PreSelect*. Door de parallelle weergave kan de gebruiker de relevante constructies van de twee zinnen in één oogopslag zien en blijft de vertaling bovendien gekoppeld aan de brontaal. Ook kunnen we, als de vertaling niet correct is of als de annotatie van de brontaal niet correct is, dat hier aangeven. Onjuiste vertalingen worden verwijderd uit de dataset.

Dutch (original) ep-10-02-09-004.xml - 923

Deze crisis is **pas** overwonnen **wanneer** degenen die hun baan zijn kwijtgeraakt weer aan de slag kunnen en ook diegenen die tot nu toe helemaal geen werk hebben gehad een baan vinden .

English (translated) ep-10-02-09-004.xml - 881

This crisis will **only** be overcome **when** those who have now become unemployed get their jobs back again , and when those who , as yet , have not had any work are able to work .

Selection tool: Word | Segment

The selected words in the original fragment do not form an instance of (a/an) *wanneer, aff, aff, to, pas, perf, perf*

This is a correct translation of the original fragment

Connector

Polarity main clause

Polarity dependent clause

Clause type

Temporal adverb

Tense main clause

Tense dependent clause

Clause order

Comments

Afbeelding 8: een voorbeeldscherm in *TimeAlign*, waarin parallel een vertaling van de zin in de brontaal wordt geannoteerd. In dit geval wordt hier de Engelse vertaling van een Nederlandse zin geannoteerd.

Het annotatieproces leverde uiteindelijk een dataset op waarin elk datapunt een vijftupel is waarin de annotaties van de vijf verschillende parallele talen (Frans, Duits, Engels, Zweeds, Nederlands) zijn opgenomen. Deze dataset van tupels bestond na het annotatieproces in totaal uit 225 voorkomens. In **afbeelding 9** is een voorbeeld te zien van een datapunt. Hierin is weergegeven wat de brontaal was en wat de vertalingen waren. Per taal is in afkortingen de annotatie weergegeven. Wat deze afkortingen betekenen, is terug te vinden in **bijlage 1**.

Source

Swedish

ep-10-03-25-011.xml fin DNA perf * ADV DNA neg förrän

Medan staterna i euroområdet under det senaste halvåret har rapporterat om att deras egna ekonomier har börjat uppvisa svarta siffror, har de första blygsamma tecknen på en ekonomisk återhämtning inte kunnat skönjas i Rumänien förrän nu.

Translations

German

erst ADV

Während einige Staaten des Euroraums in den letzten sechs Monaten wirtschaftlich wieder schwarze Zahlen geschrieben haben, werden die ersten, wenn auch bescheidenen Anzeichen einer Konjunkturbelebung in Rumänien erst jetzt sichtbar.

English

prs only aff

While states in the euro area have already been reporting a shift into the black for their own economies during the last six months, the first shoots of economic recovery, albeit fairly modest, are evident in Romania only now.

French

prs ne=que aff

Si des pays de la zone euro ont déjà signalé que leur économie n'était plus à découvert depuis les six derniers mois, les premiers signes de reprise économique, quoique relativement modeste, apparaissent que maintenant en Roumanie.

Dutch

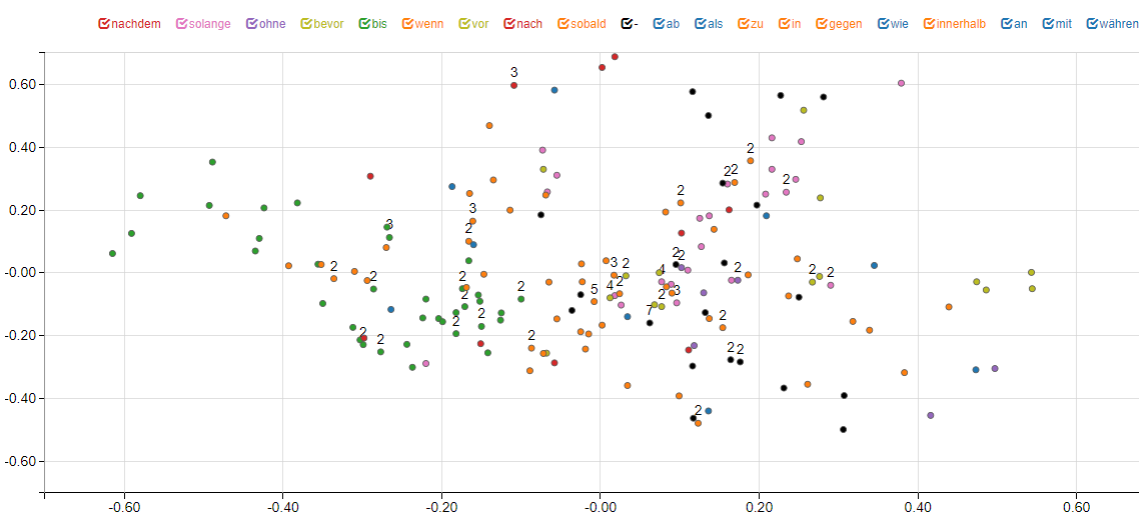
DNA prs pas ADV DNA aff

Terwijl de staten in het eurogebied weer enige groei laten zien in de afgelopen zes maanden, zijn in Roemenië nu pas de eerste timide tekenen zichtbaar van economisch herstel.

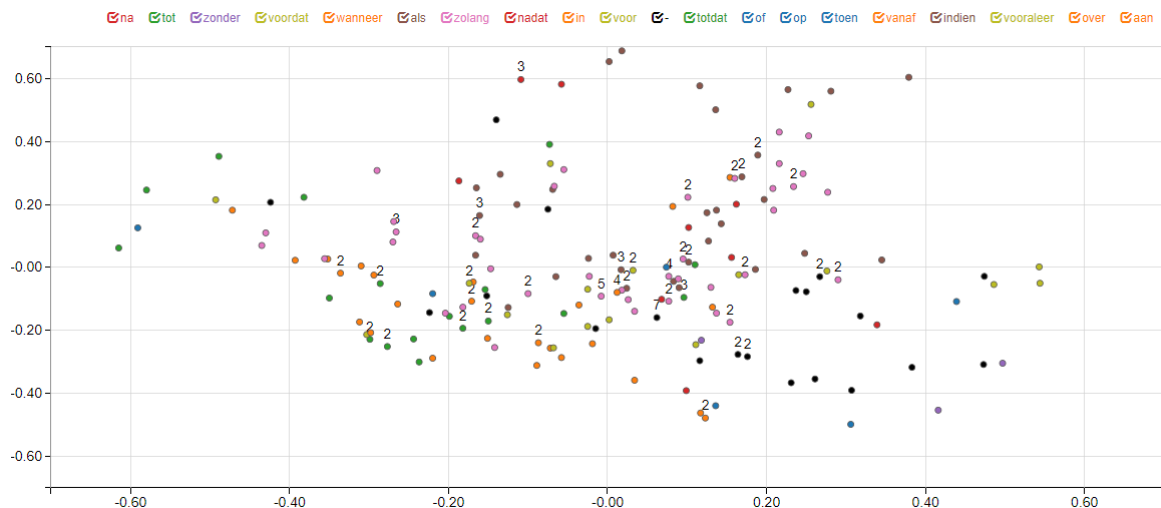
Afbeelding 9: een uitgebreid overzicht van een datapunt bestaande uit een vijf-tupel van talen. De "source" is de brontaal en laat de annotatie zien die gegenereerd is in *PreSelect*, de "Translations" zijn de vertalingen van de brontaal en laten de annotatie zien die zijn gegenereerd in *TimeAlign*.

6.4. Data Analyseren

De dataset van vijf-tupels waarin de annotaties zijn opgenomen, kan voor allerlei doeleinden gebruikt worden, waaronder MDS. Uit het vooronderzoek van het onderzoeksteam van *Time in Translation* bleek al dat er in eerdere versies van de semantische kaarten die gegenereerd zijn door MDS veel variatie zat. Ook hier is dat weer het geval. In **afbeelding 10** en **afbeelding 11** is te zien dat alle datapunten als confetti verdeeld zijn over het semantische domein. Dit maakt MDS als onderzoeksmethode voor mijn onderzoek ongeschikt omdat, door de grote spreiding van de data, de data op deze manier niet inzichtelijk gemaakt kunnen worden.



Afbeelding 10: visualisatie van de semantische ruimte van NOT.UNTIL in het Duits door *multidimensional scaling*, na toevoeging van de Nederlandse data.



Afbeelding 11: visualisatie van de semantische ruimte van NOT.UNTIL in het Nederlands door *multidimensional scaling*, na toevoeging van de Nederlandse data.

Een andere manier om de data te analyseren is het uitvoeren van een handmatige descriptieve analyse van de annotaties. Dit is dan ook de onderzoeksmethode die voor dit onderzoek gebruikt wordt. Er worden hiervoor twee zaken beschouwd. Ten eerste wordt er gekeken naar de tupelfrequentie. Dit laat zien hoe vaak een specifiek tupel van connectieven voorkwam in de dataset. Ten tweede wordt er gekeken naar spreadsheets die aan de hand van het annotatieproces zijn gegenereerd voor elke doeltaal.²

De spreadsheets die per doeltaal zijn gegenereerd aan de hand van de annotaties, bestaan uit een overzicht van de annotaties in de doeltaal met daarnaast de bijbehorende annotaties in de brontaal. De verschillende annotatiestappen die beschreven zijn, zijn over de kolommen uitgesplitst. Deze spreadsheets maken de annotaties inzichtelijk en maken het dus mogelijk om te onderzoeken welke connectieven en constructies er gebruikt worden in een taal om NOT.UNTIL uit te drukken. Door een descriptieve analyse van deze spreadsheets kunnen we de hypothesen verifiëren welke in hoofdstuk 4 geformuleerd zijn.

² De spreadsheets per brontaal zijn te bekijken via de volgende link: <https://drive.google.com/drive/u/0/folders/1HzbMT2j7YWJ3rZSGcRc2k6jgu9a0F09g> vanwege de grootte van deze spreadsheets zijn ze niet direct bijgevoegd in dit document.

7. Resultaten

In deze sectie wordt een overzicht van de resultaten gegeven. Eerst wordt er ingegaan op de tupelfrequentie. Dit laat zien hoe vaak een specifieke vijf-tupel van connectieven voorkwam in de dataset. Daarna wordt er per doeltaal een overzicht gegeven van welke constructies er in die taal voorkwamen. De resultaten worden in de hoofdstukken 8 en 9 verder geïnterpreteerd.

7.1. Tupelfrequentie

Zoals eerder benoemd, bestaat elk datapunt in de dataset bestaat uit een tupel van vijf talen. In **afbeelding 12** is een overzicht weergegeven van de meest voorkomende vijf-tupels in de dataset. Deze is gecategoriseerd op connectief, dus het kan zo zijn dat als er bijvoorbeeld 'until' is weergegeven, terwijl dit in werkelijkheid 'not...until' was. Per rij is er een vijf-tupel uit de dataset uitgesplitst over de kolommen. Elke kolom geeft dus het connectief weer die in die taal gebruikt werd om NOT.UNTIL uit te drukken. In de rechterkolom is de frequentie van precies die tupel in de dataset weergegeven.

Zoals te zien in de tabel, is de meest frequente vijf-tupel [-, -, -, -, förrän] (3,11%). Het Zweeds was hier in alle gevallen de brontaal. Dit betekent dat in 3,11% van de gevallen een Zweedse constructie met het connectief 'förrän' werd vertaald naar een zin zonder connectief. Voor deze talen is dan een '-' weergegeven in de tupel om aan te geven dat hier geen connectief werd gebruikt. De tupel [wenn, until, tant=que, zolang, förrän] kwam 2,22% voor. [solange, until, tant=que, zolang, förrän] (1,78%) en [vor, until, avant, in, förrän] (1,78%) zijn daarna het meest frequent. Het overzicht van de tupelfrequentie geeft al een eerste inzicht in hoe NOT.UNTIL wordt uitgedrukt in verschillende talen. Vanuit de hypothesen uit hoofdstuk 4 zouden we verwachten dat de tupel [wenn, until, tant=que, zolang, förrän] het meest voor zou komen. Dit is echter niet het geval. Deze tupel is wel de op één na meest frequente in de dataset, maar komt niet beduidend meer voor dan de andere tupels. Wat ook interessant is, is dat in 3,11% van de gevallen een Zweedse constructie met 'förrän' in alle andere talen vertaald wordt met een constructie zonder connectief.

Deze tupelfrequentie wordt als input gebruikt voor MDS. We kunnen al uit de uiteenlopende spreiding in het overzicht van de tupelfrequentie opmaken dat er geen duidelijke clusters aan te wijzen zullen zijn in de semantische kaarten. Er is op dit punt al duidelijk dat veel meer verschillende constructies gebruikt worden om NOT.UNTIL uit te drukken dan verwacht. De variatie in de data lijkt dus erg groot.

German	English	French	Dutch	Swedish	Count
-	-	-	-	förrän	7
wenn	until	tant=que	zolang	förrän	5
solange	until	tant=que	zolang	förrän	4
vor	until	avant	in	förrän	4
solange	until	tant=que	tot	förrän	3
nach	after	après	na	förrän=efter	3
wenn	until	tant=que	als	förrän	3
bis	until	tant=que	zolang	innan	3
wenn	when	lorsque	als	nar	3
wenn	when	lorsque	wanneer	förrän	2
wenn	when	lorsque	als	förrän	2
solange	as=long=as	aussi=longtemps=que	zolang	förrän	2
vor	until	avant	voor	förrän	2
bis	until	jusqu'à=ce=que	tot	förrän	2
bis	until	avant	in	förrän	2
bis	until	tant=que	zolang	förrän	2
vor	before	avant	-	förrän	2
-	later	-	-	förrän	2
wenn	without	tant=que	zolang	förrän	2
bis	until	avant=que	totdat	förrän	2
bis	until	tant=que	tot	förrän	2
-	-	que	-	förrän	2
bis	until	-	tot	förrän	2
in	in	en	in	förrän	2
bevor	until	tant=que	zolang	förrän	2
bis	until	jusqu'à	tot	förrän	2
bis	until	avant	voor	förrän	2
ohne	until	tant=que	als	förrän	2
wenn	until	tant=que	zolang	innan	2
ohne	without	sans	zolang	om	2
-	-	-	zolang	-	2
solange	as=long=as	tant=que	zolang	sa lange	2
wenn	if	si	als	om	2
wenn	when	quand	wanneer	nar	2
wenn	if	si	als	-	2
wenn	if	si	zolang	om	2
solange	while	tant=que	zolang	sa lange	2

Afbeelding 12: een overzicht van de tupelfrequentie van connectieven in de dataset.

7.2. Voorkomens per taal

Als we inzoomen op de data per doeltaal, dan wordt het idee van een grote taalvariatie bevestigd. In elke doeltaal werd er niet één constructie gebruikt, zoals verwacht, maar werden er zo'n 30 tot 60 verschillende constructies gevonden. Ook zien we totaal andere constructies dan we verwacht hadden. Per doeltaal volgt hier een overzicht van de gevonden constructies in die taal en hoe vaak ze voorkwamen in de dataset. De brontalen waren hier steeds het Zweeds, Nederlands of Engels. Als er 'NONE' in de tabel is weergegeven betekent dat dat er geen connectief werd gebruikt in de vertaling naar deze taal.

7.2.1. Het Zweeds

De Zweedse data bestonden uit vertalingen vanuit het Engels en het Nederlands. Dit waren in totaal 87 datapunten. In **tabel 2** is een overzicht weergegeven van de constructies in het Zweeds die het meest voorkwamen. In totaal zijn er 29 verschillende constructies gevonden ('NONE' niet meegerekend). In de tabel is te zien dat de meest gebruikte constructie in het Zweeds de constructie 'inte...förrän' was. Dit werd in 26,44% van de gevallen gebruikt. Opvallend is dat 'förrän' vrijwel altijd in een negatief polaire context voorkwam. De constructie met 'förrän' in een positief polaire context kwam immers slechts één keer voor (1,15%). In de meerderheid van de gevallen werd er een andere constructie dan 'inte...förrän' gebruikt. 'inte...innan' en 'inte så länge' zijn de op een na meest gebruikte constructies en kwamen beide in 10,43% van de gevallen voor.

Constructie	Aantal in de dataset	Percentage in de dataset
inte förrän	23	26,44%
inte innan	9	10,34%
inte så länge	9	10,34%
när	7	8,05%
inte tills	6	6,90%
inte om inte	3	3,45%
inte när	3	3,45%
om inte	2	2,30%
bara om	2	2,30%
om	2	2,30%
endast om	2	2,30%
först när	2	2,30%
inte utan	2	2,30%
för	2	2,30%
förrän	1	1,15%
inte så länge inte	1	1,15%
endast när	1	1,15%
endast	1	1,15%
NONE	5	5,75%
Overig (≤ 1)	12	13,79%

Tabel 2: overzicht van constructies die gebruikt werden om NOT.UNTIL uit te drukken in het Zweeds.

7.2.2. Het Engels

De Engelse data bestonden uit vertalingen vanuit het Zweeds en het Nederlands. Dit waren in totaal 188 datapunten. In **tabel 3** is een overzicht weergegeven van constructies die het meest voorkwamen in het Engels. In totaal zijn er 41 verschillende constructies gevonden ('NONE' niet meegerekend). In de tabel is te zien dat de meest gebruikte constructie in het Engels de constructie 'not...until' is. Dit werd in 37,77% van de gevallen gebruikt. Het is opvallend dat de constructie 'until' vrijwel alleen in een negatief polaire context voorkwam. 'Until' in een positief polaire context werd immers relatief minder vaak gebruikt (1,60%). In de meerderheid van de gevallen werd er een andere constructie gebruikt dan 'not...until'. 'Only once' en 'only when' zijn de op een na meest gebruikte constructies en werden beide in 5,32% van de gevallen gebruikt.

Constructie	Aantal in de dataset	Percentage in de dataset
not until	71	37,77%
only once	10	5,32%
only when	10	5,32%
not before	9	4,79%
only	9	4,79%
only after	8	4,26%
when	6	3,19%
not as long as	5	2,66%
only if	5	2,66%
not so long as	3	1,60%
only in	3	1,60%
until	3	1,60%
not while	3	1,60%
without	3	1,60%
not if not	2	1,06%
not as long as not	2	1,06%
not in	2	1,06%
not unless	2	1,06%
if	2	1,06%
not for	2	1,06%
only at	2	1,06%
not when	2	1,06%
not without not	2	1,06%
not without	2	1,06%
only later	1	0,53%
only by	1	0,53%
only towards	1	0,53%
only on	1	0,53%
only then	1	0,53%
only when not	1	0,53%
not only when	1	0,53%
Overig (≤ 1)	11	5,85%

Tabel 3: overzicht van constructies die gebruikt werden om NOT.UNTIL uit te drukken in het Engels.

7.2.3. Het Nederlands

De Nederlandse data bestonden uit vertalingen vanuit het Zweeds en het Engels. Dit waren in totaal 157 datapunten. In **tabel 4** is een overzicht weergegeven van de constructies die het meest voorkwamen in het Nederlands. In totaal zijn er 35 verschillende constructies gevonden ('NONE' niet meegerekend). In de tabel is te zien dat de meest gebruikte constructie in het Nederlands de constructie 'niet...zolang...niet' was (15,92%). Het is opvallend dat 'zolang' vrijwel alleen wordt gebruikt in combinatie met een negatie in de hoofd- en bijzin. Alle andere constructies met 'zolang' werden immers veel minder gebruikt: 'net zolang totdat' en 'niet zolang' kwamen beide slechts 0,64% voor. In de meerderheid van de gevallen werd er een andere constructie dan 'niet...zolang...niet' gebruikt. 'Pas' en 'niet...tot' zijn de op een na meest gebruikte constructies en werden respectievelijk in 10,19% en 7,64% van de gevallen gebruikt. Wat opvallend is, is dat de constructie 'pas wanneer' slechts 3,82% voorkomt, terwijl we hadden verwacht dat deze constructie zeer frequent zou zijn.

Constructie	Aantal in de dataset	Percentage in de dataset
niet zolang niet	25	15,92%
pas	16	10,19%
niet tot	12	7,64%
pas in	9	5,73%
niet voordat	9	5,73%
pas als	8	5,10%
pas na	8	5,10%
niet totdat	8	5,10%
niet voor	7	4,46%
pas wanneer	6	3,82%
alleen als	4	2,55%
niet als niet	3	1,91%
pas op	3	1,91%
niet vooraleer	3	1,91%
pas nadat	3	1,91%
niet zonder	3	1,91%
niet tot niet	2	1,27%
tot	2	1,27%
pas toen	2	1,27%
net zolang totdat	1	0,64%
niet zolang	1	0,64%
niet pas als	1	0,64%
pas over	1	0,64%
niet pas aan	1	0,64%
pas vanaf	1	0,64%
NONE	6	3,82%
Overig (≤ 1)	14	8,92%

Tabel 4: overzicht van constructies die gebruikt werden om NOT.UNTIL uit te drukken in het Nederlands.

7.2.4. Het Duits

De Duitse data bestonden uit vertalingen vanuit het Zweeds, Engels en Nederlands. Dit waren in totaal 220 datapunten. In **tabel 5** is een overzicht weergegeven van de constructies die het meest voorkwamen in het Duits. In totaal zijn er 46 verschillende constructies gevonden ('NONE' niet meegerekend). In de tabel is te zien dat de meest gebruikte constructie in het Duits de constructie 'nicht bis' was (15,91%). Het is opvallend dat 'bis' vrijwel altijd voorkwam in een negatief polaire context. 'Bis' in een positief polaire context kwam slechts één keer voor (0,45%). In de meerderheid van de gevallen werd er een andere constructie gebruikt dan 'nicht...bis'. 'nicht solange nicht' en 'erst' zijn de op een na meest gebruikte constructies en kwamen respectievelijk in 8,64% en 6,82% van de gevallen voor. De constructie 'erst wenn' werd maar in 3,64% van de gevallen gebruikt, terwijl de verwachting was dat deze constructie zeer frequent zou zijn. 'Nicht erst wenn' kwam in 1,82% van de gevallen voor.

Constructie	Aantal in de dataset	Percentage in de dataset
nicht bis	35	15,91%
nicht solange nicht	19	8,64%
erst	15	6,82%
erst nach	11	5,00%
wenn	11	5,00%
nicht vor	10	4,55%
nur wenn	9	4,09%
solange nicht	8	3,64%
erst wenn	8	3,64%
nicht bevor nicht	7	3,18%
erst dann nach	5	2,27%
nicht solange bis	5	2,27%
erst dann wenn	5	2,27%
nicht nach	4	1,82%
nur nach	4	1,82%
nicht erst wenn	4	1,82%
erst als	3	1,36%
nicht bevor	3	1,36%
ohne	3	1,36%
nicht ohne	3	1,36%
erst in	3	1,36%
nicht wenn nicht	3	1,36%
erst sobald	2	0,91%
nicht bis nicht	2	0,91%
nicht innerhalb	2	0,91%
nicht nach nicht	2	0,91%
erst nachdem	2	0,91%
erst zu	2	0,91%
nur nachdem	1	0,45%
nur wenn nicht	1	0,45%
nur	1	0,45%
bis	1	0,45%
NONE	7	3,18%
Overig (≤ 1)	14	6,36%

Tabel 5: overzicht van constructies die in de dataset gebruikt werden om NOT.UNTIL uit te drukken in het Duits.

7.2.5. Het Frans

De Franse data bestonden uit vertalingen vanuit het Zweeds, Engels en Nederlands. Dit waren in totaal 217 datapunten. In **tabel 6** is een overzicht weergegeven van de constructies die het meest voorkwamen in het Frans. In totaal werden er 52 verschillende constructies gevonden ('NONE' niet meegerekend). In de tabel is te zien dat de meest gebruikte constructie in het Frans de constructie 'ne...pas...tant que...ne...pas' was (21,66%). 'Tant que' lijkt altijd voor te komen in combinatie met een negatie in de hoofd- en bijzin. De voorkomens van andere constructies met 'tant que', 'ne...pas...tant que' (2,30%) en 'tant que...ne...pas' (1,38%), kwamen beduidend minder voor. In de meerderheid van de gevallen werd er een andere constructie dan 'ne...pas...tant que...ne...pas' gebruikt. 'Ne...pas...avant' en 'ne que lorsque' zijn de op een na meest gebruikte constructies en kwamen respectievelijk in 9,22% en 7,37% van de gevallen voor.

Constructie	Aantal in de dataset	Percentage in de dataset
ne pas tant que ne pas	47	21,66%
ne pas avant	20	9,22%
ne que lorsque	16	7,37%
ne que après	9	4,15%
ne que	8	3,69%
ne pas jusqu'à ce que	7	3,23%
ne pas jusqu'à	6	2,76%
ne que si	6	2,76%
ne si ne pas	6	2,76%
ne pas avant que	5	2,30%
ne pas tant que	5	2,30%
ne que à partir de	3	1,38%
ne pas avant que ne	3	1,38%
lorsque	3	1,38%
pour	3	1,38%
ne que	3	1,38%
tant que ne pas	3	1,38%
ne que une fois que	3	1,38%
ne qu'à	2	0,92%
ne pas aussi longtemps que ne pas	2	0,92%
aussi longtemps que	2	0,92%
en	2	0,92%
ne pas aussi longtemps que	2	0,92%
à	2	0,92%
dans	2	0,92%
justqu'à ce que	2	0,92%
ne pas sans	2	0,92%
sans	2	0,92%
si	2	0,92%
une fois	2	0,92%
ne que vers	2	0,92%
seulement quand	1	0,46%
NONE	7	3,23%
Overig (≤1)	21	9,68%

Tabel 6: overzicht van constructies die in de dataset gebruikt werden om NOT.UNTIL uit te drukken in het Frans.

8. Discussie I: Toetsen van hypothesen

In dit hoofdstuk worden de resultaten die zijn beschreven in hoofdstuk 7, gerelateerd aan de hypothesen die beschreven zijn in hoofdstuk 4. De hypothesen, geformuleerd aan het begin van dit onderzoek, waren als volgt:

H0: De vertaling van 'not...until' is stabiel tussen talen en daarom worden er in de talen Frans, Nederlands, Engels, Duits en Zweeds steeds met voornamelijk één (en in het geval van het Nederlands twee) constructies gebruikt om NOT.UNTIL uit te drukken:

H0.1: In het Zweeds vinden we voornamelijk de constructie 'inte...förrän' terug.

H0.2: In het Engels vinden we voornamelijk de constructie 'not...until' terug.

H0.3: In het Nederlands vinden we voornamelijk de constructies 'pas...wanneer' en 'niet...zolang...niet' terug.

H0.4: In het Duits vinden we voornamelijk de constructie 'erst...wenn' terug.

H0.5: In het Frans vinden we voornamelijk de constructie 'ne...pas...tant que...ne...pas' terug.

Uit een overzicht van de tupelfrequentie blijkt eigenlijk al dat alle bovenstaande subhypothesen van H0 te verwerpen zijn. Op basis van de hypothesen hadden we verwacht dat de tupels [wenn, until, tant=que, zolang, förrän] of [wenn, until, tant=que, wanneer, förrän] vrijwel de enige tupels zouden zijn die voorkwamen in de geannoteerde dataset. Zoals eerder genoemd was dit niet het geval en kwamen deze tupels respectievelijk in 5 van de 225 gevallen (2,22%) en in geen van de gevallen voor. Dit betekent dat er naast de constructies genoemd in de subhypothesen van H0, ook heel veel andere constructies werden gebruikt. Dit wordt bevestigd als we kijken naar een overzicht van de annotaties in de spreadsheets die per doeltaal zijn gegenereerd. In elke taal werden er zo'n 30 tot 60 verschillende constructies gebruikt. Wat opvallend is, is dat de constructies die beschreven zijn in H0.1-H0.5, vrijwel alleen voorkwamen in die specifieke vorm en ook bijna altijd het meest frequent waren in de dataset. Hieronder volgt een bespreking van de subhypothesen van H0 met betrekking tot de resultaten. Elke sectie wordt afgesloten met het verwerpen of het aannemen van de hypothese.

8.1. H0.1: Zweeds – ‘inte...förrän’

De Zweedse constructie ‘inte...förrän’ was in het Zweeds de meest frequente constructie (26,44%). Dit was ongeveer 16% meer dan de op één na meest frequente constructies ‘inte...innan’ en ‘inte...så länge’. ‘Förrän’ kwam op zichzelf slechts één keer (1,15%) voor. Dit laat zien dat als men ‘förrän’ gebruikt voor het uitdrukken van NOT.UNTIL in het Zweeds, dat in een negatief polaire context doet, dus met de constructie ‘inte...förrän’. H0.1 (de verwachting dat ‘inte...förrän’ de constructie zou zijn die voornamelijk werd gebruikt voor het uitdrukken van NOT.UNTIL in het Zweeds) is dus deels waar. Er werden, naast ‘inte...förrän’, weliswaar vele andere constructies gebruikt om in het Zweeds NOT.UNTIL uit te drukken, maar er lijkt wel een voorkeur te zijn voor het gebruik van de constructie ‘inte...förrän’. Bovendien werd ‘förrän’ vrijwel altijd in een negatief polaire context gebruikt, wat wel weer overeenkomt met de hypothese. Subhypothese H0.1. kan dus deels worden aangenomen. We zien niet voornamelijk de constructie ‘inte...förrän’ terug in het Zweeds, maar deze constructie heeft wel de voorkeur en komt bovendien voornamelijk voor in deze specifieke vorm.

8.2. H0.2: Engels – ‘not...until’

De Engelse constructie ‘not...until’ was in het Engels de meest frequente constructie (37,77%). Dit was ongeveer 32% meer dan de op één na meest frequente constructies ‘only...once’ en ‘only when’. ‘Until’ kwam op zichzelf slechts drie keer (1,60%) voor. Dit laat zien dat als ‘until’ gebruikt voor het uitdrukken van NOT.UNTIL in het Engels, men dat in een negatief polaire context doet, dus met de constructie ‘not...until’. H0.2 (de verwachting dat ‘not...until’ de constructie zou zijn die voornamelijk wordt gebruikt voor het uitdrukken van NOT.UNTIL in het Engels) is dus deels waar. Er werden, naast ‘not...until’, weliswaar vele andere constructies gebruikt om in het Engels NOT.UNTIL uit te drukken, maar er lijkt wel een voorkeur te zijn voor het gebruik van de constructie ‘not...until’. Bovendien werd ‘until’ vrijwel altijd in een negatief polaire context gebruikt, wat wel weer overeenkomt met de hypothese. Subhypothese H0.1. kan dus deels worden aangenomen. We zien niet voornamelijk de constructie ‘not...until’ terug in het Engels, maar deze constructie heeft wel de voorkeur en komt bovendien voornamelijk voor in deze specifieke vorm.

8.3. H0.3: Nederlands – ‘pas...wanneer’ en ‘niet...zolang...niet’

De Nederlandse constructie ‘niet...zolang...niet’ was in het Nederlands de meest frequente constructie (15,92%). Dit was ongeveer 5% meer dan de op twee na meest frequente constructie ‘pas’. ‘Zolang’ kwam in een andere context slechts twee keer voor (1,27%). Dit laat zien dat als er een constructie met ‘zolang’ werd gebruikt voor het uitdrukken van NOT.UNTIL in het Nederlands, dat dan zowel in de hoofdzin als de bijzin een negatie werd gebruikt. Het eerste gedeelte van H0.3 (de verwachting dat ‘niet...zolang...niet’ een van de twee constructies zou zijn die voornamelijk wordt gebruikt voor het uitdrukken van NOT.UNTIL in het Nederlands) is dus deels waar. Er werden, naast ‘niet...zolang...niet’, weliswaar vele andere constructies

gebruikt om in het Nederlands NOT.UNTIL uit te drukken, maar er lijkt wel een voorkeur te zijn voor de constructie 'niet...zolang...niet'. Bovendien werd 'zolang' vrijwel altijd gebruikt in combinatie met een negatie in zowel de hoofd- als bijzin, wat wel weer overeenkomt met de hypothese.

Het tweede gedeelte van H0.3 (de verwachting dat 'pas...wanneer' een van de twee constructies zou zijn die voornamelijk wordt gebruikt wordt voor het uitdrukken van NOT.UNTIL in het Nederlands) klopt niet. Deze constructie kwam vrij weinig voor in de dataset (3,82%) en is slechts een van de vele gebruikte constructies om NOT.UNTIL uit te drukken in het Nederlands. Wel zien we veel andere constructies met pas terug: 'pas' (10,19%), 'pas...in' (5,73%), 'pas...als' (5,10%), 'pas...na' (5,10%), 'pas...op' (1,91%), 'pas...nadat' (1,91%), 'pas...toen' (1,27%), 'pas...aan' (0,64%), 'pas...vanaf' (0,64%). De meeste van deze constructies introduceren een bijvoeglijke bepaling, zoals in een zin zoals in (14). Dit is ook een eigenschap van een constructie die NOT.UNTIL uitdrukt (vergelijk (14) en (15)), dus het is niet verrassend dat dit soort constructies opduiken.

- (14) Ik vertrek pas in december.
 (15) I will not leave until December.

Bovendien is het opvallend dat alle constructies met 'pas', inclusief 'pas...wanneer' altijd voorkwamen in een positief polaire zin, dus zonder negatie in de hoofd of bijzin, dit komt wel weer overeen met de hypothese. Subhypothese H0.3. kan dus deels worden aangenomen. We zien niet voornamelijk de constructie 'niet...zolang...niet' terug in het Nederlands, maar deze constructie heeft wel de voorkeur en komt bovendien voornamelijk voor in deze specifieke vorm. De constructie 'pas...wanneer' zien we daarentegen heel weinig terug.

8.4. H0.4: Duits – 'erst...wenn'

De Duitse constructie 'erst...wenn' kwam in het Duits in slechts 3,64% van de gevallen voor. Er werd in 2,27% van de gevallen de constructie 'erst...dann...wenn' gebruikt, wat equivalent is aan 'erst...wenn'. Het totale gebruik van 'erst...wenn' ligt dan op 5,91%. Dit is ongeveer 10% minder dan de meest frequente constructie 'nicht...bis' (15,91%). We zien als op één na meest frequente constructie 'nicht...solange...nicht' terugkomen (9,22%). Dit is niet verrassend, omdat we hadden verwacht dat de Nederlandse vertaalequivalent hiervan, 'niet...zolang...niet', een van meest voorkomende constructies zou zijn om NOT.UNTIL uit te drukken. In het Nederlands was deze constructie het meest frequent en is het niet heel verrassend dat we dit ook in het Duits veel tegen komen. Hetzelfde geldt voor de Franse vertaalequivalent van deze constructie. 'Ne...pas...tant que...ne pas' kwam in het Frans het vaakst voor (21,66%).

'Nicht...bis' is de meest gebruikte constructie is in het Duits, wat totaal niet werd verwacht. Maar, als we kijken naar bijvoorbeeld het Engels is dat ook niet een heel verrassend resultaat. In het Engels kwam de vertaalequivalent van 'nicht...bis',

'not...until', het vaakst voor (37,77%). 'Bis' komt in de uitdrukking van NOT.UNTIL vrijwel altijd in een negatief polaire context voor. Met deze resultaten moeten we H0.4. (de verwachting dat 'erst...wenn' de constructie is die voornamelijk wordt gebruikt om in het Duits NOT.UNTIL uit te drukken) verwerpen.

8.5. H0.5: Frans – 'ne...pas...tant que...ne...pas'

De Franse constructie 'ne...pas...tant que...ne...pas' was in het Frans de meest frequente constructie (21,66%). Dit is ongeveer 12% meer dan de op twee na meest frequente constructie 'ne...pas...avant.' 'Tant que' kwam in een andere context in 3,69% van de gevallen voor. Dit laat zien dat als men 'tant...que' gebruikt voor het uitdrukken van NOT.UNTIL in het Frans, dat vrijwel altijd wordt gedaan in combinatie met een negatie in zowel de hoofd- als bijzin. H0.5 (de verwachting dat 'ne...pas...tant que...ne...pas' de constructie is die voornamelijk wordt gebruikt om NOT.UNTIL uit te drukken in het Frans) is dus deels waar. Er werden, naast 'ne...pas...tant que...ne...pas', weliswaar vele andere constructies gebruikt om in het Frans NOT.UNTIL uit te drukken, maar er lijkt wel een voorkeur te zijn voor het gebruik van de constructie 'ne...pas...tant que...ne...pas'. Bovendien werd 'tant...que' vrijwel altijd gebruikt in combinatie met een negatie in de hoofd- en bijzin, wat wel weer overeenkomt met de hypothese. Subhypothese H0.5. kan dus deels worden aangenomen. We zien niet voornamelijk de constructie 'ne...pas...tant que...ne...pas' terug in het Frans, maar deze constructie heeft wel de voorkeur en komt bovendien voornamelijk voor in deze specifieke vorm.

9. Discussie II: Algemeenheden in de data

Ondanks dat geen van de subhypothesen die geformuleerd zijn in hoofdstuk 4 volledig blijken te kloppen na het toetsen aan de hand van dit corpusonderzoek, is de verdeling van connectieven in het NOT.UNTIL domein niet zo willekeurig als op het eerste gezicht lijkt. Er zijn weldegelijk tendensen in de dataset aan te wijzen. Deze worden in dit hoofdstuk besproken.

9.1. Taalvoorkeur

Ondanks dat er veel variatie bestaat tussen talen en binnen talen, hebben talen duidelijk een eigen voorkeur voor een bepaalde constructie om NOT.UNTIL uit te drukken. Ook vinden we, als we de verschillende gebruikte constructies in de talen onderling met elkaar vergelijken, dat er een aantal strategieën zijn die in alle talen worden gebruikt om NOT.UNTIL uit te drukken. In **tabel 7** is een overzicht weergegeven van de zes meest gebruikte strategieën in alle vijf talen. De naam van de strategie is steeds de Nederlandse vertaal-equivalent van de constructie, behalve in de eerste kolom. De constructie 'inte...förrän' komt namelijk alleen voor in het Zweeds, omdat deze manier van het uitdrukken van NOT.UNTIL alleen in deze taal gelexicaliseerd is. We vinden een strategie met een dergelijke constructie dus niet terug in de andere talen. Tussen haakjes is steeds weergegeven hoe vaak de constructie voorkwam in de dataset.

Zoals in de tabel te zien is hebben talen, ondanks de grote variatie, een voorkeur voor het gebruik van een bepaalde constructie voor het uitdrukken van NOT.UNTIL. Bij elke taal was er steeds één constructie die beduidend meer voorkwam dan de rest van de constructies. In het Zweeds was dat de constructie 'inte...förrän' (26,44%). 'Förrän' werd bovendien alleen gebruikt in combinatie met een negatie in de hoofdzin.

	Zweeds	Engels	Nederlands	Duits	Frans
inte förrän	inte förrän (26,44%)	-	-	-	-
niet zolang niet	inte så lange inte (1,15%)	not as long as not (1,06%)	niet zolang niet (15,92%)	nicht solange nicht (8,64%)	ne pas tant que ne pas (21,66%)
niet tot(dat)	inte tills (6,90%)	not until (37,77%)	niet tot (7,64%) niet totdat (5,10%)	nicht bis (15,91%)	ne pas jusqu'à ce que (3,23%)
niet voordat	inte innan (10,34%)	not before (4,79%)	niet voordat (5,73%)	nicht vor (4,55%) nicht bevor (1,36%)	ne pas avant (9,22%)
pas als	endast om (2,30%)	only once (5,32%)	pas als (5,10%)	nur wenn (3,64%)	seulement quand (0,46%)
pas wanneer	endast når (1,15%)	only when (5,32%)	pas wanneer (3,82%)	erst wenn (3,64%)	ne que lorsque (7,37%)
pas	endast (1,15%)	only (4,79%)	pas (10,19%)	erst (6,82%)	ne que (3,69%)

Tabel 7: overzicht van de meest gebruikte strategieën in het Zweeds, Engels, Nederlands, Duits en Frans om NOT.UNTIL uit te drukken.

Het Engels heeft een voorkeur voor de constructie 'not...until' (37,77%). 'Until' werd bovendien alleen gebruikt in combinatie met een negatie in de hoofdzin. Het Engels en het Zweeds doen in dat opzicht hetzelfde. Maar, het Engels doet wel net iets anders dan het Zweeds. Beide talen lijken dus een voorkeur te hebben voor het gebruik van een temporeel connectief met een negatie in de hoofdzin voor het uitdrukken van NOT.UNTIL, maar in het Engels kwamen ook veel constructies met 'only' voor. In het Zweeds (een constructie met 'endast') zien we dit niet zo vaak terug. 'Only...once' en 'only when' waren bijvoorbeeld veel voorkomende constructies in het Engels. Veel van deze constructies waren vertalingen van de Nederlandse constructie 'pas...als', wat net zoals deze constructies met 'only' geen scalaire constructies zijn. Het is dus logisch dat 'pas...als' vertaald werd met een dergelijke constructie. Maar, veel van deze Engelse voorbeelden waren vertalingen van Zweedse zinnen met de constructie 'inte...förrän'. Dit is een verrassend resultaat. Het Engels gebruikt hier dus twee strategieën.

Het Nederlands heeft een voorkeur voor het gebruik van een constructie met 'niet...zolang...niet' (15,92%) of een constructie met 'pas' (10,19%) in een positief

polaire context. Het Nederlands doet dus weer iets heel anders dan het Zweeds. Het gebruik van een constructie met 'pas' in een positief polaire context lijkt op de 'only'-strategie die ook in het Engels gebruikt werd. Het Frans lijkt dan weer erg op het Nederlands en heeft een erg duidelijke voorkeur voor 'ne...pas...tant que...ne...pas', de vertaalequivalent van 'niet...zolang...niet'. Deze Franse constructie kwam in 21,66% van de gevallen voor. Verder maakt het Frans, net zoals het Engels en Zweeds gebruik van een temporeel connectief met negatie in de hoofdzin, als tweede strategie. 'Ne...pas...avant' en 'ne que lorsque' kwamen ook redelijk frequent voor in het Frans en vrijwel alleen in een negatief polaire context.

Het Duits heeft een voorkeur voor de constructie 'nicht...bis' (15,91%). 'Bis' kwam hier vrijwel altijd voor in combinatie met een negatie in de hoofdzin. Dit komt dus overeen met de strategieën die het Engels en Zweeds ook gebruiken. De constructie 'nicht...solange...nicht' is de op één na meest gebruikte constructie in het Duits (8,64%). Dit komt weer overeen met de strategieën die in het Nederlands en Frans werden gebruikt.

De 'niet...zolang...niet' strategie is in het Nederlands (15,92%) en Frans (21,66%) dus de meest gebruikte strategie en is in het Duits (8,64%) de op één na meest frequente strategie. Het Engels en Zweeds gebruiken deze strategie heel weinig. Het zou logisch zijn dat er in deze talen eenzelfde constructie zou worden gevonden. Dit is echter niet het geval. Het Engelse 'not...as long as'/'not...so long as' kwam vaker voor (4,26%) dan de constructie 'not...as long as...not' (1,06%). In het Zweeds zien we hetzelfde. De constructie 'inte så lange inte' kwam maar in 1,15% van de gevallen voor, terwijl 'inte så lange' met alleen een negatie in de hoofdzin veel vaker voorkwam (10,34%).

De strategie waarbij er een negatie in de hoofdzin voorkomt zien we terug in alle talen. In het Engels ('not...until'), Duits ('nicht...bis', 'nicht...vor') en Zweeds ('inte...förrän', 'inte...innan', 'inte...så länge) is dit de meest frequente strategie. Ook in het Nederlands ('niet...tot(dat)', 'niet...voordat') en Frans ('ne que lorsque', 'ne...pas...avant') zien we deze strategie terug. De strategie dat correspondeert met het Engelse 'only' zien we veel terug in het Engels ('only', 'only...when', 'only...once' etc.) en Nederlands ('pas', 'pas...als', 'was...wanneer').

Het is dus duidelijk dat elke taal een sterke voorkeur heeft voor een bepaalde strategie, maar het verschilt per taal welke strategie dat is. Wel zijn er drie duidelijke hoofdstrategieën aan te wijzen die steeds worden gebruikt. Dit zijn het gebruik van een temporeel connectief in combinatie met een negatie in de hoofdzin ('not...until', 'inte...förrän', 'nicht...bis', het gebruik van een temporeel connectief in combinatie met een negatie in zowel de hoofd- als bijzin ('niet...zolang...niet', 'ne...pas...tant que...ne...pas') en een strategie met een (scalair) adverbium ('only', 'pas').

9.2. Een overlap met het semantische domein van de conditionele connectieven

In de data vinden we voor een groot deel constructies terug die een exclusie doen op een (temporele) schaal, wat overeenkomt met de verwachtingen vanuit de literatuur. Deze constructies zijn scalair. Maar, maar zoals in het vorige hoofdstukken al is aangetoond, gebruiken talen ook niet-scalaire constructies om NOT.UNTIL uit te drukken. Vergelijk bijvoorbeeld (16) en (17).

- (16) Ik vrees dat er alleen een einde kan komen aan de dictatuur van Zanu-PF en Mugabe als hij in zijn doodskist ligt.
- (17) We moeten echter vaststellen dat het besluit aan de late kant komt en dat het geld nog later in Portugal zal arriveren, namelijk pas eind november, begin december.

In (16) wordt de constructie 'alleen...als' gebruikt, om een exclusie aan te geven. Deze exclusie bevindt zich echter niet op een schaal. Er wordt hier meer een voorwaarde gesteld. (17) daarentegen doet hier wel een exclusie op een schaal. De maanden zijn hier geordend op een schaal en 'pas' sluit alle maanden tot november uit op deze schaal.

Er zijn nog vele andere constructies aan te wijzen, zoals (15), die niet scalair zijn en meer een voorwaarde stellen. Een duidelijk voorbeeld hiervan vinden we in het Engels, waar er veel constructies zijn gevonden met 'only' (28,72%). 'Only' geeft vaak een voorwaarde aan zoals in de constructie 'only if', waarvan een voorbeeld is gegeven in (18).

- (18) As the British would say, 'the proof of the pudding is in the eating', and I think that this plan will only be a success if we meet all the conditions .

In het voorbeeld in (18) wordt er bepaalde voorwaarde gesteld aan iets om te kunnen slagen. De constructie is hier niet scalair, het is alles of niets. In de andere talen vinden we ook voorbeelden terug van dit soort conditionele constructies. In het Frans vinden we bijvoorbeeld 'seulement quand' (equivalent aan 'alleen wanneer') terug. Deze constructie doet hetzelfde als de constructie in (18).

In het Duits vinden we verrassend veel constructies met 'nur' terug (7,27%). Dit zijn ook niet-scalaire constructies en geven ook vaak een voorwaarde aan. Een voorbeeld hiervan is gegeven in (19).

- (19) Herr Präsident , der Gipfel kann nur durch die Erzielung verbindlicher Vereinbarungen zu einem Erfolg werden.
Meneer de voorzitter, de top kan alleen een succes worden door het bereiken van bindende overeenkomsten.

In het voorbeeld in (19) wordt er door 'nur' een voorwaarde geïntroduceerd welke gesteld wordt aan hetgeen beschreven in de hoofdzin.

We vinden dus verschillende constructies terug die een uitdrukking zijn van NOT.UNTIL en een conditie introduceren. Er lijkt dus een overlap te zijn tussen het semantische domein NOT.UNTIL en het domein van de conditionele connectieven. Dit zou ook een van de verklaringen kunnen zijn voor de vele variatie die gevonden is in de dataset.

10. Conclusie

In dit hoofdstuk wordt ingegaan op de hoofdvraag die in de introductie van dit onderzoek is geformuleerd. Ook wordt er ingegaan op wat de uitkomst van dit onderzoek nu betekent voor het onderzoeksveld van KI en waarom dit onderzoek relevant is. Tot slot worden er suggesties gedaan voor vervolgonderzoek.

10.1. Beantwoorden hoofdvraag

De taalvariatie die vanuit de onderzoeken van De Swart (1996) en Wälchli (2018) werd verwacht, is niet vergelijkbaar met de grote taalvariatie die uit dit corpusonderzoek blijkt. Tussen de betrokken talen, het Frans, Zweeds, Engels, Duits en Nederlands, werd er weliswaar verwacht dat er variatie zou zijn in de manier van het uitdrukken van NOT.UNTIL, maar volgens De Swart (1996) en Wälchli (2018) zou dit stabiel zijn tussen talen en zouden we dus in elke taal maar één of soms twee constructies vinden die gebruikt werden om NOT.UNTIL uit te drukken. Volgens de subhypothesen van H0 (H0.1.-H0.5.) zouden dit de constructies 'inte...förrän' in het Zweeds, 'not...until' in het Engels, 'pas...wanneer' of 'niet...zolang...niet' in het Nederlands, 'erst...wenn' in het Duits en 'ne...pas...tant que... ne...pas' in het Frans zijn. Na annotatie van het multilinguïstische parallelle corpus Europarl bleek duidelijk dat er binnen elke taal veel meer variatie bestond dan verwacht. Er werden naast de verwachte constructies, in elke taal zo'n 30 tot 60 andere constructies gevonden. De hypothesen H0.1.-H0.5. moeten we dus in eerste instantie verwerpen. In elke taal binnen dit onderzoek kan er immers geen constructie aangewezen worden die als standaard kan worden beschouwd voor een uitdrukking van NOT.UNTIL. Er bestaat zowel variatie in de manier waarop verschillende talen NOT.UNTIL uitdrukken in vergelijking tot elkaar, als taalspecifieke variatie in de manier waarop een taal NOT.UNTIL uitdrukt.

Dat er een grote variatie binnen talen en tussen talen bestaat, impliceert niet dat deze talen compleet willekeurig uitdrukking geven aan NOT.UNTIL. Er zijn weldegelijk tendensen in de data aan te wijzen. Het Zweeds heeft een duidelijke voorkeur voor het gebruik van 'förrän' in een negatief polaire context en het Engels heeft een duidelijke voorkeur voor het gebruik van 'until' in een negatief polaire context. Dit komt dan wel weer overeen met de hypothesen waarin gesteld werd dat 'until' en 'förrän' in een negatief polaire context zouden voorkomen. Het Duits heeft een voorkeur voor het gebruik van 'bis' in een negatief polaire context. Het Nederlands heeft een duidelijke voorkeur voor de constructie 'niet...zolang...niet', wat ook in het Frans de voorkeur heeft. Hier werd vaak de constructie 'ne...pas...tant que...ne...pas' gebruikt. In beide talen worden deze constructies altijd gebruikt in

een zin waar zowel de hoofd- als bijzin beide negatief polair zijn. Ook dit komt dan weer overeen met de hypothesen waarin gesteld werd dat 'zolang' en 'tant...que' beide in combinatie voor zouden komen met een negatie in zowel de hoofd- als bijzin. Ook in het Duits is 'nicht....solange...nicht' en veelvoorkomende constructie. Er zijn dus geen specifieke regels aan te wijzen over welke taal welke constructie gebruikt, maar talen hebben wel duidelijke voorkeur voor een constructie.

Nog een tendens in de data is dat het erop lijkt dat het semantische domein van NOT.UNTIL een overlap kent met het domein van de conditionele connectieven. Een aanwijzing hiervoor was dat veel van de gevonden constructies niet scalair waren, zoals wel werd verwacht vanuit de literatuur. De niet-scalaire constructies stelden meer een voorwaarde dan dat het twee gebeurtenissen in de tijd aan elkaar koppelt. We zagen bijvoorbeeld in het Engels veel constructies terug met 'only'. Deze observatie is te rijmen met één van de andere formele analyses die De Swart (1996) in haar artikel noemde. In deze analyse werd 'not...until' gezien als iets dat ONLY betekent. We zouden dan in het Nederlands veel constructies met 'pas' moeten vinden en in het Duits constructies met 'erst', wat inderdaad het geval was.

Dan komen we terug bij de hoofdvraag: in hoeverre kloppen de hypothesen, geformuleerd aan de hand van de onderzoeken van De Swart (1996), Wälchli (2018) en *Time in Translation?* De gedachten in de subhypothesen van H0 zijn niet ver uit de richting van wat er uit de data blijkt. Ondanks dat de constructies in H0.1.-H0.5. niet de enige constructies waren die voorkwamen in de bijbehorende talen, hebben de meeste constructies die beschreven werden wel de voorkeur in die taal. Ook kwamen de constructies die werden beschreven in H0.1.-H0.5. voornamelijk voor in die specifieke vorm. Dit is per taal verschillend. De grote variatie in de data bevestigt het vermoeden van De Swart (1996) dat de verschillende formele analyses van 'not...until' zouden voortkomen uit onderlinge taalvariatie.

10.2. Relevantie voor KI en vervolgonderzoek

De resultaten van dit onderzoek dragen bij aan een beter begrip van de uitdrukking van het semantische domein NOT.UNTIL in verschillende talen en is in dat opzicht nuttige kennis voor het onderzoeksveld van kunstmatige intelligentie. Ondanks de grote gevonden variatie zijn er tendensen aan te wijzen in de manier waarop NOT.UNTIL wordt uitgedrukt in verschillende talen die ons kunnen helpen bij het verbeteren van het automatische vertalen, waar in hoofdstuk 2 over werd gesproken. Door de vele variatie in de dataset is dit echter nog niet genoeg om bijvoorbeeld als input mee te geven aan een automatisch vertaalsysteem. Hiervoor is de variatie te groot. Dit kan wel een verklaring zijn waarom gemodelleerde intelligentie soms moeite heeft met NLP op het gebied van dergelijke constructies. Het is nog niet zo eenvoudig om een één op één vertaling aan te wijzen tussen talen, omdat het vertalen van constructies in het NOT.UNTIL niet stabiel blijkt te zijn.

Het zou interessant zijn om de tendensen die nu gevonden zijn in de dataset verder te onderzoeken en in kaart te brengen. Door meer talen als brontaal toe te voegen aan de dataset en door meer constructies als uitgangspunt te nemen,

ontstaat er nog een beter en vollediger inzicht in hoe die constructies zich laten vertalen en wellicht zijn er ook, doordat er dan andere constructies als uitgangspunt worden genomen, constructies te elimineren die we nu in de dataset op zien duiken. Het is opvallend dat de resultaten van het corpusonderzoek van Wälchli (2018) en dit onderzoek zo verschillen. Dit is wellicht het resultaat van een verschil in corpus en focus. Het zou dus interessant zijn om dit onderzoek te reproduceren op nog een geheel ander corpus. Het in dit onderzoek gebruikte Europarl en het tekstcorpus dat Wälchli (2018) gebruikte in zijn onderzoek bestonden beide uit formele tekstdata. Om een vollediger beeld te krijgen is het nuttig om een corpus als uitgangspunt te nemen waarin bijvoorbeeld *spontaneous speech* in opgenomen is.

11. Referenties

Declerck, R. (1995). The problem of not. . .until. *Linguistics*, 33(1), 51–98.

<https://doi.org/10.1515/ling.1995.33.1.51>

Karttunen, L. (1974). Until. *Proceedings of the 10th Meeting of the Chicago Linguistic Society*, 284-97. <https://web.stanford.edu/~laurik/publications/archive/until.pdf>

Kortmann, B. (2012). *Adverbial Subordination: A Typology and History of Adverbial Subordinators Based on European Languages (Empirical Approaches to Language Typology [EALT] Book 18)* (1st ed.). De Gruyter Mouton.

Mittwoch, A. (1977). Negative sentences with *until*. *Proceedings of CLS*, 13, 410-17.

Norvig, P., & Russel, S.J. (2010). *Artificial Intelligence: A Modern Approach*. Prentice-Hall Inc..

Smith, S. B. (1974). *Meaning and negation*. De Gruyter Mouton.

De Swart, H. (1996). Meaning and Use of not...until. *Journal of Semantics*, 13(3), 221-263. <https://doi.org/10.1093/jos/13.3.221>

Wälchli, B. (2018). 'As long as', 'until' and 'before' clauses: Zooming in on linguistic diversity. *Baltic Linguistics*, 9, 141–236. <https://doi.org/10.32798/bl.372>

Bijlage 1: een overzicht van afkortingen gebruikt bij annotatie

Categorie	Afktorting	Betekenis
Polarity	aff	Affirmative
	neg	Negative
	DNA	Does not apply
	Neg=exp	Expletive negation
	NE	'ne' (alleen in het Frans)
Clause type	ADV	adverbium
	causalc	causal clause
	complc	complement clause
	concessive=c	concessive clause
	cond	conditional
	condc	conditional clause
	correl?	correlative
	exceptc	exceptive clause
	invers=condc	inverse conditional clause
	nom	nominal
	NP oblique	nominal phrase with oblique case marking
	PP	Prepositional Phrase
	PPmanner	Prepositional Phrase expressing manner
	tc	Temporal Clause
	tc=cleft	Cleft Temporal Clause
	tc=cleft=INF	Cleft Temporal Clause + infinitive
	tc=INF	Temporal Clause + infinitive
Tense	fut	future
	inf	infinitive
	perf	perfect
	prs	present
	pst	past
	plpq	plus-que-parfait/pluperfect
	prs?	present