
WERKWOORDSTIJDEN IN ZWEEDSE VERTALINGEN

BACHELOR SCRIPTIE ARTIFICIAL INTELLIGENCE

7.5 ECTS

C.M. DE JONG

6010717



SUPERVISOR: DR. JOS TELLINGS
TWEEDE LEZER: DR. RIANNE VAN LAMBALGEN
JUNI 2020, UTRECHT UNIVERSITY

Samenvatting

Om teksten automatisch te vertalen wordt gebruik gemaakt van Machine Translation. Google Translate is daar een voorbeeld van. In vertalingen is het belangrijk dat correct gebruik wordt gemaakt van werkwoordstijden omdat zij de temporele gebeurtenissen in een tekst coderen. Als dit niet gebeurt kan een onjuiste zinssemantiek ontstaan in de vertaling. Door het gebruik van werkwoordstijden in vertalingen te onderzoeken, kan de nauwkeurigheid van Machine Translation worden verhoogd. In deze thesis wordt onderzocht hoe werkwoordstijden in de Zweedse vertaling van *L'étranger* zich gedragen ten opzichte van de originele Franse tekst. Elke vertaalde zin van de eerste drie hoofdstukken is met het Frans vergeleken. Uit de resultaten is gebleken dat de Passé Composé grotendeels wordt vertaald met de Imperfekt. De Imparfait wordt voor het grootste deel vertaald met de Imperfekt. De Présent wordt vooral vertaald met de Presens en de Plus-que-parfait wordt vooral vertaald met de Pluskvamperfekt. Hieruit zijn probabilities voortgekomen die gebruikt kunnen worden om Rule Based Machine Translation te verbeteren. De onderzochte data kan worden gebruikt om het tweetalig literaire corpus van Statistical Machine Translation te vergroten.

Inhoudsopgave

| | | |
|----------|--|-----------|
| 1 | Inleiding | 3 |
| 1.1 | Time in Translation | 4 |
| 1.2 | Relevantie Artificial Intelligence | 4 |
| 1.3 | Onderzoeksopzet | 5 |
| 2 | Theoretische achtergrond | 5 |
| 2.1 | Machine Translation | 5 |
| 2.2 | Achtergrond Zweeds en Frans | 7 |
| 2.2.1 | Reichenbach | 7 |
| 2.2.2 | Zweeds | 8 |
| 2.2.3 | Frans | 11 |
| 2.3 | Hypothesen | 14 |
| 3 | Methode | 16 |
| 3.1 | Materiaal | 16 |
| 3.2 | Annotaties en TimeAlign | 17 |
| 3.3 | Semantische kaarten | 18 |
| 4 | Resultaten | 19 |
| 4.1 | Frans en Zweeds | 19 |
| 4.2 | Kwantitatieve resultaten | 21 |
| 5 | Discussie | 22 |
| 5.1 | Hypothesen | 22 |
| 5.2 | Implicaties | 28 |
| 6 | Conclusie | 30 |
| 6.1 | Verbeteringen | 30 |
| 6.2 | Vervolgonderzoek | 31 |

1 Inleiding

“The difference between the right word and almost the right word is the difference between lightning and a lightning bug” - Mark Twain (1888)

Met deze woorden wist de Amerikaanse schrijver Mark Twain duidelijk te maken hoe belangrijk het is om de juiste woorden te gebruiken in een tekst. Door de juiste woorden te gebruiken kunnen we precies aangeven wat we bedoelen en een pakkend stuk schrijven. Dit geldt ook voor de juiste werkwoordstijden. Wat gebeurt er met een zin die de juiste woorden bevat, maar een verkeerde werkwoordstijd krijgt toegekend? Dan verandert de zinssemantiek. Het is daarom van belang dat men ook in vertalingen de goede werkwoordstijden blijft gebruiken. Bekijk de volgende vertalingen vanuit het Frans, waarin de werkwoordstijd in het groen is gemarkeerd:

| | |
|---|---|
| J' ai habité en Suède. | [Frans, Passé Composé] |
| I lived in Sweden. | [Engels, Simple Past] |
| Ik heb in Zweden gewoond . | [Nederlands, Voltooid Tegenwoordige Tijd] |
| Ich habe im Schweden gelebt . | [Duits, Perfekt] |

In deze vertalingen worden verschillende werkwoordstijden gebruikt om de Franse zin te vertalen. In de Engelse vertaling wordt de Simple Past gebruikt om aan te geven dat de situatie voorbij is – ‘ik’ woon niet meer in Zweden. Als hier gebruik werd gemaakt van de Perfect (*have lived*), zou de situatie langere tijd duren of is de situatie permanent. In dit geval zou ‘ik’ nog steeds in Zweden wonen. In het Nederlands en het Duits is dit andersom. Hier wordt de Perfekt en de Voltooid Tegenwoordige Tijd gebruikt om aan te geven dat de situatie voorbij is. Het is een gebeurtenis uit het verleden. De Imperfekt (*lebte*) en onvoltooid verleden tijd (*woonde*) geven aan dat een actie in het verleden is begonnen, maar nog niet is afgemaakt.

1.1 Time in Translation

Deze thesis is onderdeel van het Time in Translation (TiT) project van de Universiteit Utrecht. Dit project gebruikt data van vertalingen, waarbij de nadruk ligt op werkwoordstijden. Het hoofddoel van TiT is het ontrafelen van de betekenis van werkwoordstijden in zinnen, teksten en dialogen. Er zijn diverse talen toegevoegd aan verschillende literaire parallelle corpora. Een aantal voorbeelden van deze talen zijn Engels, Duits, Nederlands en Frans. De talen zijn onder andere toegevoegd aan de literaire corpus van *L'étranger* (Camus, 1942). Dit onderzoek voegt de Zweedse vertaling van *L'étranger* toe. Door deze vertaling te vergelijken met eerdere bevindingen, wordt onderzocht of dezelfde of andere patronen worden gevonden als in eerder onderzoek naar Zweedse vertalingen (Evers, 2019; Visser, 2019). De onderzoeksvraag van deze thesis luidt als volgt:

Hoe gedragen werkwoordstijden zich in een Zweedse vertaling van *L'étranger* ten opzichte van de originele Franse tekst?

De onderzoeksvraag wordt beantwoord aan de hand van een methode die in hoofdstuk 3 wordt belicht. De resultaten kunnen worden gebruikt om de nauwkeurigheid van Machine Translation te verhogen met het doel de oorspronkelijke betekenis van de originele tekst te behouden.

1.2 Relevantie Artificial Intelligence

Natural Language Processing (NLP) is een domein in Artificial Intelligence. NLP bestaat uit onderzoeken en applicaties met het doel computers te gebruiken om tekst te begrijpen en te manipuleren. NLP is veel besproken en onderzocht. Het is een van de oudste onderzoeksgebieden naar Machine Learning en het wordt gebruikt in domeinen zoals Machine Translation (Jain et al., 2018). Verschillende onderzoekers hebben gewezen op de noodzaak om meer onderzoek te doen naar NLP (Chowdhury, 2003; Nadkarni et al., 2011).

Machine Translation is een vakgebied dat sinds de jaren vijftig wordt beoefend met het doel de nauwkeurigheid van automatische vertalingen te verhogen. Het is een van de meest belangrijke

applicaties binnen NLP (Sharma et al., 2020). Automatische vertalingen leveren in het bijzonder syntactische en semantische fouten op, waardoor automatisch vertaalde zinnen onnatuurlijk kunnen aanvoelen. Dit onderzoek richt zich vooral op de semantische fouten. Voor deze fouten zijn nog altijd geen duidelijke oplossingen (Loaiciga Sanchez, 2017).

1.3 Onderzoeksopzet

Allereerst wordt in het tweede hoofdstuk een introductie gegeven over de achtergrond van dit onderzoek. Hierin wordt het onderwerp Machine Translation en de achtergrond van het Zweeds en Frans toegelicht. Daaruit volgen een aantal hypothesen. In het derde hoofdstuk wordt de methode belicht. Hierin staat welke data op welke manier is verworven. Het vierde hoofdstuk geeft de resultaten weer, gebaseerd op de gevonden data. In het vijfde hoofdstuk staat de discussie centraal, hierin worden de resultaten besproken. De hypothesen die zijn opgesteld in het tweede hoofdstuk worden aangenomen of verworpen. In het laatste hoofdstuk zal de onderzoeksvraag worden beantwoord. Ook worden hierin eventuele verbeteringen van het onderzoek en de mogelijkheid van vervolgonderzoek besproken. De referenties die zijn gebruikt zijn weergegeven aan het eind van dit onderzoek.

2 Theoretische achtergrond

In dit hoofdstuk wordt achtergrondinformatie gegeven over Machine Translation en over de Zweedse en Franse taal. Van Machine Translation worden drie bekende methodes toegelicht.

2.1 Machine Translation

Er is niet altijd één beste vertaling van een tekst naar een andere taal. Dit komt onder andere door de flexibiliteit en ambiguïteit van taal. Dit maakt Machine Translation een uitdaging. Er is achtergrondinformatie nodig om de ambiguïteit op te lossen en de inhoud van een zin vast te stellen (Russell and Norvig, 2002). Ook kunnen verschillende woorden ongeveer hetzelfde betekenen (*data science* en *statistics*), maar dezelfde woorden kunnen ook verschillende betekenissen hebben (*to sanction*). Daarnaast zijn er nog meer elementen die taal moeilijk maken, zoals figuurlijk taal-

gebruik, ironie, sarcasme en negatie.

Er zijn methodes van Machine Translation die regels bevatten om een vertaling mogelijk te maken. Deze regels worden gemaakt door de achtergrondinformatie te bestuderen en zijn op lexicaal, syntactisch of semantisch niveau. Een van die methodes heet Rule Based Machine Translation (RBMT) (Olive et al., 2011). Een belangrijke beperking van RBMT is dat expertise is vereist om de regels op te stellen. Ook zijn er veel regels nodig om uitzonderingen op te vangen.

Een nieuwere methode is Statistical Machine Translation (SMT). SMT gebruikt statistische analyse en voorspellende algoritmen om regels te definiëren die het meest geschikt zijn voor vertalingen. Deze modellen zijn getraind met een tweetalig corpus. De taak van zo'n statistisch model wordt als volgt geformuleerd:

“Given a sentence T in the target language, we seek the sentence S from which the translator produced T . We know that our chance of error is minimized by choosing that sentence S that is most probable given T . Thus, we wish to choose S so as to maximize $Pr(S|T)$.” - (Brown et al., 1990)

Met de *target language* wordt in dit geval de vertaalde zin bedoeld. Er wordt dus gezocht naar een zin S die het meest waarschijnlijk is, gegeven dat T de vertaling is. Een beperking van SMT is de moeilijkheid om zinnen te vertalen die niet lijken op de inhoud van de tweetalige corpus.

Een nog nieuwere methode is Neural Machine Translation (NMT). NMT gebruikt neurale netwerken om te vertalen. Deze netwerken vertalen delen van een tekst tegelijk, in plaats van woord voor woord. Op deze manier kan NMT de zinnen opnieuw bekijken, waardoor bijvoorbeeld werkwoordstijden correct vertaald kunnen worden. Helaas zijn NMT-modellen computationeel erg duur en daarom (nog) niet geschikt voor grootschalig gebruik (Wu et al., 2016).

2.2 Achtergrond Zweeds en Frans

In dit onderzoek worden vertalingen van het Frans naar het Zweeds onderzocht. In dit hoofdstuk worden de twee talen met elkaar vergeleken, waarbij de nadruk ligt op de semantiek van de werkwoordstijden. Door de achtergronden en de semantiek te bestuderen kunnen voorspellingen worden gedaan over het gebruik van werkwoordstijden in vertalingen.

2.2.1 Reichenbach

Reichenbach (1947) gebruikte drie *time points* om een werkwoordstijd aan te duiden: *point of speech* (S), *point of the event* (E) en *point of reference* (R). Hij deed dit onder andere met de Engelse Perfect. De verschillende relaties tussen (S) en (R) in de werkwoordstijd categorieën zijn als volgt: Past ($R \rightarrow S$), Present (R,S) en Future ($S \rightarrow R$). In dit onderzoek heb ik deze Reichenbachiaanse structuur zelf toegepast op de Franse en Zweedse werkwoordstijden. Deze structuren zijn weergegeven in Tabel 1 en Tabel 2.

Allereerst volgen voorbeelden in het Nederlands. In de zin ‘Toen ik binnenkwam had hij al gegeten.’ is (S) het moment waarop dit wordt gezegd, (E) het moment waarop hij al had gegeten en (R) het tijdstip tussen (E) en (S), namelijk *toen* ‘ik’ binnenkwam. Deze zin wordt aangeduid met $E \rightarrow R \rightarrow S$. Dit wordt een Reichenbachiaanse structuur genoemd. De pijl geeft aan dat de *time points* na elkaar plaatsvinden. In de zin ‘Ik eet op dit moment in het restaurant.’ vinden de drie *time points* op hetzelfde moment plaats. Op hetzelfde tijdstip (R) van het spreken (S) vindt het event (E) plaats. De Reichenbachiaanse structuur wordt E,R,S. De komma geeft aan dat de *time points* op hetzelfde moment plaatsvinden. In de zin ‘Ik at gister in het restaurant.’ vinden (E) en (R) tegelijk plaats (at, gister). Het moment van spreken (S) vindt plaats nadat ‘ik’ gister at. De Reichenbachiaanse structuur wordt $E,R \rightarrow S$. In de zin ‘Ik zal morgen eten in het restaurant.’ vindt het event (E) en de referentie (R) plaats in de toekomst. De Reichenbachiaanse structuur wordt hier $S \rightarrow E,R$.

2.2.2 Zweeds

Het Zweeds is een Noord-Germaanse/Scandinavische taal met ongeveer negen miljoen sprekers over de hele wereld. Net als bij andere Scandinavische talen zijn de vervoegingen bij alle persoonsvormen hetzelfde (Andersson, 2013). De volgende werkwoordstijden worden in dit hoofdstuk behandeld: Futurum, Imperativ, Imperfekt, Konditionalis, Perfekt, Pluskvamperfekt en de Presens. Let op dat de vertaling van de voorbeeldzinnen niet altijd in dezelfde werkwoordstijd is vertaald.

Futurum

Het Futurum wordt gebruikt om de toekomstige tijd uit te drukken. De werkwoorden worden vervoegd met het hulpwerkwoord *ska* of *komma*. Bij *ska* gaat het om een plan, bedoeling, besluit, belofte of om iets dat iemand heeft gehoord. Met *komma* wordt aangegeven dat iets in de toekomst gaat gebeuren. Het gaat meestal om een voorspelling of om iets waar het onderwerp van de zin geen invloed op heeft: het gebeurt hoe dan ook (Dahl, 1995). De Reichenbachiaanse structuur van het Futurum is $S \rightarrow R,E$.

- 1 a. Här får man ingen tid på sig, man hinner knappt vänja sig vid tanken förrän man ska springa bakom likvagnen.
- b. Hier heeft men geen tijd, men heeft zich nog niet met de gedachte vertrouwd gemaakt of men moet al achter de lijkstoet aan.

Imperativ

Met de Imperativ wordt de gebiedende wijs uitgedrukt met een bevel, wens of advies. Omdat de Imperativ geen tijdsvorm of declaratieve zin is, is hier geen Reichenbachiaanse structuur aan verbonden.

- 2 a. Träna!
- b. Train!

Imperfekt

Met de Imperfekt wordt uitgedrukt dat een situatie zich op een bepaald tijdstip in het verleden heeft afgespeeld (3, *Då*, 'Toen'). Het wordt gebruikt als iets helemaal tot het verleden behoort. De Reichenbachiaanse structuur van deze zin is $E, R \rightarrow S$. Het is een event (E) dat zich op een bepaald tijdstip (R) afspeelt, nog voordat het besproken wordt (S).

- 3 a. Då tänkte jag att det där borde jag inte ha sagt.
- b. Toen bedacht ik dat ik hem dat niet had moeten zeggen.

Konditionalis

De Konditionalis heeft betrekking op een hypothetische gebeurtenis: iets dat gebeurd zou zijn als eerst iets anders was gebeurd. De Konditionalis heeft verschillende Reichenbachiaanse structuren.

- 4 a. Det skulle bli en vacker dag.
- b. Het beloofde een mooie dag te worden.

Perfekt

De Perfekt wordt gebruikt om een gebeurtenis uit het verleden weer te geven. Het is een gebeurtenis die in de toekomst afgesloten zal zijn. De Perfekt kan niet worden gebruikt als een deiktisch

bijwoord in de zin staat, in dat geval wordt de Imperfekt gebruikt (3). Als er een link met het heden is kan de Perfekt wel worden gebruikt (5).

- 5 a. Jag **har tagit** på mig att sköta det nödvändiga.
- b. Ik **heb** de maatregelen **getroffen** die daarvoor nodig zijn.

Pluskvamperfekt

De Pluskvamperfekt beschrijft een in het verleden afgesloten gebeurtenis die nog vóór een andere gebeurtenis in het verleden ligt. De Reichenbachiaanse structuur van deze zin is $E \rightarrow R \rightarrow S$.

- 6 a. Han **hade sagt** mig att de måste begrava henne mycket snabbt.
- b. Hij **had gezegd** dat men haar spoedig moest begraven.

Presens

Met de Presens wordt uitgedrukt dat iets op dit moment gebeurt. Ook wordt daarmee uitgedrukt dat men iets regelmatig doet of dat iets in de toekomst gaat gebeuren (via een tijdsaanduiding kan duidelijk worden dat het om de toekomst gaat). De Reichenbachiaanse structuur van de zin 'Ik neem de bus om twee uur.' is hetzelfde als in de zin 'Ik zal de bus om twee uur nemen.' in het Futurum, namelijk $S \rightarrow R,E$. De Reichenbachiaanse structuur van deze zin is S,R,E (8) of $S \rightarrow R,E$ (7).

- 7 a. Nu **får** ni vara ensam.
- b. Ik **laat** u alleen.

- 8 a. Jag förstår .
 b. Ik begrijp het.

Tabel 1: Reichenbachiaanse structuur van Zweedse werkwoordstijden

| Zweedse werkwoordstijd | Reichenbachiaanse structuur |
|------------------------|--|
| Futurum | $S \rightarrow R, E \ \& \ S, R \rightarrow E \ \& \ S \rightarrow R \rightarrow E$ |
| Imperfekt | $E, R \rightarrow S$ |
| Konditionalis | $R \rightarrow S, E \ \& \ R \rightarrow E \rightarrow S \ \& \ R \rightarrow S \rightarrow E$ |
| Perfekt | $E \rightarrow R, S$ |
| Pluskvamperfekt | $E \rightarrow R \rightarrow S$ |
| Presens | $S, R, E \ \& \ S \rightarrow R, E$ |

2.2.3 Frans

De volgende werkwoordstijden worden in dit hoofdstuk behandeld: Passé Composé, Imparfait, Présent, Plus-que-parfait en Futur Antérieur. Let op dat in de vertaling van de voorbeeldzinnen soms een andere werkwoordstijd is gebruikt.

Passé Composé

De Passé Composé wordt gebruikt om aan te geven dat een actie in het verleden is voltooid, maar nog verband houdt met het heden. De Passé Composé beschrijft situaties alleenstaand, alsof er geen relatie bestaat tussen de situaties (Van der Klis et al., 2018). In de Passé Composé zijn combinaties met bepaalde bijwoorden in het verleden mogelijk (9, *Aujourd'hui*). Ook is het in de Passé Composé mogelijk om zinnen die een bepaalde tijdsvolgorde uitdrukken achter elkaar te plaatsen (10). Daarnaast is het mogelijk om te praten over objecten of individuen die niet meer bestaan (11) (Ritz, 2012).

9 a. Aujourd'hui, maman **est morte**.

b. Vandaag **is** moeder **gestorven**.

10 a. Le concierge **est venu** alors de mon côté. Il s'**est assis** près de moi. Après un assez long moment, il m'**a renseigné** sans me regarder.

b. Daarna **kwam** de conciërge naast mij **staan**. Hij **nam** naast mij **plaats**. Na geruime tijd **lichtte** hij mij **in**.

11 a. Napoleon **a transformé** Paris.

b. Napoleon **transformeerde** Parijs.

Imparfait

Onderzoek laat zien dat het verschil tussen de Passé Composé en de Imparfait erg moeilijk uit te leggen is aan niet-Franssprekers (Connor, 1992; Nikolova, 2009). Ze verschillen niet op semantisch niveau, maar op aspectueel niveau. Beide werkwoordstijden worden gebruikt om een actie in het verleden te beschrijven. De Imparfait wordt gebruikt om te vertellen wat de omstandigheden waren en de Passé Composé wordt gebruikt om te vertellen wat er gebeurde of wat daarna gebeurde. Een situatie in de Imparfait is bijna nooit voltooid en wacht op een verdere verklaring in de Passé Composé (Dansereau, 1987). In (12) beschrijft de Imparfait (*avaient*, groen) de omstandigheden en de Passé Composé wat daarna gebeurde (*a dit*, geel).

12 a. Ils **avaient** tous beaucoup de peine pour moi et Céleste m'**a dit**: “On n'a qu'une mère”.

b. Zij **hadden** allen erg met mij te doen, en Céleste **zei** tegen mij: “Je hebt maar één moeder”.

Présent

De Présent wordt gebruikt voor feiten of situaties in het nu (13) en voor situaties die één keer, meerdere keren of nooit in het heden zullen plaatsvinden. Ook kan het worden gebruikt voor een toekomstige actie (14).

13 a. Vous **êtes** jeune.

b. U **bent** jong.

14 a. On ne s'est pas fait à l'idée que déjà il **faut** courir derrière le corbillard.

b. Men heeft zich nog niet met de gedachte vertrouwd gemaakt of men **moet** al achter de lijkstoet aan.

Plus-que-parfait

Wanneer men een verhaal vertelt uit het verleden en men wil iets zeggen over iets wat daarvoor gebeurde, dan gebruikt men de Plus-que-parfait. Vaak wordt de Plus-que-parfait in combinatie met de Passé Composé gebruikt (15, *a répondu*)

15 a. Immédiatement il a répondu: “Cinq ans,” comme s'il **avait attendu** depuis toujours ma demande.

b. Onmiddellijk antwoordde hij: “Vijf jaar,” alsof hij al die tijd op mijn vraag **had staan wachten**.

Futur antérieur

Het Futur Antérieur wordt gebruikt om een event te beschrijven dat in de toekomst zal gebeuren of afgelopen zal zijn.

- 16 a. Après l'enterrement, au contraire, ce sera une affaire classée et tout **aura revêtu** une allure plus officielle.
- b. Maar na de begrafenis zal het een afgehandelde zaak zijn en dan **zal** alles er veel officiëler **uitzien**.

Tabel 2: Reichenbachiaanse structuur van Franse werkwoordstijden

| Franse werkwoordstijd | Reichenbachiaanse structuur |
|-----------------------|---|
| Passé Composé | $E \rightarrow R, S$ & $E, R \rightarrow S$ |
| Imparfait | $E, R \rightarrow S$ |
| Présent | E, R, S & $S \rightarrow R, E$ |
| Plus-que-parfait | $E \rightarrow R \rightarrow S$ |
| Futur Antérieur | $S \rightarrow R, E$ |

2.3 Hypothesen

Op basis van de Reichenbachiaanse structuren en eerdere literatuur worden de Franse werkwoordstijden met het Zweeds vergeleken. In Tabel 1 en Tabel 2 zijn de Reichenbachiaanse structuren van de twee talen te zien. De Reichenbachiaanse structuur van de Passé Composé is een combinatie van de Perfekt en de Imperfekt. Ze drukken alle drie een situatie in het verleden uit. De Imperfekt kan dat doen met een deiktisch bijwoord, maar de Perfekt niet. Bij de Passé Composé kan dit allebei. Bij de Perfekt en Passé Composé kan een link met het heden aanwezig zijn. Bij de Imperfekt en Passé Composé kan de situatie volledig tot het verleden behoren. Van der Klis et al. (2018) stelt dat de Passé Composé in veel gevallen niet met de Engelse Perfect wordt vertaald, maar met de Engelse Simple Past. De Reichenbachiaanse structuur van de Engelse Simple Past is gelijk aan de Reichenbachiaanse structuur van de Zweedse Imperfekt. Ook Evers (2019) en Visser (2019) benadrukten dat de Simple Past voor het grootste deel door een Imperfekt wordt vertaald. Mijn verwachting is daarom dat de Passé Composé grotendeels zal worden vertaald met de Imperfekt. In dat kader luidt de eerste hypothese:

H_1 : De Passé Composé wordt voor het grootste deel vertaald met een Imperfekt.

De Reichenbachiaanse structuur van de Imparfait is hetzelfde als de structuur van de Imperfekt. Ook uit de data van Van der Klis et al. (2018) blijkt dat de Imparfait voor het grootste deel wordt vertaald met een Simple Past. Daarom luidt de tweede hypothese:

H₂: De Imparfait wordt voor het grootste deel vertaald met een Imperfekt.

De Reichenbachiaanse structuur van de Présent is hetzelfde als de structuur van de Presens. Ook de structuur van het Futurum is hetzelfde als de Présent. Maar omdat de Presens ook de bedoeling en betekenis van het Futurum omvangt, ga ik ervanuit dat de voorkeur uit gaat naar de vertaling met een Presens. Daarom luidt de derde hypothese:

H₃: De Présent wordt voor het grootste deel vertaald met een Presens.

Evers (2019) en Visser (2019) stellen dat de Pluskvamperfekt en Past Perfect vergelijkbaar zijn en volgens Van der Klis et al. (2018) is de Plus-que-parfait vergelijkbaar met de Past Perfect. De Reichenbachiaanse structuur van de Plus-que-parfait lijkt het meest op het schema van de Pluskvamperfekt. Daarom luidt de vierde hypothese:

H₄: De Plus-que-parfait wordt voor het grootste deel vertaald met een Pluskvamperfekt.

De Reichenbachiaanse structuur van het Futur Antérieur is hetzelfde als een van de structuren van het Futurum. Daarnaast is de structuur ook hetzelfde als een van de structuren van de Presens. Daarom luidt de vijfde hypothese:

H₅: Het Futur Antérieur wordt voor het grootste deel vertaald met een Presens en Futurum.

3 Methode

In dit hoofdstuk wordt de methode behandeld. Er wordt duidelijk welk materiaal op welke manier is onderzocht.

3.1 Materiaal

Er is gebruikt gemaakt van de eerste drie hoofdstukken van de originele novel *L'étranger* en van de vertaalde novel *Främlingen* (Lindström, 2008). Camus gebruikt opmerkelijk vaak de Passé Composé in zijn boeken (Kaplansky, 2004) en daarmee shockeerde hij de Franse lezers. Zij waren gewend dat een novel vooral in de Passé Simple werd geschreven (De Swart, 2007). Tabel 3 laat de werkwoordstijden zien uit de eerste drie hoofdstukken van *L'étranger* die in het Zweeds zijn vertaald (in aflopende volgorde van frequentie). In totaal zijn 763 verschillende werkwoordstijden vertaald. De Zweedse vertaling bevat 803 werkwoordstijden, maar 40 hiervan heb ik als incorrect of als non-target geselecteerd.

Tabel 3: Gebruik werkwoordstijden in eerste drie hoofdstukken van *L'étranger*

| Franse werkwoordstijd | Aantal |
|-----------------------|--------|
| Passé Composé | 444 |
| Imparfait | 221 |
| Présent | 64 |
| Plus-que-parfait | 33 |
| Futur Antérieur | 1 |
| totaal | 763 |

De eerste drie hoofdstukken zijn via Optical Character Recognition (OCR) software online te raadplegen. OCR-software, ook wel optische tekenherkenning, is een scanner die tekst in een afbeelding herkent door middel van patroonherkenning. Op deze manier kan de tekst als tekst worden opgeslagen, in plaats van als afbeelding. Zonder deze software wordt de tekst in de afbeelding niet herkend en kan de tekst niet worden bewerkt.

3.2 Annotaties en TimeAlign

Alle zinnen uit de vertaalde tekst zijn vergeleken met de originele tekst. Door mij is handmatig gecontroleerd of de vertaalde zin een correcte vertaling is van de originele zin. Als de vertaling juist bleek, werd ‘This is a correct translation of the original fragment’ aangevinkt. Hierna is de correcte vertaling van de geselecteerde groene woorden aangevinkt in de vertaalde zin. Dan is een juiste werkwoordstijd aan de vertaling gekoppeld en is op ‘Submit’ geklikt om naar de volgende annotatie te gaan (zie Figuur 1). Zo is per fragment een vergelijking gemaakt van het gebruik van werkwoordstijden. De volgende werkwoordstijden konden worden gekozen: Futurum, Imperativ, Imperfekt, Infinitiv, Konditionalis, Perfekt, Pluskvamperfekt, Presens, Presens Particip en het Supinum. De werkwoordstijden die invloed hebben op de temporele gebeurtenissen in een zin zijn behandeld in hoofdstuk 2.3.

Annotation

| | |
|--|---|
| <p>French (original) 1.xml - s12.4</p> <div style="border: 1px solid #ccc; padding: 5px; min-height: 20px;">Il m' a dit que non .</div> | <p>Swedish (translated) 1.xml - s12.4</p> <div style="border: 1px solid #ccc; padding: 5px; min-height: 20px;">Han svarade nej .</div> <p><input type="checkbox"/> The selected words in the original fragment do not form an instance of (a/an) <i>passé composé</i></p> <p><input checked="" type="checkbox"/> This is a correct translation of the original fragment</p> <p>Tense</p> <div style="border: 1px solid #ccc; padding: 2px;">imperfekt ▼</div> <p>Other label</p> <div style="border: 1px solid #ccc; padding: 2px; min-height: 20px;">Other label</div> <p>Comments</p> <div style="border: 1px solid #ccc; padding: 2px; min-height: 20px;">Comments</div> |
| <div style="display: flex; justify-content: space-between;">✓ Submit→ Go to another fragment</div> | |

Figuur 1: Annotatie van Frans naar Zweeds

3.3 Semantische kaarten

Na het voltooien van alle annotaties werden tuples gevormd. Als een Passé Composé is vertaald met een Imperfekt, vloeit daar een 2-tuple <Passé Composé, Imperfekt> uit. Dit zijn beschrijvende statistieken en daarom bestaat een beperkte mogelijkheid om hier conclusies uit te trekken. De tuples laten de belangrijkste trends zien, maar de tuples verklaren bijvoorbeeld niet in welke context de vertaler voor een Perfekt koos (en niet voor een Imperfekt). Daarvoor is *Multidimensional Scaling* bedacht (Van der Klis et al., 2018).

Door *Multidimensional Scaling* worden semantische mappen gegenereerd en dat gaat als volgt: de 2-tuples zijn gebruikt als input voor een algoritme. Dit algoritme berekent de afstand tussen alle paren van de tuples. De afstand wordt gedefinieerd door twee 2-tuples met elkaar te vergelijken. Als de 2-tuples hetzelfde zijn, hebben de 2-tuples afstand nul. Als één werkwoordstijd afwijkt, dan wordt de afstand $\frac{1}{2}$. Het kan dus zijn dat veel tuples afstand nul tot elkaar hebben. Dit is te zien op de semantische kaart en dit wordt in hoofdstuk 5.1 besproken. De semantische kaart is een visuele representatie van de tuples. Daarnaast wordt de afstand ook bepaald naar de variatie in betekenis van het werkwoord. In het Frans kan in de ene context *prêter* ('lenen') worden gebruikt en in een andere context *emprunter* ('lenen'). Naar alle waarschijnlijkheid dekken die contexten een andere betekenislanding (van der Klis et al., 2017). Zie (17) en (18) waar twee keer *ai dû* is gebruikt en het woord twee keer met een ander Zweeds werkwoord (*har varit tvungen, måste*) is vertaald.

- 17 a. Maintenant il est trop grand pour moi et j' **ai dû** transporter dans ma chambre la table de la salle à manger.
- b. Nu är den för stor för mig och jag **har varit tvungen** att flytta in matbordet i mitt sovrum.
- c. Maar nu zijn zij te groot voor mij, zodat ik de huiskamertafel naar mijn eigen kamer **heb moeten** brengen.

- 18 a. J' ai dû avoir l' air fatigué parce que Raymond m' a dit qu' il ne fallait pas se laisser aller.
- b. Jag måste ha sett trött ut för Raymond sa att man fick inte ge upp.
- c. Ik moet er vermoeid hebben uitgezien, want Raymond zei mij dat ik de moed niet moest verliezen.

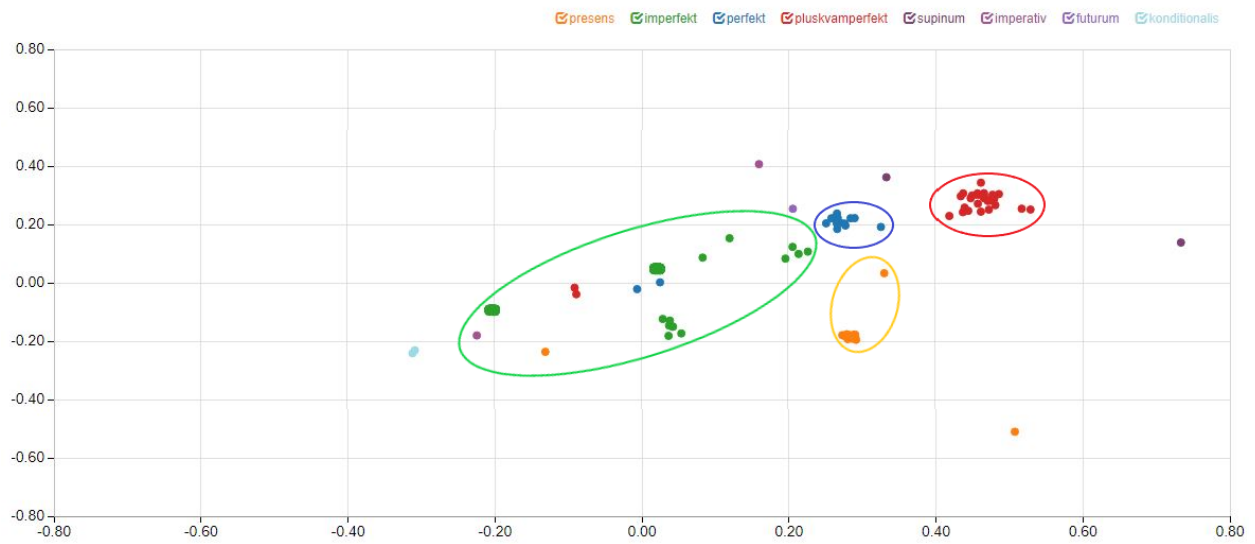
Op alle semantische kaarten in dit onderzoek heeft de x-as dimensie 1 en de y-as dimensie 2. Op deze manier zijn de punten het best van elkaar te onderscheiden. De verschillende semantische kaarten hebben dezelfde datapunten. De kleuren van de datapunten verschillen van elkaar, als de datapunten onder een andere werkwoordstijd vallen. De kleuren zijn hetzelfde voor werkwoordstijden die vergelijkbaar zijn. De clusters zijn aangegeven met gekleurde ovals. Deze gekleurde ovals zijn naar mijn eigen interpretatie getekend.

4 Resultaten

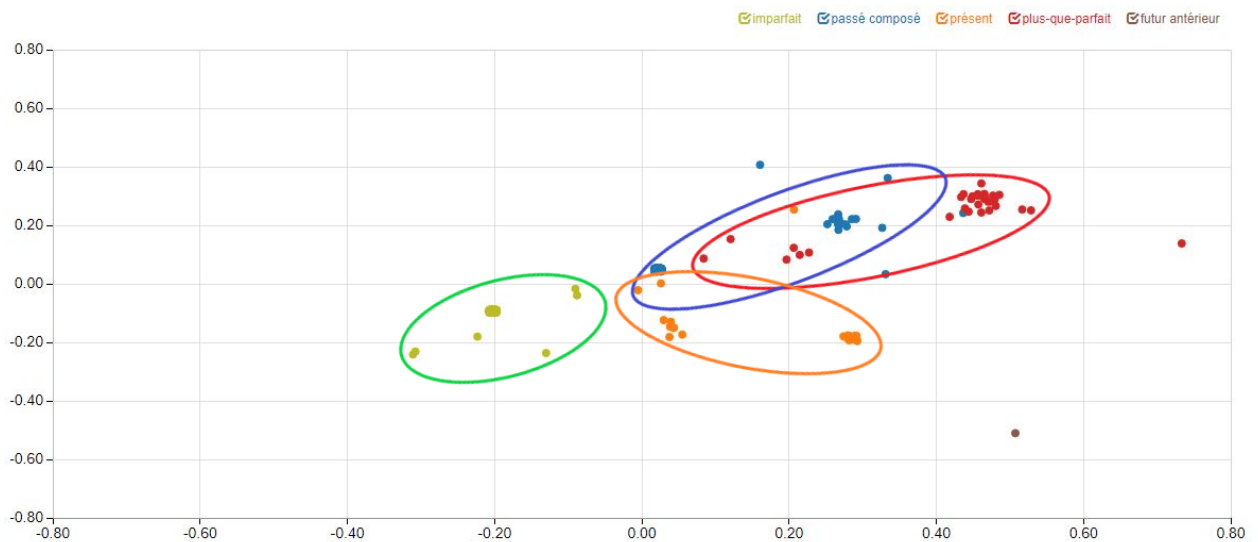
In dit hoofdstuk worden de resultaten getoond. De resultaten zijn gebaseerd op de eerste drie hoofdstukken van *L'étranger*. Allereerst volgen de resultaten van het Frans en Zweeds, daarna volgen de kwantitatieve resultaten.

4.1 Frans en Zweeds

In Tabel 4 staan alle 2-tuples die zijn gevonden in de 763 annotaties. Figuur 2 en Figuur 3 laten de semantische kaarten zien van het Zweeds en Frans.



Figuur 2: Semantische kaart gebaseerd op het gebruik van werkwoordstijden in het Zweeds ten opzichte van het Frans



Figuur 3: Semantische kaart gebaseerd op het gebruik van werkwoordstijden in het Frans ten opzichte van het Zweeds

Tabel 4: Alle 2-tuples in *L'étranger*

| Zweeds | | | | | | | | | |
|--------|--------------------------|--------------|----------------|----------------|--------------------|---------|---------------------------|---------|--------------|
| | Tense | Futu- rum | Impera- tiv | Imper- fekt | Konditio- nalis | Perfekt | Plus- kvam- perfekt | Presens | Supi- num |
| Frans | Passé Composé | | 1 | 427 | | 13 | 1 | 1 | 1 |
| | Impar- fait | | 1 | 215 | 2 | | 2 | 1 | |
| | Présent | 1 | | 7 | | 2 | | 54 | |
| | Plus- que- parfait | | | 6 | | | 26 | | 1 |
| | Futur Antérieur | | | | | | | 1 | |
| | | | | | | | | | |

4.2 Kwantitatieve resultaten

In Tabel 4 is te zien dat werkwoordstijden in het Frans over het algemeen het meest worden vertaald door de Imperfekt, daarna de Presens, de Pluskvamperfekt en dan de Perfekt. Over de rest van de Zweedse werkwoordstijden kunnen geen uitspraken worden gedaan vanwege de geringe aantal datapunten. Het gaat namelijk over 7 datapunten verdeeld over 4 verschillende werkwoordstijden.

In de semantische kaarten valt op dat het Presens-cluster (geel) in Figuur 2 kleiner is dan het Présent-cluster (geel) in Figuur 3. Tevens valt op dat het Imparfait-cluster (groen) in Figuur 2 groter is dan het Imperfekt-cluster in Figuur 3.

Omdat de Passé Composé én de Imperfait allebei worden vertaald met een Imperfekt, is de Imperfekt (groen) in Figuur 3 groter dan de Imperfait in Figuur 2. Daardoor beheert de Imperfekt een groter deel van de semantische kaart.

Tot slot blijkt dat de Passé Composé (blauw) in Figuur 3 erg groot is. Reden hiervoor is dat de Passé Composé onder andere vertaald wordt door de Perfekt en Imperfekt. De tuple <Passé Composé, Imperfekt> komt 427 keer voor, maar omdat de tuples semantisch overeenkomen hebben ze

afstand nul. Dit wordt in hoofdstuk 5.1 nader besproken.

5 Discussie

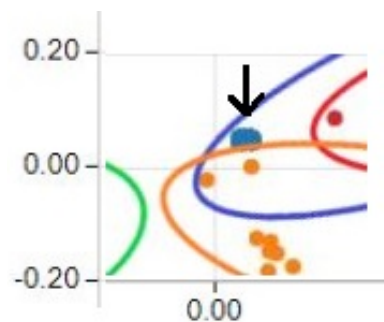
Met behulp van de resultaten uit het vorige hoofdstuk worden in dit hoofdstuk de vijf hypothesen uit hoofdstuk 2.3 uiteengezet en in dat kader aangenomen of verworpen.

5.1 Hypothesen

De voorbeeldzinnen in dit hoofdstuk komen uit de eerste drie hoofdstukken van *L'étranger*. De volgende hypothesen zijn in hoofdstuk 2.3 beschreven:

H_1 : De Passé Composé wordt voor het grootste deel vertaald met een Imperfekt.

De Passé Composé is 444 keer vertaald. Uit Tabel 4 blijkt dat de Passé Composé in 427 gevallen met de Imperfekt is vertaald en in 13 gevallen met de Perfekt. In Figuur 4 zijn de gevallen die met de Imperfekt zijn vertaald met een pijl aangegeven. Dit zijn 427 gevallen en de tuples hebben ongeveer afstand nul, waardoor de punten op elkaar staan.



Figuur 4: Franse semantische kaart: 427 annotaties waarbij de Passé Composé is vertaald met de Imperfekt

In een aantal gevallen wordt de Passé Composé vooral vertaald met de Imperfekt en niet met de Perfekt. Allereerst worden in *L'étranger* veel deiktische bijwoorden gebruikt (De Swart, 2007).

Zoals eerder benoemd, kan de Perfekt niet worden gebruikt als er een deiktisch bijwoord in de zin staat. De Imperfekt kan daarentegen wel worden gebruikt (19, 20, deiktische bijwoord in cursief). Omdat er deiktische bijwoorden in de zin voorkomen, staat de *speech time* (S) na de *reference time* (R) én na de *event time* (E). Hierdoor krijgen de zinnen een structuur die gelijk is aan de Zweedse Imperfekt (E,R → S) en niet aan de Perfekt (E → R,S). Het voorgaande is te zien in (19) en (20). In (19) is (E) het moment waarop de conciërge achter mij binnen kwam, (R) refereert naar ‘op dat ogenblik’ en (S) is het moment waarop de zin wordt gezegd. Als er een deiktisch bijwoord in de zin staat, refereert (R) hiernaar.

- 19 a. *A ce moment*, le concierge **est entré** derrière mon dos.
b. *Just då* **kom** portvakten in bakom ryggen på mig.
c. *Op dat ogenblik* **kwam** achter mij de conciërge **binnen**.

- 20 a. *Aujourd’hui* j’**ai** beaucoup **travaillé** au bureau.
b. *I dag* **arbetade** jag hårt på kontoret.
c. *Vandaag* op kantoor **heb** ik hard **gewerkt**.

L’étranger is in verhalende vorm geschreven. Er wordt daarom vaak gebruik gemaakt van de werkwoorden *dire* (‘zeggen’, 53 keer), *demander* (‘vragen’, 20 keer) en *répondre* (‘antwoorden’, 13 keer). Deze werkwoorden zijn altijd vanuit de Passé Composé naar een Imperfekt vertaald (21, 22).

- 21 a. J’**ai dit** “oui” pour n’avoir plus à parler.
b. Jag **sa** ja för att inte behöva prata mer.
c. Ik **zei** “ja”, om niet tegen hem te hoeven praten.

- 22 a. Mais le concierge m' a dit qu'il fallait que je rencontre le directeur.
- b. Men portvakten sa att jag måste träffa föreståndaren.
- c. Maar de conciërge zei mij dat ik mij eerst bij de directeur moest melden.

In 7 van de 13 gevallen waarin de Passé Composé is vertaald met een Perfekt, gebeurt dit in de dialoogvorm (23, 24). De Imperfekt is in 15 van de 427 gevallen vanuit de Passé Composé vertaald in dialoogvorm. Het zou kunnen dat de Perfekt vaker wordt gebruikt wanneer wordt geschreven in dialoogvorm. Er is echter te weinig data om dit aan te nemen.

- 23 a. "J' ai lu le dossier de votre mère."
- b. "Jag har läst er mors papper."
- c. "Ik heb uw moeders map doorgelezen."

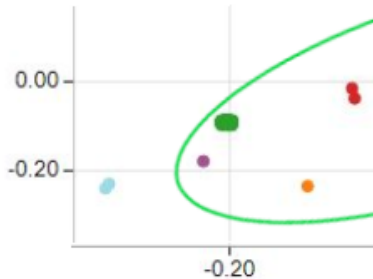
- 24 a. "J' ai bien agi avec toi et tu me le rends mal."
- b. "Jag har varit hygglig mot dig och du ger inte mycket tillbaka."
- c. "Ik ben goed voor je geweest, maar jij betaalt het mij slecht terug."

Naar mijn mening moet H_1 worden aangenomen. Uit de door mij onderzochte data blijkt dat de Passé Composé voor het grootste deel wordt vertaald met de Imperfekt.

H_2 : De Imparfait wordt voor het grootste deel vertaald met een Imperfekt.

De Imparfait is 221 keer vertaald. In 215 gevallen waarin de Imparfait wordt gebruikt, wordt het vertaald met de Imperfekt. Het cluster in Figuur 3 omvat alle annotaties waarbij de Imparfait is gebruikt. Ook hier staan de punten op elkaar, waardoor moeilijk is af te lezen om hoeveel punten het gaat. In Figuur 5 staan alle punten die vanuit een Imparfait zijn vertaald. De Imperfait is 215 keer met de Imperfekt vertaald (groen), twee keer met een Konditionalis (blauw), twee keer met

een Pluskvamperfekt (groen), één keer met een Presens (oranje) en één keer met een Imperativ (paars).



Figuur 5: Zweedse semantische kaart: Alle annotaties vertaald vanuit een Imparfait

H_2 moet wat mij betreft worden aangenomen nu de onderzochte data aantoont dat de Imparfait in 215 van de 221 gevallen met de Imperfekt is vertaald.

H_3 : De Présent wordt voor het grootste deel vertaald met een Presens.

De Présent is 64 keer vertaald. De Présent wordt in 54 gevallen met de Presens vertaald. De Présent is in 7 gevallen met de Imperfekt vertaald. De 7 vertalingen bevatten een vervoeging van het Franse werkwoord *être*. De Présent wordt vertaald met een Imperfekt wanneer in het Frans iets wordt uitgelegd op het moment zelf, met referentie naar het verleden (25, 26). Het gaat om iets dat nu nog steeds geldt. Zie de volgende twee voorbeelden:

- 25 a. C'est un peu pour cela que dans la dernière année je n'y suis presque plus allé.
- b. Det var mest därför som jag nästan aldrig for dit längre det sista året.
- c. Dat is een van de redenen waarom ik er het laatste jaar bijna niet meer ben geweest.

- 26 a. C' **est** à ce moment que les amis de maman sont entrés.
- b. Det **var** då mammans vänner kom in.
- c. Op dat ogenblik kwamen de vrienden van moeder binnen.

In (25) wordt uitgelegd dat er in het verleden een reden was waarom 'ik' er het laatste jaar bijna niet meer ben geweest. Deze reden geldt nu nog steeds. In (26) wordt uitgelegd dat in het verleden een moment bestond waarop de vrienden van moeder binnenkwamen. Dat moment is nu nog hetzelfde. Ook is *à ce moment* een deiktisch bijwoord, waardoor de Imperfekt wordt gebruikt in plaats van een andere verleden tijd.

De Presens is 2 keer vertaald met een Perfekt – beide gevallen staan in (27). Opvallend is dat deze vertaalde Perfekt wederom in dialoogvorm staat.

- 27 a. “Il y **a** longtemps que vous **êtes** là?”
- b. “**Har** ni **varit** här länge?”
- c. “Hoe lang **bent** u al hier?”

H₃ moet wat mij betreft worden aangenomen omdat uit de onderzochte data blijkt dat in 54 van de 64 gevallen de Présent met een Presens is vertaald.

H₄: De Plus-que-parfait wordt voor het grootste deel vertaald met een Pluskvamperfekt.

De Plus-que-parfait is 33 keer vertaald. De Plus-que-parfait is in 26 gevallen met de Pluskvamperfekt vertaald, één keer met een supinum en 6 keer met een Imperfekt. Bij de vertalingen met de Imperfekt is in de helft van de annotaties sprake van een bijzin en hoofdzin, waarbij de hoofdzin ook met een Imperfekt is vertaald (28, hoofdzin in groen, bijzin in geel). In het Frans bevat de hoofdzin een Pluskvamperfekt en de bijzin een Imparfait.

- 28 a. L'un des hommes qui entouraient la voiture s' **était laissé** dépasser aussi et **marchait** maintenant à mon niveau.
- b. En av männen som gick bredvid vagnen **kom** också efter och **gick** nu i jämnhöjd med mig.
- c. Een van de mannen die naast de wagen liepen **bleef** ook achter, zodat hij nu op gelijke hoogte **was** als ik.

In 8 gevallen waarin de Plus-que-parfait is vertaald met een Pluskvamperfekt is de zin opgedeeld in verschillende hoofdzinnen. De verschillende hoofdzinnen staan in de Plus-que-parfait en zijn vertaald met een Pluskvamperfekt (29). Hieruit zou kunnen worden opgemerkt dat er meer kans is op een vertaling van een Pluskvamperfekt wanneer een zin meerdere hoofdzinnen heeft met een Plus-que-parfait erin. Alle zinnen die vertaald zijn met een Pluskvamperfekt zijn in het eerste hoofdstuk van *L'étranger* gevonden. Ook hieruit kan worden opgemaakt dat meerdere Pluskvamperfekts volgen wanneer in de vertaling gebruik wordt gemaakt van een Pluskvamperfekt.

- 29 a. Il y avait longtemps que j' **étais allé** la campagne et je sentais quel plaisir j' aurais pris à me promener s' il n' y **avait** pas **eu** maman.
- b. Det var längesen jag **hade varit** på landet och jag kände hur fint det hade varit att promenera om det inte **hade varit** för mamma.
- c. Het was lang geleden sinds ik op het land **was geweest** en ik voelde hoe heerlijk het zou zijn een wandeling te maken wanneer moeder er niet **was geweest**.

H_4 moet wat mij betreft worden aangenomen omdat uit de door mij onderzochte data blijkt dat de Plus-que-parfait in 26 van de 33 gevallen met een Pluskvamperfekt is vertaald.

H_5 : Het Futur Antérieur wordt voor het grootste deel vertaald met een Presens en Futurum.

Het Futur Antérieur is in totaal één keer vertaald met de Presens. De hoeveelheid data is echter te

gering om hier een conclusie uit te trekken. Omdat er te weinig data is, wordt H_5 niet aangenomen en evenmin verworpen.

5.2 Implicaties

Zoals eerder gesteld is het belangrijk om de semantische betekenis in een vertaalde zin te behouden. Door het juiste gebruik van werkwoordstijden wordt de nauwkeurigheid van de vertaalde zin verhoogd. In dit hoofdstuk worden de voorkeuren voor werkwoordstijden in vertalingen behandeld. Verschillende talen hebben verschillende voorkeuren voor werkwoordstijden. Door de voorkeuren van een taal te onderzoeken kan de nauwkeurigheid van automatische vertalingen worden verhoogd. In hoofdstuk 2.1 zijn drie verschillende methodes besproken van Machine Translation. De nieuwste methode Neural Machine Translation is computationeel erg duur en daarom zijn de andere methodes ook besproken in het genoemde hoofdstuk. In dit hoofdstuk worden daarom regels opgesteld voor Rule Based Machine Translation. De data waaruit de regels worden opgesteld kan worden gebruikt om de corpus van Statistical Machine Translation te vergroten.

Gebaseerd op de resultaten uit Tabel 4 worden *probabilities* opgesteld voor Machine Translation. De *probability* ('kans') op de vertaling van elke werkwoordstijd wordt berekend. In Tabel 4 staat de frequentie van de werkwoordstijden. Middels deze frequenties en de *Theorema van Bayes* (30) kan de kans worden berekend. De kans ligt tussen de 0 en 1. De formule van Bayes ziet er als volgt uit:

$$30 \quad P(\mathbf{A}|\mathbf{B}) = \frac{P(\mathbf{B}|\mathbf{A})P(\mathbf{A})}{P(\mathbf{B})}$$

\mathbf{A} en \mathbf{B} worden *eventen* genoemd. De kans dat \mathbf{A} waar is (gegeven dat \mathbf{B} waar is) is gelijk aan de kans dat \mathbf{B} waar is (gegeven dat \mathbf{A} waar is), vermenigvuldigd met de kans dat \mathbf{A} waar is, gedeeld door de kans dat \mathbf{B} waar is (Stone, 2013). Gebaseerd op de uitkomsten van de hypothesen volgen hieronder mijn conclusies over de kansen.

$$31 \ P(\text{Imperfekt}) = \frac{655}{763} = 0.8585$$

$$32 \ P(\text{Imperfekt}|\text{PC}) = \frac{\frac{427}{444} \cdot \frac{655}{763}}{\frac{655}{763}} = 0.9617$$

$$33 \ P(\text{Imperfekt}|\text{IMPF}) = \frac{\frac{215}{221} \cdot \frac{655}{763}}{\frac{655}{763}} = 0.9728$$

De kans op een vertaling in de Imperfekt staat in (31). Uit (32) blijkt de kans op een Imperfekt in de vertaling, gegeven dat de Passé Composé (PC) in de originele tekst staat. In (33) staat de kans op een Imperfekt, gegeven dat de Imparfait (IMPF) in de originele tekst staat. De *probabilities* kunnen worden uitgebreid door een extra event **C** toe te voegen. Bijvoorbeeld wanneer gebruik is gemaakt van een deiktisch bijwoord in de tekst. Gezien de resultaten zal de kans zich dan (nog) meer richting de 1 bewegen. De formule zal er dan als volgt uit zien:

$$34 \ P(A|B \cap C) = \frac{P(B|A \cap C)P(A|C)}{P(B|C)}$$

In (35) staat de kans op een Presens. In (36) staat de kans op een Presens, gegeven dat de Présent (PS) in de originele tekst staat.

$$35 \ P(\text{Presens}) = \frac{57}{763} = 0.075$$

$$36 \ P(\text{Presens}|\text{PS}) = \frac{\frac{54}{64} \cdot \frac{57}{763}}{\frac{57}{763}} = 0.8438$$

In (37) staat de kans op een Pluskvamperfekt. In (38) staat de kans op een Pluskvamperfekt, gegeven dat de Plus-que-parfait (PQP) in de originele tekst staat. Ook hier kan gebruik worden gemaakt van event **C**, bijvoorbeeld wanneer in dezelfde zin een ander werkwoord ook in de PQP staat.

$$37 \ P(\text{Pluskvamperfekt}) = \frac{29}{763} = 0.038$$

$$38 \ P(\text{Pluskvamperfekt}|\text{PQP}) = \frac{\frac{26}{33} \cdot \frac{29}{763}}{\frac{29}{763}} = 0.7879$$

6 Conclusie

De onderzoeksvraag van deze thesis luidt als volgt:

Hoe gedragen werkwoordstijden zich in een Zweedse vertaling van *L'étranger* ten opzichte van de originele Franse tekst?

Uit dit onderzoek is gebleken dat in de eerste drie hoofdstukken van *L'étranger* de Passé Composé merendeels wordt vertaald met de Imperfekt. Er wordt veel gebruik gemaakt van deiktische bijwoorden, waardoor de Reichenbachiaanse structuur van de Imperfekt (en Passé Composé) vaak terugkomt. In de gevallen waarin de Passé Composé met een Perfekt is vertaald, gebeurt dit meestal in dialoogvorm. De Imparfait wordt merendeels vertaald met de Imperfekt. Omdat *L'étranger* in verhalende vorm is geschreven, wordt vaak gerefereerd naar een bepaald tijdstip in het verleden. Een vertaling met een Imperfekt als werkwoordstijd past dan het best. De Présent wordt merendeels vertaald met een Presens. In sommige gevallen werd de Présent met de Imperfekt vertaald. In deze gevallen werd vaak een situatie uit het verleden beschreven die nog steeds geldt. De Plus-que-parfait wordt merendeels vertaald met een Pluskvamperfekt. Als de Plus-que-parfait werd vertaald met een Imperfekt, was meestal sprake van een hoofd- en bijzin.

De door mij onderzochte data kan worden gebruikt om het corpus van Statistical Machine Translation te vergroten. De probabilities in hoofdstuk 5.2 kunnen worden gebruikt voor Rule Based Machine Translation.

6.1 Verbeteringen

In dit onderzoek is de Zweedse vertaling *Främlingen* van *L'étranger* onderzocht. De data is echter gebaseerd op één vertaling van één auteur, terwijl vaak meerdere correcte vertalingen van een zin bestaan. Daarnaast zijn Frans en Zweeds niet mijn moedertalen, waardoor de resultaten mogelijk minder betrouwbaar zijn.

6.2 Vervolgonderzoek

Om meer zicht te krijgen op hoe de werkwoordstijden zich in het Zweeds gedragen, zal meer data moeten worden geanalyseerd. In dit onderzoek is een literatuurwerk toegevoegd waarin vaak de Passé Composé is gebruikt. Voor vervolgonderzoek is het nodig om een ander soort corpora te gebruiken, zodat meer nadruk kan worden gelegd op andere werkwoordstijden. Daarnaast is uit dit onderzoek te weinig data van het Futur Antérieur naar voren gekomen. Om uitspraken te doen over het Futur Antérieur zal meer data onderzocht moeten worden. Ook zou het interessant zijn om niet alleen de werkwoordstijden in de literatuur te onderzoeken, maar ook in spreektaal of in (in)formele situaties. Bovendien zou het Zweeds met meerdere talen kunnen worden vergeleken. Tot slot kunnen vertalingen uit een originele Zweedse tekst worden onderzocht om een vollediger beeld te krijgen van het gebruik van werkwoordstijden in het Zweeds.

Referenties

- Andersson, E. (2013). 9 Swedish. *The Germanic Languages*, page 271.
- Brown, P. F., Cocke, J., Pietra, S. A. D., Pietra, V. J. D., Jelinek, F., Lafferty, J. D., Mercer, R. L., and Roossin, P. S. (1990). A Statistical Approach to Machine Translation. *Comput. Linguist.*, 16(2):79–85.
- Camus, A. (1942). *L'étranger*. Gallimard.
- Chowdhury, G. G. (2003). Natural language processing. *Annual review of information science and technology*, 37(1):51–89.
- Connor, M. (1992). A Processing Strategy Using Visual Representation to Convey the “passé composé/imparfait” Distinction in French. *IRAL: International Review of Applied Linguistics in Language Teaching*, 30(4):321.
- Dahl, Ö. (1995). *The tense system of Swedish*.
- Dansereau, D. (1987). A discussion of techniques used in the teaching of the passé composé/imparfait distinction in French. *French Review*, pages 33–38.
- De Swart, H. (2007). A cross-linguistic discourse analysis of the perfect. *Journal of pragmatics*, 39(12):2273–2307.
- Evers, N. M. (2019). The Perfect in Swedish Dialogue. B.S. thesis, Utrecht University.

- Jain, A., Kulkarni, G., and Shah, V. (2018). Natural language processing. *Int. J. Comput. Sci. Eng*, 6(1).
- Kaplansky, J. (2004). Outside The Stranger? English Retranslations of Camus' *L'étranger*. *Palimpsestes. Revue de traduction*, pages 187–198.
- Lindström, S. (2008). *Främlingen*. Bonnier. Origineel van Camus.
- Loaiciga Sanchez, S. (2017). *Pronominal Anaphora and Verbal Tenses in Machine Translation*. PhD thesis, University of Geneva.
- Nadkarni, P. M., Ohno-Machado, L., and Chapman, W. W. (2011). Natural language processing: an introduction. *Journal of the American Medical Informatics Association*, 18(5):544–551.
- Nikolova, O. (2009). Difficulty components of a French grammar test. In *Poster session presented at the Conference for Scholarship of Teaching and Learning, Statesboro, GA*.
- Olive, J., Christianson, C., and McCary, J. (2011). *Handbook of natural language processing and machine translation: DARPA global autonomous language exploitation*. Springer Science & Business Media.
- Reichenbach, H. (1947). The Tenses of Verbs. *Time: From Concept to Narrative Construct: a Reader*, pages 1–12.
- Ritz, M.-E. (2012). Perfect tense and aspect. *The Oxford handbook of tense and aspect*, pages 881–907.
- Russell, S. and Norvig, P. (2002). *Artificial Intelligence: A Modern Approach*.
- Sharma, V., Panchal, A., and Rane, V. Y. (2020). An Analysis on Current Research Trends and Applications of Natural Language Processing. *Advance and Innovative Research*, pages 63–69.
- Stone, J. V. (2013). *Bayes' rule: A tutorial introduction to Bayesian analysis*. Sebtel Press.
- van der Klis, M., Le Bruyn, B., and de Swart, H. (2017). Mapping the Perfect via Translation Mining. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 497–502, Valencia, Spain. Association for Computational Linguistics.
- Van der Klis, M., Le Bruyn, B., and de Swart, H. (2018). A multilingual corpus study of the competition between past and perfect in narrative discourse.
- Visser, M. (2019). The Perfect in Swedish. The Perfect in narrative and dialogue compared in Swedish, English, German, and Dutch, for improvement of Natural Language Processing. B.S. thesis, Utrecht University.
- Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., et al. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*.