A Corpus Study on Verb Tenses in Dutch and English Conditionals

Koen Klomps

First Supervisor: Jos Tellings

Second Supervisor: Rick Nouwen

July 2, 2021

BSc Kunstmatige Intelligentie

7,5 ECTS

Utrecht University

Abstract

The goal of this research is to examine and discover the differences in verb tense usage of English and Dutch conditional sentences. For this analysis the Europarl Corpus is used, together with tools from the Time in Translation project. Fragments containing conditional sentences were extracted from the corpus and then annotated with the correct verb tense. Results show that English type zero and type one conditionals often get directly translated into a corresponding similar verb tense in Dutch, e.g. simple present → onvoltooid tegenwoordige tijd. English type two conditionals get translated into a variety of verb tenses. This is possibly due to the dissimilarities in semantics between the different potential Dutch translated verb forms being minimal. These findings could prove useful when working with Rule Based Machine Translation systems.

Contents

1 Introduction

In this research the differences in tense use between English and Dutch in conditional sentences will be linguistically analyzed.

Conditionals have been studied for a long time by logicians, philosophers and linguistics. This paper aims to add to this research by looking at the way English conditional types are translated into Dutch. A conditional is a sentence often of the form "if... then..." (Dutch: "*als... dan...*"), with the "if" clause being the antecedent and the "then" clause being the consequent. To classify types of conditionals categorization will be done based on form/verb tense. Categorization makes it easier to directly compare English and Dutch conditional sentences. To obtain annotated conditionals tools and data from the Time in Translation[1] project will be used together with the Europarl Corpus[2]. The goal of this analysis of verb tenses in conditionals is the improvement of machine translation.

The research question of this paper is: "How does the tense use in conditional sentences compare between the languages English and Dutch?"

1.1 Time in Translation

Time in Translation (TinT) is a project that initially focused on the perfect tense. The project uses corpora that contain multiple languages, to compare tense use between these languages. Previously, the TinT project has annotated English conditional sentences and their respective Dutch translations. Schematically shown, this was in the direction EN→NL. This research will annotate Dutch conditional sentences and the English translations of these sentences, NL→EN. Bidirectional analysis of the verb tenses is important to create a full overview of the verb tenses in English and Dutch. These data can then be used to investigate Dutch conditionals and the English translations, on top of the already existing annotations in the EN→NL direction.

This paper is structured in the following way: following this chapter is a segment on the relevance of this research. The third chapter of this paper contains a theoretical background regarding the characteristics and types of conditional sentences. The methodology chapter talks about the methods that will be used to answer the research question. The chapter that follows after looks at the results. Finally, I will look at the implications of this research in the discussion and attempt to answer my research question in the conclusion.

[1] https://time-in-translation.hum.uu.nl
[2] https://www.statmt.org/europarl/

2 Scientific relevance

This research is written in the field of natural language processing (NLP). NLP is a subdiscipline of artificial intelligence concerning the processing of natural language by computers. In this paper two natural languages will be examined (English and Dutch). Further understanding of the verb tenses of conditional sentences could help with the improvement of different tools in the NLP field. For instance, it might aid with Rule Based Machine Translation (RBMT). RBMT works with dictionaries and grammars to attempt to translate a sentence. For example, the following Dutch sentence (1) gets translated into English, first by directly translating words using a dictionary (2) and then by applying the correct grammar (3):

(1) Deze zin is vanuit het Nederlands naar het Engels vertaald.
(2) This sentence is from the Dutch to the English translated.
(3) This sentence is translated from Dutch to English.

Normally a RBMT system will contain many more layers and be much more complex. RBMT systems could provide a better translation if there is improved knowledge about the verb tenses used in the input and output language. In other words, they would profit from improved grammar. This research hopes to contribute by providing more knowledge about the rules concerning verb tenses for conditional sentences.

3 Theoretical Background

This chapter covers the theoretical background of this research using relevant literature.

3.1 Components of a Conditional Sentence

Conditional sentences contain two parts: the antecedent (protasis) and the consequent (apodosis)(von Fintel, 2009). The truth of a conditional is dependent on the truth value of the antecedent. On the condition that the antecedent holds, the consequent will also be true:

(4) If you don't study, then you will fail the exam.

That is to say, on the condition that you don't study, the consequence will be a bad grade. The following paragraphs will provide a further look into the two parts that form a conditional sentence. It is important to understand the characteristics of a (Dutch) conditional sentence, as they will be extracted from the Europarl Corpus in this research.

3.1.1 The Antecedent

Most of the time, Dutch conditionals begin with a conditional conjunction (*voorwaardelijk voegwoord*): *als*, *indien* or *wanneer*. More rarely used conjunctions are *al*(van Bree and van der Hee, 2002), *mits*(Loosen, 2015), *zolang*(de Jonge, 2011) and *stel* (*dat*)(Boogaart and Verheij, 2013). Consider the following conditional with the conjunction *mits*:

(5) Mits je je kamer opruimt, dan mag je naar buiten.

But it is also possible for a conditional sentence to start without a conjunction:

(6) Had je dat eerder gezegd, dan kon ik het nog begrijpen.

Conditionality in (6) is still present, because the consequent is dependent on the antecedent clause. Additionally, the sentence would still be proper if *als* is placed in front of the antecedent.

This is not to say that every sentence containing a conjunction contains is a conditional sentence. Most of the conditional conjunctions in the Dutch language have multiple uses, the word *als* for instance. Two examples of sentences containing the word *als* can be compared in (7) and (8), with only (8) being a conditional. In (7), *als* is used as a correlative conjunction together with *zo*, with the goal of comparing two items (*vergelijkend voegwoord*).

(7) Het verslag van de heer von Wogau is lang niet zo onomstreden als de stemming in de Economische en Monetaire Commissie doet vermoeden.

(8) Een definitieve beoordeling kan pas volgen als de Commissie haar wetsvoorstellen openbaar maakt.

3.1.2 The Consequent

The consequent is the main clause of the sentence, it is dependent on the truth value of the antecedent clause. If the consequent in Dutch is announced with a conjunction, for the majority of the time it starts with the *dan*. This can be seen in (6). Although often the conjunction *dan* in the consequent can be omitted:

(9) Had ik dat geweten, had ik het nooit gedaan.

3.2 Types of Conditionals & Verb Tenses

There are numerous ways to categorize conditional sentences. For example, it is possible to make a categorization based on the meaning of the conditional, or the relation between the antecedent and the consequent. Four common categories in which conditionals are semantically sorted are: biscuit/speech act

conditionals, factual/premise conditionals, indicative conditionals and subjunctive/counterfactual conditionals (von Fintel, 2009).

Narayanan et al. (2009) proposed a different way of categorization. Categorization can also be done based on the form of the conditional. This is done by looking at the tense patterns of conditionals. The result of categorization in this manner is four types of conditionals with different verb pairs for the antecedent and consequent:

Type	Verb Tenses
0	Simple present → simple present
1	Simple present → modal + infinitive (future)
2	Past tense → modal + infinitive
3	Past perfect → modal + have + past participle

Table 1: English Conditional Types (Narayanan et al., 2009)

This type of categorization is more useful for this research since the verb tense of conditionals is the main focus. Also, categorization based on form/verb tenses instead of semantically is a lot easier when working with large amounts of data. To classify a conditional semantically requires more precision and takes more time compared to looking at the verb in a conditional and correctly annotating the sentence. More on this will be given in the methodology section.

The following paragraphs give examples of the different types of conditionals.

3.2.1 Type Zero Conditionals

In zero conditionals both the antecedent and the consequent are in the simple present in English. These conditionals mostly state facts.

(10) If you cool water to zero degrees, it freezes.

3.2.2 Type One Conditionals

Type one conditionals are used for events that can happen in the future. These conditionals usually express an intention or a possibility.

(11) If you go now, then you will still be on time.

In these conditionals it is not certain that the consequent is to happen, but it is likely. The consequent uses a modal verb (e.g., might, will and could) and an infinitive. The antecedent is in the simple present in English.

3.2.3 Type Two Conditionals

These conditionals usually express a wish or preference, that has a lower chance of being fulfilled. Second conditionals also contain the category of counterfactual conditionals. These conditionals describe a situation that is achievable, but unlikely to happen or something that could happen under different conditions:

(12) If I had a lot of money, I would donate some.

The consequent uses a modal verb (could, would or might) and most of the time talks about a hypothetical event. The antecedent is in the past tense. The usage of the past tense here is remarkable, as it does not signify an event that happened in the past. Fragment (12) contains the verb "had" in the antecedent, but it does not mean that the subject had a lot of money (in the past). The past tense is not used for semantically talking about the past.

This usage of the past tense is described in literature as the "fake past" or "fake tense" (Iatridou, 2000; Schulz, 2014; Mackay, 2015; Schulz, 2017).

3.2.4 Type Three Conditionals

Consider if the following sentence was spoken after finishing a football match:

(13) If you had scored, we would have been able to win.

The sentence outlines the possibility was there for the consequent to happen, in other words to win the game. But because the event in the antecedent did not happen, scoring a goal in this case, winning is no longer feasible. The antecedent is in the past perfect and the consequent consists of a modal verb + have + past participle in English.

3.3 Hypotheses

The categorization based on tense use is applied to form the following hypotheses:

Type zero conditionals in English have the verb tenses <simple present, simple present> (<antecedent, consequent>). They mostly portray facts and certainties. In Dutch, the simple present is called the *onvoltooid tegenwoordige tijd* (ott).

> H_0: Type zero conditionals are of the form <onvoltooid tegenwoordige tijd (ott), ott> in Dutch.

Type one conditionals in English are of the form <simple present, modal + infinitive>. They are used for events that will likely happen in the future.

> H_1: Type one conditionals are of the form <ott, modal verb + infinitief> in Dutch.

Type two conditionals in English contain the verb tenses <simple/perfect past, modal + infinitive>. They are usually express a wish or preference. In Dutch, the simple past is called the *onvoltooid verleden tijd* (ovt) and the perfect past is called the *voltooid verleden tijd* (vvt).

> H_2: Type two conditionals are of the form <ovt/vvt, modal verb + infinitief> in Dutch.

Type three conditionals in English include the verb tenses <past perfect, modal + "have" + past participle>. The Dutch translation of past participle is *voltooid deelwoord*.

> H_3: Type three conditionals are of the form <voltooid verleden tijd (vvt), modal verb + hebben + voltooid deelwoord> in Dutch.

These hypotheses result in the following table for the categorization of Dutch conditionals:

Type	Verb Tenses
0	ott → ott
1	ott → modaal werkwoord + infinitief (future)
2	ovt/vvt → modaal werkwoord + infinitief
3	vvt → modaal werkwoord + hebben + voltooid deelwoord

Table 2: Possible Dutch conditionals types

4 Methodology

This section focusses on the techniques that I used to find answers to my research question. First, I will talk about the corpus that I used, next I will talk about the extraction of Dutch conditionals from that corpus.

4.1 Europarl Corpus

The corpus from which I will be extracting conditionals from is the Europarl Corpus. This corpus was created by Philipp Koehn (2004). It contains proceedings from the European Parliament, which have been translated into several European languages. All these proceedings have been scraped from the European Parliament site. The data for a single day and language is stored in a separate file. Next, sentences in these files were split and tokenized. Finally, sentences were aligned using the Gale and Church method(Gale and Church, 1993), which looks at the length of sentences for alignment.

The corpus contains two different types of files: sentence alignment files and tokenized corpus files. Tokenized corpus files contain the proceedings from a

```
<s id="15">
 <chunk type="NP" id="c5-1">
  <w hun="PRP" tree="PP" lem="I" pos="PRP" id="w15.1">I</w>
 </chunk>
 <chunk type="VP" id="c5-2">
  <w hun="MD" tree="MD" lem="would" pos="MD" id="w15.2">would</w>
  <w hun="VB" tree="VV" lem="like" pos="VB" id="w15.3">like</w>
 </chunk>
 <chunk type="NP" id="c5-3">
  <w hun="PRP$" tree="PP$" lem="your" pos="PRP$" id="w15.4">your</w>
  <w hun="NN" tree="NN" lem="advice" pos="NN" id="w15.5">advice</w>
 </chunk>
 <chunk type="PP" id="c5-4">
  <w hun="IN" tree="IN" lem="about" pos="IN" id="w15.6">about</w>
 </chunk>
 <chunk type="NP" id="c5-5">
  <w hun="NNP" tree="NP" lem="Rule" pos="NNP" id="w15.7">Rule</w>
  <w hun="CD" tree="CD" lem="@card@" pos="CD" id="w15.8">143</w>
  <w hun="VBG" tree="VVG" lem="concern" pos="NN" id="w15.9">concerning</w>
  <w hun="NN" tree="NN" lem="inadmissibility" pos="NN" id="w15.10">inadmissibility</w>
 </chunk>
 <w hun="." tree="SENT" lem="." pos="." id="w15.11">.</w>
</s>
```

Figure 1: Tokenized sentence from the Europarl Corpus

single day in Parliament in a single language. The sentence alignment files contain information that is needed to align sentences between two languages in the corpus files. An example of a sentence from a tokenized corpus file can be seen in Figure 1.

4.2 PerfectExtractor

To extract conditional sentences from the Europarl Corpus a script called the PerfectExtractor[3] was used. This was initially created by the Time in Translation project to extract sentences in the perfect tense from corpora. Adjustments to the code have made it possible to also extract conditionals from the corpus. For this research, any sentence containing an "als" from the corpus was extracted. Extracting sentences containing "als" provides enough data to work with. Therefore, it is not necessary to search for any of the other conditional conjunctions that were discussed in 3.1.1. The extraction resulted in a list of sentences containing "als", but not all these sentences were conditionals. An example of such a case was also shown in 3.1.1. I manually filtered out and deleted any sentences that were not conditionals. The filtering out of conditional sentences could have further been improved by e.g. looking for a "dan" further in the sentence or removing any sentences that do not have at least two clauses.

It is important to note that not every conditional in Dutch is translated into a conditional in English:

[3]https://github.com/UUDigitalHumanitieslab/perfectextractor

Figure 2: Example of an annotation in TimeAlign

> (14) Als wij als enige indicator het bruto binnenlands product per hoofd van de bevolking nemen, kunnen wij ons vergissen.

Fragment (14) is a Dutch sentence from the Europarl corpus containing a conditional. The English translation of this sentence is:

> (15) We might be mistaken in using the gross domestic product per inhabitant as the sole indicator.

As you can see, the English translation does not contain a conditional. Cases where the translation did not contain a conditional or where there was no English translation available were removed.

4.3 TimeAlign

The data is generated using the sentences that were annotated in TimeAlign. An example of a single annotated sentence can be seen in Figure 2. In this particular example, the antecedent of the Dutch sentence is in the ott and the consequent is an infinitive, with *kunnen* as a modal verb. All this information is manually annotated, and the antecedents are highlighted in green. In English, the antecedent is in the simple present and the consequent is an infinitive, in this case "may" is the modal verb.

5 Results

In this part of the paper, I will display my results. There are results for both directions (NL→EN, EN→NL).

5.1 Dutch → English

These next figures display an overview of all the annotated sentences, in the direction Dutch to English. This means that Dutch conditional sentences were

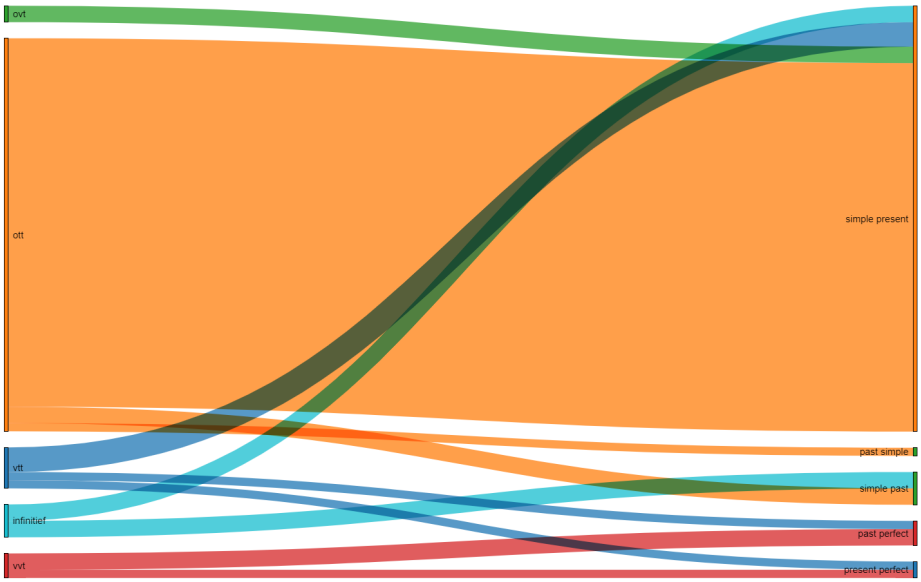

Figure 3: Sankey diagram of Dutch to English verb tenses of antecedents

extracted from the corpus and the sentences together with their English translations were used to generate the following data. The total amount of sentences annotated in this direction is 62.

Figure 3 shows the antecedent tenses of the annotated Europarl sentences. On the left the tense of the Dutch sentences is shown and, on the right, the corresponding English tense is displayed. A large part, around 80%, of the antecedents is in the *onvoltooid tegenwoordige tijd* ott) and translated into the simple present. Only a small part of the antecedents contains a different tense. In Figure 4 both parts of the conditionals can be seen in the form of a tuple: <antecedent, consequent>. This also shows the high number of antecedents in the ott and translated into the simple present.
Figure 5 displays the tuple frequencies for the data. On the left the Dutch tuples can be seen and tuples for the English translations are on the right.
In Figure 6 a semantic map of the tenses used is shown. Each dot represents a tuple pair, with the first tuple being for a Dutch conditional and the second for an English conditional into which the sentence was translated. For example: {<simple present, simple present>, <ott, ott>}.

These maps are interactive when accessed on the web, and when hovering over a dot more information is shown. For example, the bottom orange dot represents 8 sentences (fragments) where the tense of the antecedent is ott and the consequent tense is infinitief. The dot has been given the color orange because when it was translated into English, the tenses <simple present, simple present> were

Tables per language

Dutch

Label	Count
<ott,infinitief>	30
<ott,ott>	17
<vtt,infinitief>	5
<infinitief,infinitief>	3
<infinitief,ott>	1
<ott,vtt>	1
<vvt,ovt>	1
<vvt,vvt>	1
<ovt,infinitief>	1
<vvt,infinitief>	1
<ovt,ott>	1

English

Label	Count
<simple present,infinitive>	33
<simple present,simple present>	17
<simple past,infinitive>	3
<simple present,simple past>	1
<past perfect,simple present>	1
<past perfect,infinitive>	1
<present perfect,simple present>	1
<simple past,simple present>	1
simple present	1
<past perfect,perfect>	1
<past simple,infinitive>	1
<present perfect,infinitive>	1

Figure 4: Total count of possible verb combinations shown in tuples

Tuple frequencies

Off Show colors Off Show categories On Show labels

Search:

Dutch	English	Count
<ott,infinitief>	<simple present,infinitive>	18
<ott,ott>	<simple present,infinitive>	11
<ott,infinitief>	<simple present,simple present>	8
<ott,ott>	<simple present,simple present>	6
<vtt,infinitief>	<simple present,infinitive>	2
<infinitief,infinitief>	<simple past,infinitive>	2
<infinitief,ott>	<simple present,simple present>	1
<ott,vtt>	<simple present,simple past>	1

Figure 5: Total count of possible tuple combinations

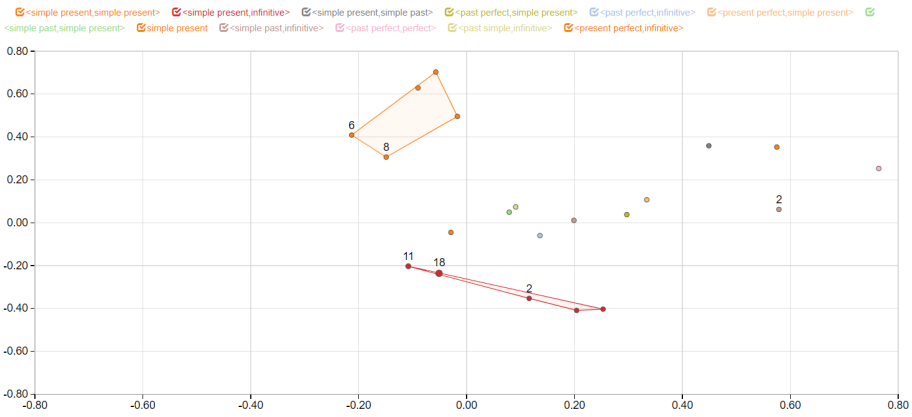

Figure 6: Semantic map of Dutch conditionals

used. The legend for the meaning of the other colored dots is visible above the graph.

In Figure 6 two groups of dots a clearly visible. The orange cluster representing sentences that were translated into <simple present, simple present> and the red cluster representing sentences translated into <simple present, infinitive>. The lines between the dots form an area which covers the contexts that are translated into a tuple pair. A larger distance between any two dots means a greater difference in the context in which these tenses are used.

When looking at the multiple orange dots, we can see that the simple present is used to translate several types of Dutch tenses, this was also visible in Figure 3. For instance, take the following two sentences:

> (16) Het zou eerlijker zijn als de Commissie de Staalsteuncode strikt naleefde.
>
> (17) In the interests of equality, there are grounds to either apply the Steel Aid Code strictly or modify the Code if the Commission wishes to authorise types of aid other than those legally acceptable at the moment.

The Dutch sentence in the consequent uses the ovt, a past tense, while the English sentence uses the simple present.

In Figure 6, the orange area contains the Dutch sentences that were in the <infinitief, ott>, <vtt, infinitief>, <ott, infinitief>, <ott, ott> and <ovt, ott> and translated into <simple present, simple present>.

5.2 English → Dutch

These next figures display an overview of the annotated sentence in the direction English to Dutch, so in the opposite direction of the previous results. These sentences were not annotated or extracted by me but were already available from the Time in Translation website. The total amount of annotated sentences in this direction is 746.

Figure 7 displays a Sankey diagram of the antecedent verb tenses in English and to which Dutch tense they have been translated to. Around 50% of the verb tenses are in the simple past, 43% in the past perfect and 6% in the simple present. Most antecedents in the simple present get translated into the ott in Dutch. In the past perfect most tenses get translated into the vvt. Table 3 shows the distribution of the translations of verb tenses for the simple past.

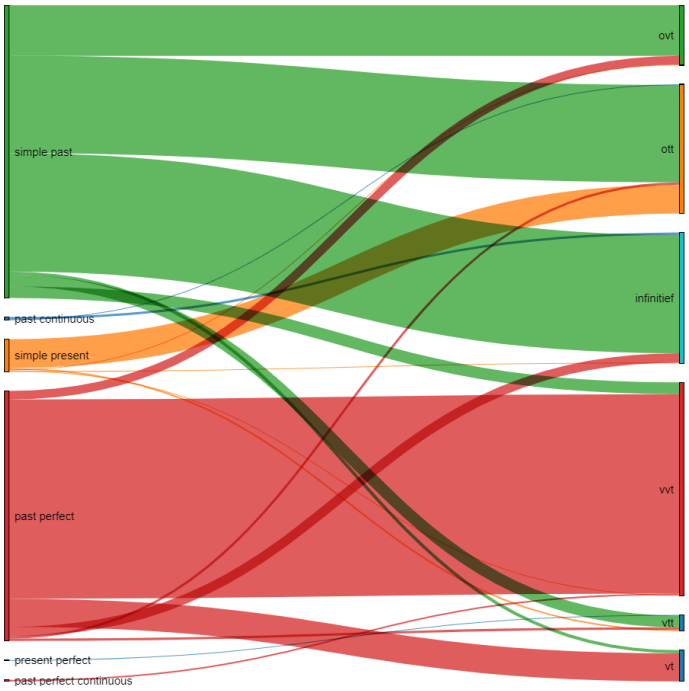

Figure 7: Sankey diagram of English to Dutch verb tenses of antecedents

In Figure 8 the total count for each combination of tenses can be seen in tuples. Figure 9 displays the tuple frequencies for the data.
Finally, a semantic map similar to Figure 6 can be seen in Figure 10, but for the direction EN→NL.

Verb tense	Percentage
ovt	17%
ott	33%
infinitief	40%

Table 3: Distribution of simple past translations

Tables per language

English

Label	Count
<simple past,infinitive>	332
<past perfect,perfect>	260
<past perfect,infinitive>	52
<simple present,infinitive>	19
<simple past,simple present>	17
<simple present,simple present>	16
<simple past,simple past>	12
<simple past,perfect>	11
<past perfect,simple present>	8
<simple present,simple past>	5
<simple past,present perfect>	3

Dutch

Label	Count
<infinitief,infinitief>	143
<vvt,vvt>	126
<vvt,vt>	118
<ott,infinitief>	82
<ott,ott>	75
<ovt,infinitief>	53
<vt,vt>	27
<vvt,infinitief>	27
<infinitief,ott>	14
<vt,infinitief>	9
<ovt,vt>	8

Figure 8: Total count of possible verb combinations shown in tuples

Tuple frequencies

Off Show colors Off Show categories On Show labels

Search:

English	Dutch	Count
<simple past,infinitive>	<infinitief,infinitief>	131
<past perfect,perfect>	<vvt,vvt>	106
<past perfect,perfect>	<vvt,vt>	102
<simple past,infinitive>	<ott,infinitief>	63
<simple past,infinitive>	<ovt,infinitief>	49
<simple past,infinitive>	<ott,ott>	49
<past perfect,perfect>	<vt,vt>	23
<past perfect,infinitive>	<vvt,infinitief>	20

Figure 9: Total count of possible tuple combinations

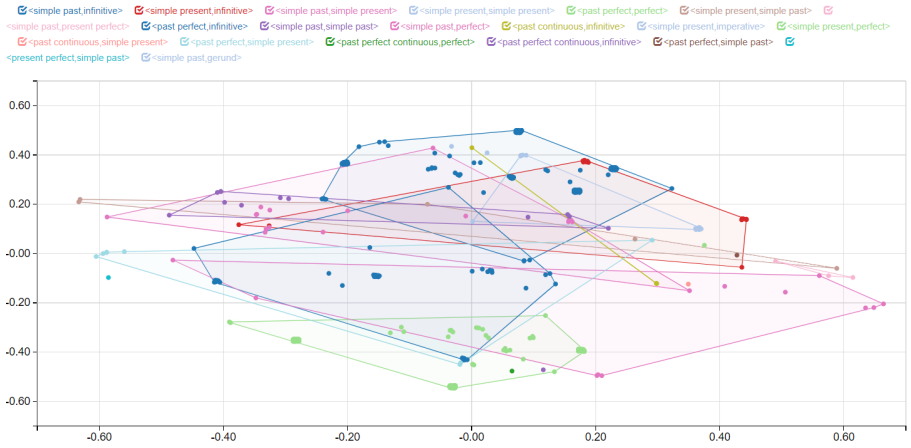

Figure 10: Semantic map of English conditionals

6 Discussion

6.1 Hypotheses

6.1.1 H_0

To draw a conclusion concerning the research question I will first go back to the hypotheses.

> H_0: Type zero conditionals are of the form <onvoltooid tegenwoordige tijd (ott), ott> in Dutch.

The tuple frequency tables from the previous chapter show that most type zero conditionals in English (of the form <simple present, simple present>) get translated into Dutch conditionals of the form <ott, ott> or <ott, infinitief>. The form <ott, infinitief> contains a modal verb in the consequent.

Translation of a <simple present, simple present> conditional into a <ott, ott> conditional is the most popular choice, and seems to be the most sensible.

> (18) If, in addition, we consider the competition which came from television during the 1970s, it is not hard to see why hundreds, thousands of cinemas have been closed for so long.
>
> (19) Als wij voorts kijken naar het effect van de concurrentie van de televisie vanaf de jaren zeventig, dan begrijpen wij meteen waarom duizenden bioscopen zo lang dicht zijn gebleven.

Both the English sentence (18) and its Dutch counterpart (19) give the impression that the consequent of the sentence is a foregone conclusion. As has been

mentioned before, certainties, factual statements and rules are characteristic type zero conditionals. That is not to say that these conditionals are genuine facts, sometimes the speaker merely wishes to present the statement in such a way.

More noteworthy are the <simple present, simple present> conditionals which are translated into <ott, infinitief>. This tuple pair of Dutch verb tenses corresponds with the English tuple pair <simple present, infinitive> rather closely, which is a type one conditional. The question is whether type zero conditionals in Dutch can also be of the form <ott, infinitief> or if the translation is of a different conditional type.

> (20) This is an extremely dangerous method which risks seriously damaging the surrounding environment if the dam bursts, as was the case in Romania.
>
> (21) Dat is een uiterst gevaarlijke methode die het omringende milieu ernstige schade kan toebrengen als de dam rond het bassin doorbreekt, wat in Roemenië ook is gebeurd.

When strictly looking at the verb forms, the Dutch sentence (20) is translated in a notably different manner when compared to the English sentence (21). "Kan toebrengen" is a combination of an infinitive and a modal verb which translates directly to "could inflict" in English. In the English sentence no such modal verb is present. At first glance, it seems like the Dutch consequent clause contains more uncertainty than the English consequent, where the simple present is used. This is possibly not be the case though. The modality may be hidden in the word "risks" in (20).

All translations are made by professional translators, but translating is not an exact science. Translation is done with the goal of keeping the meaning of the original sentence the same across the different languages spoken in the European Parliament. In the case of sentences (20) and (21) only the Dutch sentence contains a modal verb, but both sentences seem to portray a similar level of uncertainty. This creates a difficult situation where if conditionals are strictly separated by form/verb tense these conditionals would be in different categories. However, when looked at semantically the sentences could fit in the same category.

When analyzing the tuple frequency table for the other direction (NL→EN) the same findings can be concluded. Most of the English sentences of the form <simple present, simple present> (which are type zero conditionals) are either translated into <ott, ott> or <ott, modal verb + infinitief>.

To conclude, it is difficult to give a definitive answer. As has been explained earlier, classification is done in this paper based on the form/verb tenses used in conditionals. When looking at the tuple tables of the results section it can

be concluded that most English type zero conditionals correspond with Dutch sentences of the form <ott, ott>, but not all.

The differences in verb tenses between English and Dutch may mean that type zero conditionals can also be of the form <ott, infinitief> besides the <ott,ott> form. It is also possible that the translators chose to add a modal verb + infinitief in the consequent because this translation stays closer to the original semantics of the sentence, as can be seen in (20).

The overlap of type zero conditionals being translated into type one conditionals and vice versa is not totally unsurprising, as the meaning of these conditionals and the contexts in which they are used are more similar compared to type two or type three conditionals.

The common use of the verb forms <ott, ott> for Dutch type zero conditionals supports the hypothesis H_0, with other forms only being of limited presence.

6.1.2 H_1

Now I will look at my second hypothesis:

> H_1: Type one conditionals are of the form <ott, modal verb + infinitief> in Dutch.

To judge the validity of this hypothesis the tuple pair <simple present, infinitive> in the tuple frequency tables is important, as this is the form of a type one conditional in English. Most type one conditionals in English get translated into the Dutch verb forms <ott, infinitief> or <ott, ott>. This is also true for the direction NL→EN.

The same situation as in 6.1.1 appears to occur here. Often a sentence with a modal verb is translated directly into a sentence containing a modal verb:

> (22) If this comes to nothing, then the government has to step in.
>
> (23) Als die inzet faalt, moet een overheid ingrijpen.

But other times the translator has made a choice to either omit an explicit modal verb in the translation, or in the other way around, added a modal verb to the translation. The following English sentence (22) contains the modal verb "can" while no direct modal verb can be found in its Dutch translation:

> (24) But the report before the House tonight is a prime example of how, if we are not very careful, we can produce very grandiose-sounding ideas that lack the substance to make them relevant to the people who benefit directly from them.

(25) Als wij echter niet oppassen, dreigt het onderhavige verslag neer te komen op de zoveelste opsomming van prachtige denkbeelden, die echter niet voldoende houvast bieden om concrete resultaten op te leveren voor de mensen die ervan zouden moeten profiteren.

Although, in this case, the verb "dreigt" (threatens) could be the word that conveys the modality that is missing from the Dutch consequent clause. The same phenomenon was seen in (20) and (21).

Concluding, it looks like the hypothesis H_1 holds. But other verb forms may also be possible when the modality is not explicitly wrapped in a modal verb, or at least not a verb that is marked a modal verb during the annotation progress.

6.1.3 H_2

H_2: Type two conditionals are of the form <verleden tijd (vt), modal verb + infinitief> in Dutch.

This hypothesis requires us to look at the type two conditionals in English, which are defined as having the following form: <past tense, modal + infinitive>. The "past tense" here can either be the verb tense simple past or past perfect. For the direction EN→NL these two tenses are by far the most prevalent.

The most common translation of these sentences is <infinitief, infinitief> in Dutch, with the translation of the antecedent into ott or ovt also being popular. The reason for this spread of verb tenses used in the translations is that the English sentences contain the "fake past" tense. The fake past tense was also covered in the Theoretical Background. An example will make the problem more apparent:

(26) If that were the case, we would again be making incredible fools of ourselves today.

(27) Als dat namelijk het geval zou zijn, dan zouden we ons vandaag opnieuw onsterflijk belachelijk maken.

In English the past subjunctive is used, of which the conjugation rules are almost the same as the simple past except for the verb "to be", it uses "were" instead of "was". In Dutch, a modal verb (*zou*) is used along with an infinitive. The following two sentences are variances of (27), but these in these sentences the verb tense is in the ott and ovt:

(28) Als dat namelijk het geval is, dan zouden we ons vandaag opnieuw onsterflijk belachelijk maken.

(29) Als dat namelijk het geval was, dan zouden we ons vandaag opnieuw onsterflijk belachelijk maken.

The semantic meaning of these sentences lies very close to the original meaning of (27). This explains why there are a good number of antecedents translated into ott or ovt.

Some sentences were also translated into the form <ott, ott>, removing the modal verb from the consequent. Again, the meaning of this variance is close to the sentence containing a modal verb in the consequent (27).

> (30) Als dat namelijk het geval is, dan maken we ons vandaag opnieuw onsterflijk belachelijk.

All these different counterfactuals have similar, but not identical semantic values. Most of these differences are pointed out by Nieuwint (1984).

To sum up, the type two conditionals in English get translated into mostly <infinitief, infinitief>, <ott, infinitief>, <ovt, infinitief> and <ott, ott> in Dutch. This is not inline with the hypothesis H_2.

6.1.4 H_3

> H_3: Type three conditionals are of the form <voltooid verleden tijd (vvt), modal verb + "hebben" + infinitief> in Dutch.

It is difficult to single out the type three conditionals in English, because the consequent has the form <modal + have + past participle>, and any clause containing a modal verb is annotated as an infinitive.

Unfortunately, it is impossible to give a clear judgement on this hypothesis because of the limited number of sentences containing a type three conditional. This might be because the Europarl corpus contains proceedings that are spoken in a similar formal context. More on this is in the points of improvement below.

6.2 Points of Improvement

The sentences spoken in the European Parliament tend to be more complex and formal than regular sentences used in everyday dialogue. This fact, together with the manual annotation of the fragments may have caused some of the sentences to be annotated incorrectly.

Another point where this research could be improvement is with the addition of other corpora, not just the Europarl corpus. Europarl contains a specific formal type of dialogue spoken in the European Parliament. Including different will give more variety in the type and tone of the fragments. This could improve the number of conditionals for each of the conditional types which can be researched, this was for example a limiting factor in 6.1.4.

Lastly, translators might have made mistakes or at least could have been less precise in some translations. This could have influenced some of the results.

7 Conclusion

In this paper research has been done to answer the following research question: "How does the tense use in conditional sentences compare between the languages English and Dutch?"

The Europarl corpus was used to analyze the verb tenses of conditionals between English and Dutch. The results of this analysis show that for some types of conditionals English verb tenses are directly translated into the corresponding Dutch tense, for example for type zero conditionals. The simple present in the antecedent and consequent is directly translated into the *onvoltooid tegenwoordige tijd*. Type one conditionals are frequently of the form <ott, modal verb + infinitief> in Dutch. Type two conditionals, which includes the category of counterfactuals, can be translated into several verb forms. This is most likely because of semantic differences between the forms being minimal. It is too difficult to give a satisfying conclusion about type three conditionals because of the used dataset.

References

Boogaart and Verheij (2013). "Als dát geen insubordinatie is! De pragmatiek van zelfstandige conditionele zinnen".

de Jonge, Alma (2011). "Zolang-ie maar conditioneel is! Een onderzoek naar het conditionele gebruik van zolang".

Gale, William A. and Kenneth W. Church (1993). "A Program for Aligning Sentences in Bilingual Corpora". *Computational Linguistics* 19.1, pp. 75–102. URL: https://aclanthology.org/J93-1004.

Iatridou, Sabine (2000). "The Grammatical Ingredients of Counterfactuality". *Linguistic Inquiry* 31.2, pp. 231–270. ISSN: 00243892, 15309150. URL: http://www.jstor.org/stable/4179105.

Koehn, Philipp (Nov. 2004). "EuroParl: A parallel corpus for statistical machine translation". 5.

Loosen, Gert (2015). "'Mits': oud en stout : een nieuw leven voor voorwaardelijk 'mits'". 29, pp. 35–42. URL: https://digilib.phil.muni.cz/bitstream/handle/11222.digilib/134983/1_BrunnerBeitratgeGermanistikNordistik_29-2015-2_5.pdf.

Mackay, John (Sept. 2015). "Actuality and fake tense in conditionals". *Semantics and Pragmatics* 8.12, pp. 1–12. DOI: 10.3765/sp.8.12.

Narayanan, Ramanathan, Bing Liu, and Alok Choudhary (2009). "Sentiment analysis of conditional sentences". English (US). Copyright: Copyright 2020 Elsevier B.V., All rights reserved.; 2009 Conference on Empirical Methods in Natural Language Processing, EMNLP 2009, Held in Conjunction with ACL-IJCNLP 2009 ; Conference date: 06-08-2009 Through 07-08-2009, pp. 180–189. DOI: 10.3115/1699510.1699534.

Nieuwint, P.J.G.M. (1984). "Werkwoordstijden in Nederlandse 'counterfactuals'". *De Nieuwe Taalgids. Jaargang 77*, pp. 542–555.

Schulz, K. (2014). "Fake Tense in conditional sentences: a modal approach". *Natural Language Semantics* 22.2, pp. 117–144. ISSN: 0925854X, 1572865X. URL: http://www.jstor.org/stable/23748494.

— (2017). "Fake Perfect in X-marked Conditionals". *Semantics and Linguistic Theory* 27, pp. 547–570.

van Bree, Cor and Jeroen van der Hee (2002). "Conditioneel al". 15–16, pp. 158–183. URL: https://www.aup.nl/journal-downloads/taal-en-tongval/jrg_2002-2003_thema_1516_-_conditioneel_al.pdf.

von Fintel, Kai (2009). "Conditionals". URL: http://web.mit.edu/fintel/fintel-2009-hsk-conditionals.pdf.