REPORTING VERBS: A STUDY ON THE BEHAVIOUR OF DUTCH, ENGLISH, GERMAN AND FRENCH PERFECTS.

Aron Theunissen (4062876)
Supervisors: Henriëtte de Swart, Martijn van der Klis
Second Corrector: Onbekend
Date: 21 November 2017
Bachelor: Kunstmatige Intelligentie, Universiteit Utrecht
Bachelorthesis 7,5 ECTS

Contents

1 Introduction

Artificial Intelligence is a broad area of study with many possible subjects. Having undertaken several linguistic courses and a minor in linguistics, I had acquired a knowledge in the linguistic field and a fascination with the subjects covered in these courses. These influenced my decision to undertake research on a linguistic subject for this thesis.

The difficulty faced was in selecting one specific subject out of all of interest. My decision to investigate computational linguistics stemmed from my discussion with the degree program advisor, in which I was told Henriëtte de Swart was doing a project on said subject. Upon communication with the project leaders, this was found to be a suitable subject for research.

At the initial meeting, I was updated on the current status of the project and I had to be assigned a fitting subject. This led us to reporting verbs, which we believe to be an interesting aspect for the project. My contribution to the project is thus by researching reporting verbs from the EUROPARL corpus in English, Dutch, French and German. Reporting verbs are verbs used to report that something was stated, like 'say', 'ask', 'tell' and 'promise'. The perfect form of verbs in these languages is formed in a manner that resembles one another to a large degree, with each having an auxiliary verb (to be/to have) with a past particle. For example, in the sentence "I have known him for a long time.", 'have' is the auxiliary verb and 'known' is the past particle.

In addition, this research follows from the Time in Translation project in which reporting verbs were not examined. An objective of this study is thus to prove that there are differences between reporting verbs and verbs in general and to compare the results to other studies conducted within formal and informal language use.

In this research, several programs were employed to collect, connect and compare reporting verbs from the EUROPARL corpus. Artificial Intelligence is an essential and crucial tool in accurately processing data collected. Algorithms were used to extract the perfects from the EUROPARL corpus and to link them up with corresponding contexts from the other languages so as to create the semantic maps. Results uncovered an aspect of the mechanism behind the perfect form that can lead to a better understanding of and possible better algorithms for translating verbs and their tenses, thus making machine translations more intuitive, precise and effective.

2.1 Previous research

My research is an expansion of the recently started Time in Translation project (http://time-in-translation.hum.uu.nl/). This project analyses the behavior of the perfect within and between different languages. Different languages differ in their use of the perfect. For example:

1) John has lived in Maastricht for 7 years now.
2) Jan woont nu 7 jaar in Maastricht.
3) The votes are in. Donald Trump has been elected as the new president.

In sentence (1), a present perfect is used in English, but the Dutch translation in sentence (2) uses an *onvoltooid tegenwoordige tijd (ott*). This means that for a continuative action, English can use a perfect where Dutch cannot. The perfect can be used in different ways, at the level of a sentence as well as at the level of conversations. The perfect can be used to indicate the past as well as the present, so there is a difference in reference time of the perfect. In sentence (3), for example, the perfect used refers to something that happened in the past, but has relevance in the present. While sentence (1) refers to a continuous period starting in the past and lasting till the present. This has been researched before, but mostly in qualitative research. In Time in Translation they use a combination of both qualitative and quantitative research to investigate the perfect across languages.

The Time in Translation project has started with the behavior of the perfect in the formal usage of languages at the European parliament. All their meetings are digitalized in every European language. These digitalized meetings are registered, part-of-speech-tagged and lemmatized in the EUROPARL corpus (Tiedemann, 2012). This means that every word used in this corpus has been labeled according to their lexical category and their lemma. Thanks to this information, finding a perfect in these texts is manageable. More information about the algorithm that makes this possible is given in the section "Methodology". The interesting part about this corpus is that even though translations are supposed to have the same meaning, the tenses of the used verbs across translations don't have to correspond with each other. This has been useful for the project Time in Translation. The project managed to find out that the perfect is mostly used when referring to an event in the past, as shown in sentence (3). This proved their hypothesis, since most meetings at the European parliament are about problems that took place in the past.

This research will focus on reporting verbs. It will research the tenses of several reporting verbs used in the EUROPARL corpus and how these tenses are used in different languages.

2.2 Literature

Van der Klis et al. (2007) researched the perfect form using data extracted from multilingual parallel corpora to generate semantic maps. This was done in five European languages, namely German, English, French, Dutch and Spanish. Firstly, perfects were extracted from the EUROPARL corpus by using an algorithm. A human annotator then marked the corresponding verb phrases in the aligned fragments in the languages before tenses were assigned to these verb phrases. This tense attribution process returned five-tuples of aligned tense attributions, every tuple representing one language. These tuples were made into a similarity matrix, which was turned into a plot by using multidimensional scaling. Using these methods, it was confirmed, for example, that English and Spanish require a past form with a locating time adverbial, whereas German, Dutch and French tolerate a perfect form. One of their findings was that the French *passé composé* had a wide range of perfect uses. German and English might use a simple past for various contexts whereas French would use the *passé composé*. Dutch had less contexts in which the simple past was used.

Building on these findings, Swager (2017) has carried out research on the usage of the perfect in Dutch, French, Germand and English literature. The perfect of these languages is formed in the same manner, an auxiliary verb + a part particle. For instance: "*John has arrived."* 'has' is the auxiliary verb and 'arrived' is the past particle. This is the *present perfect*, the perfect for English. In Dutch this would be the *voltooid tegenwoordige tijd (vtt)*, in French the *passé composé*, in German the *Perfekt* and in Spanish the *pretérito perfecto compuesto*.

The text used in this research is L'Étranger by Albert Camus. Occurrences of the French *passé composé* were extracted and the verbs linked with one another as in the Time in Translation research. Semantic maps were made from this data and analyzed. The hypotheses were that the *passé composé* would be translated into a *Perfekt* in German. In Dutch, it would be either a *vtt* or an *ovt*. In English it would be either a *present perfect* or a *simple past*. There was no expectation for Spanish, since the knowledge before the research was too little to make any assumptions. These hypotheses were right with the addition of Spanish behaving mostly like English.

Schaden (2009) claims that in languages like Spanish and English, using the *simple past* tense is almost always possible, and sometimes you must. Languages like German and French differ in that using the *present perfect* is almost always possible, and sometimes mandatory. This signals that the usage of the perfect would be similar between Spanish and English, and between German and French.

Nishiyama and Koenig (2010) explain that there are several different types of perfects. Consider these following examples and the examples given in sentence (1) and (3) from the previous section.

4) I won't be at your presentation tomorrow. I've caught the flu.
5) - Where is Naomi? – Naomi has been hit by a car.

Sentence (4) is called a *Resultative perfect*. This means that something from the past results in something in the present. In sentence (4), the flu is causing the person's inability to attend the presentation. Catching the flu happened in the past, while the moment this sentence is uttered is the present. The flu is also still present. Most of these types are *entailed resultative perfects.* This means that the cause and the result are given in the sentence. Sentence (5) is called an *Existential perfect*. This kind of sentence only shows that something did happen in the past and bears current relevance. Sentence (1) is a *Continuative perfect*. A continuative perfect is used to show that a certain event started

Commented [ST1]: Is it a Dutch thing to use active voice in thesis (such as this sentence), cuz I would suggest using passive voice instead :) So it becomes:
Occurrences of the French *passé composé* were extracted and the verbs linked with one another as in the Time in Translation research.

also note:
"each other" is used when only two things are referred to; more than that use "one another"

in the past and is still going on. As in sentence (1), John has lived in Maastricht for seven years and still does. This means that this kind of sentence show a certain interval starting in the past and is still ongoing. The last type of perfect is shown in sentence (2), this is called a *'hot news' perfect.* These are the least common type of perfects, only being used in sentences which reflect 'hot news'. This is a perfect used for an event in the recent past. 'Hot news' perfects can be reduced to existential perfects according to Nishiyama and Koenig, so they won't be further explained here.

The biggest difference between the three types of perfects we focus on is the reference time they give information about and if there's any information about the resultant state. While not every study agrees on the difference between a resultative and an existential perfect, a continuative perfect is defined in more or less the same manner in every study. The definition of continuative perfects is that the input state of the sentence is continuing. Most perfects are either entailed resultative perfects or continuative perfects, but there are more kinds. For instance, speech-act/epistemic perfects.

There are two kinds of subtypes for this speech-act/epistemic perfects. The first subcategory is for evidential uses and the second subcategory is for topic negotiation. The first one is used by speakers or authors to communicate that what follows the reporting verb presently holds or is likely to hold in the future, as seen in sentence (6).

6) Donald Trump has promised that he will build a wall between Mexico and the U.S.A.

Promised is a reporting verb and, if we would consider Donald Trump to be trustworthy, the rest of the sentence is true. This is something that holds for all the reporting verbs. If somebody says or promises something, it is likely to become true or has been true, and therefore it is likely to hold.

7) a. Mark loved games
   b. Mark has said that he loved games.
   c. Mark has forgotten that he loved games.
   d. Mark said that he loved games.

When changing direct speech into reported speech, a reporting verb is used, as seen in the transition from sentence (7a) to (7b). This can be done using the verbs 'say', 'tell', 'ask' or 'promise' for example. However, these are not the only reporting verbs possible. Every verb that can be used to report what someone said is a reporting verb. 'Warn' and 'advise' are two examples of that. There are dozens of verbs that would fit into this category, some more clearly than others. For example, 'forget' could be used in a reporting manner, but that might be a little bit vague. As in sentence (7c), whether this state is reported by Mark is not clear.

As can be seen in sentence (7d), reporting verbs can be used with a simple past as well. The difference between this sentence and sentence (7b) is explained by the differences between a present perfect and a past simple in general. In sentence (1), a present perfect is used to show that an unfinished action started in the past and continues in the present. If a simple past was used in this sentence, the action would have been finished. In sentence (5), a present perfect is used to show that an action finished during someone's life and the person is still alive. If this sentence had a simple past, the person would be dead. In sentence (4), a present perfect is used to show that an action is finished with a result in the present. A simple past would have implied that the finished action had no result in the present. The fourth and final reason a present perfect could be chosen over a simple past is that an unfinished time

word is used, like this week or today. Last week and yesterday would imply that the time the event occurred in is finished, and then a simple past is used.

Nishiyama and Koenig their monosemous account of the semantics of the perfect, meaning that they only have one meaning, leaves a significant part of its interpretation to pragmatics To fill the gap within the constraints of their inability to assess the types of rules, speakers may or must typically use pragmatic inferences relevant to the interpretation of the perfect. A corpus study of over 600 English *present perfect* examples from a diverse range of genres (narrative texts, discussions, newspapers and conversations) was conducted. Most of the examples were either entailed resultative perfects or continuative perfects, which determined which inference rules must have been used.

2.3 Research questions

This study attempts to examine the differences in perfect usage between the languages through answering the following questions:

- Will the perfects of reporting verbs in the corpus used in English, Dutch, German and French translate to perfects in other languages?
- After getting semantic maps, will the data cluster in these semantic maps?
- Will the semantic maps from the EUROPARL corpus done by van der Klis et al. be similar to the results from this research?

2.4 Hypotheses

Research hypotheses were formulated using information from the literature review. Firstly, French and German are expected to use the perfect in the same manner in most of the sentences. This is according to the theory from Schaden that German and French have similar usage for the perfect. This is also backed up by the finding of Van der Klis et al. that German and French have more similarities in using perfects. Since English is not constrained to using the perfect, a lower number of perfects is expected. Dutch lies somewhere between these two combinations. As stated in Swager's thesis, Dutch is often incapable of using a perfect in a story, but on the other hand it can use a perfect in other situations whereas English couldn't, as seen in sentence (8a) and (8b). As we will use the EUROPARL corpus, which doesn't contain a lot of stories, the expectation is that Dutch will be closer to German and French than to English.

8) a. Ik heb iets gevonden
   b. I found something.

Since the expectation is that the Dutch *vtt,* the French *passé composé* and the German *Perfekt* are similar in use across the largest portion of sentences not told in a story, the expectation is that these languages will have similar results. As English has such different rules for when the *present perfect* is used, English will most likely be the most different.

All the sentences with a specific time determination will be translated to a *passé composé* in French, a vtt in Dutch and a *present perfect* in English, thus it is most likely these three will cluster together as well. Since the French *passé composé* can also be translated to a Dutch *ovt* and an English *simple past*, these are expected to similarly cluster at the French semantic maps. Other clusters might also appear. The biggest difference between the languages is hypothesized to be between English and the other three languages.

Since there is no real reason to think that the results from this experiment would differ from the results from Van der Klis et al., the expectations are that the results will be the same.

2.5 Relation to Artificial Intelligence

Knowledge and skills gained in the Bachelor of Artificial Intelligence program were befitting of this research, allowing for accurate processing of data collected. Firstly, the expansive exposure to linguistic elements in courses attended imparted relevant and necessary knowledge on content as well as research techniques. These increased the ease with which research articles were understood and facilitated research procedures. Secondly, the acquired skills from working with several computer programs and how algorithms work made it easier to understand the steps to efficient data gathering for this research. The combination of these two subjects were of great importance for this study, which are both essential in Artificial Intelligence.

3.1 Methodology

The definition and function of reporting verbs were first obtained on two sites that explicitly explained what some reporting verbs are and when they are used. Firstly, Perfect English Grammar (http://www.perfect-english-grammar.com/reporting-verbs.html) wrote down fifteen reporting verbs and when they are used in a reporting manner. Since these were explained well and are often used verbs, these fifteen were used as reporting verbs for this experiment. Secondly, Education First (http://www.ef.com/english-resources/english-grammar/reporting-verbs/) has a list of about 50 reporting verbs among which are the fifteen already defined. Among the others, a lot were not usable. For instance, the verb 'see' is on this list. Using this verb as a reporting verb seems possible, but in most circumstances it is not. This can be seen in sentences (9a) and (9b)

9) a. I will see if that's true.
   b. I have seen a movie.

10) The prime minister has doubted that this bill will pass.

The above type of verbs was not added to our list for the above-mentioned reason. A second problem lies within the truth value of a report. In sentence (10), the truth value of the report can be either false or true, whether the bill will pass or not. But when we would use a verb like 'doubt', no such thing can be said. For instance, sentence (10) shows use no truth value for the report even if it was wrong. This is since this verb shows doubt itself. Such a verb might influence the use of its perfect, so these will be eliminated for now. Considering these two problems, only three more were added from this list, namely 'mention', 'deny' and 'warn', which means the total list consists of 18 verbs.

For the other languages, the 18 English reporting verbs were translated and checked whether they could be used in a reporting manner. This resulted in 18 verbs for the German corpus and 17 verbs for the Dutch and French corpus. This is because 'agree' is not easily translated into Dutch and 'argue' is not a reporting verb in French. Spanish is excluded from this research, because in our group we didn't have someone with native speaker intuitions. Whether a verb is a reporting verb or not can't be determined because these intuitions lack. In addition, the translations would be hard to understand and would not be examined effectively. The total list of reporting verbs is as follows:

English	Dutch	German	French
Say	Zeggen	Sagen	Dire
Tell	Vertellen	Erzählen	Raconter
Ask	Vragen	Fragen	Demander
Advise	Adviseren	Beraten	Conseiller
Agree		Zustimmen	Enregistrer
Apologize	Verontschuldigen	Sich entschuldigen	S'excuser
Decide	Besluiten	Entscheiden	Decider
Encourage	Aanmoedigen	Ermutigen	Encourager
Explain	Uitleggen	Erklären	Expliquer
Insist	Aandringen	Darauf bestehen	Pretendre
Promise	Beloven	Versprechen	Promettre
Recommend	Aanbevelen	Empfehlen	Signaler
Remind	Herinneren	Erinnern	Se souvenir de
Suggest	Voorstellen	Vorschlagen	Proposer
Warn	Waarschuwen	Warnen	Avertir
Mention	Aankaarten	Erwähnen	Parler
Deny	Ontkennen	Verweigern	Nier
Argue	Beweren	Argumentieren	

These verbs were recognized in their respective perfect tenses from the EUROPARL corpus. That is, verbs that follow the rules for these tenses. For English this means have/has (been) + a past particle, Dutch is zijn/hebben + voltooid deelwoord, or voltooid deelwoord + zijn/hebben, German is sein/haben + Partizip or Partizip + sein/haben and French is être/avoir + participle passé. In Dutch and German, it is allowed to have other words between these verbs, whereas in English this can only happen with adverbs of frequency or adverbs of indefinite time as said by Van der Voort (2009), and in French it can only happen when it is an adverb giving information about the verb. After recognizing the perfects from the text, we could see them individually in fragments. These fragments were compared to the same fragments in the other languages. The algorithm used for this extraction is made by Van der Klis et al. in their previous research on the EUROPARL corpus. The code for the algorithm can be found on the website of the project: http://time-in-translation.hum.uu.nl/

With TimeAlign, a program used for comparing two texts with each other, verbs could be selected manually. This is used to select a matching construction of the highlighted construction of the source language. For all the sentences found with a *present perfect* in English, the corresponding verbs had to be selected for Dutch, German and French. For every *Perfekt* found in German, the corresponding verbs had to be selected for Dutch, English and French. For every *passé composé* found in French, the corresponding verbs had to be selected for Dutch, English and German. For every *vtt* found in Dutch, the corresponding verbs had to be selected for English, French and German. Figure 1 shows an example of corresponding verbs in all three languages.

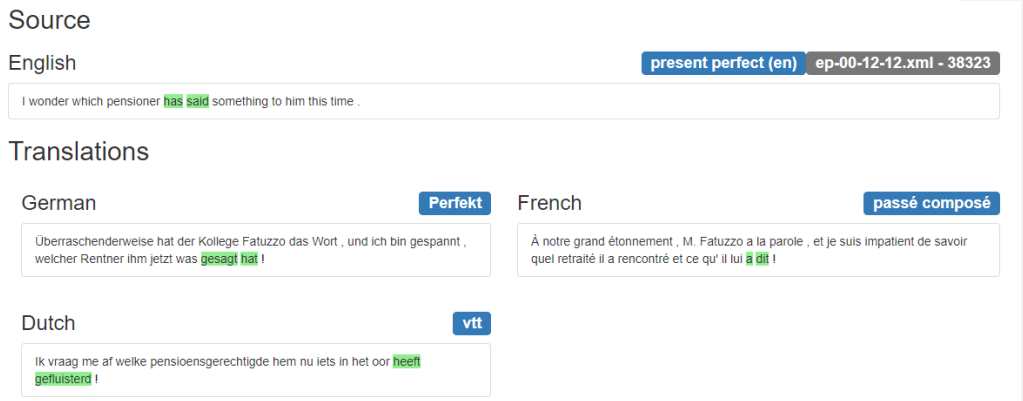


Figure 1: An example of an English sentence as the source language compared to the contexts in the other three languages. All the contexts were translated correctly. The top context is the source of the fragment. The highlighted green verbs are a perfect of a reporting verb. The other languages are the corresponding translations of this sentence. The highlighted verbs are the one picked, representing the translation of the perfect from the source language.

When a context was marked wrongly as a perfect, one of the check boxes had to be marked. If this happened, it was most likely a Dutch sentence. This makes sense according to the theory behind forming a *vtt* in Dutch. These sentences have been marked as not forming a *vtt*, eliminating them from the data. Figure 2 and figure 3 are examples for this kind of rejection. Another reason to eliminate sentences from the data would be when they were translated wrongly. This usually happened when a verb was translated into an adverb, adjective or a noun. Figure 4 is an example for this kind of rejection. This happened often with German and Dutch sentences. After getting rid of the sentences that can't be used, we tried to end with 500 contexts.

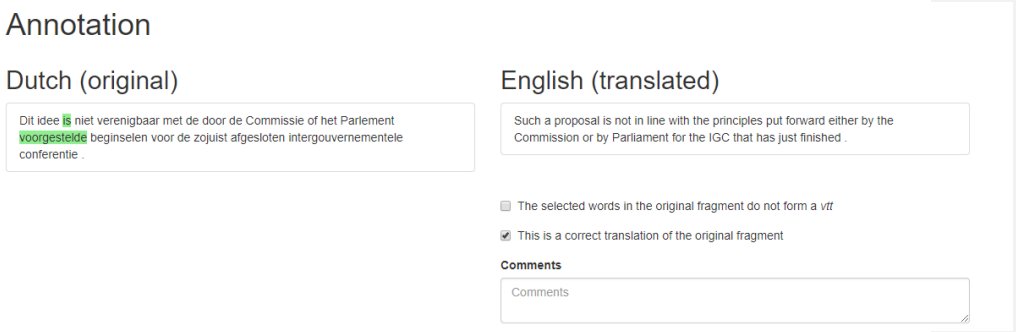


Figure 2: An example of the marked verbs in the source language not being a perfect. The two verbs marked are not linked to another.

Figure 3: An example of the marked verbs in the source language not being a perfect. The two verbs marked are not linked to another.

Figure 4: An example of the text of the source language not being translated rightly. The perfect in English ‘have agreed’ is translated to a noun ‘overeenstemming’.

We wanted to compare tenses with each other, so we have to add tenses to the marked verbs. This was done manually in an excel sheet containing all the information of the sentences and verbs from the previous step as seen in figure 5. The context in which these verbs were found are located more to the right in this file and were used to determine the tense as well as previously said. The original sentence is even more to the right, making it able to double check the given verbs.

id	tense	source/target	w1	w2	w3	w4	w5	pos1	pos2	pos3
24907	vtt	target	heb	genoemd				verbpressg	verbpapa	
24919	vtt	target	hebben	laten	weten			verbprespl	verbinf	verbinf
24839	ott	target	dringen	op	aan			verbprespl	prep	adv
24935	vtt	target	is	gezegd				verbpressg	verbpapa	
24889	vtt	target	is	gezegd				verbpressg	verbpapa	
24867	vtt	target	hebben	gesloten				verbprespl	verbpapa	
24881	vtt	target	heeft	gezegd				verbpressg	verbpapa	
24905	vtt	target	is	bereikt				verbpressg	verbpressg	
24884	vtt	target	overeengekomen	is				verbpapa	verbpressg	
24883	vtt	target	heb	gezegd				verbpressg	verbpapa	
27348	vtt	target	heeft	voorgesteld				verbpressg	verbpapa	
27247	vtt	target	hebben	voorgesteld				verbprespl	verbpapa	
27249	ott	target	motiveren					verbprespl		
27349	vtt	target	heeft	voorgesteld				verbpressg	verbpapa	
27434	vtt	target	heeft	voorgesteld				verbpressg	verbpapa	
27386	vvt	target	werd	gesteld				verbpastsg	verbpapa	
27431	ott	target	dringen					verbprespl		
27244	vtt	target	is	gezegd				verbpressg	verbpapa	
27223	vtt	target	heeft	gevraagd				verbpressg	verbpapa	
27248	ovt	target	zei					verbpastsg		

Figure 5: A part of the excel sheet for the Dutch verbs. The second column is the one filled in manually, the fourth, fifth and sixth column contained the verbs which had to be categorized.

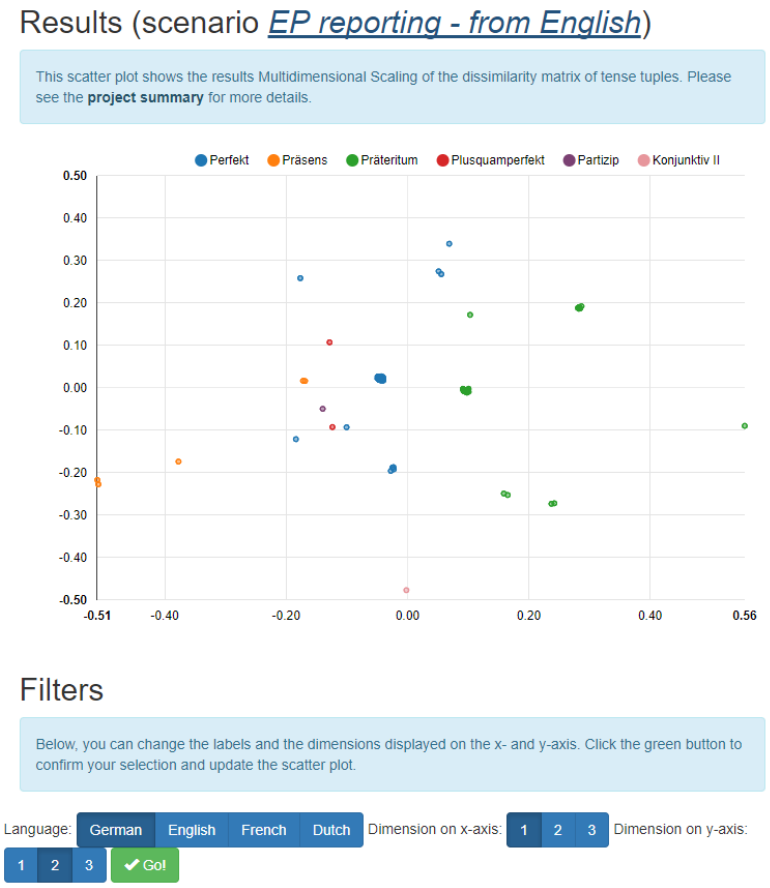
For Dutch the tenses that were found were *voltooid tegenwoordige tijd* (*vtt), onvoltooid tegenwoordige tijd (ott), voltooid verleden tijd (vvt)* and *onvoltooid verleden tijd* (*ovt)*. For German the tenses that were found were *Präsens, Perfekt, Präteritum, Plusquamperfekt, Partizip, Konjunktiv II, Futur I* and *Futur II*. For French the tenses that were found *were passé composé, présent, plus-que-parfait, subjonctif passé, passé récent, imparfait, infinitive* and *participe passé*. For English the tenses that were found were *simple present, simple past, infinitive, present participle* and *present perfect*.

All the tenses were connected if they belonged to the same context from the Dutch, German, French and English sentences. With this connection 4-tuples arise, giving all the tenses for a certain context. The 4-tuples have a fixed order, namely <*French, English, Dutch, German*>. An example of such a tuple is <*passé composé, present perfect, vtt, Perfekt>,* which is the tuple that corresponds with the contexts in Figure 1. These combinations will differ in tenses as said in the previous paragraph. The distance is defined by comparing two tuples with each other. These distances are the input for the *multidimensional scaling (MDS)* which will be the next step in the methods.

When the tenses in the two tuples are the same, the distance is zero. If one tense differs between the two tuples, the distance is one, divided by the length of the tuple. This is four in this research. This means the distance is ¼ when one tense differs. This way of calculating the differences between two 4-tuples is done between all the combinations of 4-tuples found from the corpus. These distances were put in a distance matrix, showing all the distances between possible tuples. These distances are

visualized with the use of MDS, in the same way that Wälchli and Cysouw (2012) did, resulting in the semantic maps seen later.

Every tense has its own label, making it easy to see how a certain tense behaves. Every tense has its own color as well, which make them visible in the semantic maps. Every color is its own category. For example, the *Perfekt* in figure 6 is blue. Blue means that the corresponding tense is a perfect in every language. So, a *vtt* would be blue as well. This way the patterns across languages can be visualized easier. These semantic maps can be found on the site from the Time in Translation project.



4

Figure 6: One of the semantic maps following the results of the study. This semantic map are the reporting verbs from English compared to the connected verbs in German. The points resemble the distance calculated with the use of Multidimensional Scaling. Every color corresponds to a specific tense. The distances visualized on both the y- as the x-axis. The further a point is away from the [0,0] point, the more different the tenses are to a perfect.

As shown in figure 6, one semantic map shows all data points for one language compared to another language. The location of the data is determined by the distance of the tuple each point represents. This means that the points are the same for all the languages compared to the same source languages, but the colors differ. The colors are the most important aspect for such a map, showing that the tenses compare to the tenses of the other languages if the points would cluster. In this case, the German tenses are compared to the English tenses. All the points are divided in two dimensions. If you put your mouse on of the points, you can see which 4-tuple corresponds with that point. At the bottom you can choose which language to compare against with the blue buttons. After you press '*Go!*', the semantic map for that language is shown.

3.2 Results

As mentioned in the Methodology section, semantic maps were formed with the TimeMapping algorithm. Every semantic map shows the distances between the points and the used tenses for one language. Because all four languages were compared to each other, every language has three semantic maps, for example [English, French], [English, Dutch] and [English, German]. This means that there are twelve semantic maps in total. Firstly, the results for the French perfects will be presented.

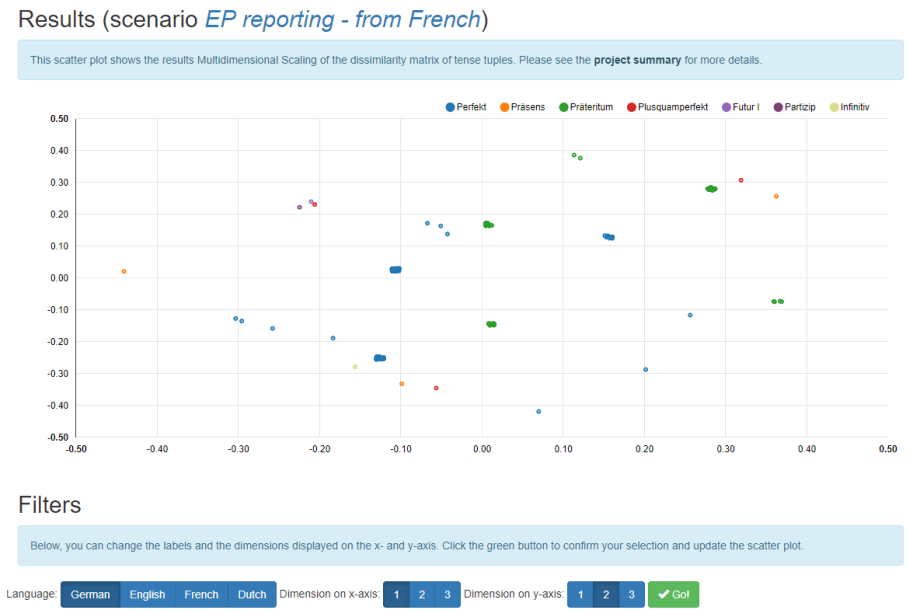


Figure 7: The semantic map for the German tenses compared to the French perfects. The different colors show different tenses, as shown right above the graph. No clusters can be found in this semantic map.

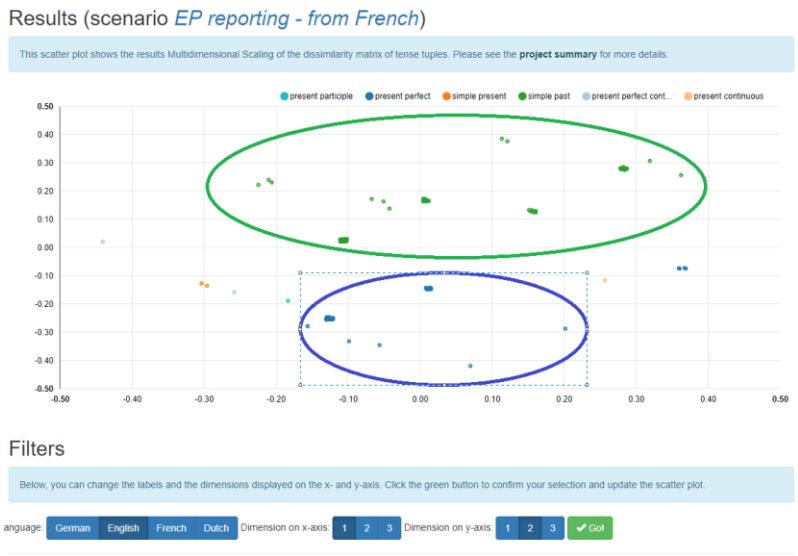


Figure 8: The semantic map for the English tenses compared to the French perfects. The different colors show different tenses, as shown right above the graph. A cluster was found for the *simple past* (green) and for the *present perfect* (blue).

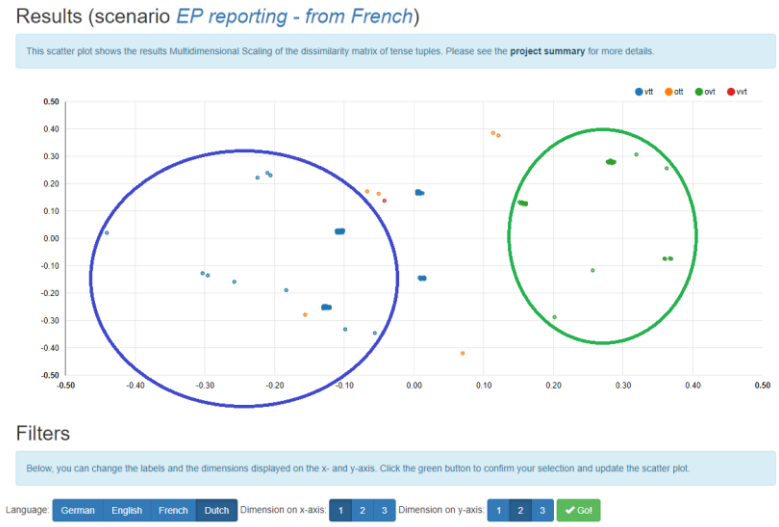


Figure 9: The semantic map for the Dutch tenses compared to the French perfects. The different colors show different tenses, as shown right above the graph. A cluster was found for the *vtt* (blue) and for the *ovt* (green).

Totals per language

German

Tense	Count
Perfekt	80
Präteritum	57
Plusquamperfekt	3
Präsens	3
Partizip	1
Infinitiv	1
Futur I	1

English

Tense	Count
simple past	85
present perfect	55
present perfect continuous	2
simple present	2
present continuous	1
present participle	1

French

Tense	Count
passé composé	146

Dutch

Tense	Count
vtt	104
ovt	35
ott	6
vvt	1

Figure 10: The division of the tenses of all four languages of the French perfects.

For every source language three maps were made. These maps are figure 7, 8 and 9 for French. For every three maps there will always be a fourth graph showing the descriptive statistics of the maps. This would be Figure 10 for French. Figure 10 shows how the elements of the tuples in these maps are divided. For instance, 80 of the tuples contain a *Perfekt*.

Figure 7 shows that German has a wide range of possible tenses when compared to French perfects. Not only do the tenses vary a lot, they are scattered all over the graph as well, meaning no clusters could be made. As figure 10 shows, 80 out of the 146 French perfects are translated to a *Perfekt* and three to a *Plusquamperfekt*, meaning that the remaining 63 are not translated to a perfect tense. Most of them are translated to a *Präteritum*, which translates to *simple past*. Figure 10 shows that seven different tenses are used.

Figure 8 shows that English doesn't have a wide a range of possible tenses as German, but there are still a couple. The *simple past* and the *present perfect* are clustered in the top and bottom of the graph. The *simple past* is more common, as can be seen in figure 10 as well. 55 Are *present perfects*, two are *present perfect continuous* and one is a *present continuous*. This makes 58 out of the 146 French perfects translated to an English perfect. 88 Are not translated to a perfect. Figure 10 shows that six different kind of tenses are used in total.

Figure 9 shows that Dutch has a small range of possible tenses. The *vtt* and *ovt* are clustered in the left and right of the graph. The left cluster is not perfect, meaning it doesn't contain all the *vtt*'s in the graph and contains one other tense as well. It is still close to containing all of them. *Vtt*'s are the most common tense in this graph, shown by figure 10. Figure 10 demonstrates that 104 sentences contain a *vtt*, and one *vvt*, meaning that 105 out of 146 French perfects are translated to a Dutch perfect. 41 French perfects are translated otherwise, mostly into an *ovt*. Figure 10 shows that four different kinds of tenses are used in total.

Figures 7, 8, 9 and 10 together show a big difference among the languages. Dutch has the highest number of perfects in the translation, whereas English has the least. Dutch has the lowest number of possible tenses, whereas German has the most. English and Dutch show clusters, and German does not. Even though both English and Dutch show clusters, the overlap between the *ovt* and *simple past* is very small, and so is the overlap between the *vtt* and the *present perfect.*

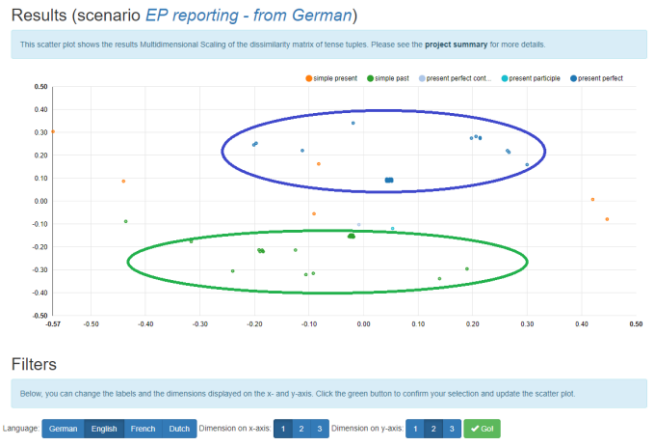


Figure 11: The semantic map for the English tenses compared to the German perfects. The different colors show different tenses, as shown right above the graph. A cluster was found for the *present perfect* (blue) and for the *simple past* (green).

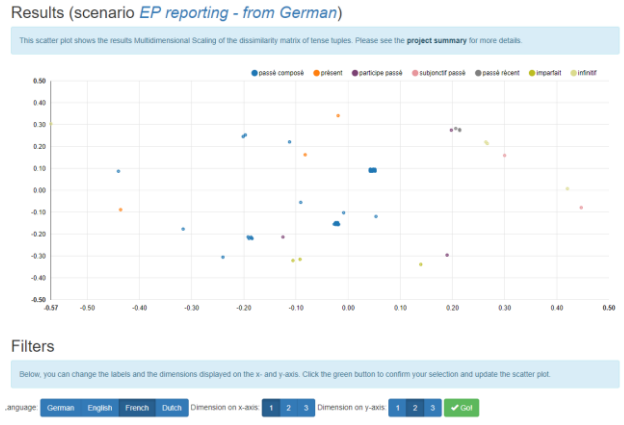


Figure 12: The semantic map for the French tenses compared to the German perfects. The different colors show different tenses, as shown right above the graph. No clusters can be found in this semantic map.

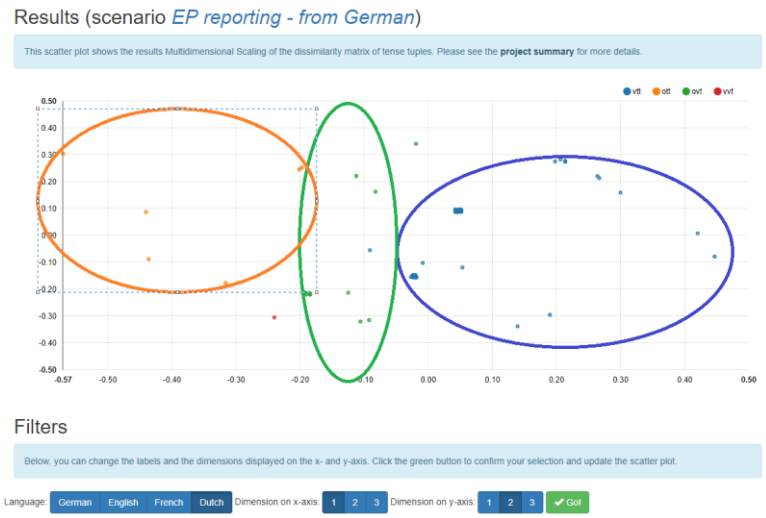


Figure 13: The semantic map for the Dutch tenses compared to the German perfects. The different colors show different tenses, as shown right above the graph. A cluster was found for the *vtt* (blue), for the *ovt* (green) and for the *ott* (orange).

Totals per language

German

Tense	Count
Perfekt	81

English

Tense	Count
present perfect	42
simple past	31
simple present	6
present participle	1
present perfect continuous	1

French

Tense	Count
passé composé	63
infinitif	4
participe passé	3
présent	3
imparfait	3
passé récent	3
subjonctif passé	2

Dutch

Tense	Count
vtt	63
ovt	11
ott	6
vvt	1

Figure 14: The division of the tenses of all four languages of the German perfects.

Figure 11 shows that English has a decent range of possible tenses when compared to German perfects. The *present perfect* and the *simple past* are clustered in the top and bottom of the graph. The *present perfect* is more common, as seen in figure 14. 42 Sentences contain a *present perfect* and one contains a *present perfect continuous*. This makes 43 out of the 81 German perfects translated to an English perfect. 38 Are not translated to a perfect and most of these are a *simple past*. Figure 14 shows that five different kind of tenses are used in total.

Figure 12 shows that French has a wide range of possible tenses when compared to German perfects. Not only are there a lot of different tenses, they are spread across the graph as well, meaning no clusters could be found. The *passé composé* is the most common tense found, as seen in figure 14. 63 Sentences contain a *passé composé*, meaning that 63 out of 81 German perfects translated to a French perfect. 18 Are not translated to a perfect and these are divided almost evenly over the other tenses. Figure 14 shows that seven different kind of tenses are used in total.

Figure 13 shows that Dutch has a small range of possible tenses when compared to German perfects. The *ott, ovt* and *vtt* are clustered in the left, middle and right of the graph. The *vtt* is the most common tense, as seen in figure 14. 63 Sentences contain a *vtt* and one sentence contains a *vvt*. This makes 64 out of the 81 German perfects translated to a Dutch perfect. 17 Are not translated to a perfect and most of them are an *ovt*. Figure 14 shows that four different kind of tenses are used in total.

Figures 11, 12, 13 and 14 together show some differences between the languages. Dutch and French both have a high number of perfects in the translation, whereas English has less. But French also has the highest number of different possible tenses, whereas Dutch has the fewest. Dutch has three clusters, which is the highest, followed by two clusters for English and none for French. Once again, the overlap between tenses is very small between the corresponding tenses in English and Dutch. The French perfects can be found partially among the Dutch perfects, partially among the English perfects and partially among neither of the perfects.

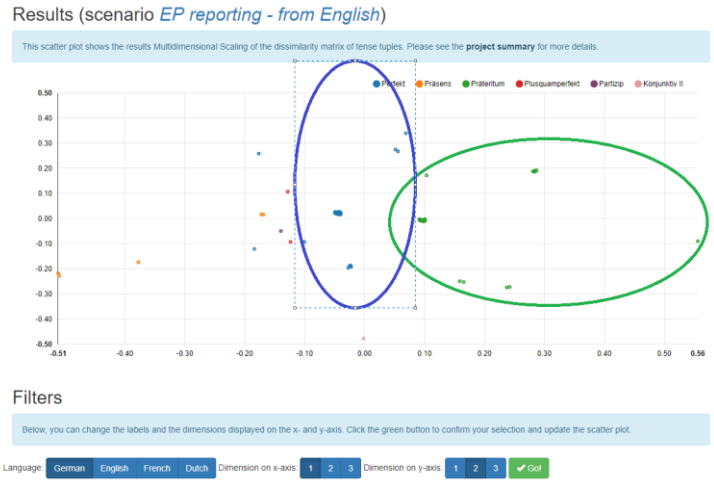


Figure 15: The semantic map for the German tenses compared to the English perfects. The different colors show different tenses, as shown right above the graph. A cluster was found for the *Perfekt* (blue) and the *Präteritum* (green).

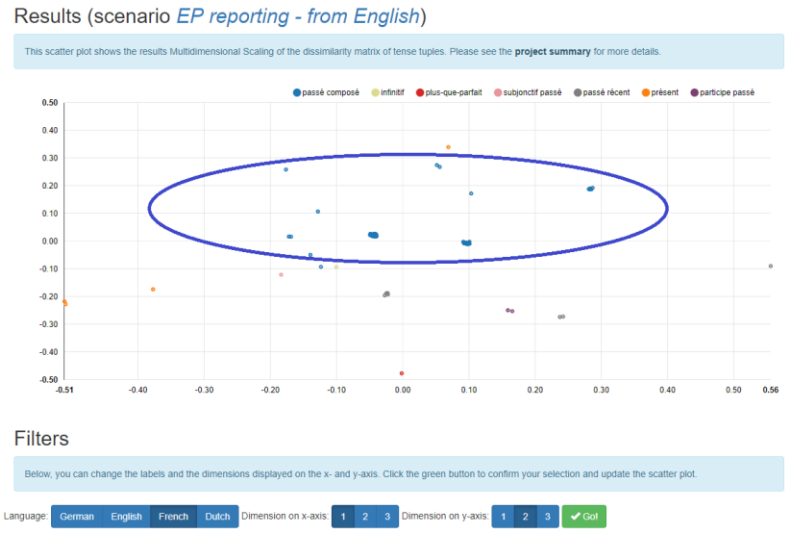


Figure 16: The semantic map for the French tenses compared to the English perfects. The different colors show different tenses, as shown right above the graph. A cluster was found for the *passé composé* (blue).

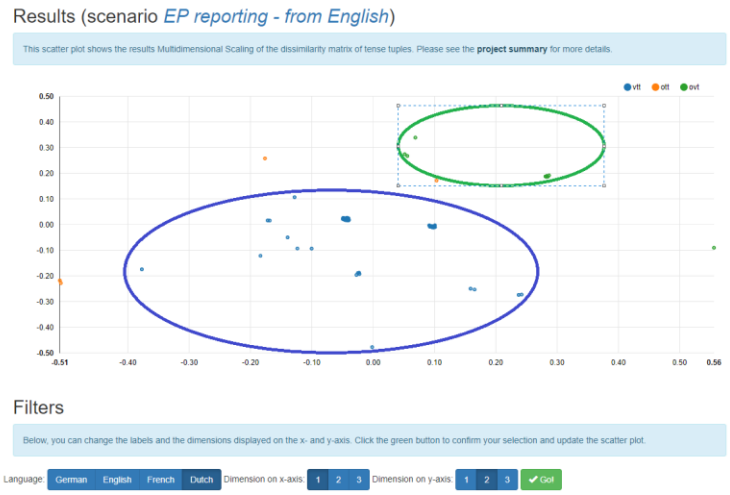


Figure 17: The semantic map for the Dutch tenses compared to the English perfects. The different colors show different tenses, as shown right above the graph. A cluster was found for the *vtt* (blue) and *ovt* (green).

Totals per language

German

Tense	Count
Perfekt	57
Präteritum	29
Präsens	5
Plusquamperfekt	2
Partizip	1
Konjunktiv II	1

English

Tense	Count
present perfect	95

French

Tense	Count
passé composé	79
passé récent	7
présent	4
participe passé	2
plus-que-parfait	1
infinitif	1
subjonctif passé	1

Dutch

Tense	Count
vtt	82
ovt	9
ott	4

Figure 18: The division of the tenses of all four languages of the English perfects.

Figure 15 shows that German has a wide range of possible tenses when compared to English perfects. The *Perfekt* and *Präteritum* are clustered in the middle and right of the graph. The P*erfekt* is more common, as seen in figure 18. 57 Sentences contain a *Perfekt* and two contain a *Plusquamperfekt*. This makes 59 out of the 95 English perfects translated to a German perfect. 36 Are not translated to a perfect and most of these are a *Präteritum*. Figure 18 shows that six different kind of tenses are used in total.

Figure 16 shows that French has a wide range of possible tenses when compared to English perfects. The *passé composé* is clustered in the middle of the graph. The *passé composé* is the most common tense, as seen in figure 18. 79 Sentences contain a *passé composé*, meaning that 79 out of 95 English perfects translated to French perfects. 16 Are not translated to a perfect and most of these are a *passé récent*. Figure 18 shows that seven different kind of tenses are used in total.

Figure 17 shows that Dutch has a small range of possible tenses when compared to English perfects. The *vtt* and *ovt* are clustered in the bottom and top of the graph. The *vtt* is the most common tense, as seen in figure 18. 82 Sentences contain a *vtt*, meaning that 82 out of 95 English perfects are translated to a Dutch perfect. 13 Are not translated to a perfect and most of these are an *ovt*. Figure 18 shows that three different kind of tenses are used in total.

Figures 15, 16, 17 and 18 together show some differences between the languages. Dutch and French both have a high number of perfects in the translation, whereas German has less. But French has the highest number of different possible tenses, whereas Dutch has the fewest. Dutch and German have two clusters, which is the highest, and French has the least with one cluster. The German clusters don't correspond with any of the other clusters. However, this time, the French and Dutch cluster for respectively the *passé composé* and the *vtt* have a decent amount of overlap between the two clusters.

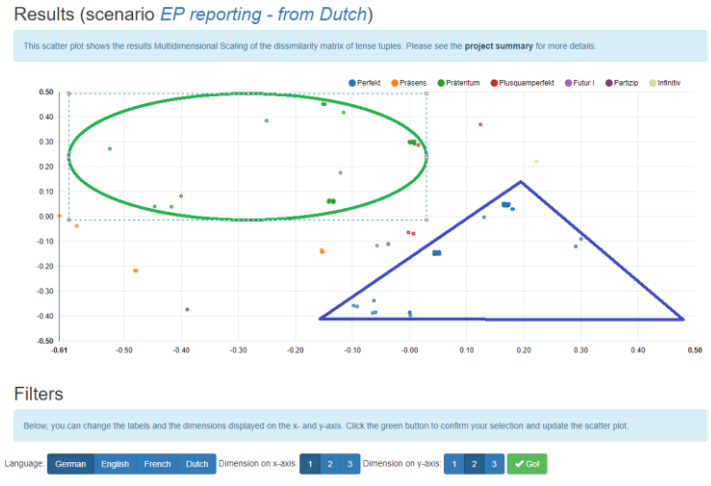


Figure 19: The semantic map for the German tenses compared to the Dutch perfects. The different colors show different tenses, as shown right above the graph. A cluster was found for the *Perfekt* (blue) and *Präteritum* (green).

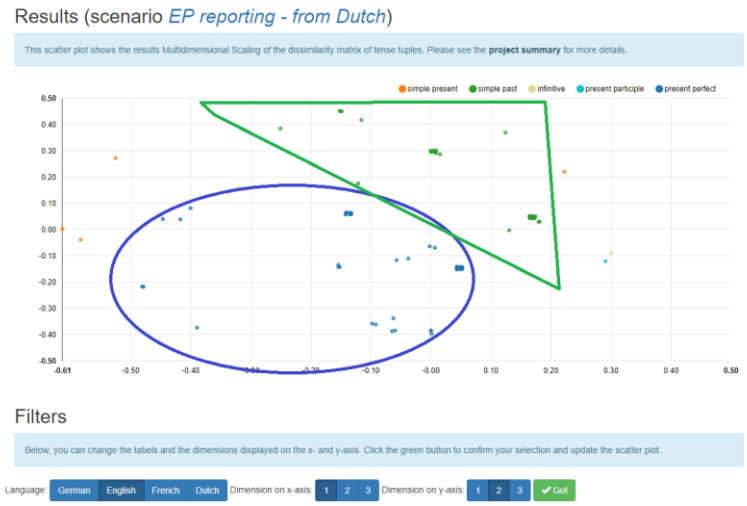


Figure 20: The semantic map for the English tenses compared to the Dutch perfects. The different colors show different tenses, as shown right above the graph. A cluster was found for the *present perfect* (blue) and *simple past* (green).

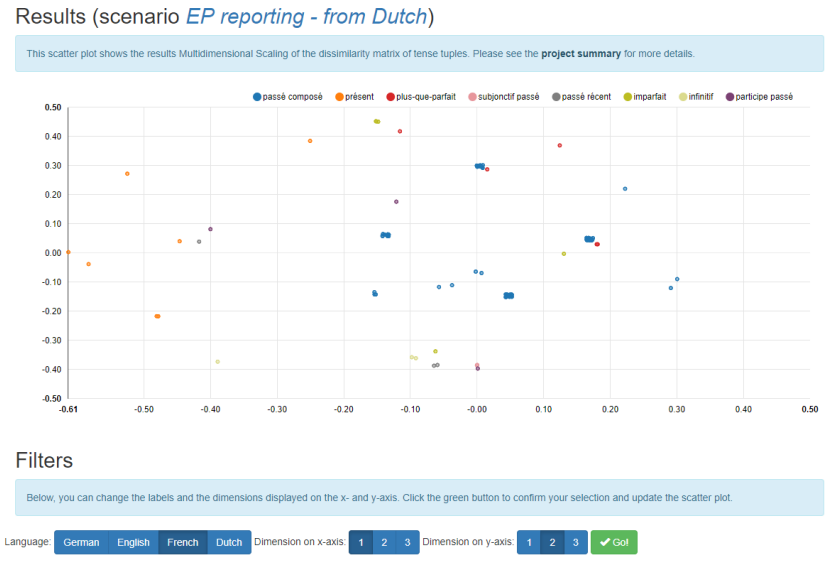


Figure 21: The semantic map for the French tenses compared to the Dutch perfects. The different colors show different tenses, as shown right above the graph. No clusters can be found in this semantic map.

Totals per language

German

Tense	Count
Perfekt	79
Präteritum	38
Präsens	7
Plusquamperfekt	4
Partizip	2
Futur I	2
Infinitiv	1

English

Tense	Count
present perfect	73
simple past	54
simple present	4
present participle	1
infinitive	1

French

Tense	Count
passé composé	105
présent	7
imparfait	5
plus-que-parfait	5
participe passé	3
passé récent	3
infinitif	3
subjonctif passé	2

Dutch

Tense	Count
vtt	133

Figure 22: The division of the tenses of all four languages of the Dutch perfects.

Figure 19 shows that German has a wide range of possible tenses when compared to Dutch perfects. The *Perfekt* and *Präteritum* are clustered in the right and left of the graph. The *Perfekt* is the most common, as seen in figure 22. 79 Sentences contain a *Perfekt* and four contain a *Plusquamperfekt*. This makes 83 out of the 133 Dutch perfects translated to a German perfect. 50 Are not translated to a perfect and most of these are a *Präteritum*. Figure 22 shows that seven different kind of tenses are used in total.

Figure 20 shows that English has a decent range of possible tenses when compared to Dutch perfects. The *present perfect* and *simple past* are clustered in the bottom and top of the graph. The *present perfect* is more common, as seen in figure 22. 73 Sentences contain a *present perfect*, meaning that 73 out of the 133 Dutch perfects translated to an English perfect. 60 Are not translated to a perfect and most of these are a *simple past*. Figure 22 shows that five different kind of tenses are used in total.

Figure 21 shows that French has a wide range of possible tenses when compared to Dutch perfects. Not only are there a lot of different tenses, they are spread across the graph as well, meaning no clusters could be found. The *passé composé* is the most common, as seen in figure 22. 105 Sentences contain a *passé composé* and five sentences contain a *plus-que-parfait*. This makes 110 out of the 133 Dutch perfects translated to a French perfect. 23 Are not translated to a perfect and these are divided quite evenly over the other tenses. Figure 22 shows that eight different kind of tenses are used in total.

Figures 19, 20, 21 and 22 together show some differences between the languages. French has a high number of perfects in the translation, whereas German and English have less. But French has the highest number of different possible tenses, whereas English has the fewest. English and German have two clusters, which is the highest, and French has the least with no clusters. The overlap between the English and German clusters is minimal.

German has four clusters in total, six or seven tense possibilities per language and has the lowest number of perfects three times. English has six clusters in total, five or six tense possibilities per language and the lowest number of perfects three times. French has one cluster in total, seven or eight tense possibilities per language and has the highest number of perfects three times. Dutch has seven clusters in total, three or four tense possibilities per language and the highest number of perfects three times.

The overlap between the clusters in this research is low. The only clusters that seemed to correspond a decent amount was the French *passé composé* and the Dutch *vtt* when translated from the English perfects. This shows that the languages choose a perfect in their translations for different reasons. This will be discussed further in the Conclusion section.

3.3 Conclusion

The first research question was what the translations of the perfects would look like. As can be seen in the results section, Dutch and French had the most perfects in the translations and German and English had the least. The hypothesis was that German, French and Dutch would score high and English would score low. The only language that does not fit this hypothesis is German. The differences seem to be consistent, being present in every language translations. Only the English translations of the French perfects seem not to have a perfect in most of the translations.

That French and Dutch would score high, was expected with the theory of Schaden. Dutch was expected to score high since the EUROPARL corpus is not a corpus where stories are told. The same goes for the fact that English scored low. The fact that German scored low might be because this corpus has a lot of sentences which refer to the recent past. An observation about this recent past is that German does not behave according to the rules in all circumstances, as can be seen in figure 23. It may have had an influence on the translations in this corpus, explaining the low frequency of perfects.

Figure 23: A sentence in which all the languages use a perfect, besides German. In this sentence, what 'mentioned' refers to is something in the recent past.

The possible tenses in which the contexts could be translated into, differ a lot among the languages. French scores the highest with seven or eight different tenses, followed by German with six or seven different tenses, English with five or six and Dutch with three or four. This might show that some languages have difficulty with translating perfects from different languages. However, it could be that this is directly linked to the number of tenses the languages have in total and how well they seem to perform on their translations. French has many possible tenses, being able to be more precise about the moment the referenced action has been executed. An example would be the passé récent, which expresses something that happens just before the current event. German, English and Dutch all have less tenses, but only Dutch has a high amount of translations into a perfect.

The next research question was about eventual clusters in the semantic maps. For both Dutch and English, the *simple past* and the *present perfect* clustered in all the semantic maps. German almost did the same, it only misses clusters in the French translations. This observation shows clearly that German does different things than expected. German and French should have been similar to reflect the hypothesis. The French perfect only clustered in the Dutch contexts. The expectation was that the

French perfect would cluster in all the translations, because a high number of perfects was expected. This might be explained with the large number of tenses that the French has plus the fact that English and German showed a low number of perfects. MDS might be responsible for this, preferring languages with less tenses to make them cluster.

Then there was the observation of the clusters not corresponding with each other a lot. In the translations of the French perfects, the Dutch and English clusters showed little resemblance. In the translations of the German perfects, once again the Dutch and English clusters showed little resemblance. In the translations of the English perfects, German clusters showed no resemblance to the clusters of French and Dutch. However, the one cluster found for French does resemble the *vtt* cluster of the Dutch translations for the most part. In the translations of the Dutch perfects, English and German did not resemble each other once again.

In general, this means that the only clusters that showed resemblance were the French and Dutch ones. This is since French only has one cluster in this research. So once again French and Dutch pair up. This time though, English and German do not. This might be because the rules for using a perfect are still very different between these two languages. However, these results once more show that German acted differently than French in this research, which was contrary to what was expected.

The final research question was about the results of Van der Klis et al. and if those results would be like the results of this research. Both English and German show a high rate of resemblance concerning the *simple past* and *present perfect*. Dutch and French show a much higher rate of perfects in the translations than the semantic maps of Van der Klis et al. showed. This means that when a language used a perfect, it will most likely be translated to a perfect in both French and Dutch. This since the rate of perfects translating into a perfect is much higher than the overall rate off perfects in these languages.

4 Discussion

First, the results showed us that the German was less translated into a perfect than expected. The given reason for this is that German might behave differently than expected when a verb refers to the recent past, something reporting verbs do often. This is certainly not a proven theory and other characteristics of reporting verbs might have influenced the results. The present knowledge and intuitions about German certainly isn't enough to give a solid theory behind the low number of perfects in this research.

Our theory about the low frequency of German perfects in this research is far from proven. To support this theory, research should be done on the recent past. This might not be a correct theory, so more research in reporting verbs and their characteristics will help. This is especially interesting since German seemed to behave like the theories in all other researches in the Time in Translation project.

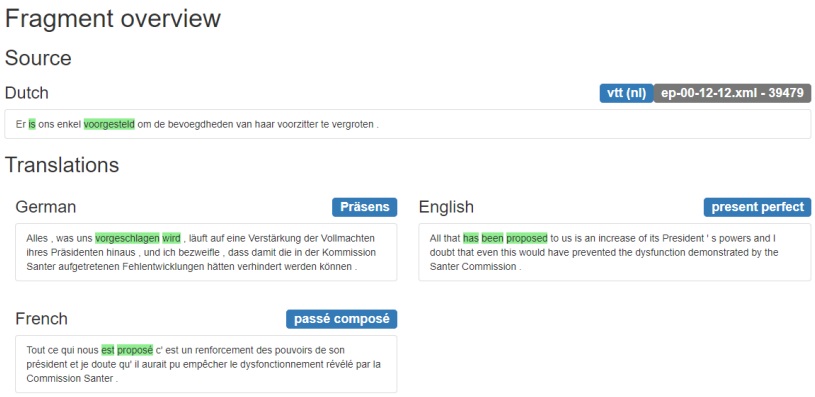


Figure 24: A fragment overview with a passive context from the source language. Dutch, French and English have a present perfect, where German has a simple present.

Another influence might have been passive contexts. For instance, the example in figure 24, where German is translated to a simple present instead of a perfect. It could be that passive contexts are translated into different tenses across languages. Since this research has not taken this into account, the results might have been influenced slightly. Dutch passive contexts seem to appear quite often, making a theory about this is essential for further knowledge about the perfect.

11) a. Mark has said that he loved games.
b. Mark has asked if he loved games.

Most of the reporting verbs, as said in the literature section, has a truth value dependent on the trustworthiness of the speaker or author. As can be seen in sentence (11a), if Mark is deemed trustworthy, he must have loved games. However, some of the reporting verbs don't follow this general rule. For instance sentence (11b), where it is not possible to say whether this sentence is true of false just looking at the trustworthiness of Mark. Moreover, since this is a question, we do not even know the truth value. 'Recommend' and 'suggest' have the same problem, since the action of what is recommended or suggested is possibly executed by the one who got the recommendation or the

suggestion, but that person doesn't have to follow it. Whether this had any influence on the results is unclear. Since they did not turn up a lot in the contexts used for this research it would not have a big influence even if it did.

A possible future research direction would be examining the differences between reporting verbs like 'promise' and 'say' versus reporting verbs like 'ask'. They differ in rules determining the truth value of the sentence and might influence the number of perfects in some languages. This could be done in a more broader way. For instance, by doing this kind of research and looking for differences between questions and regular sentences. To cover another problem found in the discussion, the same research could be done between passive and active sentences.

Another thing about this research is that it is supposed to be about reporting verbs in general, but our data consists of sentences from the EUROPARL corpus only. There might be differences in spoken and written language and formal and informal language for example. These differences have not been covered by this research due to limitations in time and data. Still, these differences might have different results on reporting verbs. Researching reporting verbs in the same manner as this research has done, but on different kind of texts would make the conclusions of this research more solid. If all the possible differences are researched, perfects will be better understood, and thus used, than ever before.

References

Van der Klis, M., Le Bruyn, B., & De Swart, H. (2017). Mapping the Perfect via Translation Mining. *EACL 2017*, 497

Nishiyama, A., & Koenig, J. (2010). What is a Perfect State? *Language, 86*(3), 611-646.

Schaden, G. (2009). Present perfects compete. *Linguistics and Philosophy, 32*(2), 115–141. https://doi.org/10.1007/s10988-009-9056-3

Swager, K. (2017). Camus in Kaart. (Bachelor thesis, University Utrecht)

Tiedemann, J. (2012). Parallel Data, Tools and Interfaces in OPUS. *LREC*, *2012*, 2214-2218.

Van der Voort, P. (2009). *Core Grammar: For higher education.* Groningen: Noordhoff.

Wälchli, B. & Cysouw, M. (2012) Lexical typology through similarity semantics: Toward a semantic map of motion verbs. *Linguistics, 50*(3), 671-710.

Reporting verbs - http://www.ef.com/english-resources/english-grammar/reporting-verbs/

Reporting verbs - http://www.perfect-english-grammar.com/reporting-verbs.html

Project - http://time-in-translation.hum.uu.nl/