THE PERFECT IN FARSI

Name Student:
Marziah Bijani
5717515

Supervisors:
prof. dr. Henriette de Swart
Martijn van der Klis

Second reviewer:
dr. Sjoerd Stuit

Bachelor Cognitive Artificial Intelligence
Utrecht University
7,5 ECTS

Utrecht University

Table of Contents

1. Introduction

Finishing my bachelors in AI by doing a thesis had always excited me. This would be my chance to finally have a taste of what AI research would be like, after seeing all the different perspectives on the broad field of AI through the courses I had followed. Therefore it was important for me to do my thesis on something which would be relevant and something which would genuinely interest me. When choosing a subject, I tried to focus on supervisors who were doing their own project, so I could experience the reality of AI research and not face my own expectations by proposing my own idea for research. Not saying that coming up with your own idea isn't the reality of research, but knowing the research questions of people who are more experienced and educated, gives a more realistic and up to date perspective on AI research.

What interested me the most in my bachelors were the problems one can face when using AI related methods. Problems like responsibility, ethics and context. In fact, these are the problems which need research in the hope to make systems effortless. Topics which interested me in this notion were philosophy, the perception of language and ambiguous visual stimuli. First I had approached a project about responsibility in AI. I didn't end up doing my thesis in that project, because I had to propose my very own idea which would put me back in my safe comfort zone of my own expectations of AI research. Plus I had to do a philosophical literature review, which lacked the practical side of AI I wanted to see.

I looked further and came across the 'Time-in-translation' project of prof. dr. Henriëtte de Swart. It looked like the Perfect combination of language perception, context and computational linguistics. However, I have to admit that the aim of the project and its relation with AI was unclear for me until I had a meetup with dr. de Swart. During our meetup it was explained to me what the project is exactly about.

In short, the project is about considering the semantics of the 'Perfect' in translations from French to other languages, according to the context. The aim is to investigate when the Perfect in French will be translated as a Perfect in the other language. This will be done by using a French novel and its translations in different languages. The final purpose of this analysis is to define a general meaning of the Perfect tense. Examples of the application of the results of this project, can be writing more appropriate grammars for education purposes or for example for online translating engines like Google translate, so it will be able to produce more natural translations which are esthetically more appealing. In sections 2.2 and 3.3 I will discuss some studies our results could contribute to.

During our conversation dr. de Swart guessed that I could speak Farsi, and that is indeed true. Farsi or Persian is the language which is spoken in Iran. She asked if I would like to include Farsi in the analysis and I immediately got excited. She explained that lots of European languages are included in the project. Farsi belongs to the family of Indo-European languages, which would offer a very interesting comparison. I accepted her suggestion for my

research topic with full satisfaction. I had finally succeeded at finding a relevant project where my contribution would be unique and appreciated too!

The fact that I read and speak Farsi enables me to involve Farsi in the project. Also my background knowledge on linguistics and its computational applications make me very suitable for effortlessly contributing to this project. The following steps will be undertaken: First I have to find an online Farsi translation of the novel and convert it to an appropriate format so it can be parsed by the parser which is developed by the PhD candidate Martijn van der Klis. Next I have to find the corresponding verb of the French *passé compos*é in the Farsi translation. After this, I have to assign the verb tense. Lastly semantic maps can be produced. Using these maps, comparisons can be made with other languages using multilingual parallel corpus of literary texts. Therefore, the research question is how the use of the Perfect in the Farsi translation behaves compared to the other languages. The other languages will be the set of five languages used in the study of Martijn van der Klis, Bert le Bruyn and Henriëtte de Swart (2017); German, English, Spanish, French and Dutch.

I am excited to take you with me during this process! In the following sections I will provide sufficient background information about the project and Farsi and comprehensively explain each step I have taken.

2. Background

In this section I will provide a more specific explanation of the Time in Translation project. I will discuss why the Perfect is interesting for research and give a short overview of the already done research by the team of Time in Translation.

2.1 Time in Translation

The Perfect is a verb form which indicates that an action or circumstance occurred earlier than the time under consideration, often focusing on the resulting state rather than on the occurrence itself. There can be many interpretations on the Perfect which cannot be easily categorized. It is unclear what the function of the Perfect actually is, for example why authors choose it over a Simple Past form (Nishiyama & Koenig, 2010). However, linguists commonly distinguish the three core Perfect meanings in the following example given by Van Der Klis, Le Bruyn & de Swart (2017):

(1) a. Mary has visited Paris.
(her past visit is relevant now) [experimental]
b. Mary has moved to Paris.
(she currently lives in Paris) [resultative]
c. Mary has lived in Paris for five years (now).
(she moved there five years ago) [continuative]

Next to each option in (1), the label for the interpretation of the Perfect is written. Below each example, the interpretations are written in Italic. The resultative meaning in (1b) is thought to constitute the core meaning of the Perfect. However Van Der Klis, Le Bruyn & de Swart (2017) have shown that the same meaning of a past event and a result with current relevance, can be conveyed by Past, Perfect of Present.

Although various interpretations of the English Present Perfect are distinguished, there is still a wide disagreement on how to categorize those interpretations and whether they correspond to different meanings. The use of the Perfect at the sentence and discourse level also varies across languages, and it competes with past and present tenses. The Time in Translation project does not avoid this variation, rather it embraces it in order to unveil the meaning of the Perfect, using a 'smart' integration of quantitative and qualitative methodology in a data intensive approach (see Time in Translation website). Multilingual corpora are an excellent source for this, because they provide variation across languages in contexts where the meaning is stable.

In an earlier study of Van Der Klis, Le Bruyn and de Swart (2017) the focus was on a set of five European languages (German, English, Spanish, French, Dutch) using the EuroParl corpus. The current research which I will contribute to is using the novel L'Étranger written by Albert Camus. This novel is chosen because of the noticeable use of the *passé composé*. The use of the *passé composé* in the translated versions is investigated to see how this form is used in another language but in the same context.

The switch from the EuroParl corpus to a novel had different reasons. The initial reason to choose for a multilingual parallel corpora like EuroParl, was because translation equivalents provide us with form variation across languages in contexts where the meaning is stable (Van Der Klis, Le Bruyn & De Swart, 2017). The problem with EuroParl was that it was hard to track what the source language was and what of the translation. Also the subjects which were discussed were typical parliament talk using parliamentary vocabulary and speech style. This made this source of information limited to formal talk and limited to the few languages only used in the parliament. Therefore the novel of Camus, L'Étranger was chosen as the source language with parallel translations. Using a novel makes it possible to include many different languages; every language which the book is translated in, also Farsi!

2.2 Relationship with Artificial intelligence

One of the aspects which makes an entity intelligent is language. It is indeed language which separates humans from other animals. Language is even one of the main aspects what separates humans from computers (Jurafsky & Martin, 2014)! In the field of natural language processing, researches are trying to change this by getting computers to perform useful tasks involving human language. Tasks like enabling human-machine communication, improving human-human communication or simply doing useful processing of text or speech (Jurafsky & Martin, 2014).

A more familiar example is from the field of machine translation. A translation engine like Google Translate or any other comparable engine used for educational purposes. Artificial Intelligence technologies are widely used in developing educational applications for learning languages. Lewis Johnson and Andre Valente (2008) have discussed the several ways Artificial Intelligence is used to teach foreign languages and cultures.
We can also think about amount of data which is now available through the internet and digital libraries. About 80 percent of the internet and digital library resources today are available in English, while 55% of the internet users are non-English. (Bian & Chen, 2000). Researchers and engineers are thinking about ways on how to manipulate the large volume of multilingual data. There are several ways of doing this in the field of natural language processing, which you can get more information about through Chowdhury's interesting article (2003).

All of the above examples can have benefit from the results of projects like Time in Translation. This small thesis, which investigates the linguistic behavior of the Perfect in Farsi, can have a modest contribution on technologies wanting to use Farsi. The contribution can be realized by forming grammars and rules based on the context, which can help translation engines to make more human like translations, like how Heshaam Feili and Gholamreza Ghassemi (2004) have done by using Tree Adjoining Grammars (TAG) for machine translation from English to Persian.

3. Farsi

In this section I will present the grammar and make clear why it is interesting to investigate the Perfect and how this is a relevant topic in Farsi too. Giving this information, I will be able to clarify my research question, which I have shortly mentioned in the introduction.

The information about Farsi and its grammar comes mainly from the website www.jahanshiri.ir . The owner of this website is Ali Jahanshiri. At the age of 15 he had learned 20 languages by himself. He has explored different areas of linguistics and has a master degree in computer engineering. He is also interested in Artificial intelligence and computational linguistics. His website provides a Persian verb conjugator, which I will explain more in the method section. He has also presented the Persian grammar with good equivalent English examples.

3.1 About Farsi

Farsi, also known as Parsi or Persian, is one of the oldest languages of the world which remained in continues use after several thousands of years. It is the most widely spoken member of the Iranian branch of Indo-Iranian languages, a subfamily of Indo-European languages. Historically it was a more widely understood language in an area ranging from the Middle East to India. The history of the Persian language can be divided into three distinct periods: Old Persian, Middle Persian and New Persian. Old Persian dates back to about 3000 years ago. It took about 200 years to transform Middle Persian to New Persian. Hereby the starting point of New Persian is about 1200 years ago. Farsi is written in a variety of the Arabic script called Perso-Arabic, which has some innovations to account for Persian phonological differences. It is written from left to right and its alphabets contains 32 letters.

Persian is known for the world-famous poets such as Ferdowsi, Rumi, Khayyâm, Hâfez and Sa'di. New Persian is not very different from Middle Persian in grammar. The main difference is in vocabulary because after the Arab invasion of Persia, many Arabic words entered Persian.

Today, Persian is spoken and written primarily in Iran (Persia), Tajikistan and parts of Afghanistan. It has more than 130 million speakers. Persian was for a long time the lingua franca of the western parts of Islamic world and of Indian subcontinent. It also has been a medium for literary and scientific contributions to the Islamic world as well as the Western. The status of Persian was comparable to Latin in the field of literature.

3.2 Grammar

In this section I will explain the grammar. I will mainly explain the forms which have appeared in the Farsi translation, because that information is the most relevant for understanding the results. Also there are lots of forms which aren't in the scope my research question. For further information about other tenses which aren't relevant for my thesis, you can visit the website www.jahanshiri.ir

I will explain the grammar using the tense-modality-aspect, TMA. This grammatical system covers the expression of the tense, aspect and mood. The terms you will see later on in section 5.3 are of the following form: (mood + tense + aspect). In section 3.2.1 I will explain the grammatical moods and in section 3.2.2 the grammatical tense and aspect.

3.2.1 Grammatical moods

In linguistics grammatical mood is used to show the use of verbal inflections that allow speakers to express their attitude toward what they are saying (Palmer, 2001). Attitudes can be based on a truth or thought or else. So it expresses the realis or irrealis nature of an assertion and generally takes 3 forms: indicative, subjunctive and imperative. I will give a short explanation for each.

1. The **indicative** mood (Gozareshi – گزارشی) is used for factual statements and opinions. The indicative mood appears in the past, present and future. Most of the statements in Farsi are in the indicative mood. Farsi has an verbal inflection for indicating the moods (Hajatollah Taleghani, 2006). The indicative mood is shown for most of the tenses by the prefix *mi-*. In (1) you will see the indicative mood for the past simple with a case without the prefix.

   (1)

   دیروز به مدرسه نرفتم.

dirooz	be	madrese	na-raft-am
yesterday	PREP	school	not-go-IND.PAST.SIMP

   ‘I did not go to school yesterday.’

   In (2) you can see an indicative form with the prefix *mi-*.

   (2)

   سارا می تواند.

sara	*mi-tavan-ad*
Sara	IND-can-PRS.IMP

   ‘Sara can.’

2. The **subjunctive** mood (eltezami – التزامی) is the opposite of the indicative mood. This mood refers to unreal hypothesized events. It is used to express thoughts, possibility, desire etc. The prefix which indicates the subjunctive mood is *be-, bi-* or *bo-*. It is worth noting that in contrast to the English subjunctive, which is morphologically neutral with respect to the tense and aspect, the subjunctive in Farsi has both a present, and a past narrative. Present subjunctive indicates an action, in the present time or future, which may or may not happen (Hajatollah Taleghani, 2006). This is illustrated in (1).

   (1) شما می توانید این کار را بکنید.

shoma	mi-tavan-id	in	kar	ra	be-kon-id
you	IND-can-PRES.IMP	this	work	PREP	SUB-do-PRES.SIMP

   ‘You can do this (work).’

In (2) you see an example for the past narrative form, which shows an action started in the past, but the speaker is not certain whether it is completed or not (Hajatollah Taleghani, 2006).

(2) سارا شاید به مدرسه رفته باشد.

sara	shayad	be	madrese	rafte	ba-shad
sara	may	to	school	gone-PASTPARTICIPLE	SUB-be-PAST.NAR

'Sara may have gone to school.'

3. The **Imperative** mood (dastoori – دستوری) expresses commands, requests or warnings. It is used to signal a prohibition, permission or any other kind of exhortation. In Persian all persons have an imperative form. In English it limited to the second person. The imperative only occurs in the present simple. The imperative mood is made from the subjunctive mood with the difference that the ending for the second person singular is omitted. The prefix which indicates the imperative mood is *be-*.

(1) سیب بخور.

sib	be-khor.
apple	IMP[1]-eat-PRES.SIMP

'Eat the apple.'

3.2.2 Grammatical tense and grammatical aspect

The **grammatical tense** generally expresses time relative to the moment of speaking. Basic tenses are for example the past, present and future.
The **grammatical aspect** of a verb defines its temporal flow. For example the difference between 'I read', 'I am reading' and 'I have read' is in the aspect. Aspect expresses how an action, event or state denoted by a verb, extends over time. For example distinctions can be made on states which are ongoing, repetitive or aren't relevant anymore.
I will give a short explanation followed with some examples of the most appeared tenses in the translation.

1. Gozashte sade - *(past simple)*
This tense is used to talk about an action that was started and finished at some point in the past and occurs with the indicative mood. In (1) we see the simple past tense

(1) دیروز یک فیلم دیدم.

dirooz	yek	film	did-am.
yesterday	one	movie	see-IND.PAST.SIMP.

'I saw a movie yesterday.'

2. Gozashte naghli – *(past narrative)*
The narrative aspect narrates a state resulting from a previous action, therefore it is only possible in the past. There narrative aspect isn't found in English. There are different types of

[1] In all the other examples I use IMP as an abbreviation for imperfect. In this case IMP is the imperative mood.

the narrative, and some have an equivalent in English. The indicative past narrative which I will explain in a moment is similar to the Present Perfect in English.
But English doesn't have an equivalent of any other narrative tense of Farsi (there are a total of 10 types of the narrative, which I won't discuss in this thesis). For example:

(1)
آن گفت همه ی اتاق ها را گشته بوده.

Ann	goft	hameye	otagh ha	ra	gashte	boode
Ann	say.IND.PAS.SIMP	all	room.PLURAL	PREP	search.IND.PAST.PREC.NAR	

'Ann said she had searched all the rooms.'

In (1) we see that the Farsi sentence is expressed in the indicative past precedent narrative, whereas it is expressed with the Past Perfect in English, which is just the same as the English equivalent for the indicative past precedent.

Back to the variant which does have an equivalent in English: the indicative past narrative with the equivalent Present Perfect. The following is an example of the past narrative:

(2) من به فرانسه رفته ام.

man	be	faranse	raft-eam.
I	PREP	France	go-IND.PAST.NAR

'I have been to France.'

3. Gozashte natamam – *(past imperfect)*
This tense is also known as *'Gozashte estemrari'* and has like any other Imperfect tense no equivalent in English. It can be translated with different English tenses based on the context. However in most cases its perceived as the Past Continuous or Past Simple. It is used to talk about an action that was taking place at a given time in the past. In the context, the subject will be in the middle of doing the action at the given time in the past. The formation of the past imperfect is as the following: *mi* + past stem + past ending.
To show the different ways the Past Imperfect in Farsi can be translated into English, I will give two examples:

(1) دیشب من کتاب می خواندم.

dishab	man	ketab	mi-khand-am.
last night	I	book	IND-read-PAST.IMPF

'Last night I was reading a book.'

In (1) we see that the Farsi sentence is translated by using the Past Continuous in English and in (2), the Past Simple is used.

(2) همیشه می خواست انگلیسی یاد بگیرد.

hamishe	mi-khast	engelisi	yad be-girad
always	IND-want-PAST.IMP	English	SUB-learn-PRES.SIMP

'He always wanted to learn English.'

4. Hal-e na tamam – *(present imperfect)*

This tense was a bit tricky for me to understand. According to the grammar of Ali Jahanshiri, a distinction is made between the indicative present simple and the subjunctive present simple. However, when asking my parents, they told me that it is true that there is a difference in mood, but in usage, the indicative present simple is never used (so the distinction made by Ali Jahanshiri was a bit unusual to them). But this distinction, has caused a vague termination. The verb tense he calls a Indicative Present Imperfect, should actually be the indicative Present Simple. The semantics of the Present Imperfect, is actually what he explains for the Indicative Present Progressive. I will show what I mean in figure 1:

INDICATIVE PRESENT

~~SIMPLE~~		SIMPLE ~~IMPERFECT~~		IMPERFECT ~~PROGRESSIVE~~	
man	ravam	**man**	miravam	**man**	dâram miravam
to	ravi	**to**	miravi	**to**	dâri miravi
u	ravad	**u**	miravad	**u**	dârad miravad
mâ	ravim	**mâ**	miravim	**mâ**	dârim miravim
shomâ	ravid	**shomâ**	miravid	**shomâ**	dârid miravid
ânhâ	ravand	**ânhâ**	miravand	**ânhâ**	dârand miravand

Figure 1: What a Farsi speaker would expect ofr the terminology.

Luckily, he does say that the English equivalent is the Present Simple. But as a Farsi speaker, it was very confusing to see this as an Imperfect!

Going back to the definition of Ali Jahanshiri, this tense is not technically equivalent to the English Present Simple, but its usage covers it. For example ‘mi-ravam’ which has a Present Imperfect tense in Farsi, can mean ‘I go’ and ‘I will go’. I show this through examples:

(1) هر روز به آنجا می ریوم.

har rooz be anja mi-ravam.

everyday PREP there IND-go-PRES.IMP

‘I go there every day.’

In (1), ‘mi-ravam’ is translated as ‘I go’, so as the Present Simple.

(2) فردا به تو زنگ می زند.

farda be anja mi-ravam

tomorrow PREP there IND-go-PRES.IMP

‘I will go there tomorrow.’

And in (2), ‘mi-ravam’ is translated as ‘I will go’, so as the Indicative Future in English. So we’ve seen how one Present Imperfect verb can be translated in two different tenses in English based on the context. However, as I have already mentioned, it usually covers the English Present Simple.

5. Hal-e sade – *(present simple)*
This tense is used to talk about a state of being or action being performed at the present time. Or about repeated or habitual actions, general truths, or any other action that you are going to do in the future. The equivalent in English is also the Present Simple.

(1) میز سنگین است.
miz sangin ast.
table heavy is.IND.PRES.SIMP.
‘The table is heavy.’

6. Gozashte pishin – *(past precedent)*
The precedent aspect is used to refer to an event that has completed before another past action. Or to describe an event that has not yet happened but which is expected or planned to happen before another stated occurrence. In both cases it precedes another action and that is why this aspect is called *precedent* (gozashte) in Farsi. The auxiliary verb ‘boodan’ (to be) is used in the Past Simple after the past participle of the main verb.

The past precedent occurs in two moods: indicative and subjunctive. With the indicative, next to the plain past precedent, there are lots of other options which I will only mention: past precedent imperfect, -progressive, -narrative, -narrative imperfect and -narrative progressive. For the subjunctive mood, there are two options: subjunctive pas precedent narrative and the subjunctive past precedent narrative imperfect. For more information about each type, you can look it up on the website of Ali Jahanshiri.

Only the plain indicative past precedent has appeared in the translation, so I will only provide an example for this form in (1).

(1) فراموش کرده بودم.
faramoosh karde bood-am
forget do-PASTPARTICIPLE have-IND.PAST.PREC
‘I had forgotten.’

We see that the English equivalent for this tense is the Past Perfect.

3.3 Literature

So far you've been introduced to some key elements of the grammar of Farsi. Also several possible uses of this research investigation have been discussed in section 2.2. To be able to form a more specified expectation of what the results might be and mean, I find it useful too look at some studies done on Farsi, which might be relevant for my research topic. This may be from papers which are written in Farsi.

To start off, I find it refreshing to look at the problem from an interhuman perspective, where a translation error occurs for a person who is learning a new language. Falk (1978) discusses that when learning a new language, the learner will compare the new grammar to the grammar of the language he/she already knows. This process of learning will not be without errors. As the comparisons will cause the learner to not be able to distinguish between cases were a specific tense can or cannot be used. By looking at the most made mistakes by learners, linguists can work on those grammars to optimize the learning process.

Ayati and Manouchehri (2013) have done this for Farsi speakers who are learning French whilst knowing English. They took a survey where equivalent French and English sentences were presented with a gap in the sentence. It was known which verb had to be filled in the gap. The participant had to choose the verb tense.
Results show that Farsi speakers often used the *imparfait* instead of the *passé compos*é. They suggest that the interaction of the Farsi grammar was the reason for this choice. I will present an example:

> *Hier à l'examen, je (ne pas savoir).....répondre à la question n°3.*

The answer which has to be filled in, is 'n'ai pas su', a *passé composé*. But all the participants chose the *imparfait.* The participants were also asked to fill in the gap for the English equivalent for this sentence:

> *Yesterday, at the exam session, I (not to know).. how to respond to question number 3.*

All the participants filled in the correct answer: 'didn't know'. So they had used a Past Simple. According to the paper, the Past Simple is one of the equivalents for the *passé compos*é in English. Another equivalent is the Present Perfect. (This diversity in equivalents shows, again the ambiguity of the semantics of the Perfect in each language!)

After the survey was taken, the examinators asked a few participants to explain their choice for the *imparfait* in this sentence. Most of them thought considering the Farsi translation made them choose for the *imparfait*.

In the following example we see the equivalent of the French sentence in Farsi. We see that the verb is translated to an Indicative Past Imperfect.

دیروز در امتحان، جواب سؤال 3 را نمی دانستم.

dirooz	dar	emtehan	javab	soal	3	ra	ne-mi-danest-am
yesterday	at	exam	answer	question	3	PREP	not.IND-know-PAST.IMP

Ayati and Manouchehri suggest that considering the Farsi translation caused the participants to make the incorrect choice of using the *imparfait.* If the students had considered the English sentence as their 'example', they would fill in the right answer, using the *passé composé.* Since an equivalent of the *passé compos*é was used in the English example.

The influence of already known grammars on forming translations is shown very well in this study. Because this study is done between French and Farsi, it is a good tool to understand the relationship of Farsi and French in the setting of translations. Furthermore, it also shows that the ambiguity of the semantics of the perfect is relevant among the Farsi speakers. This is an example how including Farsi in the project, can contribute to educational purposes!

In section 2.1 I had presented some examples on how the perfect is ambiguous in English. It is interesting to know that this ambiguity also occurs in Farsi. Partly inspired by de Swart, Rahimifar, Rezai and Motavalian (2017) have done a comprehensive analysis on the perfect in Farsi, which considers its different behaviors based on the tense-modality-aspect (TMA) which I have already introduced in section 3.2. Because the analysis is mainly based on the tense-modality aspect and is more an issue among the linguists only considering the semantics of the grammar of Farsi, I won't discuss it further.

4. Hypothesis

So far I have tried to seek a deeper understanding on the problem by demonstrating some research done on this subject, both in English and Farsi. I have also explained the grammar relevant to my dataset. As you may have noticed, most of the information I have found, is about French and English. Therefore I will form my hypotheses on the comparison of Farsi with English and French, and gradually compare the other European languages with it too.

1. In section 3.2.2 I have mentioned the English equivalents of the Farsi verb tenses. We should consider those as assumptions, because the actual aim of the project is to reconsider the semantics of the tenses. Therefore I will use the equivalents Ali Jahanshiri has provided, as my first hypothesis. I have listed the Farsi tenses, both with the original name as the translated names, paired with the English equivalent in table 1. I have to mention that I made a choice for the equivalent of the Indicative Past Imperfect. I had mentioned in section 3.2.2 that this form has 2 equivalents in English; the Past Continuous and the Past simple. In my opinion the Past Continuous is more likely to be its equivalent, because it includes the ongoing aspect of the form more.
   However the Present Perfect is our subject of investigation, I have colored them blue in table 1. So I expect the equivalent of the Indicative Past Narrative to be the Present Perfect in English.

Translated name	Farsi name	English equivalent
Indicative past simple (IPAS)	Gozareshi gozashte sade	Past simple
Indicative past narrative (IPAN)	Gozareshi gozashte naghli	**Present perfect**
Indicative past precedent (IPAP)	Gozareshi gozashte pishin	Past perfect
Indicative past imperfect (IPAI)	Gozareshi gozashte natamam	Past continuous
Indicative present imperfect (IPRI)	Gozareshi hale natamam	Present simple
Indicative present simple (IPRS)	Gozareshi hale sade	Present simple

Table 1: A table with the Farsi tenses with the equivalent English tense..

2. Knowing the novel L'Étranger is chosen because of its generous use of the *passé composé*, I don't expect that the perfect will appear as much as it does in the novel. Based on the literature I have read, I know the perfect is ambiguous, so I do expect a variety of tenses used for the *passé composé*. Also considering the many tenses Farsi has, the probability that a variety of tenses will show, isn't unlikely.

3. Reflecting on the study of Ayati and Manouchehri (2013) on the often made mistakes in French by Farsi speakers (which I have explained in section 3.3), we can understand that a *passé composé* sentence, is translated by Farsi speakers as an Imperfect in Farsi. However, their research was about recognizing the right tense in the French sentence based on comparisons made with Farsi. And here I am backtracking the understanding of a *passé composé* in Farsi, which may be an Imperfect! We will see!

4. I expect most of the tenses to have the indicative mood. That is because L'Étranger is a narrative novel, which means that it talks about events in the past. And the indicative mood is the mood denoting statements. Additionally, Nishiyama and Koenig (2010) have mentioned that it is unclear what the function of the perfect actually is, for example why authors choose it over a Simple Past form. This statement which I had already mentioned in section 2.1, makes me think that the (indicative) Past Simple will be used in the Farsi translation. Especially because Ayati and Manouchehri (2013) had mentioned too that the equivalent for the *passé composé* is the Past Simple..

5. Method

In the following sections I will describe each step I have taken for my analysis. Along with explaining my steps, I will also talk about my experiences an show some examples on cases I found hard to handle.

5.1 Corpus: Farsi translation

The analysis of the Perfect, is being done by using a French novel which includes lots of Perfect forms. The novel is called L'Étranger and is written by Camus. First of all I had to find whether the novel was translated in Farsi and if I could get my hands on an online version. I found a translation which is done by Jalal Al-e-Ahmad who is very well-known among Farsi speakers. He was a Iranian novelist, short story writer, translator, philosopher, socio-political critic, sociologist as well as an anthropologist (Ahmad, 1984). So I am sure I got a reliable translation. Reading the translation I noticed that the translation style was quite formal, while the French original is known to be very easy to read.

5.2 Paragraph alignment

I had to format the e-book so the parser could accept it as input. First I converted the pdf to a .txt document using an online converter. During this process some words missed out, which I added by hand.
After I had made sure the text was complete and all the missed words were added, the next step was to paragraph-align the text with the original French version. This means wherever the original has a paragraph break, there should be a paragraph break in the translation too.

Doing this was a bit hard at first, because Farsi is written from right to left. It took me a bit to figure out that I had to add Farsi as a language to word in order to be able to type directly in the word document and get the right response when using the backspace, enter and space key.

5.3 Time-Align

Assigning the verb tense was done using the time-Align software the Time in Translation website provides. The time-Align program puts fragments of the original text next to the corresponding fragment of the translation. In the original fragment, the *passé composé* is marked green. I had to mark the same verb in the Farsi translation. The French *passé composé* is found by using an algorithm which is designed to recognize this verb tense (Van Der Klis, Le Bruyn & De Swart, 2017). Annotation by hand for the Farsi verbs is a good strategy, because algorithms for verb detection in Farsi aren't that well developed yet (Iranpour Mobarakeh & Minaei-Bidgoli, 2009). You can see this setup in figure 2. There I've already figured out what the translation of '*a dit*' is, and I have marked the corresponding verb in Farsi green. When clicking on *submit* the annotation gets saved.

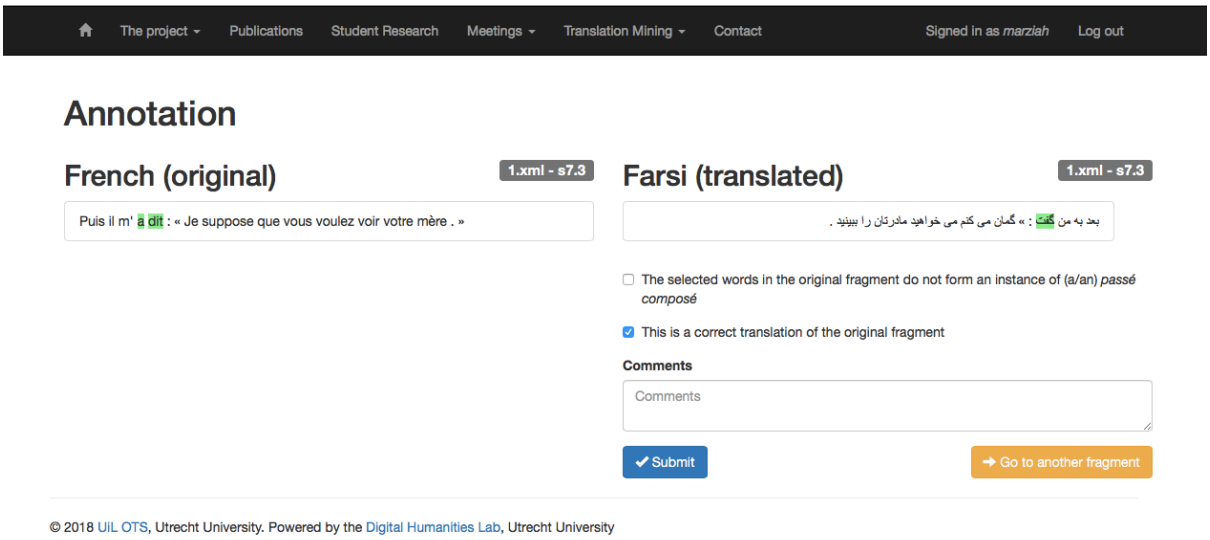

Figure 2: Screenshot of the time-align setup, French to Farsi

I found it quite hard to find the right verb in Farsi. Most of the Farsi sentences were translated in a different manner by choosing different words or using Iranian expressions. Or some French expressions had to be translated in Farsi. To understand the French expressions I used a context converter which was very helpful. As a non-French speaker, translating the sentence word by word wasn't a good strategy to fully understand the sentence. That's why the context converter helped me a lot to understand the sentence fully in order to be able to mark the right verb in Farsi. Because in some cases, more verbs seemed to represent the marked *passé composé*.

Some other difficulties I had faced during this process was the fact that I only could mark words, not letters. In Farsi a negation is included in the word, which I was not able to extract. I also came across fragments were the manner in which it was translated excluded a verb. In these cases I gave a short explanation in the comments. All the comments I have made are saved in the Time in Translation website and can be retrieved. An example is the following sentence from chapter one of the novel:

(1)
a. French:
…et ma joie quand l' autobus **est entré** dans le nid de lumières d' Alger et que j' ai pensé que j' allais me coucher et dormir pendant douze heures .

b. English:
..and my joy when the bus **entered** the nest of lights which was Algiers and I knew I was going to go to bed and sleep for a whole twelve hours.

c. Farsi:
...و خوشحالی من هنگام ورود اتوبوس به روشنائی های الجزیره و فکر اینکه دوازده ساعت تمام خواهم خوابید
'…and my joy during the entering of the bus in the nest of lights of Algiers, and the thought that I will sleep for whole twelve hours.'

The verb which I had to find in the Farsi translation is in bold. The sentence under the Farsi sentence in (1.c) is a translation in English of how the original text is translated in Farsi. We can see that the verb 'enter' is translated as a noun, 'the entering'. In these cases the translation was right, but I couldn't assign the verb. As you can see in figure 2 there are 2 checkboxes. In the just described case, the Farsi fragment was *'a correct translation of the original fragment'*. When checking this box, you **have to** mark a word in the Farsi fragment, otherwise you will not be able to submit your answer. This shortcoming of the program forced me to mark some random word, just to be able to submit my answer. This random marked words would be extracted from the analysis in the next step, but it was just an unnecessary step which could have been avoided.
There were also cases where the wrong sentence was matched with the original fragment. Sometimes only a part of the correct sentence was missing and sometimes the wrong sentence was matched. In these cases I unchecked the second checkbox. There was one case where the Farsi translation was incorrect. A verb missed in a sentence where the grammar didn't allow it to be absent. This mistake appeared in chapter one in the following sentence:

(2)
a. French:
Je me suis retourné et j' **ai vu** le vieux Pérez à une cinquantaine de mètres derrière nous.

b. English:
I turned round and **saw** old Pérez about fifty yards behind us.

c. Farsi:
به عقب برگشتم و پرز را که پشت سر به فاصله پنجاه متری می آمد.
Be aghab bargashtam wa perez ra ke poshte sar be fasele panjah metri mi amad.
'I turned around and Pérez which was coming about fifty meters behind us.'

In the English equivalent of the Farsi translation in (1.c) it is shown that the verb 'see' is missing. This is not a style choice or an adaption of the sentence, because according to the grammar after '*ra*' a verb has to follow. You can compare '*ra*' with 'which'. When reading the Farsi translation, the question will immediately arise for the reader: 'Pérez what?'.

5.4 Assigning verb tense

After I finished the time-Align, my annotations were added to an excel document, which you can see in figure 3. I dealt with the shortcoming of being forced to mark a word in a sentence which is the right translation but the right verb is excluded due to the way of translating, by checking whether I had left comments. However I also noticed immediately by seeing the random word as a match for the French verb. For cases where the wrong fragment from the translation was matched with the French fragment, you see a 'No' in column E. In column B the verb tense of the annotated Farsi verb were added. In figure 3 I have already done that.

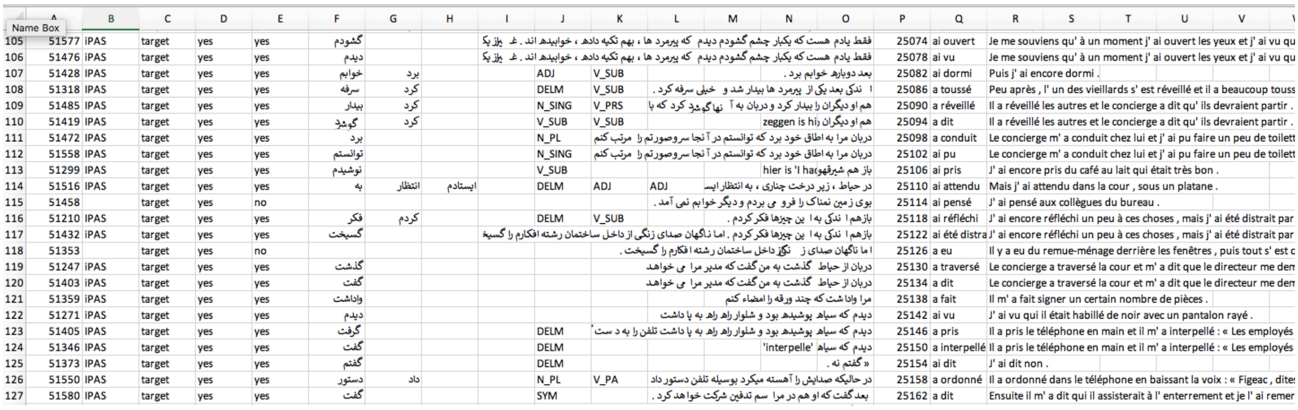

Figure 3: Excel document used to assign the verb tense.

The verb conjugator of Ali Jahanshiri helped me a lot with doing the time-align fast and correct. First I had to subtract the verb to the infinite form. Giving that as an input to the verb conjugator, it gave all the possible conjugations for different combinations of tense aspect and tense mood. In figure 4 you can see the layout of the verb conjugator.
Using a conjugator reduces the chance on mistakes and errors. Plus it gives a consistent and reliable output, based on the definition of the grammar I have used as the source. What I found very handy was that you could choose to type the Farsi word with Arabic letters or with Latin letters. This made it easier for me because I am not used to typing Farsi on a computer at all.

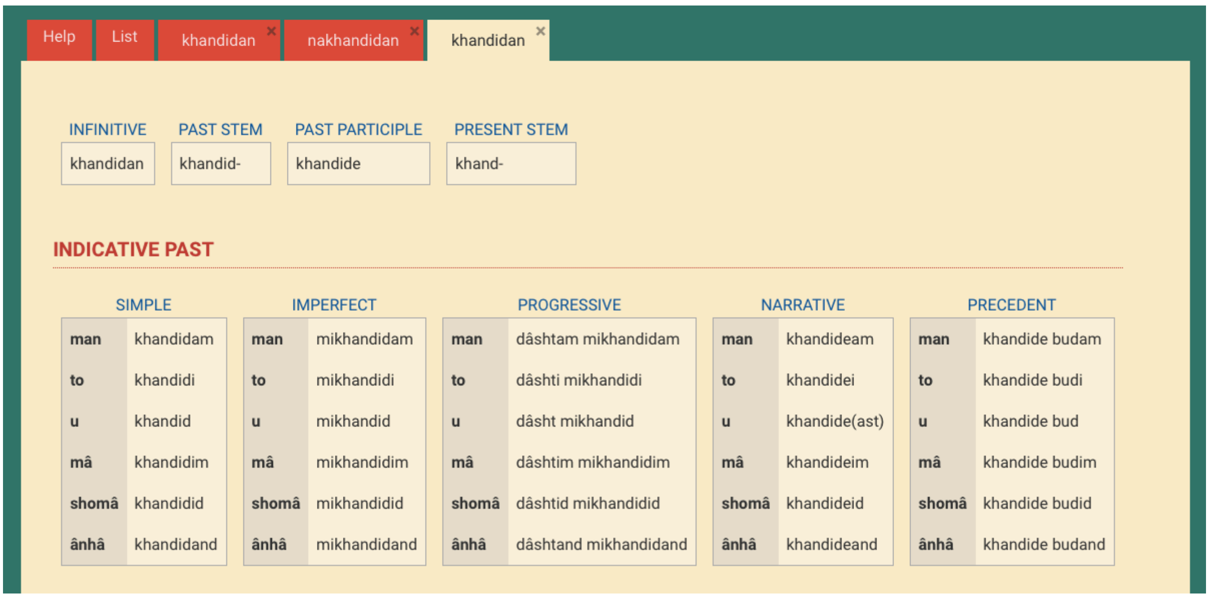

Figure 4: The verb conjugator of Ali Jahanshiri in English. Using the conjugator for the verb 'khandidan' which means laughing, retrieved from www.jahanshiri.ir .

After finding the right verb tense, I abbreviated the name of the tense and added it to the excel file. In the following table you can see the tenses which have appeared. The color labels for each tenses remain the same among the different languages. So for example the perfect tense in each language is colored blue. In Farsi that is IPAN. This consistency in color labels enables multilinguistic comparisons.

Abbreviation	Translated name	Farsi name	color
IPAS	Indicative past simple	Gozareshi gozashte sade	Dark green
IPAN	Indicative past narrative	Gozareshi gozashte naghli	Blue
IPAP	Indicative past precedent	Gozareshi gozashte pishin	Red
IPAI	Indicative past imperfect	Gozareshi gozashte natamam	Yellow
IPRI	Indicative present imperfect	Gozareshi hale natamam	Orange
IPRS	Indicative present simple	Gozareshi hale sade	Orange

Table 2: The tenses which have appeared in the corpus, with the abbreviation used in the excel files and semantic maps .

In total there were 64 cases the parser had failed to find a match with the right Farsi sentence. For all these cases I found the right sentence by hand and added the Farsi sentence to the excel file. This took quite a while to do, but it was necessary, because the more data points included in the semantic map, the more accurate the result. After all, there were only 15 cases, where no corresponding verb in the Farsi sentence existed. So from the 391 *passé composé* occurrences in the corpus, I found 376 corresponding verbs in the Farsi translation

I found it unhandy that there wasn't a similar key for a sentence in the excel file and in the time-align program. In the time-align program it was indicated from which chapter the sentence is from. In the excel file each sentence is given a code, which isn't in line with the code a sentence is given in the time-align program. When I wanted to find the right sentence by hand, I could easily search for it in the right chapter. But when adding it in the excel file, I had to search for the right sentence again. It wasn't a very big problem though because with the ctrl+f operation its easier, but it would be more convenient if both excel and time-align shared the same key for a sentence.

5.5 Semantic maps

Results will be visualized using semantic maps. After the tense annotation, a set of six languages (including Farsi) are chosen through which a multilinguistic comparison of the behavior of the perfect can be made. In order to do this six tuples are generated for the six chosen languages. An example of a 5-tuple is shown in figure 6. When the verb tense is translated to the same tense in all the languages, the distance is set to 0 using a simple distance function (Van Der Klis, Le Bruyn & de Swart, 2017). If the tense isn't the same among the 5 languages, for each mismatch 1 is added and the sum is divided by five. This distance function is used to create a matrix. The decision is made by the authors to only include tuples which have a translation in each language. This is also one of the reasons I decided to find the missing sentences the parses couldn't find, by hand! Hence as many tuples could be included as possible.

The dissimilarity matrix can be plotted using the multidimensional scaling resulted in semantic maps. In figure 5 an example of a semantic map is shown.

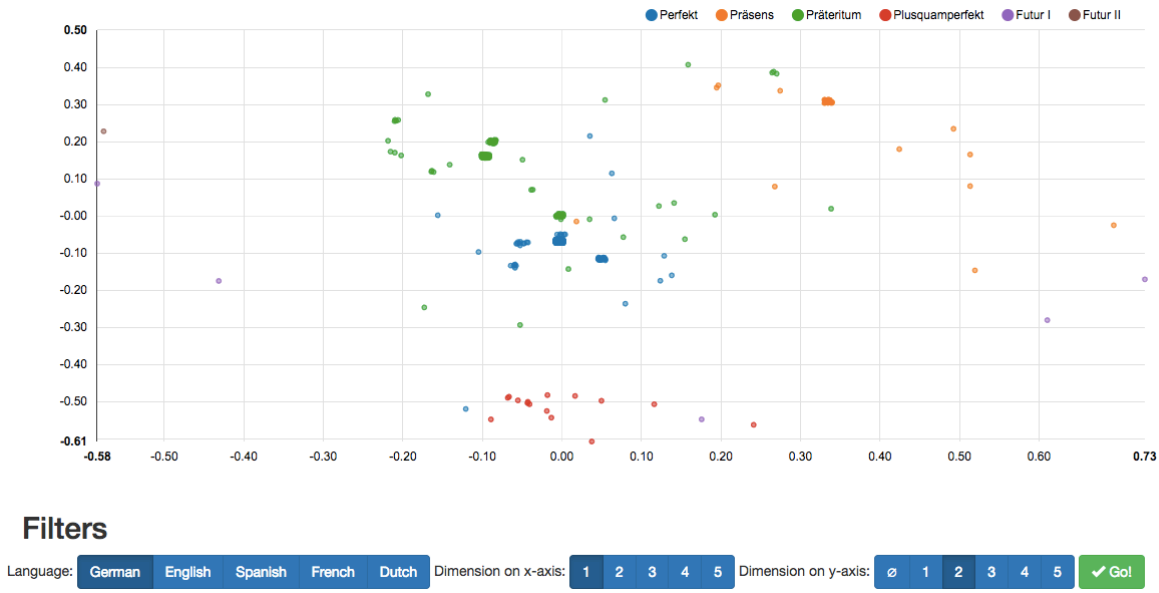

Figure 5: A semantic map. The points are labelled using the tenses of the selected language. Users can also change the dimensions shown.

In figure 5 the space the various tenses occupy on the map is shown. This way of visualization makes it possible to get a feel how tenses interact with language in one glance. The color labeling on the right top of figure 5 remains constant between languages. For example the label of the German *perfekt* which is colored blue, will also remain blue for the English perfect. This allows for comparison between languages. By clicking on a label, you can remove the set of tenses it stands for. So clicking on the blue label will delete all the blue points from the plot. This feature can be useful when you want to focus on a specific tense. Manipulating the dimensions which you can see at the bottom of the plot, facilitates interpretation. Holding the mouse on a point in the plot, will show the 5-tuple the data point is based on. Next, clicking on that point will show a fragment overview. You can see these two actions respectively in figure 6 and 7.

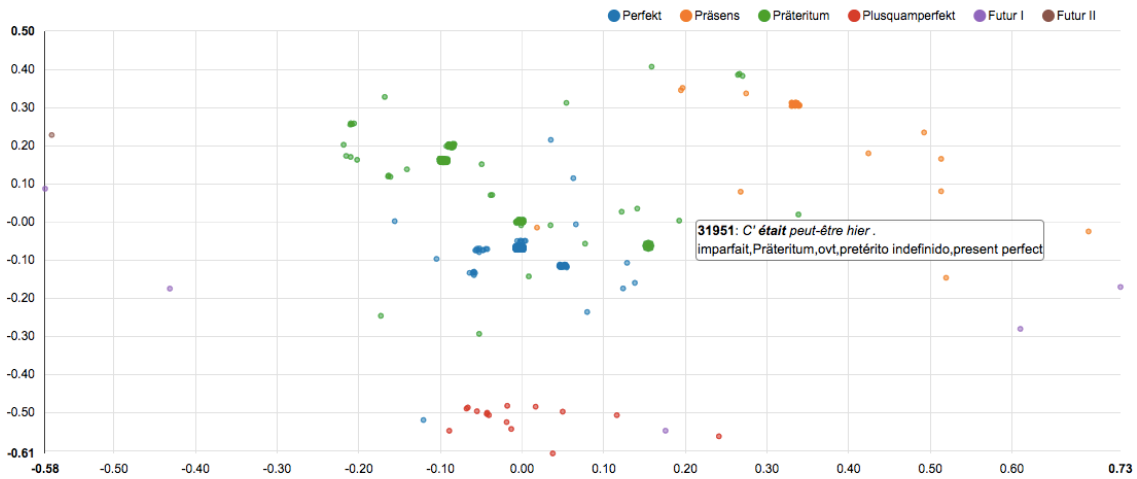

Figure 6: How the interface looks like when you rest the mouse on a point. The 5-tuple of that data point is shown.

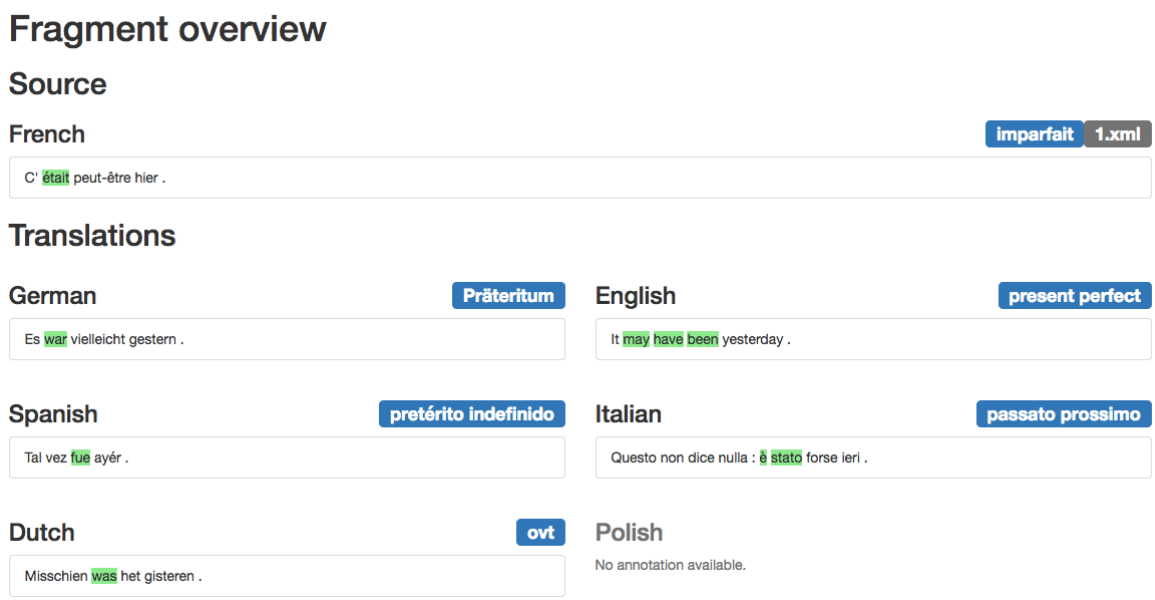

Figure 7: Detailed view of the 5-tuple after you click on a point. The Perfect is marked in green.

6. Results

Now it's finally time to show the results after all the steps I have taken. The results are produced by the TimeMapping algorithm on the first three chapters of L'Étranger. I've decided to compare Farsi next to French and English, to German, Spanish and Dutch too. So this analysis contains in total 6 languages and will produce 6-tuples.

In the following figures you will see the semantic map for each language. All the maps are shown with the x-axis on 1 and y-axis on 2. Sometimes a group of dots will appear as a cluster. That is because corresponding the 6-tuples have distance close to 0 and will come together.

I will first start off with a quantitative analysis of the results. In figures 8 up to and including 13, you see the semantic maps of all the six languages. I have marked the existent clusters using circles. With each map I have also provided a table with the frequencies of the tenses. I have placed the table on an empty place on the maps, so there are no data-points hidden behind the tables! In table 3, which you will find after the semantic maps, the most appeared 6-tuples are listed.

I have to note that I have added the semantic map of French just to help the reader to create a visualization of the problem. Otherwise it is obvious that all the points in French are colored blue, because we're investigating the translations of the *passé composé!*

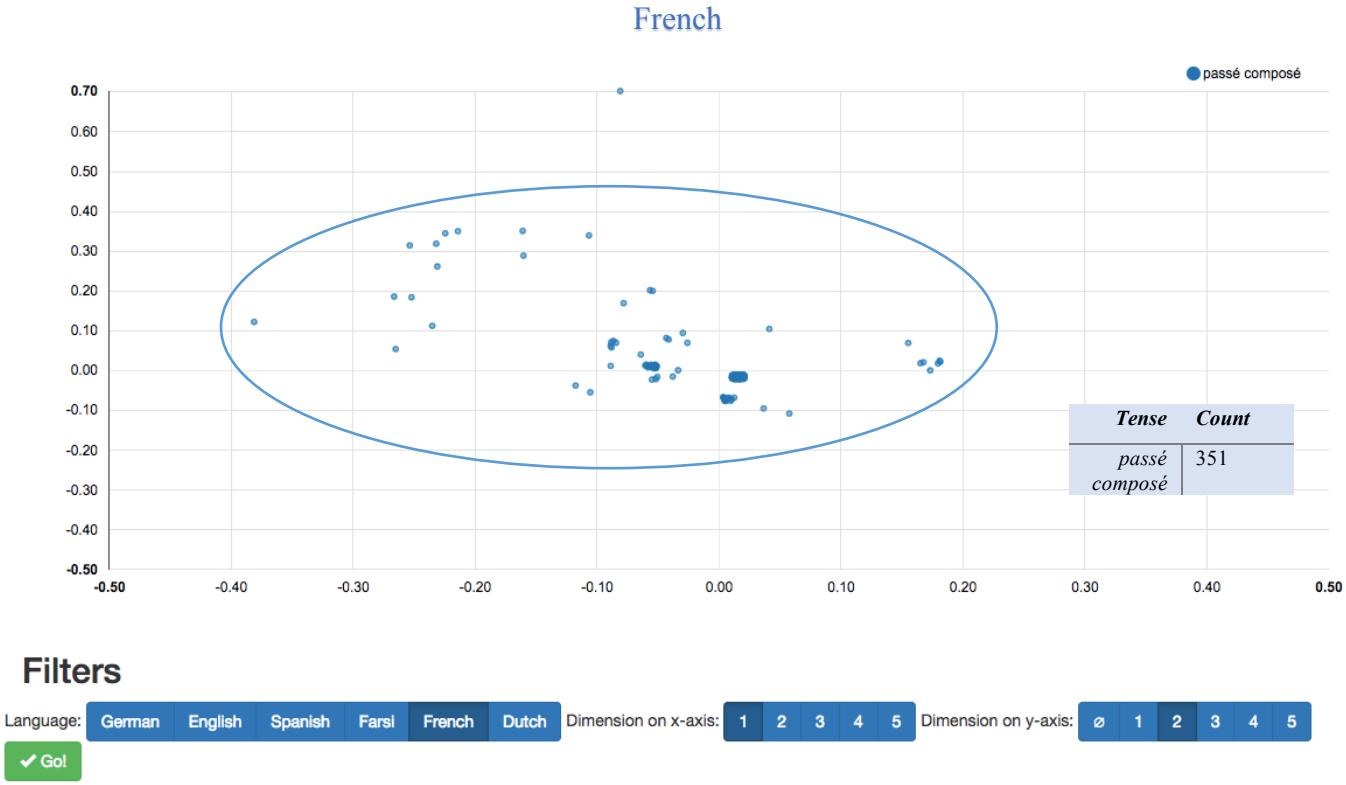

Figure 8: Semantic map for French.

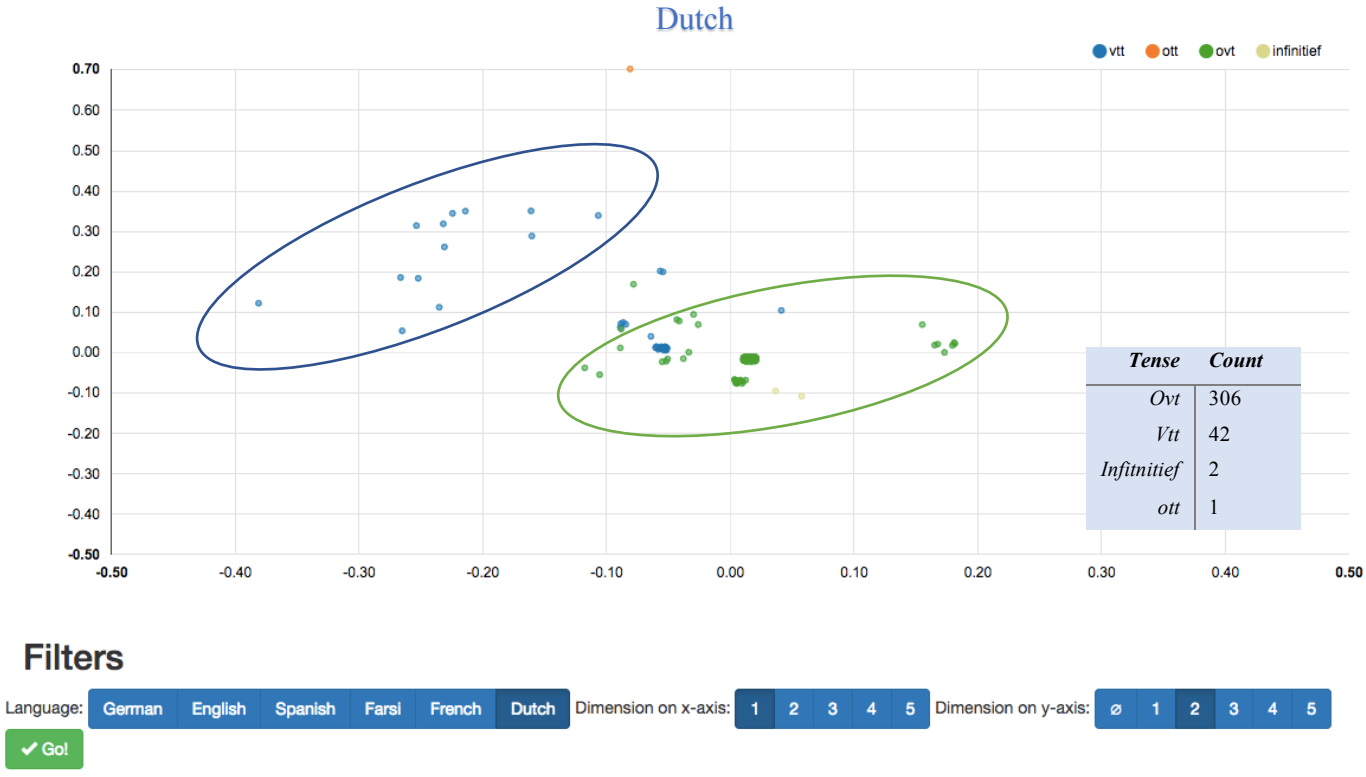

Figure 9: Semantic map for Dutch.

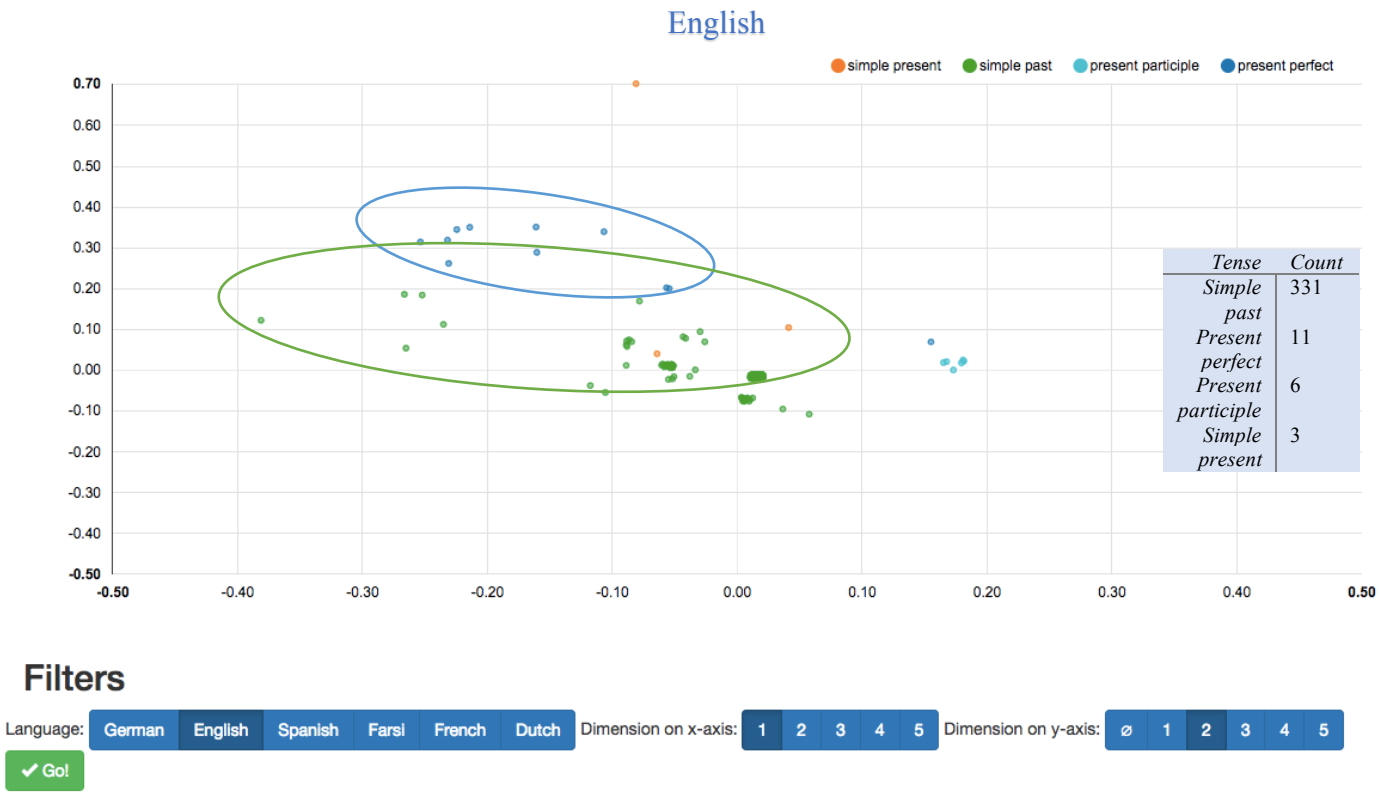

Figure 10: Semantic map for English.

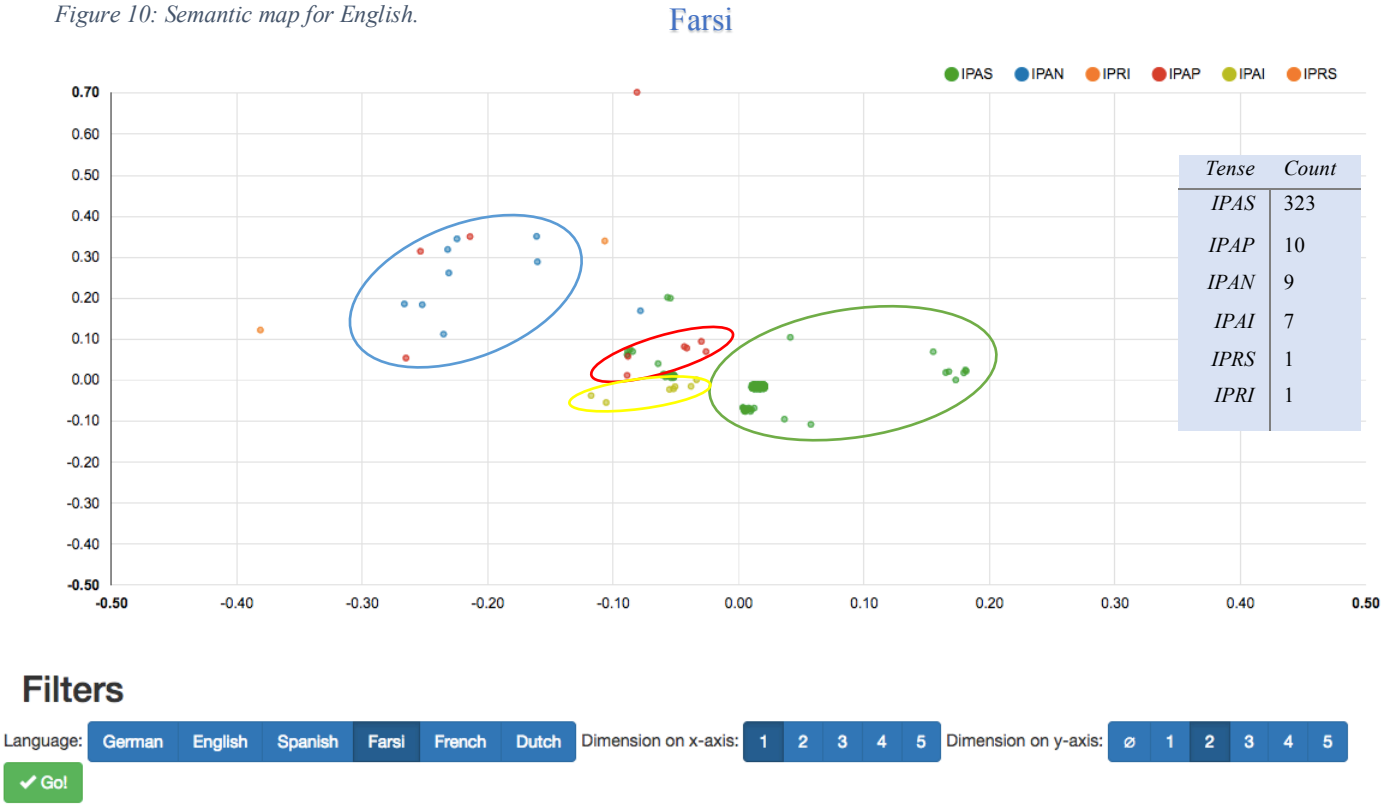

Figure 11: Semantic map for Farsi.

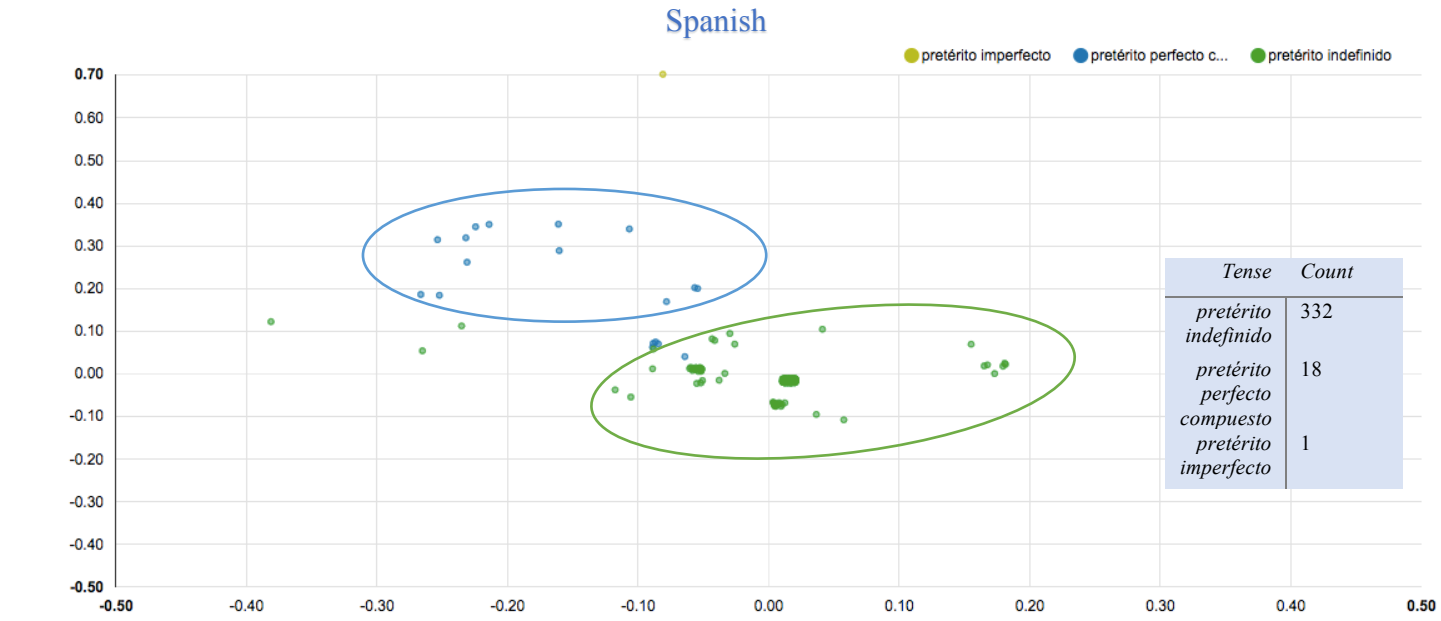

Figure 12: Semantic map for Spanish.

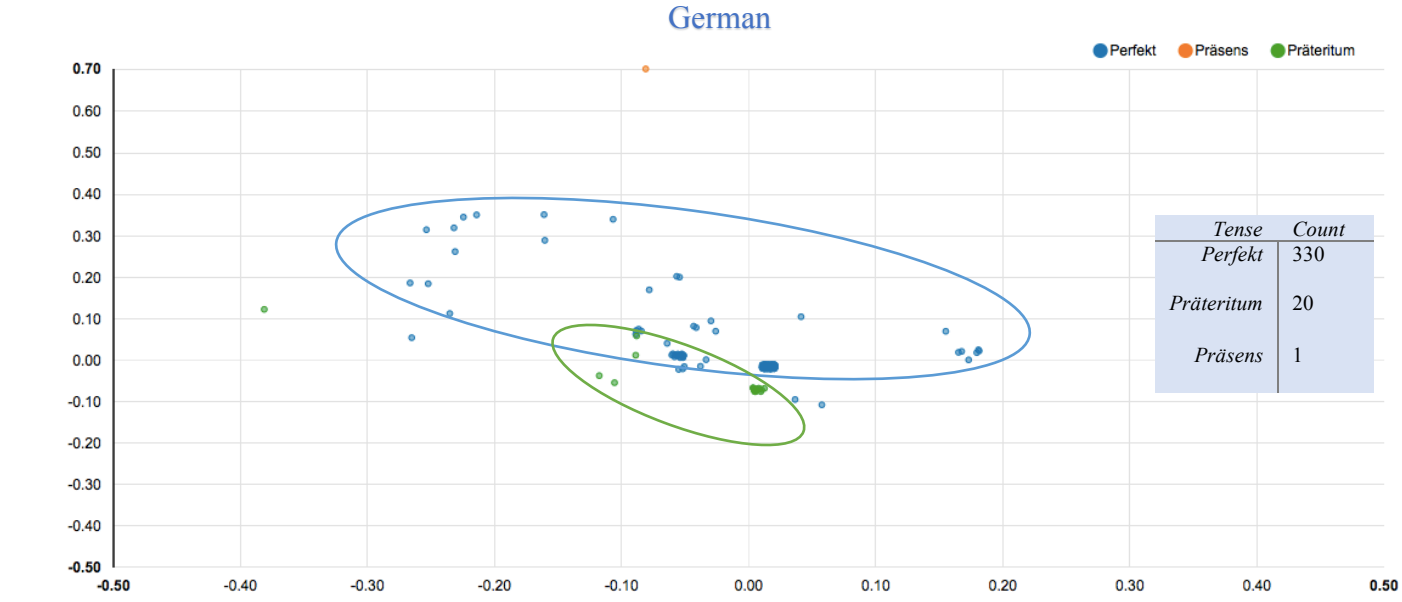

Figure 13: Semantic map for German..

Tuple (German, English, Spanish, French, Dutch, Farsi)	**Count**
Perfekt, simple past, pretérito indefinido, passé composé, ovt, IPAS	270
Perfekt, simple past, pretérito indefinido, passé composé, vtt, IPAS	21
Präteritum, simple past, pretérito indefinido, passé composé, ovt, IPAS	15
Perfekt, present participle, pretérito indefinido, passé composé, ovt, IPAS	6
Perfekt, simple past, pretérito indefinido, passé composé, ovt, IPAI	5
Perfekt, present perfect, pretérito perfecto compuesto, passé composé, vtt, IPAN	5
Perfekt, simple past, pretérito indefinido, passé composé, ovt, IPAP	4

Table 3: The 7 most appeared 6-tuples. The complete table is accessible from the Time in Translation website.

7. Discussion

In this section I will come back to my main research question and discuss the expectations based on the results. The aim of my research was to compare the behavior of the *passé composé* translated in Farsi to a set of 5 Indo-European languages, with my main focus on French and English. In section 7.1 I will analyze and compare the plots in general and mention what I find remarkable. In section 7.2 I will come back at the hypothesis I had made in section 4 and discuss each of them.

I find it clarifying to mention again, that only the cases will be used, where the *passé composé* is translated into a verb in each of the including languages. Because only in those cases 6-tuples can be formed. As mentioned in section 5.2, I was able to find 376 corresponding verbs in the Farsi translation. As you can see at the frequency tables of each semantic map, the total amount of verbs included, is 351. So among the six languages, it was possible to form 6-tuples in 351 cases.

7.1 Overall look

I will compare the semantic maps by looking for overlap or a clear distinction between clusters. At the first glance, we see that all the languages mostly have blue and green in the map, indicated by the circles. It is only Farsi which shows a variety of colors. For all the languages except German, green is the most occurred color.

In between the blue and green circle in the map for Farsi in figure 11, we see a kind of an interzone, where different tenses occur. In this zone, 2 clusters can be distinguished; IPAP and IPAI. Paying attention to this zone for Dutch, Spanish and German, we see that the green and blue dots alternate and don't form a clear cluster like in the blue and green circles.

Also when looking at the distribution of the points on the map, we see an outlying point at the top. Let's look at the 6-tuple of this point in figure 14, to understand its distribution better:

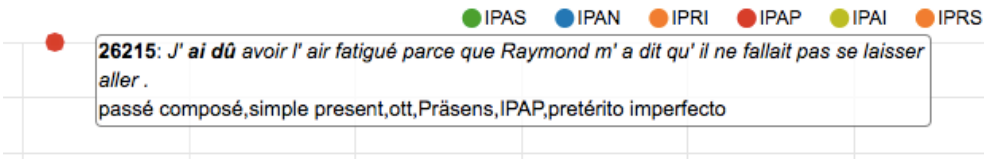

Figure 14: The 6-tuple of the outlying point shown in the semantic map for Farsi..

We see that '*ai dû'* is translated variety of tenses. In Farsi its translated into IPAP. When looking at figure 11, we see that IPAP is also distributed along the cluster for IPAN, and shows similarity in behavior to IPAN.

Looking closer at the semantic map for Farsi, we can see that the Perfect forms, indicated by the blue dots, aren't that frequent. It is remarkable that only in 2 cases a Present tense is used, indicated by the two orange dots. The other points are all in the Past tense. We see that the most used tense is IPAS, with an occurrence of 323 with the green label. After this, the 3 most used tenses are the IPAP (occurred 10 times), IPAN (occurred 9 times) and the IPAI (occurred 7 times). The first two tenses have the perfect as an equivalent in English, see table 1. IPAP has the Past Continuous as equivalent and IPAN the Present Perfect. In section 7.2 I will find out whether these assumptions are in line with my findings.

7.2 Back to the hypothesis

Let's go back to my expectations of section 4. I will briefly summarize them here first:

1. A variety of tenses will occur in Farsi.
2. a. The equivalent for IPAN is the Present Perfect in English.
   b. The equivalent for IPAS is the Past Simple in English.
   c. The equivalent of IPAI is the Past Continuous in English.
3. The *passé composé* will be translated as an Imperfect in Farsi a considerable amount of time.
4. a. The indicative mood will appear the most.
   b. The Past Simple will be the most used tense in the Farsi translation.

The **first** hypothesis is an easy one to consider. Only by looking at the semantic map for Farsi, we can see the variety of tenses it has compared to the other languages. In the maps of the other languages we barely see any other color than green and blue. But in Farsi we see red, yellow and orange too!

There are two ways to consider whether the assumptions in the **second** hypothesis are correct. First we can just look at the maps and see how frequently the color for the given tense is used for the same dots in both languages. To illustrate (2.a) you have to look how frequent the dots, which are colored blue in the map for Farsi, are blue in English too. The second method, which is less visual, is to look at the frequency table of the 6-tuples. You just sum the counts of the 6-tuples containing the Present Perfect and IPAN. Table 3 shows the 7 most frequent 6-tuples, but you can find the complete table on the website of Time in Translation.

Coming back to (2.a) we see that IPAN occurs 9 times in the dataset (see the table on figure 11). I have listed the English equivalents of IPAN with their frequencies in table 4. We see that in half of the cases, the equivalent of IPAN is the Present Perfect. Hence it's not distinct what the equivalent for IPAN in English is.

Equivalent for IPAN	**Frequency**
Present Perfect	5
Past Simple	4

Table 4: The English equivalents for IPAN.

To see whether (2.b) is true, we do the same. IPAS appears 323 times in the data-set. The distribution of the equivalents is shown in table 5. Out of the 323 times, IPAS has 312 times the Past Simple as equivalent. In total 9 cases weren't in line with the assumption. So we can say with confidence that the assumption in (2.b) holds.

Equivalents for IPAS	**Frequency**
Past Simple	312
Present Perfect	3
Present Simple	2
Present Participle	6

Table 5: The English equivalents for IPAP.

IPAI has occurred 7 times in the dataset. In table 6 the occurrences of IPAI are shown. We see that in all the cases the IPAI is translated as a Past Simple. So the hypotheses in (2.c) isn't supported at all based on the data.

Equivalent for IPAI	**Frequency**
Past Simple	7

The **third** hypothesis was inspired by Ayati and Manouchehri (2013). Looking at the frequency table in figure 11, we see that IPAI (the imperfect) has appeared 7 times. That is 2 times less than IPAN, which was assumed to be the equivalent for the Present Perfect by Ali Jahanshiri. In comparison to the 351 *passé compos*é cases, these 7 appearances of the Imperfect aren't that impressive. However, it is interesting to see that the frequency doesn't differ that much from IPAN, which is seen as an equivalent for the present perfect.

Last but not least, I had expected that most of the tenses in Farsi would have the indicative mood and be a past simple and that is indeed the case. The indicative mood is the only appeared mood! And the Indicative Past Simple is the most occurred tense, as you can see at the table in figure 11.

8. Conclusion

In this project, which is a part of the Time in Translation project, I included a new language into the database. More specifically I focused on the translations of the Perfect from French to Farsi. After all the preparational steps, I compared Farsi to other languages of the Indo-European family which they all come from, by using semantic maps.

My main aim was to consider how the *passé compos*é is translated to Farsi, compared to Dutch, English, German and Spanish. We have seen that all of these languages, except for German, the *passé compos*é is translated to the Past Simple. As discussed in section 7, we have seen some similar and different patterns among these languages. What made Farsi different in comparison to the other languages, was its variety of tenses and its quite irregular distribution of the tense IPAP. Next to the general comparison to the other languages, I also questioned the assumptions made in the literature on equivalents for the Farsi tenses in English and French. I can confirm the assumption that the closest equivalent of IPAS is the 'Past Simple', from the results. However I can't say with certainty that the Present Perfect is the equivalent of IPAN. It is remarkable to see that even in Farsi, a tense which is considered to be a Perfect, shows ambiguity in interpretation. So this finding may be odd, but it is in line with the whole idea that the Perfect is ambiguous.

Furthermore, the distribution of IPAP and IPAI are unique, because these categories have only occurred in Farsi. To understand why these tenses only occur in the Farsi translation, it is helpful to retrieve the information about these tenses in section 3.2. Ali Jahanshiri had already mentioned in his grammar that the English equivalent of IPAI is based on the context. He provided some examples on his website, which were either in the Past Continuous or in the Past Simple. In our dataset, every occurrence of IPAI is translated with a Past Simple. The presence of IPAI in the dataset can be explained by the aspect of IPAI. I had assumed that the closest equivalent of IPAI would be the Past Continuous in English, but the other equivalent, the Past Simple only appeared in the data set. This could be due to the narrative character of the novel.
On the other hand, I find it hard to find an explanation for why the closest equivalent of IPAP is the Past Simple, while it should be Past Perfect according to Ali Jahanshiri. I guess that the ambiguity for the Perfect exists in general, even when it is in the past tense.

Besides the grammatical hardships, I also have to mention that the parser didn't work that well for Farsi. I saved 67 cases by hand which did take quite a long time. So there is room for improvement in that aspect of the project. Also, as van der Klis, Le Bruyn and the Swart (2017) had mentioned, the distance function might be too simplistic. The difference between IPAP which is comparable to the Past Perfect, is equal to the IPAS.

To sum up, also for Farsi the Past Simple was the most used translation for the *passé compos*é. Compared to the other included languages, Farsi has shown more variation in its tense use. However I believe that Farsi can show more complexity, when investigated in a

bigger research program and when paid more attention to the context. When doing that, appearances of IPAP and IPAI can be explained in more detail, clarifying their usage. Taking In to consideration, the studies I have mentioned in section 2.3 and 3.3, I hope my contribution through this project will help fellow interested researchers in Farsi and its usage in computational linguistics.

9. Journal

Day	Activity
8-03-18	Emailed dr. De Swart whether I could contribute to the project.
20-03-18	**First meeting:** talked about the project, including Farsi.
22-03-18	Found a Farsi translation of L'Étranger and added it to the group drop box.
17-04-18	**Second meeting:** talked about the steps I have to take.
20-04-18	Converted the pdf to a txt. Document.
26-04-18	Completed the paragraph alignment and emailed it to Martijn.
30-04-18	Received my log in for the TimeAlign .
32-04-18 until 8-05-18	Finished the annotation.
8-05-18	**Third meeting:** discussed the challenges of annotation and talked about assigning the verb tense.
9-05-18	Started writing the introduction.
10-05-18	Read about the present perfect in Farsi and other literature.
11-05-18 until 14-05-18	Assigned the verb tense for 180 sentences.
13-05-18	Separated the sentences which I had trouble doing the annotation so I could discuss it later on.
14-05-18	Finishing the introduction and reading some articles.
16-05-18	Finished assigning the verb tense and rechecked half of them. Wrote the table of contents and did some layout work.
19-05-18 / 20-05-18	Writing the thesis
22-05-18	**Fourth meeting:** evaluated the draft of the thesis. Also discussed some problems I faced during the verb-annotation and time-align.
26-05-18	Adjusting the verb terms in the excel file. Annotated 34 cases the parser couldn't find by hand
27-05-18	Annotating the remaining by hand, total of 63 cases!
01-06-18	Added the cases I had annotated by hand to the excel file plus some annotations we had discussed in our last meeting. Also did the time align of these cases.
10-06-18	Writing the grammar part of the thesis, and reading about Farsi grammar
11-06-18	Finishing the grammar part, adding the linguistic theory part.
13-06-18	**Fifth meeting:** Talked about all the questions I had and discussed some linguistic glossary.
15-06-18 till 03-07-18	Writing, writing and more writing!

10. References

Ahmad, J. A. E. (1984). *Occidentosis: A plague from the west* (pp. 135-136). Berkeley, CA: Mizan Press.

[Akram, A & Manouchehri (2013)] اکرم, آیتی, & منوچهری. (2013). تجزیه و تحلیل خطا در استفاده از زمان دستوری زبان فرانسه توسط فارسی زبانان با دانش زبان انگلیسی. *مطالعات زبان و ترجمه*, *44*(1).

Chowdhury, G. G. (2003). Natural language processing. *Annual review of information science and technology*, *37*(1), 51-89.

Feili, H., & Ghassem-Sani, G. (2004, August). An application of lexicalized grammars in English-Persian translation. In *Proceedings of the 16th European Conference on Artificial Intelligence* (pp. 596-600). IOS Press.

Hojatollah Taleghani, A. (2006). The interaction of modality, aspect and negation in Persian.

Iranpour Mobarakeh, M., & Minaei-Bidgoli, B. (2009). Verb detection in persian corpus. *International Journal of Digital Content Technology and its Applications*, *3*(1), 58-65.

Jurafsky, D., & Martin, J. H. (2014). *Speech and language processing* (Vol. 3). London:: Pearson.

Johnson, W. L., & Valente, A. (2008, March). Tactical Language and Culture Training Systems: Using Artificial Intelligence to Teach Foreign Languages and Cultures. In *AAAI* (pp. 1632-1639).

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

Palmer, F. R. (2001). *Mood and modality*. Cambridge University Press.

Rahimifar, M., Rezai, V., & Motavalian, R. (2017). A Functional Analysis of Present Perfect in Persian. *Theory and Practice in Language Studies*, *7*(12), 1337-1349.

Van Der Klis, M., Le Bruyn, B., & De Swart, H. (2017). Mapping the PERFECT via translation mining. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers* (Vol. 2, pp. 497-502).

Wiklund, A. L. (2007). *The syntax of tenselessness: tense/mood/aspect-agreeing infinitivals* (Vol. 92). Walter de Gruyter.

Website of Time in Translation project: https://time-in-translation.hum.uu.nl/

Website of Ali Jahanshiri, providing the grammar and verb conjugator: http://www.jahanshiri.ir

Website of the context reverser: http://context.reverso.net/translation/french-english/a+paru

Website for additional information on the grammar: http://persian.sdsu.edu/Verbs/story.html