Time in Translation: Past, Present & Future
Write a semantics of the PERFECT.

> Five languages (3 Germanic, 2 Romance)

> Account for distribution at the sentence level, in narration and in dialogue

> Methodology: back and forth between theory and data

Translation Mining (corpus-based)
Time in Translation

Henriëtte de Swart
Bert Le Bruyn
Martijn van der Klis (PhD)
Anja Goldschmidt (Post-Doc)
Jos Tellings (Post-Doc)

Year 1: Experimenting with Translation Mining
Translation Mining

Data from translation corpora allow us to compare tense/aspect systems of languages and study them in detail. Example: translations of Camus’ *L’Étranger*, Chapter 1
DUTCH

Perfect

Simple Past
GERMAN

Perfect

Simple Past
Translation Mining

Using data from translation corpora to compare tense/aspect systems of languages and study them in detail. Example: translations of Camus’ *L’Étranger*, Chapter 1

> Bigger picture but also strong grip on the data
J' ai lu le dossier de votre mère.

Translating:

**French:** passé composé (fr) 1.xml - 24774

**German**
Ich habe die Akte Ihrer Mutter gelesen.

**English**
I've read your mother's file.

**Spanish**
He leído el expediente de su madre.

**Italian**
Ho letto la pratica di vostra madre.

**Dutch**
Ik heb uw moeders map doorgelezen.
FRENCH
MDS visualization (scenario Camus - all tenses - chapter 1)

Stress:

This scatter plot shows the results Multidimensional Scaling of the dissimilarity matrix of tense tuples. Please see the project summary for more details.

Filters

Below, you can change the labels and the dimensions displayed on the x- and y-axis. Click the green button to confirm your selection and update the scatter plot.
MDS visualization (scenario Camus - all tenses - chapter 1)

This scatter plot shows the results of Multidimensional Scaling of the dissimilarity matrix of tense tuples. Please see the project summary for more details.

Filters

Below, you can change the labels and the dimensions displayed on the x- and y-axis. Click the green button to confirm your selection and update the scatter plot.
MDS visualization (scenario *Camus - all tenses - chapter 1*)

Stress:

This scatter plot shows the results of Multidimensional Scaling of the dissimilarity matrix of tense tuples. Please see the project summary for more details.

**Filters**

Below, you can change the labels and the dimensions displayed on the x- and y-axis. Click the green button to confirm your selection and update the scatter plot.
MDS visualization (scenario Camus - all tenses - chapter 1)

Stress:

This scatter plot shows the results Multidimensional Scaling of the dissimilarity matrix of tense tuples. Please see the project summary for more details.

Filters

Below, you can change the labels and the dimensions displayed on the x- and y-axis. Click the green button to confirm your selection and update the scatter plot.
Fragment overview

Source
French
j'ai dormi pendant presque tout le trajet.

Translations
German
Ich habe fast während der ganzen Fahrt geschlafen.

English
I slept almost all the way.

Spanish
Dormí durante casi todo el trayecto.

Italian
Ho dormito quasi tutto il percorso.

Dutch
De hele rit bijna slecht geslapen.

Polish
No annotation available.

Russian
Когда проснулся, то оказалось, что голова моя лежит на плече какого-то военного, моего соседа; он мне углобил и спросил, издалика ли я еду.

Mandarin
我于上车一角，

Hebrew
כשכון את המדר.

Greek


Estonian
Taas olen laulunud ka varem.

All other languages have no annotation available.
Translation Mining

Using data from translation corpora to compare tense/aspect systems of languages and study them in detail.
Example: translations of Camus’ *L’Étranger*, Chapter 1

> Bigger picture but also strong grip on the data
> Sliding scale
English < Spanish < Dutch < German < French
I’ve read your mother’s file.

We’ve transferred her to our little mortuary.

I’ve taken it upon myself to make the necessary arrangements.

I saw them more clearly than I’ve ever seen anyone.

In eight years they haven’t changed their route.

**Classical perfects**

Past event

State holding at the present / current relevance
English < Spanish < Dutch < German < French
Mother died today.

Hoy, mamá ha muerto.

green simple past

blue present perfect
We covered her up.
La hemos cubierto.

Look mommy, I drew a picture of you!
Look mommy, I’ve drawn a picture of you!
English < Spanish < Dutch < German < French
After that everything **happened** so quickly...

Todo **pasó** después con tanta precipitación...

Daarna **is** alles zo snel... **gegaan**...

Danach **ist** alles so überstürzt... **abgelaufen**...

I **slept** almost all the way.

**Dormí** durante casi todo el trayecto.

De hele rit bijna **heb ik** **geslapen**.

Ich **habe** fast während der ganzen Fahrt **geschlafen**.
After that everything happened so quickly...

Todo pasó después con tanta precipitación...

Daarna is alles zo snel... gegaan...

Danach ist alles so überstürzt... abgelaufen...

I'm very fond of white coffee, so I accepted and he came back a few minutes later with a tray.

Como me gusta mucho el café con leche, acepté, y, al cabo de un momento, volvió con una bandeja.

Aangezien ik dol op koffie ben nam ik zijn aanbod aan, en een ogenblik later kwam hij terug met een blad.

Da ich Milchkaffee sehr gern mag, habe ich angenommen, und er ist nach einer Weile mit einem Tablett zurückgekommen.
Reflections

We have experimented with several parallel corpora (EuroParl, DCEP, OpenSubtitles, ...) We find that translations of novels provide us with the best control over the data (one writer per novel, one translator per novel, clear source and target language, ...), especially because we assume we don’t know the linguistic variables yet.

EuroParl has become our corpus of choice to check predictions as soon as we think we have identified a relevant variable.
Reflections

We are currently looking into the (statistical) status of findings in novel-based corpora and the way our findings relate to those in typological work that uses a similar methodology.

Translation Mining has been about surveying ingredients. The analysis still has to come :-)

Next to a corpus component, the project also contains a native speaker component :-)
Year 2: Exploring dialogue
Two insights from the literature

Portner (2003)

The use of the perfect presupposes that the proposition it is in is a (partial or complete) answer to the question under discussion.

I’ve been diagnosed with cancer. How are you?

I was diagnosed with cancer. What happened?

Mom died. What happened?
Two insights from the literature

Nishiyama & König (2010)

Speakers sometimes use a perfect at the beginning of a conversation to set up a topic.

Have you seen *Dances with Wolves*?

Did you see *Dances with Wolves*?

Where were you on 9/11?
Back to the data
<table>
<thead>
<tr>
<th></th>
<th>en</th>
<th>de</th>
<th>es</th>
<th>fr</th>
<th>it</th>
<th>nl</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PRESENT</strong></td>
<td>6</td>
<td>6</td>
<td>4</td>
<td>3</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td><strong>PERFECT</strong></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td><strong>PAST</strong></td>
<td>430</td>
<td>461</td>
<td>336</td>
<td>313</td>
<td>316</td>
<td>465</td>
</tr>
<tr>
<td><strong>IMPERFECT</strong></td>
<td>0</td>
<td>0</td>
<td>126</td>
<td>130</td>
<td>112</td>
<td>0</td>
</tr>
<tr>
<td><strong>PAST PERFECT</strong></td>
<td>29</td>
<td>27</td>
<td>18</td>
<td>30</td>
<td>24</td>
<td>32</td>
</tr>
</tbody>
</table>

Tab. 1: Most frequently attributed tense categories in the **narrative** parts of the novel.

<table>
<thead>
<tr>
<th></th>
<th>en</th>
<th>de</th>
<th>es</th>
<th>fr</th>
<th>it</th>
<th>nl</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PRESENT</strong></td>
<td>126</td>
<td>154</td>
<td>122</td>
<td>129</td>
<td>129</td>
<td>162</td>
</tr>
<tr>
<td><strong>PERFECT</strong></td>
<td>27</td>
<td>78</td>
<td>23</td>
<td>99</td>
<td>100</td>
<td>60</td>
</tr>
<tr>
<td><strong>PAST</strong></td>
<td>120</td>
<td>64</td>
<td>88</td>
<td>0</td>
<td>2</td>
<td>84</td>
</tr>
<tr>
<td><strong>IMPERFECT</strong></td>
<td>0</td>
<td>0</td>
<td>38</td>
<td>50</td>
<td>39</td>
<td>0</td>
</tr>
<tr>
<td><strong>FUTURE</strong></td>
<td>20</td>
<td>15</td>
<td>13</td>
<td>11</td>
<td>23</td>
<td>9</td>
</tr>
</tbody>
</table>

Tab. 2: Most frequently attributed tense categories in the **dialogue** parts of the novel.
60% is present oriented  36% is present oriented
The Man with Two Faces

It was Quirrell.
‘You!’ gasped Harry.
Quirrell smiled. His face wasn’t twitching at all.
‘Me,’ he said calmly. ‘I wondered whether I’d be meeting you here, Potter.’
‘But I thought – Snape –’
‘Severus?’ Quirrell laughed and it wasn’t his usual quivering treble, either, but cold and sharp. ‘Yes, Severus does seem the type, doesn’t he? So useful to have him swooping around like an overgrown bat. Next to him, who would suspect p-p-poor stuttering P-Professor Quirrell?’
Harry couldn’t take it in. This couldn’t be true, it couldn’t.
‘But Snape tried to kill me!’

‘No, no, no. I tried to kill you. Your friend Miss Granger accidentally knocked me over as she rushed to set fire to Snape at that Quidditch match. She broke my eye contact with you. Another few seconds and I’d have got you off that broom. I’d have managed it before then if Snape hadn’t been muttering a countercurse, trying to save you.’

Storyline in dialogue

Splitting up dialogue further

141 verbs in ‘storyline’

77 verbs in ‘pure dialogue’
And what can we say about dialogue?

**English**
77 verbs
59 simple pasts
18 perfects

Same contexts as in Camus (= those known from the literature on sentence-level uses)

**Dutch**
77 verbs
12 simple presents
33 simple pasts
30 perfects
And what can we say about dialogue?

**Dutch**

12 simple presents  
33 simple pasts  
30 perfects

Hij wist alles al. Hij zei alleen: 'Harry *is* achter hem aan, hè?' en rende direct naar de derde verdieping."

"Maar ik hoorde u een paar dagen geleden snikken - ik *dacht* dat Sneep u bedreigde ... "

"*Hebt* u die trol *binnengelaten*? 'You *let* the troll in?'

He already knew - he just said, “Harry *'s* *gone* after him, hasn 't he?” and hurtled off to the third floor.

‘But I heard you a few days ago, sobbing - I *thought* Snape was threatening you ... ’
Reflections

These are just the initial stages...

Relevant factors we’re investigating (next to aspectual class): rhetorical relations, sequence of tense, modality, syntax, etc. In parallel, we’re looking into building tense/aspect extensions of conversation frameworks like inquisitive semantics.

The data already show that sentence, discourse and dialogue interact. Our final analysis will therefore – by necessity – incorporate all three levels.

See Jos’ SALT talk in May ☺
Beyond

More languages
More areas of grammar
SLA
Beyond

More languages
More areas of grammar
SLA
Example: translations of Harry Potter, Book 1, Chapter 1

FRENCH
Simple Past
Imparfait
Present
Perfect
Pluperfect
Translation Mining and SLA

GenSL
A

Usage-based approaches
GenSLA

80s-90s

> Basic question: is there UG access in SLA?
> Role of Poverty of the Stimulus Argument and Transfer.

More recent developments

> Interface issues.
> Interpretable and uninterpretable features.
> Role of input and how it relates to languages that are already present in the learner’s mind.
Role of input and how it relates to languages that are already present in the learner’s mind.

Learner starts with a coarse-grained hypothesis on how the new language compares to those that s/he already knows. Learner refines the hypothesis on the basis of input.

Based on what? When? When? How?

Basic issue: what, when, why, how do we generalize (i.e. learn a language)?
Usage-based approaches

> Frequency, contingency, prototypicality, ...
> Clear awareness of role of languages that are already present in the learner’s mind but little hypothesizing about interaction with input.
Basic issue: what, when, why, how do we generalize (i.e. learn a language)?

Basic data we need:

Mapping data similar to those we get through Translation Mining
Envoi
Thanks for your attention!

And visit us at https://time-in-translation.hum.uu.nl/