

Using multidimensional scaling to map the PERFECT

Martijn van der Klis, Bert Le Bruyn, Henriëtte de Swart

Kick-off workshop, Utrecht, 23 June 2017



The PERFECT – the basics

- Morpho-syntax: *auxiliary (HAVE/BE) + past participle*
- Appears a.o. in German, English, Spanish, French, Dutch
- Core meaning: *past event with current relevance.*

- 1) Mary **has visited** Paris. [experiential]
(her past visit to Paris is relevant now)
- 2) Mary **has moved** to Paris. [resultative]
(she currently lives in Paris)
- 3) Mary **has lived** in Paris for five years now. [continuative]
(she is still living in Paris)



The PERFECT – current approaches

- Current approaches tend to look at the PERFECT in **isolation**:
 - They ignore variation between languages
 - Same meaning can be conveyed by other tense-aspect forms
 - They ignore competition within a language
 - Competition with other tense-aspect forms in the grammar (in particular PAST and PRESENT)



The PERFECT – our data

- Multilingual parallel corpora
 - Translation equivalents provide us with **form variation** across languages in contexts where the **meaning is stable**.
- ... for various genres
 - EuroParl (Tiedemann 2012)
 - OpenSubtitles (Lison&Tiedemann 2016)
 - Literary corpus (following de Swart 2007)

The PERFECT – our proposed analysis

- **Semantic maps**
(Haspelmath 1997/2003)
 - Allows to see variation between and competition within languages
- **Idea:** generate maps directly from data
(Wälchli&Cysouw, 2012)
- **Our twist:** grammatical instead of lexical level:
Translation Mining

	tree	wood (stuff)	firewood	small forest	large forest
German	_____	_____	_____	_____	_____
Danish	<i>Baum</i>	<i>Holz</i>		<i>Wald</i>	
French		<i>trac</i>		<i>skov</i>	
Spanish	<i>arbre</i>		<i>bois</i>		<i>forêt</i>
	<i>árbol</i>	<i>madera</i>	<i>leña</i>	<i>bosque</i>	<i>selva</i>

Fig. 19:
A semantic map for 'tree'/'wood', and the boundaries of four languages

Step 1) Extraction of PERFECTS

- Simple principle:
 - Look for an auxiliary verb (HAVE/BE)
 - Find a neighboring past participle
- Takes care of words in between, lexical restrictions, reversed order, passive present perfects.
- Stand-alone application: **PerfectExtractor**
 - Source code on GitHub:
<https://github.com/UUDigitalHumanitieslab/time-in-translation>

Step 2) Word-level alignment

- **Manual** word-level alignment of translations of PERFECTS.
- Web application: **TimeAlign**
 - Source code on GitHub:
<https://github.com/UUDigitalHumanitieslab/timealign>

Annotation

Dutch (original)

We zijn allemaal blij dat de rechter hem heeft **vragesproken** en duidelijk heeft gemaakt dat ook in Rusland het recht op toegang tot informatie over het milieu door de grondwet wordt gegarandeerd .

English (translated)

It is the case of Alexander Nikitin .

All of us here are pleased that the courts **have acquitted** him and made it clear that in Russia , too , access to environmental information is a constitutional right .

The selected words in the original fragment do not form a present perfect

This is a correct translation of the original fragment

© 2016 Digital Humanities Lab, Utrecht University

Step 3) Tense attribution

- We assign every (verbal) translation a grammatical tense (so e.g. *passé composé* or *Präteritum*)

German	English	French	Spanish	Dutch
Heute ist Mama gestorben .	Mother died today.	Aujourd' hui, maman est morte .	Hay, mamá ha muerto .	Vandaag is moeder gestorven .
Perfekt	Simple past	Passé composé	Pretérito perfecto compuesto	Voltooid tegenwoordige tijd

- We discard 'other' translations (nominalizations, periphrastic constructions, etc.)

Step 4) Dissimilarity matrix

- ➔ **Distance function:** we define a tuple of source and translations to be maximally similar ($d=0$) if all tenses match up.

#	German	English	French	Spanish	Dutch
1	Perfekt	Present perfect	Passé composé	Pretérito perfecto compuesto	Voltooid tegenwoordige tijd
2	Präterium	Simple past	Passé composé	Pretérito perfecto compuesto	Voltooid tegenwoordige tijd
3	Perfekt	Present perfect	Passé récent	Pasado reciente	Voltooid tegenwoordige tijd

- ➔ $d(1, 2) = 2/5$, $d(1, 3) = 2/5$, $d(2, 3) = 4/5$

Step 4) Dissimilarity matrix

- ▶ Applying our defined distance function, we can create a dissimilarity matrix.

	#1	#2	#3
#1	-	2/5	2/5
#2	2/5	-	4/5
#3	2/5	4/5	-

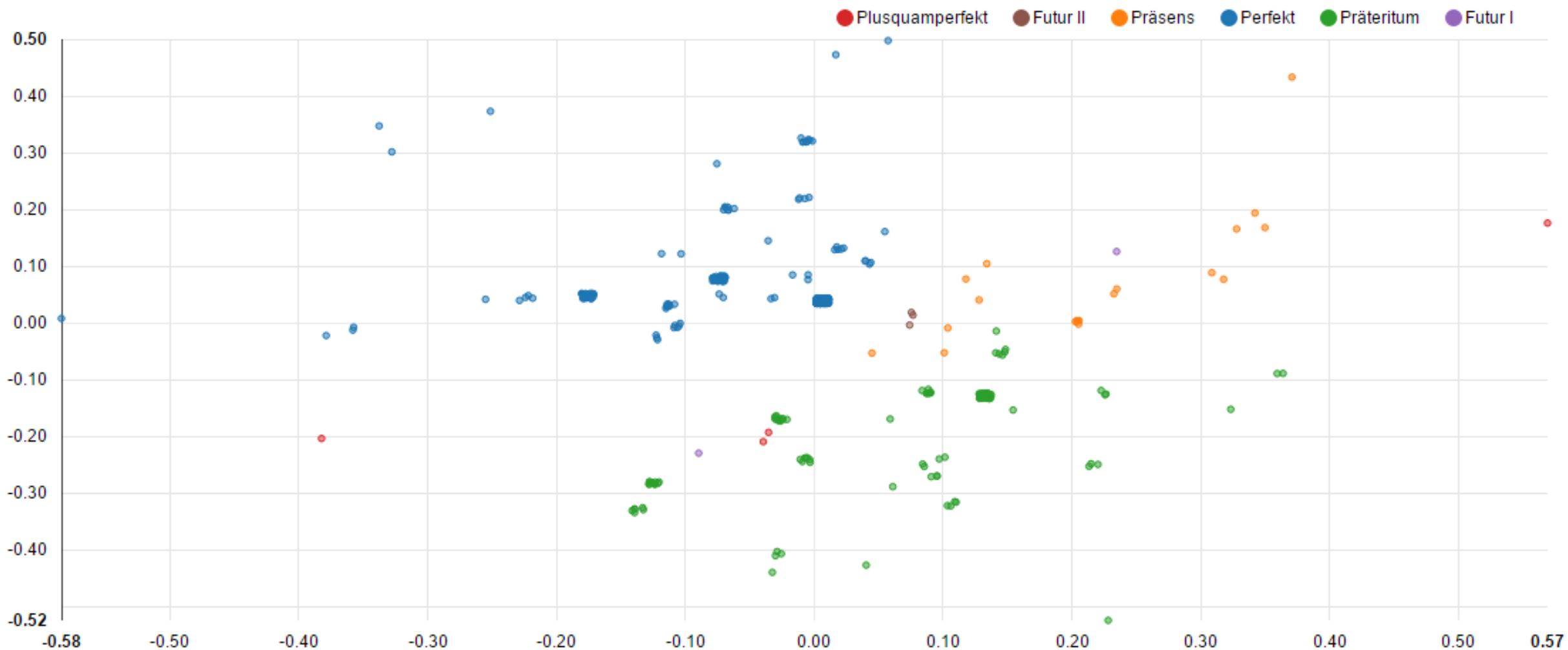


Step 5) Multidimensional Scaling

- **Multidimensional scaling (MDS)** tries to create a low-dimensional representation of the data, while respecting distances in the original high-dimensional space.
- Advantage: drilldown to original data points possible

Step 5) Multidimensional Scaling

- Web application: **TimeMapping**
 - Visualizes the results of MDS on a scatter plot, we use the attributed tenses as labels.
- Friendly user interface:
 - Click/unclick tenses for filtering
 - Choose which dimensions of the MDS to show
 - Choose which language to show
 - Drill-through to raw data
- Source code on GitHub:
<https://github.com/UUDigitalHumanitieslab/timealign>



Filters

Language: German English Spanish French Dutch Dimension on x-axis: 1 2 3 4 5 Dimension on y-axis: 1 2 3 4 5 Go!



Interpreting the maps

- Switching languages reveals two main contrasts:
 - Aspectual perspective: PERFECT vs. SIMPLE (x-axis)
 - Temporal orientation: PRESENT vs. PAST (y-axis)
- We are working on statistical analysis using ANOSIM (Clarke 1993), see BA Thesis Woelk.

Qualitative research (1)

- The maps show more **central** and more **peripheral** uses of the PERFECT
- We can confirm earlier research...
 - E.g. confirm that English requires a PAST with a *locating time adverbial*, whereas German, Dutch and French tolerate a PERFECT in this configuration. Spanish patterns with English (see Schaden 2009). Example on next page.

Dutch

ep-00-12-11.xml - 3552

Mevrouw de Voorzitter , op 4 december hebben wij hierover gestemd .

Translations

German

Perfekt

Frau Präsidentin , wir haben am 4. Dezember abgestimmt .

English

simple past

Madam President , we voted on 4 December .

Spanish

presente

Señora Presidenta , votamos el pasado 4 de diciembre .

French

passé composé

Madame la Présidente , nous avons voté le 4 décembre .



Qualitative research (2)

- ▶ ... but also look at **special verb forms**.
- ▶ E.g. English *Present Perfect Continuous* has been claimed to translate as a PRESENT in French (cf. Nishyama & Koenig 2010) and Dutch.
 - ▶ But: German maintains a PERFECT in sentences that contain a *seit* ('since') adverbial. Note the special construction in Spanish.
 - ▶ More: see BA thesis 't Hoen (2017)

Seit Monaten **haben** die Nachrichten von der Elfenbeinküste eine begründete Unruhe in der europäischen Öffentlichkeit und entsprechende Besorgnis in unserem Parlament **hervorgerufen** .

Translations

English

present perfect continuous

Mr President , for months the news from Côte d ' Ivoire **has been causing** justified alarm within European public opinion and the corresponding concern in our Parliament .

Spanish

pretérito perfecto compuesto

Señor Presidente , desde hace meses las noticias de Costa de Marfil **han venido causando** justificada alarma en la opinión pública europea y la correspondiente preocupación en nuestro Parlamento .

French

présent

Monsieur le Président , les nouvelles de la Côte-d'Ivoire **suscitent** depuis des mois l' inquiétude justifiée chez les citoyens européens et l' inquiétude correspondante au sein de notre Parlement .

Dutch

ott

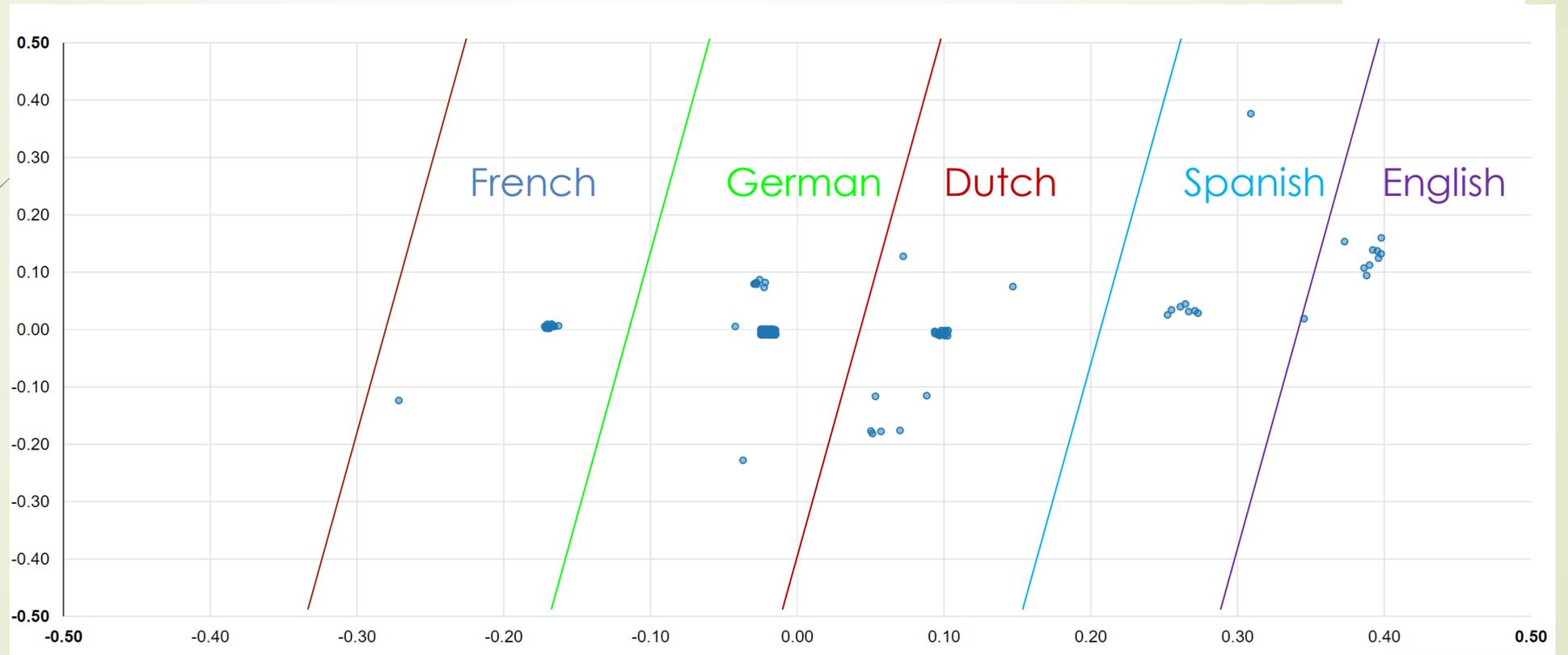
Mijnheer de Voorzitter , sinds enkele maanden **zorgen** de berichten uit Ivoorkust terecht voor onrust in de Europese publieke opinie en voor overeenkomstige bezorgdheid in ons Parlement .



No data or perfect data?

- OpenSubtitles (BA thesis Verkleij&Wimmers) is far less clear than our EuroParl data.
- New idea: start from one specific form of the PERFECT, **the narrative PERFECT**, in a small literary corpus: Camus' L'Étranger (BA thesis Swager)

The narrative PERFECT



State verbs

Narrative contexts

'classical'
Perfect contexts



Problems with our methodology

- Requires quite some manual labor
- Distance function is naive
- Interpreting dimensions in MDS is hard
- Lack of corroborating evidence
 - (but: web surveys in year 3 of the project)
- Any other?



Ideas for future research

- PERFECT in languages that do not have a PERFECT
 - Chinese, Russian, Polish, Hebrew
- PERFECT in competition (BA Thesis Broekhoff)
- *passé composé* vs. *imparfait* for state verbs (BA Thesis Orsel)
- L2 acquisition of the Perfect
- More?



Thanks for your attention!

- Any questions or feedback?
- More on our website: <http://bit.ly/timeintranslation>