Stative verbs as edge cases in the PERFECT construction

Martijn van der Klis Bert Le Bruyn Henriëtte de Swart

UiL OTS, Utrecht University

August 31, 2018

Universiteit Utrecht

Intro - the PERFECT construction

Morpho-syntax: *auxiliary* (HAVE) + *past participle*
e.g. Mary **has read** Camus.

- Appears in most West-European languages
- Core meaning: past event with current relevance

Universiteit Utrecht

Intro - the PERFECT construction

Morpho-syntax: *auxiliary* (HAVE) + *past participle*
e.g. Mary **has read** Camus.

- Appears in most West-European languages
- Core meaning: past event with current relevance

But: distribution of PERFECT differs between languages, e.g.:

(1) Was **hast** du **gemacht**?
lit. What have you done?

Universiteit Utrecht

Intro - the PERFECT construction

Morpho-syntax: *auxiliary* (HAVE) + *past participle*
e.g. Mary **has read** Camus.

- Appears in most West-European languages
- Core meaning: past event with current relevance

But: distribution of PERFECT differs between languages, e.g.:

(1) Was **hast** du **gemacht**?
lit. What have you done?
What **did** you **do**?

Universiteit Utrecht

Intro -stative verbs in the PERFECT

From English, we know states are special: in the PERFECT, they raise a (potential) continuative reading. However, this is language-specific (de Swart, 2016):

Universiteit Utrecht

Intro -stative verbs in the PERFECT

From English, we know states are special: in the PERFECT, they raise a (potential) continuative reading. However, this is language-specific (de Swart, 2016):

(2) Mary **has lived** in Tallinn for five years.
lit. Marie hat seit fünf Jahren in Tallinn gelebt.

Universiteit Utrecht

Intro -stative verbs in the PERFECT

From English, we know states are special: in the PERFECT, they raise a (potential) continuative reading. However, this is language-specific (de Swart, 2016):

(2) Mary **has lived** in Tallinn for five years.
lit. Marie hat seit fünf Jahren in Tallinn gelebt.
Marie **lebt** seit fünf Jahren in Tallinn

Universiteit Utrecht

Intro - stative verbs in the PERFECT

Stative verbs also appear less frequent in the PERFECT per se. Evidence comes form a collostructional analysis on the British National Corpus, fiction section (van der Klis, 2018):

Universiteit Utrecht

Intro - stative verbs in the PERFECT

Stative verbs also appear less frequent in the PERFECT per se. Evidence comes form a collostructional analysis on the British National Corpus, fiction section (van der Klis, 2018):

attracted verbs

- hear
- see
- happen
- change
- finish

repelled verbs

- feel
- want
- know
- think
- say

Universiteit Utrecht

Intro - cross-linguistic variation

Schaden (2009) postulates a dichotomy for the PERFECT in the European languages:

- English and Spanish are alike:
  - Not licensed with past time adverbial
    (* I **have read** a book *yesterday*)
  - Not licensed in narrative context
    (* ...and *then* I **have seen** Mary)

Universiteit Utrecht

Intro - cross-linguistic variation

Schaden (2009) postulates a dichotomy for the PERFECT in the European languages:

- English and Spanish are alike:
  - Not licensed with past time adverbial
    (* I **have read** a book *yesterday*)
  - Not licensed in narrative context
    (* ...and *then* I **have seen** Mary)
- French and German are alike:
  - Can appear with past time adverbial
  - Can appear in narrative context

Universiteit Utrecht

Intro - cross-linguistic variation

Schaden (2009) postulates a dichotomy for the PERFECT in the European languages:

- English and Spanish are alike:
  - Not licensed with past time adverbial
    (* I **have read** a book *yesterday*)
  - Not licensed in narrative context
    (* ...and *then* I **have seen** Mary)
- French and German are alike:
  - Can appear with past time adverbial
  - Can appear in narrative context

Main questions in this talk:

- Are this indeed the important differences? Do we find differences between these language pairs?
- Is this a dichotomy, a scale, or something else? What about other languages?

Universiteit Utrecht

Our data

We use parallel corpora: translation equivalents provide us with **form variation** across languages in contexts where the **meaning is stable**.

Universiteit Utrecht

Our data

We use parallel corpora: translation equivalents provide us with **form variation** across languages in contexts where the **meaning is stable**.

While there are several alternatives available, we opt to use translations of novels.

Universiteit Utrecht

Our data

We use parallel corpora: translation equivalents provide us with **form variation** across languages in contexts where the **meaning is stable**.

While there are several alternatives available, we opt to use translations of novels.

Advantages:

- more down-to-earth register (as opposed to Bible/Europarl)
- high quality translations (as opposed to OpenSubtitles)
- allows to study phenomena in dialogue vs. discourse

Universiteit Utrecht

Our data

We use parallel corpora: translation equivalents provide us with **form variation** across languages in contexts where the **meaning is stable**.

While there are several alternatives available, we opt to use translations of novels.

Advantages:

- more down-to-earth register (as opposed to Bible/Europarl)
- high quality translations (as opposed to OpenSubtitles)
- allows to study phenomena in dialogue vs. discourse

Disadvantages:

- small datasets (so frequent phenomena required)
- possible translator effects
- copyright issues

Universiteit Utrecht

Our data - L'Étranger parallel corpus

We use the first three chapters of Albert Camus' novel **L'Étranger** as our data. Why?

- internal monologue: *passé composé* can be used in French, but not in most other languages: PERFECT stretched to its max
- confirmation of earlier research (e.g. de Swart (2007))

Universiteit Utrecht

Our data - L'Étranger parallel corpus

We use the first three chapters of Albert Camus' novel **L'Étranger** as our data. Why?

- internal monologue: *passé composé* can be used in French, but not in most other languages: PERFECT stretched to its max
- confirmation of earlier research (e.g. de Swart (2007))

Languages in L'Étranger corpus:

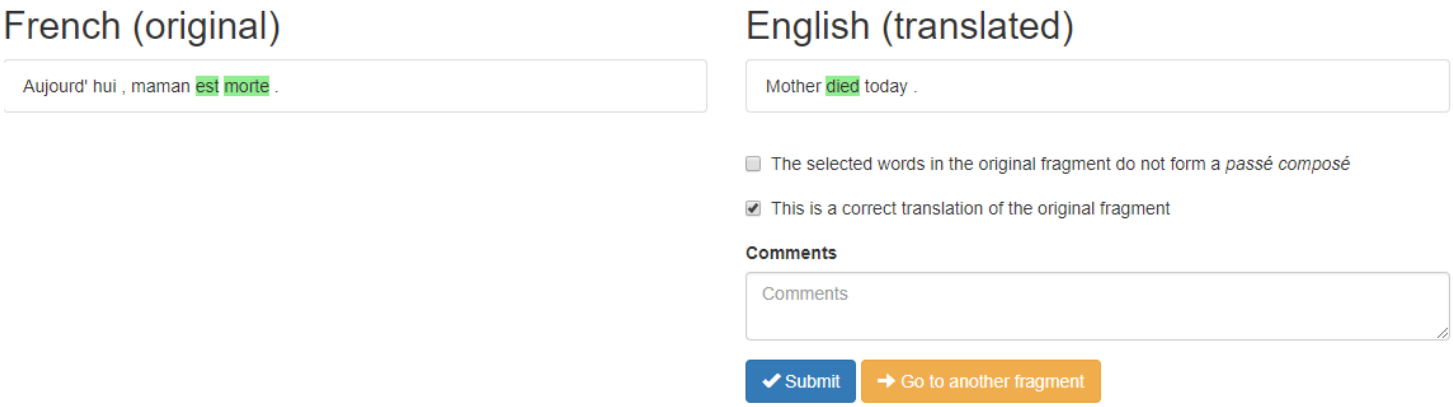
Romance	French, Italian, Spanish
Germanic	German, Dutch, English
other	Breton, Estonian, Farsi, Greek, Hebrew, Mandarin, Russian

Universiteit Utrecht

Annotation

Annotation is done in a web application (dubbed *TimeAlign*). Steps:

1. algorithm extracts all *passé composé* forms automatically
2. annotators select corresponding verb forms in translation
3. annotators assign tense-aspect labels (so e.g. *simple past* or *Perfekt*)



Universiteit Utrecht

Annotation - example

language	fragment	TA-label
fr	Aujourd'hui, maman **est morte**.	*passé composé*
de	Heute **ist** Mama **gestorben**.	*Perfekt*
nl	Vandaag **is** moeder **gestorven**.	*voltooid tegenwoordige tijd*
es	Hoy, mamá **ha muerto**.	*pretérito perfecto compuesto*
en	Mother **died** today.	*simple past*

Universiteit Utrecht

Annotation - example

language	fragment	TA-category
fr	Aujourd'hui, maman **est morte**.	PERFECT
de	Heute **ist** Mama **gestorben**.	PERFECT
nl	Vandaag **is** moeder **gestorven**.	PERFECT
es	Hoy, mamá **ha muerto**.	PERFECT
en	Mother **died** today.	PAST

We refer to a TA-labeling for a single fragment as a **tuple**.

Universiteit Utrecht

Results - descriptive statistics

Descriptive statistics for all *passé composé*-forms in the first three chapters:

Tense	fr	de	nl	es	en
PERFECT	375	351	45	19	12
PAST	-	23	325	355	354
PRESENT	-	1	2	1	3
other tenses	-	-	3	-	6

Note:

- We discarded 'other' translations (nominalizations, periphrastic constructions, etc.) and only considered complete tuples

Universiteit Utrecht

Results - tuple frequencies

As we are dealing with parallel data, we can also count tuple frequencies:

de	nl	es	en	#
Past	Past	Past	Past	20
Perfect	Past	Past	Past	297
Perfect	Perfect	Past	Past	25
Perfect	Perfect	Perfect	Past	6
Perfect	Perfect	Perfect	Perfect	10

All other possible combinations: less than 5 occurrences. This hints at a **subset relation** rather than a dichotomy.

Universiteit Utrecht

Alternative technique - multidimensional scaling

Wälchli & Cysouw (2012) provide a technique to generate **semantic maps** directly from parallel corpus data using **multidimensional scaling** (MDS). We showcase this technique on our data.

Universiteit Utrecht

MDS - distance function

We define a tuple to be maximally similar ($d = 0$) if all tenses match up.

#	fr	de	nl	es	en
1	Perfect	Perfect	Perfect	Perfect	Perfect
2	Perfect	Perfect	Perfect	Past	Past
3	Perfect	Past	Past	Perfect	Perfect

In this table, $d(1, 2) = 0.4$, $d(1, 3) = 0.4$ and $d(2, 3) = 0.8$.

Universiteit Utrecht

MDS - dissimilarity matrix

Applying our defined distance function, we can create a dissimilarity matrix:

	#1	**#2**	**#3**	$\cdots$
#1	-	0.4	0.4	
#2	0.4	-	0.8	
#3	0.4	0.8	-	
$\vdots$				$\ddots$

Universiteit Utrecht

MDS - dissimilarity matrix

Applying our defined distance function, we can create a dissimilarity matrix:

	#1	**#2**	**#3**	$\cdots$
#1	-	0.4	0.4	
#2	0.4	-	0.8	
#3	0.4	0.8	-	
$\vdots$				$\ddots$

We use **multidimensional scaling** (MDS) to visualize this dissimilarity matrix. MDS tries to create a low-dimensional representation of the data, while respecting distances in the original high-dimensional space.

Universiteit Utrecht

MDS - demonstration

Demo time!

Universiteit Utrecht

MDS - conclusions

Conclusions from multidimensional scaling:

- **subset relation** - no dichotomy - between Western European languages
- **clear distinctions** between language pairs that were presumed to be close together

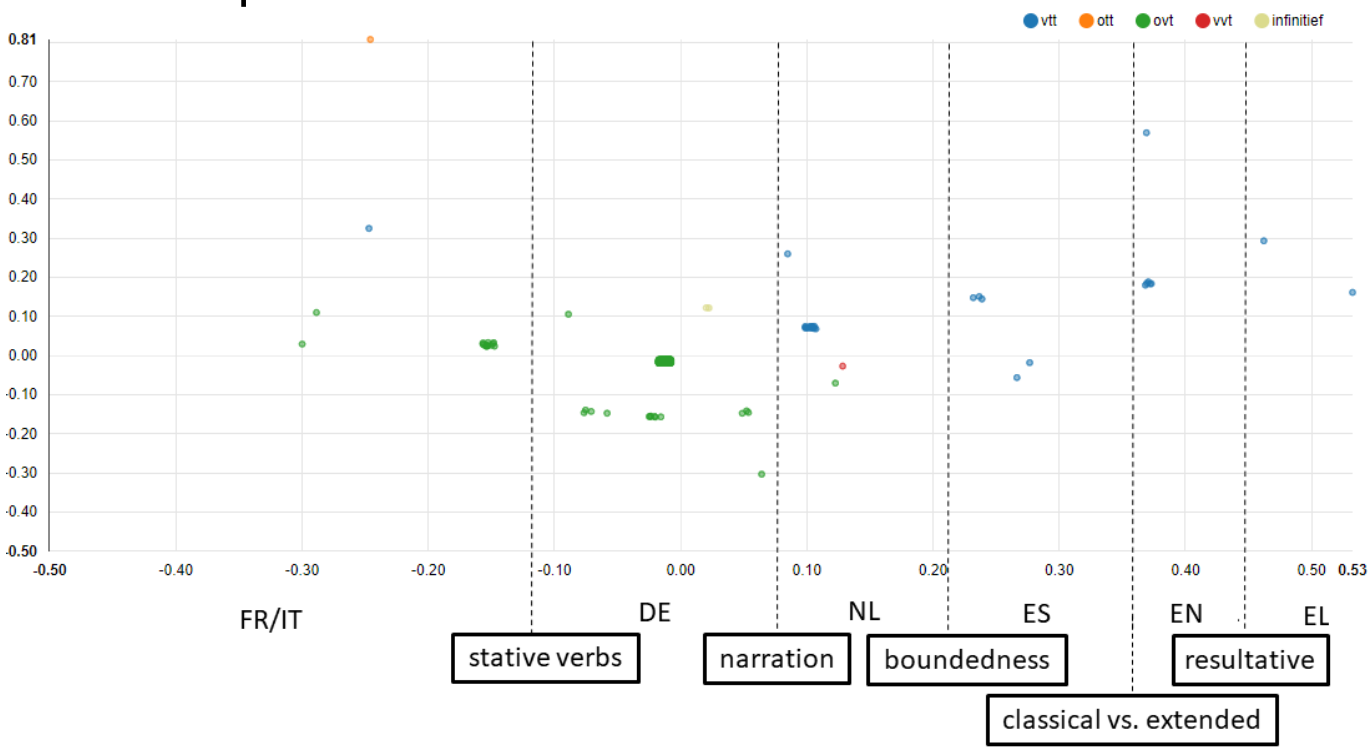
Universiteit Utrecht

MDS - conclusions

Conclusions from multidimensional scaling:

- **subset relation** - no dichotomy - between Western European languages
- **clear distinctions** between language pairs that were presumed to be close together

Our interpretation of the distinctions:



MDS - zooming in

Let's zoom in on the difference between French/Italian and German:

- French can have stative verbs in the PERFECT tense
- (cognitive) stative verbs go with a PAST tense in German (and nl/es/en)

Example:

(3) fr J'**ai voulu** voir mamam tout de suite.

de Ich **wollte** sofort zu Mama.

nl Ik **wilde** moeder meteen zien.

es Yo **quería** ver a mamá inmediatamente.

en I **wanted** to see mother straight away.

Universiteit Utrecht

Zooming in - Association analysis

Let's annotate all verbs for stativity (Maienborn, 2015); and compare with the tense choice in German:

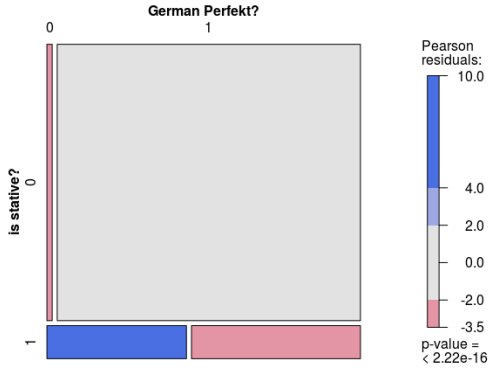
	Präteritum	*Perfekt*
dynamic	6	345
stative	19	23

Universiteit Utrecht

Zooming in - Association analysis

Let's annotate all verbs for stativity (Maienborn, 2015); and compare with the tense choice in German:

	Präteritum	*Perfekt*
dynamic	6	345
stative	19	23

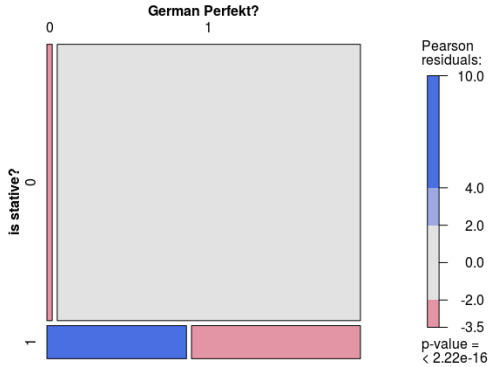


Universiteit Utrecht

Zooming in - Association analysis

Let's annotate all verbs for stativity (Maienborn, 2015); and compare with the tense choice in German:

	Präteritum	*Perfekt*
dynamic	6	345
stative	19	23



Typical state verbs in the *Präteritum*:

- *être* 'to be'
- *avoir* 'to have'
- *paraître* 'to seem'
- *falloir* 'to must'
- *comprendre* 'to understand'
- *vouloir* 'to want'
- *croire* 'to believe'
- *trouver* 'to find'

Universiteit Utrecht

Results - comparing all tenses

We repeated the annotation procedure for all indicative VPs in the first chapter of L'Étranger.

Conclusions:

- Clear dimensions: dimension 1: PAST vs. PRESENT, dimension 2: PERFECT vs. unmarked
- Variation between languages mainly in the competition between PERFECT and PAST
- Only one new PERFECT introduced: continuative use with stative verbs in English (as expected)

(4) en **Have** you **been** here long?

fr Il y a longtemps que vous **êtes** là?

de **Sind** Sie schon lange hier?

nl Hoe lang **bent** u al hier?

es ¿Hace mucho tiempo que **está** usted aquí?

Universiteit Utrecht

Conclusion

We improved Schaden (2009):

- Clear differences between languages
- No dichotomy, but rather a spectrum of PERFECT use
- Stative verbs lead to differences on both ends of the spectrum

Universiteit Utrecht

Conclusion

We improved Schaden (2009):

- Clear differences between languages
- No dichotomy, but rather a spectrum of PERFECT use
- Stative verbs lead to differences on both ends of the spectrum

Advantages of using MDS:

- hypothesis generation (and confirmation via additional annotation)
- white box method

Disadvantages:

- difficult to interpret dimensions
- overlapping contexts

Universiteit Utrecht

On our to-do list

- Analyse languages that do not have a PERFECT (e.g. Mandarin, Russian)
- More annotation layers, regression analysis of factors deciding between PERFECT and PAST
- Analyse competition between PERFECT and PRESENT (novel written in present tense)
- Repeat analysis for different genres (e.g. news articles, subtitles etc.)

Universiteit Utrecht

On our to-do list

- Analyse languages that do not have a PERFECT (e.g. Mandarin, Russian)
- More annotation layers, regression analysis of factors deciding between PERFECT and PAST
- Analyse competition between PERFECT and PRESENT (novel written in present tense)
- Repeat analysis for different genres (e.g. news articles, subtitles etc.)

- Your idea here?

Universiteit Utrecht

On our to-do list

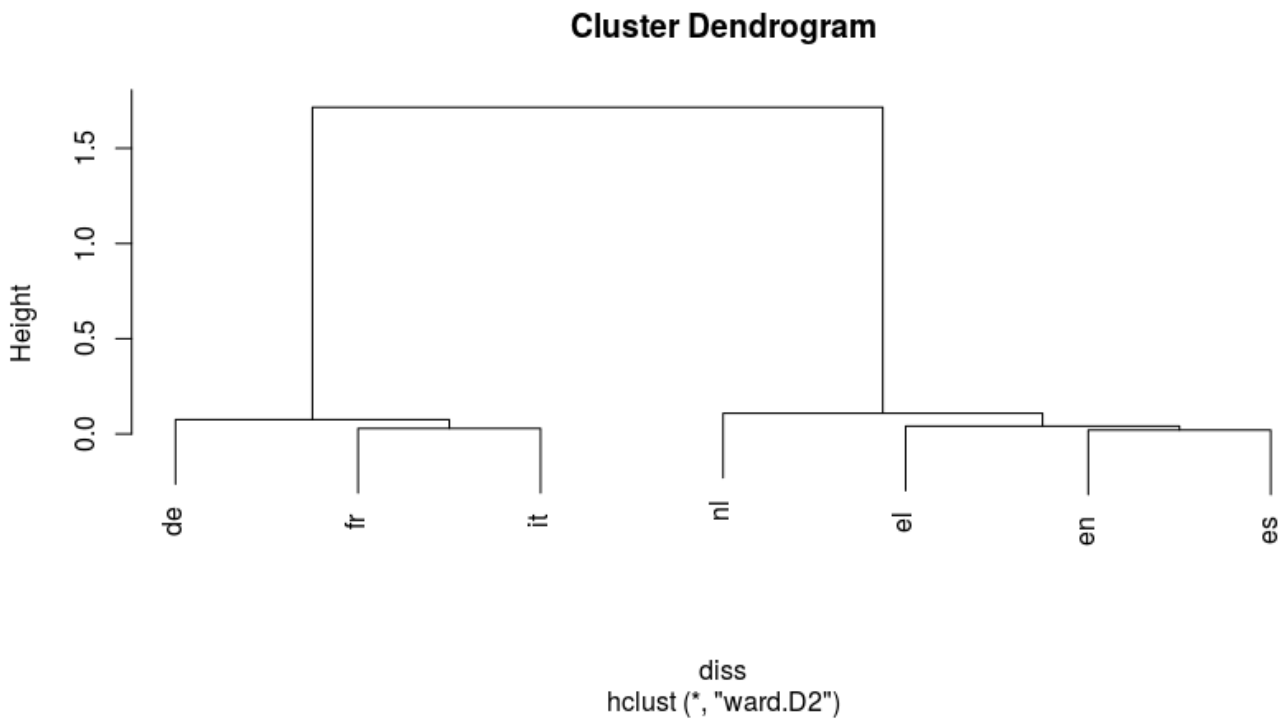
- Analyse languages that do not have a PERFECT (e.g. Mandarin, Russian)
- More annotation layers, regression analysis of factors deciding between PERFECT and PAST
- Analyse competition between PERFECT and PRESENT (novel written in present tense)
- Repeat analysis for different genres (e.g. news articles, subtitles etc.)

- Your idea here?

Thanks! Stay tuned via `time-in-translation.hum.uu.nl`

Universiteit Utrecht

Bonus slide: hierarchical analysis



Universiteit Utrecht