

# Integrating parallel corpora and experimental research in crosslinguistic tense-aspect semantics: a multi- pronged approach to the Perfect

Martín Fuchs  
Utrecht University

Anglistik, Amerikanistik und Anglophone Kulturen  
Universität des Saarlandes  
06 July 2022



Universiteit Utrecht



# Collaborative work (Time in Translation)



Martijn van der Klis



Henriëtte de Swart (PI)



Bert Le Bruyn (PI)

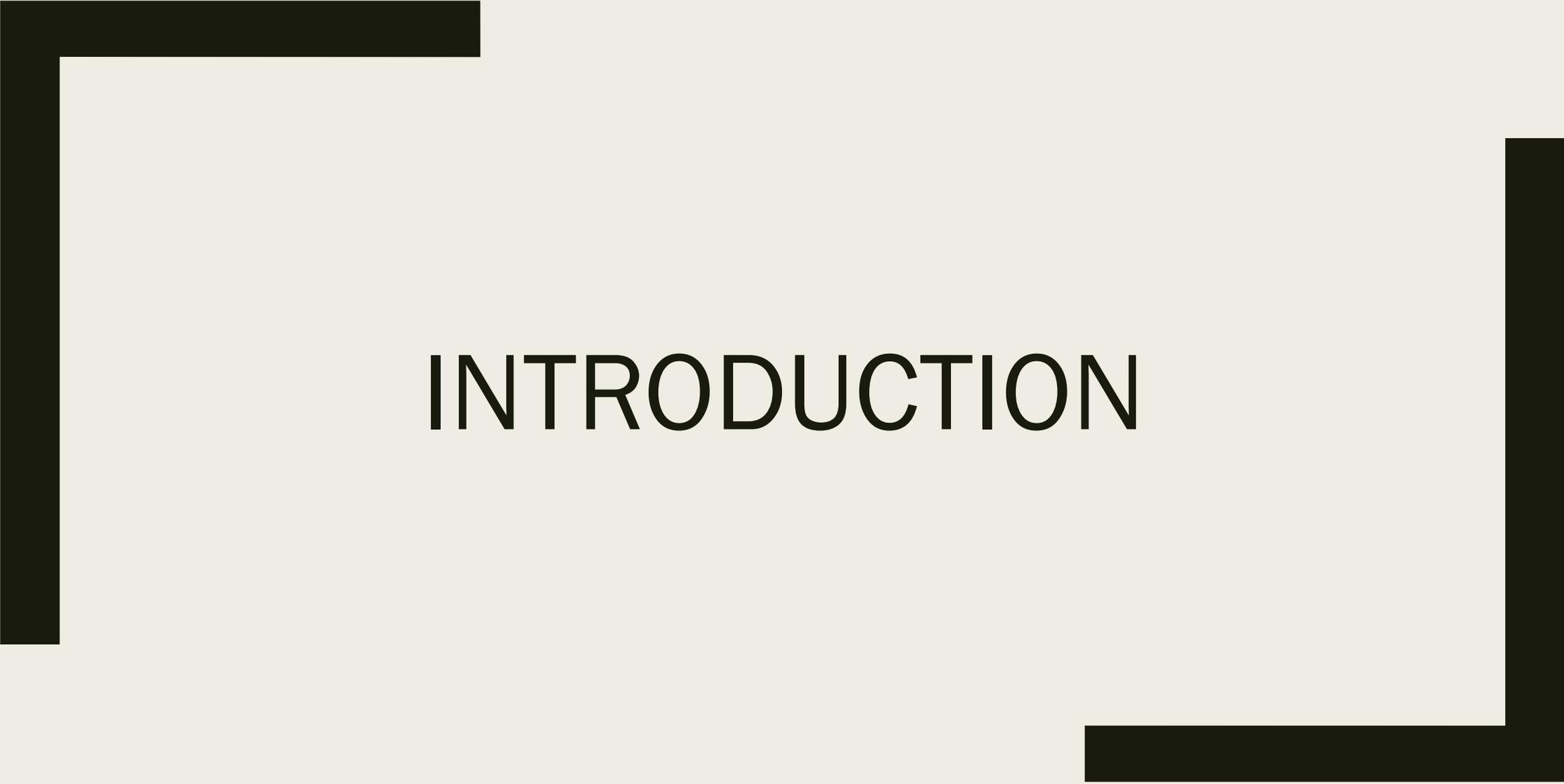
# Time in Translation NWO-project (2017-2022)



- Write a crosslinguistically valid semantics of the **PERFECT** *marker* : HAVE + Past Participle.
- Based (originally) on five Western European languages (Dutch, German, French, Spanish, English).
- Ingredients at the lexical, sentential, discourse and extralinguistic level.
- Main methodology: **Translation Mining** (parallel-corpora)
- Corpora: Camus (1942) *L'Etranger*, and J.K. Rowling (2012) *Harry Potter and the Philosopher's stone*.

# Roadmap

- 1. Introduction & Research Questions: The semantics of the Perfect, its competition with the Past, and its crosslinguistic variation.
- 2. Methods: Translation Mining and *L'étranger*: subset relation across languages.
- 3. Testing for the reliability and granularity of parallel-corpora methods.
- 4. A case study: the *Present Perfect puzzle* in English and Spanish (and Dutch) through an acceptability judgment task.
- 5. General conclusion (tl;dr): triangulate your data for better generalizations and wider empirical coverage.

The image features two large, thick black L-shaped brackets. One is positioned in the top-left corner, and the other is in the bottom-right corner. They are oriented towards each other, framing the central text.

# INTRODUCTION

# Introduction

## Perfect/Past in English

Variation between Perfect and (Perfective) Past forms responds to a set of constraints not clearly understood. (1) and (2) are both possible:

- |  |                                    |
|--|------------------------------------|
| 1) Mary <i>ate</i> breakfast at that café.       | (Perfective) <b>Past</b> $E-R < S$ |
| 2) Mary <i>has eaten</i> breakfast at that café. | (Present) <b>Perfect</b> $E < R,S$ |

but constraints in (3) and (4) only allow the Simple Past and bans Perfect use:

- |  |  |
|--|--|
| 3) Mary ( <i>#has</i> ) <i>left</i> work at 5pm.                                   | <b>R=S restricts past-time adverbs</b> |
| 4) When Mary ( <i>#has</i> ) <i>left</i> , she ( <i>#has</i> ) <i>waved</i> at me. | <b>R=S blocks anaphors/narration.</b>  |

# Introduction

## The Perfect

- The PERFECT: auxiliary (*have*) + past participle (i.e., *HAVE* perfect, Dahl & Velupillai 2013).
- **Core** meaning: past event and resulting state with *current relevance*.
- The Perfect is said to display at least four different readings (e.g., Iatridou et al. 2003, Portner 2003, i.a.):

**Resultative:** Mary has moved to Utrecht (now she lives there).

**Experiential:** Mary has visited Utrecht (and she might go again).

**Continuative:** Mary has lived in Utrecht for 5 years (and she still does)

**'Hot news' (hodiernal):** Mary has won the lottery (just now)!

# Introduction

## Lack of crosslinguistic coverage

- However, most research has been done in English (e.g., McCawley 1981, Michaelis 1994, Portner 2003, Nishyama & Koenig (2010), where these descriptions (and partially those analyses) work.
- However, different languages allow their corresponding Perfect (and Past) marker to express these readings to different extents.
  - *Marie a pris / (#prit) le petit-déjeuner dans ce café*
  - *Maria heeft (in dat café) ontbeten / (ontbeet) in dat café*
  - *Maria hat gefrühstückt / (frühstückte) in diesem Café*
  - *María (?ha desayunado) / desayunó en ese café*
- There are some **language-specific analysis** other than English (Bertinetto 1986, Vet 1992, Boogaart 1999, Löbner 2002, Howe 2006) but only more recently some **awareness of cross-linguistic variation** (e.g., de Swart 2007, Schaden 2009, Dahl & Velupillai 2013).

# Introduction

## Constraints in other languages

- French, Dutch, German PERFECT: compatible with past time adverbials.
  - *Marie est sortie du travail à 17 heures.* [French/German]
  - *Maria is om 17 uur van werk weggegaan.* [Dutch]
- French, German PERFECT: allow narrative use, but Dutch does not.
  - *Quand Marie est partie, elle m'a fait signe.* [French/German]
  - *#Toen Marie is weggegaan, heeft zij naar me zwaaiden.* [Dutch]
- Three-way division of languages (de Swart 2007):
  - *German/French can use PERFECT in narration, or w/past-time adverbs;*
  - *Dutch is ok with past-time adverbials, but not with narration*
  - *English (and Spanish?) cannot do either.*

# Introduction

## Diachronic and synchronic variation

- PERFECT is diachronically unstable (“aoristic drift”) (Harris 1982, Dahl 1985, Bybee et al. 1994).
- Condoravdi & Deo (2014): Resultative aspect markers often develop into perfect markers, which then end up as **perfect + perfective**.
- Schaden (2009): to understand the Perfect, we need to look at its competition with the PAST: a dichotomy where Spanish/English pattern together, as opposed to German/French.
- Variation across dialects is also prevalent (e.g., for Spanish, Howe 2006, Fuchs & González 2022).
- (Not) Surprisingly: crosslinguistic variation has had little effect on semantic accounts of the PERFECT (focused on English).

# Introduction

## Diachronic and synchronic variation

- The Perfect is still *not* a Past, since it allows present and future-time adverbials (e.g., German; Musan 2001: 361):
  - PAST: *Hans schrieb **gestern** den Brief*
  - PAST: *\*Hans schrieb **jetzt/morgen** den Brief*
  
  - PERF: *Hans hat **gestern** den Brief geschrieben*
  - PERF: *Hans hat **jetzt/morgen** den Brief geschrieben*

# Introduction

## Our methods

- Research question:
  - *Which features differentiate the PAST and the PERFECT in languages that have two markers? Are there only constraints on past-time adverbials and narration? Is there something more at play?*
- How to tackle the problem:
  - *Parallel corpus : translation (same context/meaning, different forms).*
  - *Data-driven: extract all finite verb forms in one language, align with translations, analyze similarities/differences in marker use.*
  - *Look at data patterns and extract generalizations.*
  - *Experimentally check the reliability (and the granularity) of the results.*

The image features two large, thick black L-shaped brackets. One is positioned in the top-left corner, and the other is in the bottom-right corner, framing the central text. The text is centered between these brackets.

# TRANSLATION MINING

# In a nutshell...

- **Convert** texts + translations into electronic documents.
- **Extract all finite forms** from multilingual corpora (e.g., *Passé Composé*)
- **Align** sentences (based on VPs) between original and translations.
- **Mark TMA forms** in the translations (with their language-specific labels: *Present Perfect, Simple Past, Pretérito Indefinido, Ontvoloid Verleden Tijd...*)
- **Create a dissimilarity matrix** to compare forms
- **Use multidimensional scaling** to visualize variation in a semantic map
- Go back to the **actual data** and extract generalizations.

# 1. VPs Extraction

- Application that is able to extract all finite forms across languages. For instance, for the Perfect:
  - *Looks for an auxiliary verb (HAVE/BE)*
  - *Finds a neighboring past participle*
- Takes care of words in between, lexical restrictions, reversed order, passive present perfects.

## 2. Verb/Sentence alignment

- TimeAlign Application
- Allows for manual alignment of translations.

### Annotation

<b>French (original)</b> <span>1.xml - s5.12</span> 	<b>Rioplatense Spanish (translated)</b> <span>1.xml - s5.13</span>
J' ai cru qu' il me reprochait quelque chose et j' ai commencé à lui expliquer .	Creí que me reprochaba alguna cosa y empecé a darle explicaciones .
	Selection tool: <span>Word</span> <span>Segment</span>
	<input type="checkbox"/> The selected words in the original fragment do not form an instance of (a/an) <i>passé composé</i>
	<input checked="" type="checkbox"/> This is a correct translation of the original fragment
	<b>Tense</b> pretérito indefinido (R)
	<b>Other (camus)</b> -----
	<b>Comments</b> Comments
	<input type="button" value="✓ Submit"/>

# 3. Marker assignment

- To calculate “distances” between fragments, we assign a TMA marker to every selected translation.

## Annotation

<b>French (original)</b> <span>1.xml - s5.12</span>	<b>Rioplatense Spanish (translated)</b> <span>1.xml - s5.13</span>
J' ai cru qu' il me reprochait quelque chose et j' ai commencé à lui expliquer .	Creí que me reprochaba alguna cosa y empecé a darle explicaciones .
	Selection tool: <span>Word</span> <span>Segment</span>
	<input type="checkbox"/> The selected words in the original fragment do not form an instance of (a/an) <i>passé composé</i>
	<input checked="" type="checkbox"/> This is a correct translation of the original fragment
	<b>Tense</b> pretérito indefinido (R)
	<b>Other (camus)</b> -----
	<b>Comments</b> Comments
	<input type="button" value="✓ Submit"/>

## 4. Dissimilarity matrix

- **Distance function:** a pair of source and translations is maximally similar ( $d=0$ ) if all tenses match up across languages. We compute  $+1/\#\text{languages}$  for every ‘mismatch’.

#	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 simple	Pretérito Indefinido	Voltooid tegenwoordige tijd

$$d(1, 2) = 2/5$$

$$d(2, 3) = 4/5$$

## 4. Dissimilarity matrix

- On the basis of the distance functions for each pair, we can create a dissimilarity matrix.

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

# 5. Multidimensional Scaling

- Multidimensional Scaling (MDS; Wälchli & Cysouw 2012) takes the dissimilarity matrix and creates a low-dimensional representation of it.
- It generates a cartographic visualization of groups of tense uses in context.
- It creates a two-dimensional map (a scatter plot), based on the comparison of all contexts in all the languages.
- Each dot represents a context (with the specific configuration/tuple across languages). Clusters are dots that are equal in their configuration.
- We use the tenses per language as labels with color coding.
  - *We can filter the tenses to see specific ones.*
  - *We can switch between languages.*
- We can click on individual points to see the source and translations.

# Corpus: *L'étranger* and its translations (van der Klis et al. 2020 et seq.)

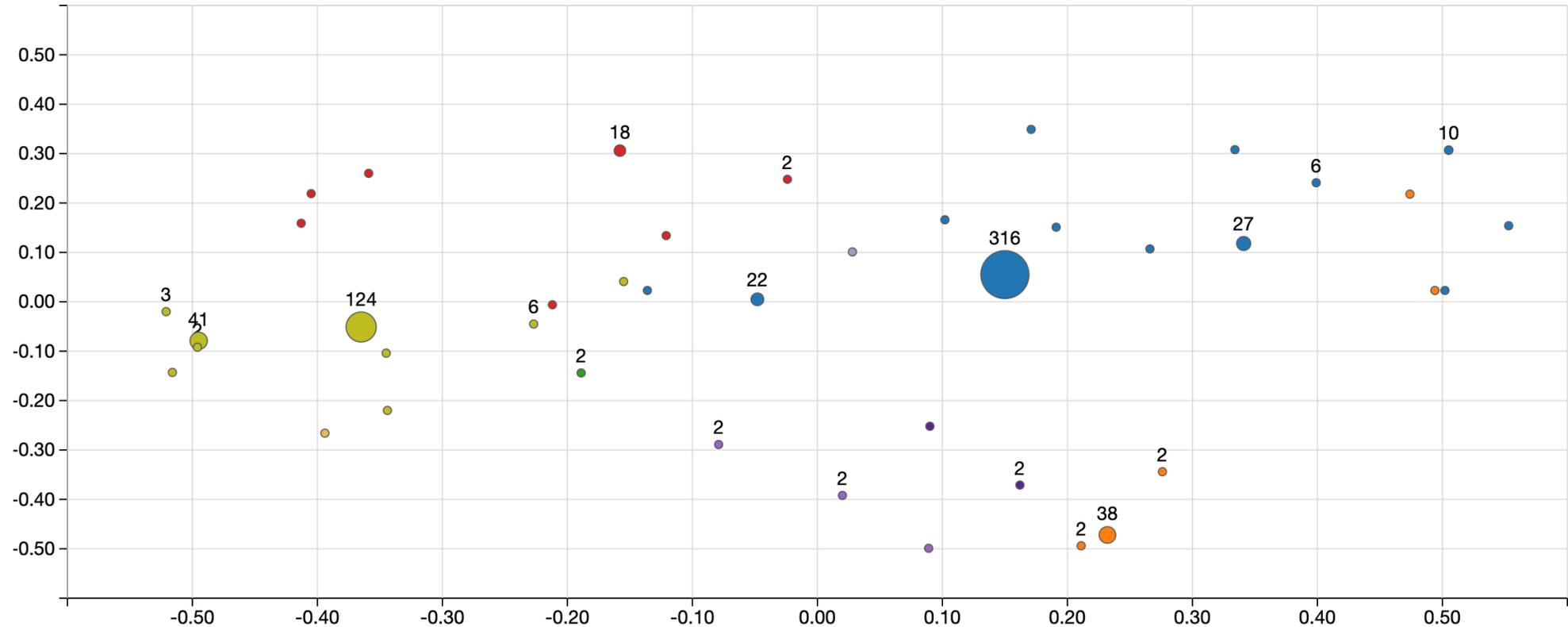
- Classical narration in French literature: *Passé Simple* + *Imparfait* (= PERFECTIVE/IMPERFECTIVE PAST  $\approx$  sequence of events/states).
- *L'Etranger*: sequence of *Passé Composé* + *Imparfait* (= PERFECT + IMPERFECTIVE PAST). This was not regular narrative style (in 1942).
- This literary style raises translation problems in languages that have a less liberal/more restricted use of the PERFECT.
- Approach: temporal maps showing the competition between PERFECTIVE PAST and PERFECT.
- From language use to grammar: determine the criteria that PERFECT use is sensitive to in French, German, Dutch, Spanish and English.

# French

## MDS visualization (scenario *Camus - finite indic. forms - KU Leuven*)

Normalized stress: 0.027 (0 indicates perfect fit, 0.025 excellent, 0.05 good, 0.1 fair and 0.2 poor)

📄 passé composé 📄 présent 📄 futur simple 📄 imparfait 📄 futur antérieur 📄 plus-que-parfait 📄 futur proche 📄 conditionnel présent 📄 passé simple



### Filters

Language: German English Spanish French Dutch  Show clusters

Dimension on x-axis: 1 2 3 4 5 Dimension on y-axis: ∅ 1 2 3 4 5

✓ Go!

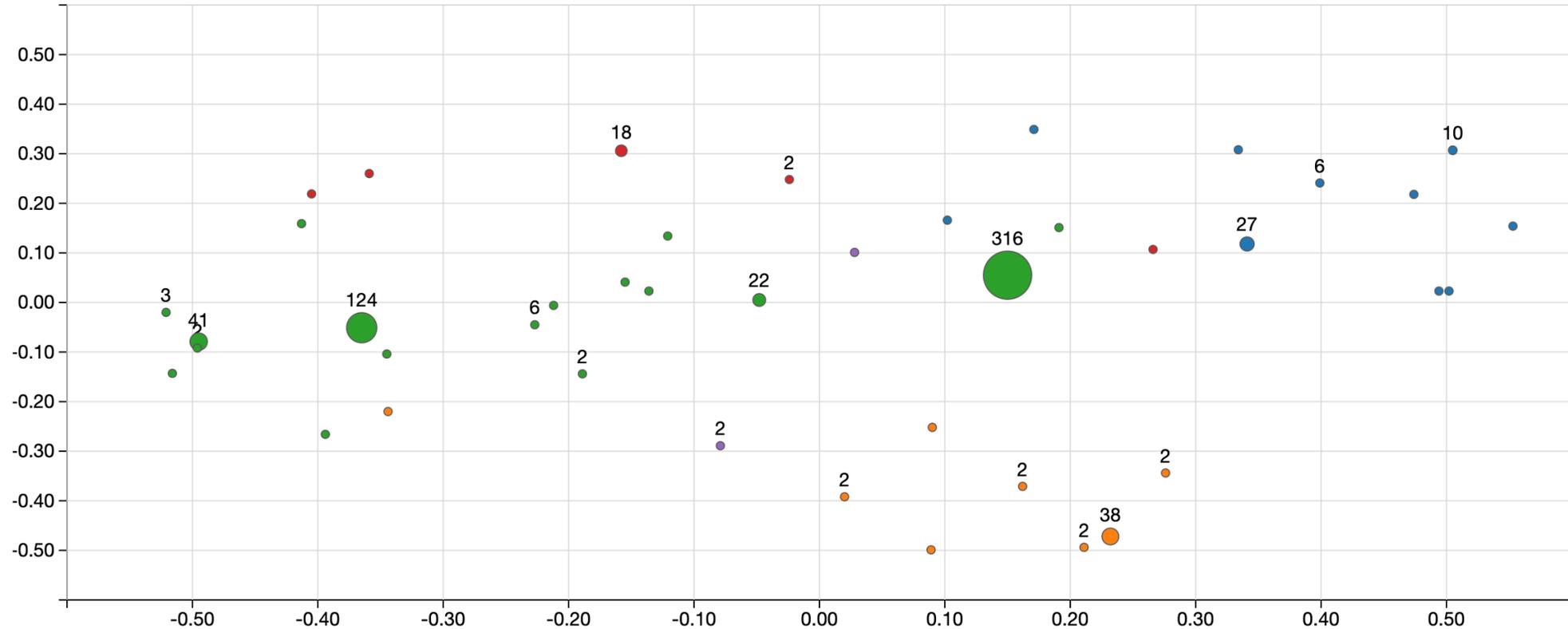


# Dutch

## MDS visualization (scenario *Camus - finite indic. forms - KU Leuven*)

Normalized stress: 0.027 (0 indicates perfect fit, 0.025 excellent, 0.05 good, 0.1 fair and 0.2 poor)



### Filters

Language: German English Spanish French Dutch  Show clusters

Dimension on x-axis: 1 2 3 4 5 Dimension on y-axis: ∅ 1 2 3 4 5

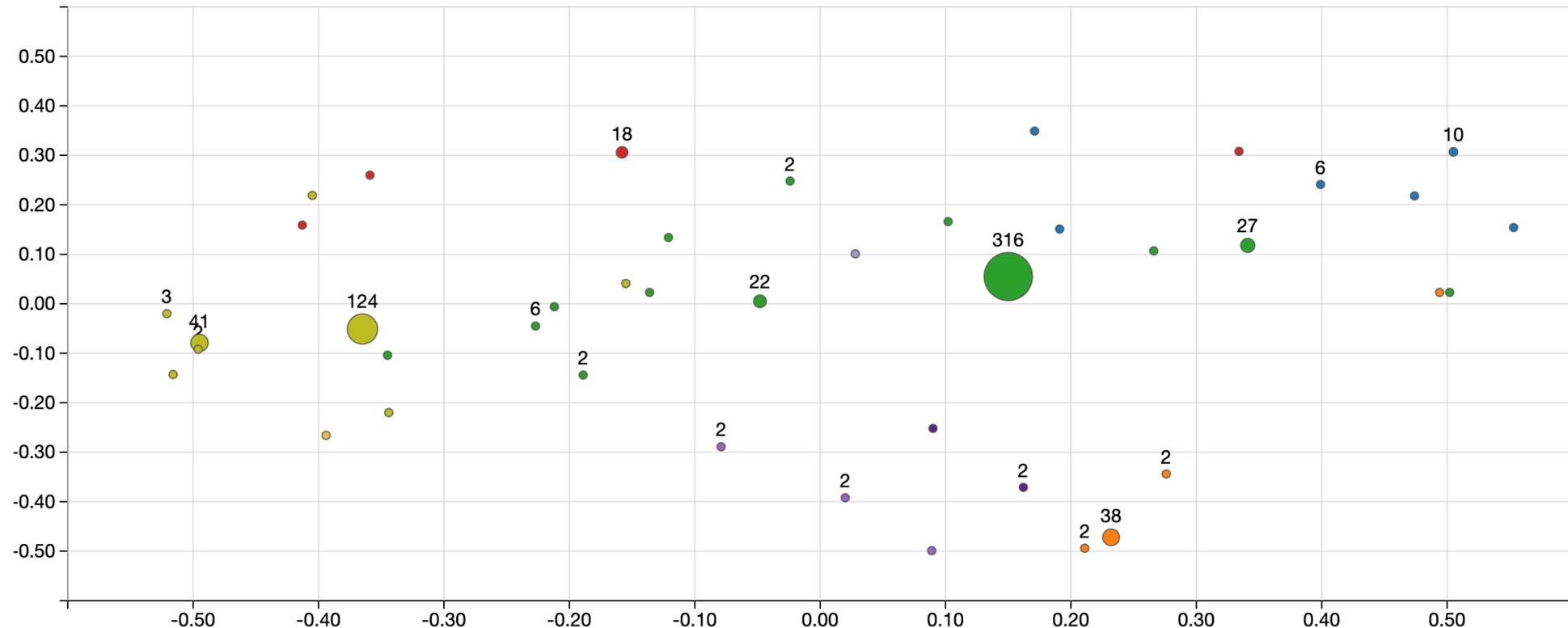


# Spanish

## MDS visualization (scenario *Camus - finite indic. forms - KU Leuven*)

Normalized stress: 0.027 (0 indicates perfect fit, 0.025 excellent, 0.05 good, 0.1 fair and 0.2 poor)

pretérito perfecto compuesto  pretérito indefinido  presente  futuro imperfecto  pretérito imperfecto  futuro perfecto  pretérito pluscuamperfecto  futuro próximo   
condicional simple



### Filters

Language:  German  English  Spanish  French  Dutch  On  Show clusters

Dimension on x-axis:  1  2  3  4  5 Dimension on y-axis:  0  1  2  3  4  5

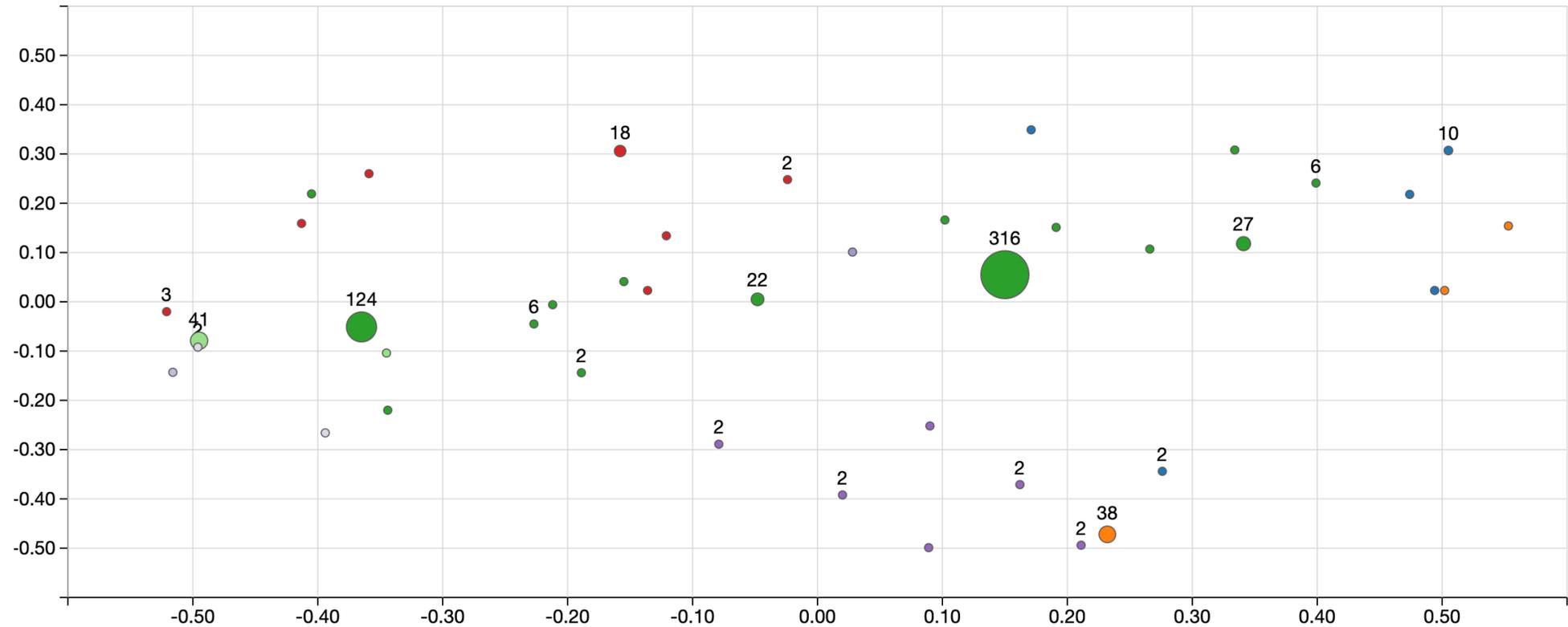
Go!

# English

## MDS visualization (scenario *Camus - finite indic. forms - KU Leuven*)

Normalized stress: 0.027 (0 indicates perfect fit, 0.025 excellent, 0.05 good, 0.1 fair and 0.2 poor)

simple past present perfect simple present simple future future perfect past continuous future in the past continuous past perfect future in the past



### Filters

Language: German English Spanish French Dutch  Show clusters

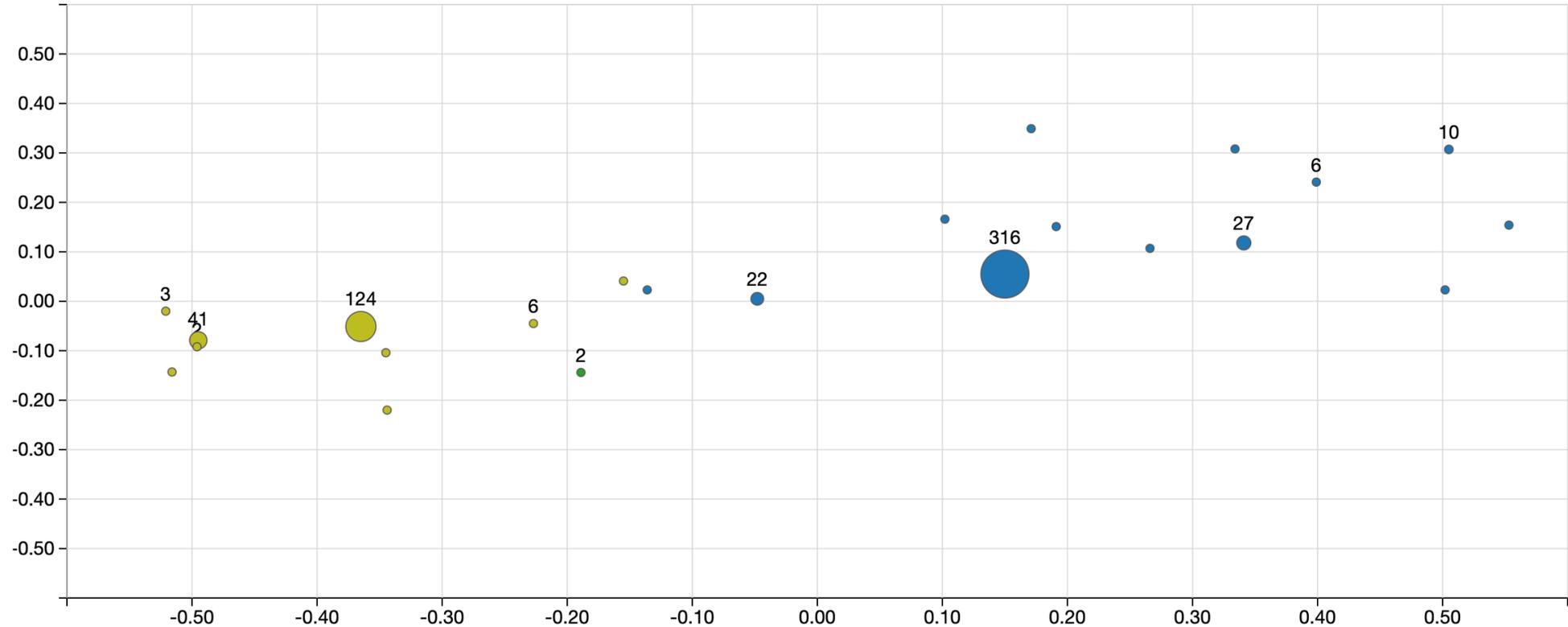
Dimension on x-axis: 1 2 3 4 5 Dimension on y-axis: ∅ 1 2 3 4 5

# French - PERFECT and PAST

## MDS visualization (scenario *Camus - finite indic. forms - KU Leuven*)

Normalized stress: 0.027 (0 indicates perfect fit, 0.025 excellent, 0.05 good, 0.1 fair and 0.2 poor)

☑ passé composé   □ présent   □ futur simple   ☑ imparfait   □ futur antérieur   □ plus-que-parfait   □ futur proche   □ conditionnel présent   ☑ passé simple



### Filters

Language: German English Spanish French Dutch   On  Show clusters

Dimension on x-axis: 1 2 3 4 5   Dimension on y-axis: ∅ 1 2 3 4 5

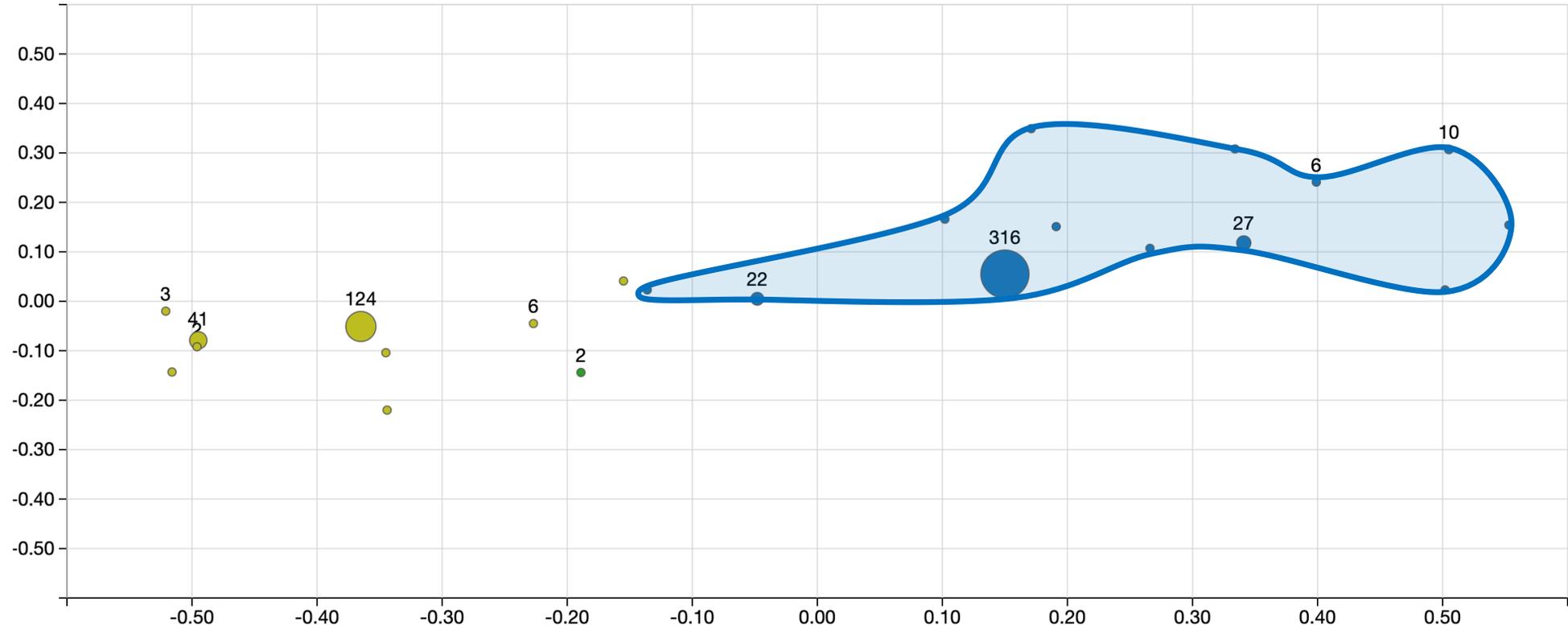
Go!

# French - PERFECT and PAST

## MDS visualization (scenario *Camus - finite indic. forms - KU Leuven*)

Normalized stress: 0.027 (0 indicates perfect fit, 0.025 excellent, 0.05 good, 0.1 fair and 0.2 poor)

passé composé  présent  futur simple  imparfait  futur antérieur  plus-que-parfait  futur proche  conditionnel présent  passé simple



### Filters

Language:  German  English  Spanish  French  Dutch  On  Show clusters

Dimension on x-axis:  1  2  3  4  5 Dimension on y-axis:  ∅  1  2  3  4  5

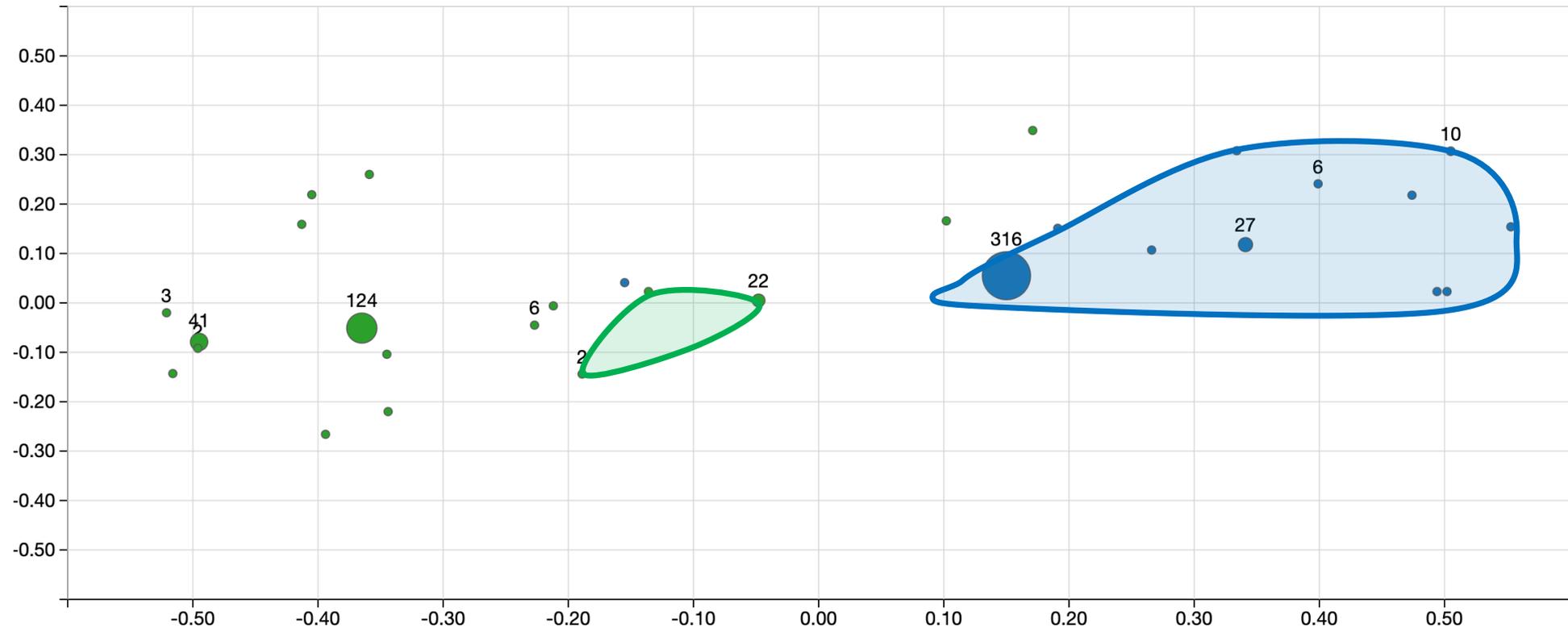
Go!

# German - PERFECT and PAST

## MDS visualization (scenario *Camus - finite indic. forms - KU Leuven*)

Normalized stress: 0.027 (0 indicates perfect fit, 0.025 excellent, 0.05 good, 0.1 fair and 0.2 poor)

Perfekt  Präteritum  Präsens  Futur I  Futur II  Plusquamperfekt



### Filters

Language:  German  English  Spanish  French  Dutch  On  Show clusters

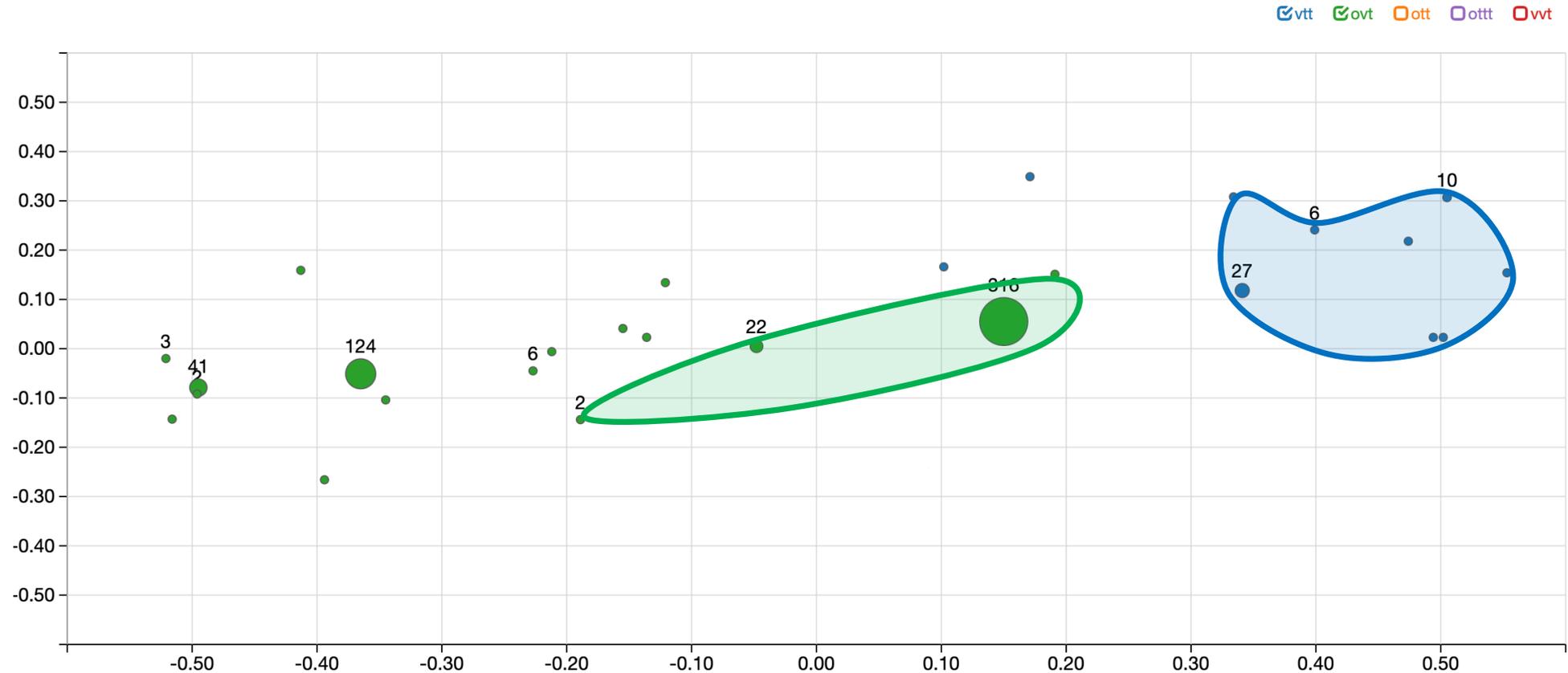
Dimension on x-axis:  1  2  3  4  5 Dimension on y-axis:  ∅  1  2  3  4  5

Go!

# Dutch - PERFECT and PAST

## MDS visualization (scenario *Camus - finite indic. forms - KU Leuven*)

Normalized stress: 0.027 (0 indicates perfect fit, 0.025 excellent, 0.05 good, 0.1 fair and 0.2 poor)



### Filters

Language: German English Spanish French Dutch  Show clusters

Dimension on x-axis: 1 2 3 4 5 Dimension on y-axis: ∅ 1 2 3 4 5

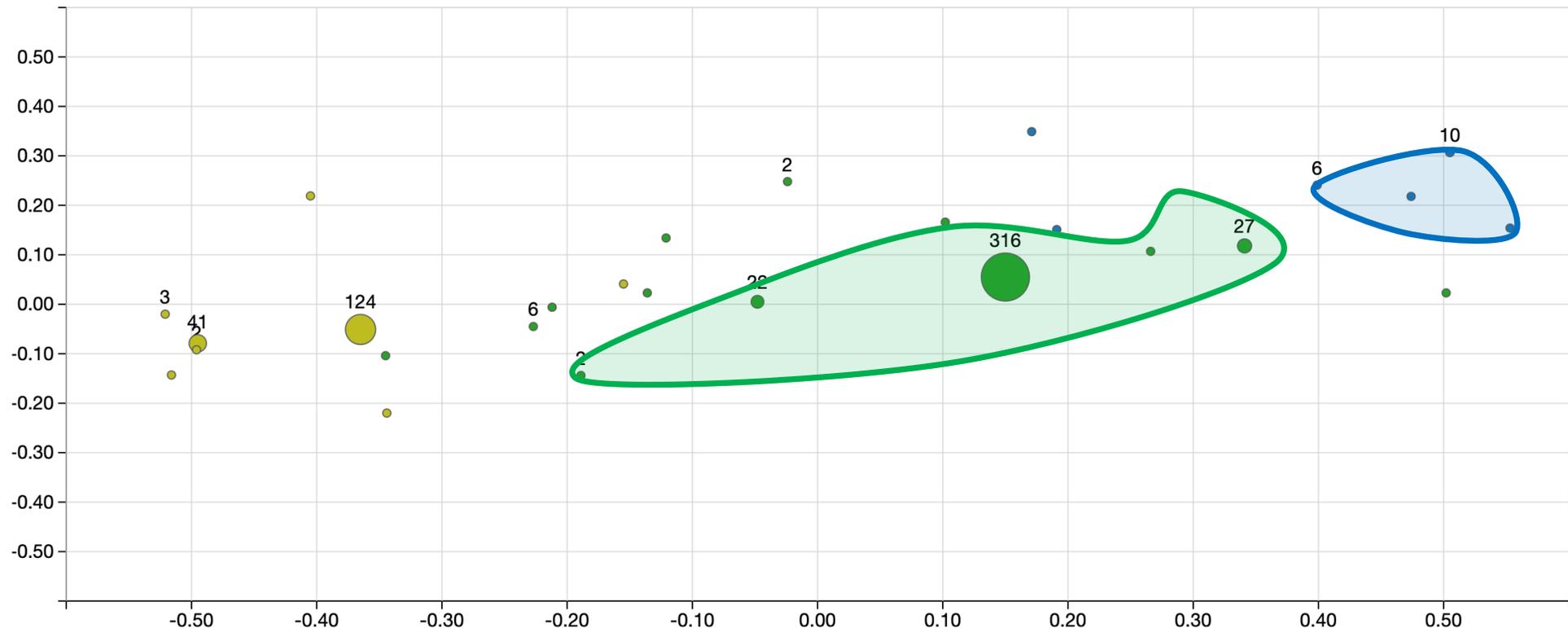
Go!

# Spanish - PERFECT and PAST

## MDS visualization (scenario *Camus - finite indic. forms - KU Leuven*)

Normalized stress: 0.027 (0 indicates perfect fit, 0.025 excellent, 0.05 good, 0.1 fair and 0.2 poor)

pretérito perfecto compuesto  pretérito indefinido  presente  futuro imperfecto  pretérito imperfecto  futuro perfecto  pretérito pluscuamperfecto  futuro próximo  condicional simple



### Filters

Language:  German  English  Spanish  French  Dutch  On  Show clusters

Dimension on x-axis:  1  2  3  4  5 Dimension on y-axis:  0  1  2  3  4  5

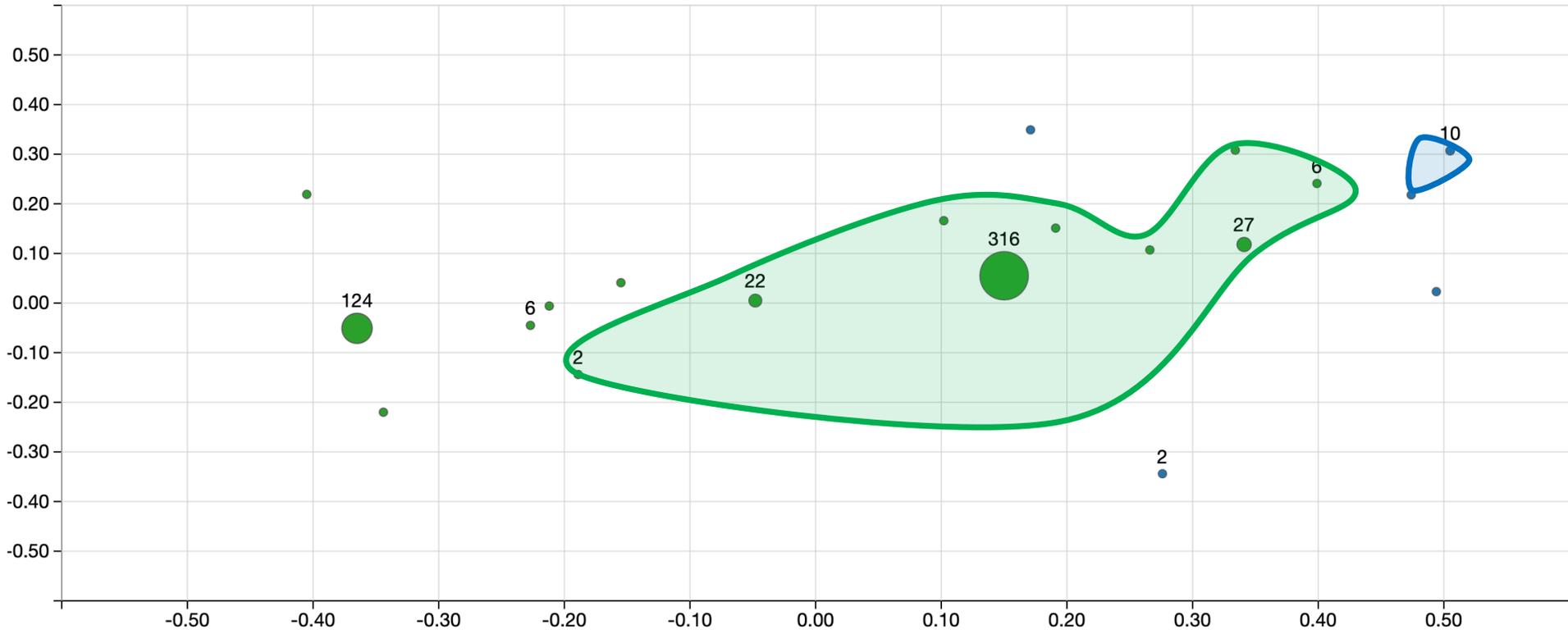
Go!

# English - PERFECT and PAST

## MDS visualization (scenario *Camus - finite indic. forms - KU Leuven*)

Normalized stress: 0.027 (0 indicates perfect fit, 0.025 excellent, 0.05 good, 0.1 fair and 0.2 poor)

simple past    present perfect    simple present    simple future    future perfect    past continuous    future in the past continuous    past perfect    future in the past



### Filters

Language:  German    English    Spanish    French    Dutch    On   Show clusters

Dimension on x-axis:  1    2    3    4    5   Dimension on y-axis:  0    1    2    3    4    5

Go!

# Descriptive statistics

## Descriptive statistics for *Camus - finite indic. forms - KU Leuven*

[Charts](#)
[Tables](#)
[Totals per category](#)
[Tuple frequencies](#)

### Tuple frequencies

Show colors
  Show categories
  Show labels

Search:

French	German	English	Spanish	Dutch	Count
passé composé	Perfekt	simple past	pretérito indefinido	ovt	316
imparfait	Präteritum	simple past	pretérito imperfecto	ovt	124
imparfait	Präteritum	past continuous	pretérito imperfecto	ovt	41
présent	Präsens	simple present	presente	ott	38
passé composé	Perfekt	simple past	pretérito indefinido	vtt	27
passé composé	Präteritum	simple past	pretérito indefinido	ovt	22
plus-que-parfait	Plusquamperfekt	past perfect	pretérito pluscuamperfecto	vvt	18
passé composé	Perfekt	present perfect	pretérito perfecto compuesto	vtt	10
passé composé	Perfekt	simple past	pretérito perfecto compuesto	vtt	6
imparfait	Präteritum	simple past	pretérito indefinido	ovt	6
imparfait	Präteritum	past perfect	pretérito imperfecto	ovt	3
futur simple	Futur I	simple future	futuro imperfecto	ottt	2

# Descriptive statistics

## Descriptive statistics for *Camus - finite indic. forms - KU Leuven*

Charts Tables Totals per category **Tuple frequencies**

### Tuple frequencies

On Show colors  Off Show categories  On Show labels

Search:

4. not Dutch

3. not Spanish

5. not German

1. All

2. not English

French	German	English	Spanish	Dutch	Count
passé composé	Perfekt	simple past	pretérito indefinido	ovt	316
imparfait	Präteritum	simple past	pretérito imperfecto	ovt	124
imparfait	Präteritum	past continuous	pretérito imperfecto	ovt	41
présent	Präsens	simple present	presente	ott	38
passé composé	Perfekt	simple past	pretérito indefinido	vtt	27
passé composé	Präteritum	simple past	pretérito indefinido	ovt	22
plus-que-parfait	Plusquamperfekt	past perfect	pretérito pluscuamperfecto	vvt	18
passé composé	Perfekt	present perfect	pretérito perfecto compuesto	vtt	10
passé composé	Perfekt	simple past	pretérito perfecto compuesto	vtt	6
imparfait	Präteritum	simple past	pretérito indefinido	ovt	6
imparfait	Präteritum	past perfect	pretérito imperfecto	ovt	3
futur simple	Futur I	simple future	futuro imperfecto	ottt	2

# Descriptive statistics

- The *Passé Composé* is generally translated by a **PERFECT** in German, and by a **PAST** in Dutch/Spanish/English.
- The most frequent combination: < *Passé Composé*, *Perfekt*, *Simple Past*, *Pretérito Indefinido*, *ovt*>.
- German patterns with French, Spanish pattern with English: variation within the family of Romance/Germanic languages (Schaden 2009)

# From statistics to grammar

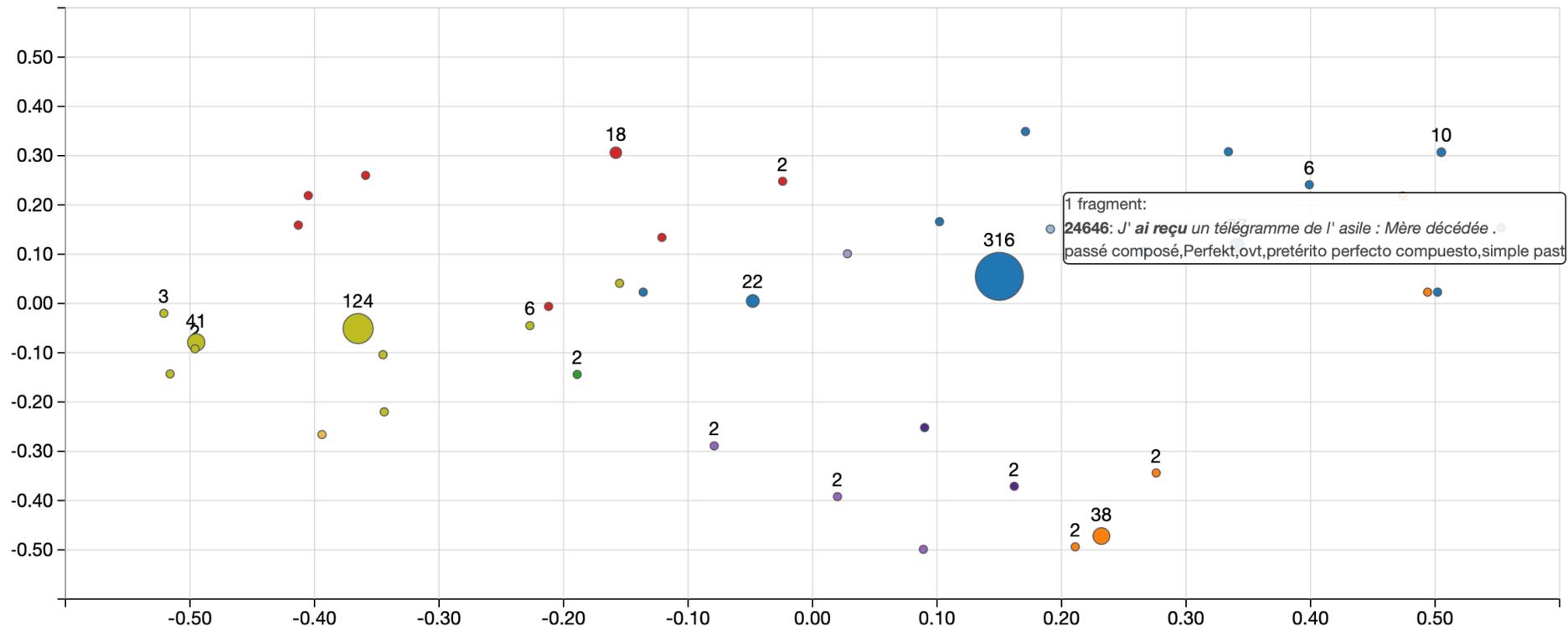
- Limits of descriptive statistics: global tendencies at the level of the grammar.
- Principle of isomorphism (Haiman 1985): variation in form reflects variation in meaning.
- Our semantic maps allow us to dig deeper and extract qualitative generalizations!

# From the maps to the data

## MDS visualization (scenario *Camus - finite indic. forms - KU Leuven*)

Normalized stress: 0.027 (0 indicates perfect fit, 0.025 excellent, 0.05 good, 0.1 fair and 0.2 poor)

☑ passé composé ☑ présent ☑ futur simple ☑ imparfait ☑ futur antérieur ☑ plus-que-parfait ☑ futur proche ☑ conditionnel présent ☑ passé simple



### Filters

Language: German English Spanish French Dutch On Show clusters

Dimension on x-axis: 1 2 3 4 5 Dimension on y-axis: ∅ 1 2 3 4 5

Go!

# Dynamic Interface: back-and-forth between maps and data

## Fragment overview

### Source

French

passé composé (fr) 1.xml - 24646

J' ai reçu un télégramme de l' asile : Mère décédée .

### Translations

German

Perfekt

Ich habe ein Telegramm vom Heim bekommen : « Mutter verstorben .

English

simple past

I had a telegram from the home : ' Mother passed away .

Spanish

pretérito perfecto compuesto

He recibido un telegrama del asilo : « Madre fallecida .

Italian

passato prossimo

Ho ricevuto un telegramma dall' ospizio : " Madre deceduta .

Dutch

ovt

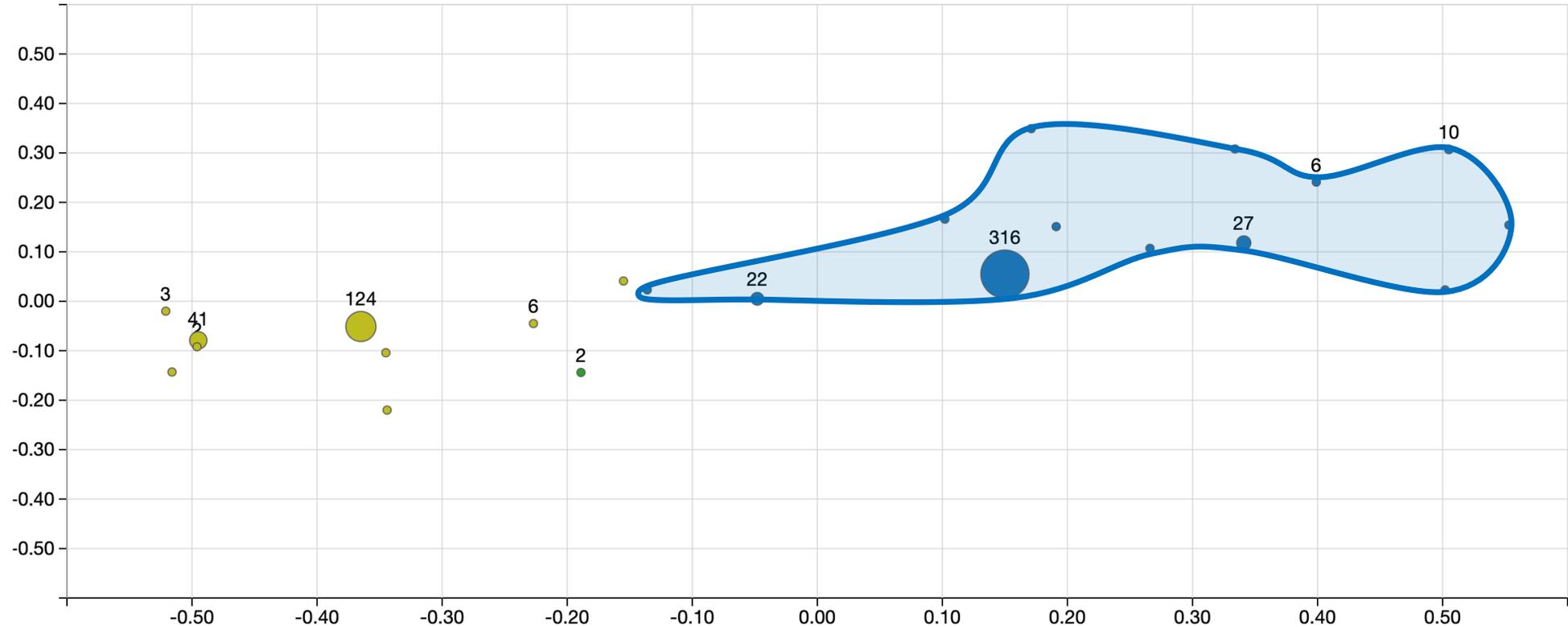
Ik ontving een telegram uit het gesticht : ' Moeder overleden .

# French - PERFECT and PAST

## MDS visualization (scenario *Camus - finite indic. forms - KU Leuven*)

Normalized stress: 0.027 (0 indicates perfect fit, 0.025 excellent, 0.05 good, 0.1 fair and 0.2 poor)

passé composé  présent  futur simple  imparfait  futur antérieur  plus-que-parfait  futur proche  conditionnel présent  passé simple



### Filters

Language:  German  English  Spanish  French  Dutch  On  Show clusters

Dimension on x-axis:  1  2  3  4  5 Dimension on y-axis:  ∅  1  2  3  4  5

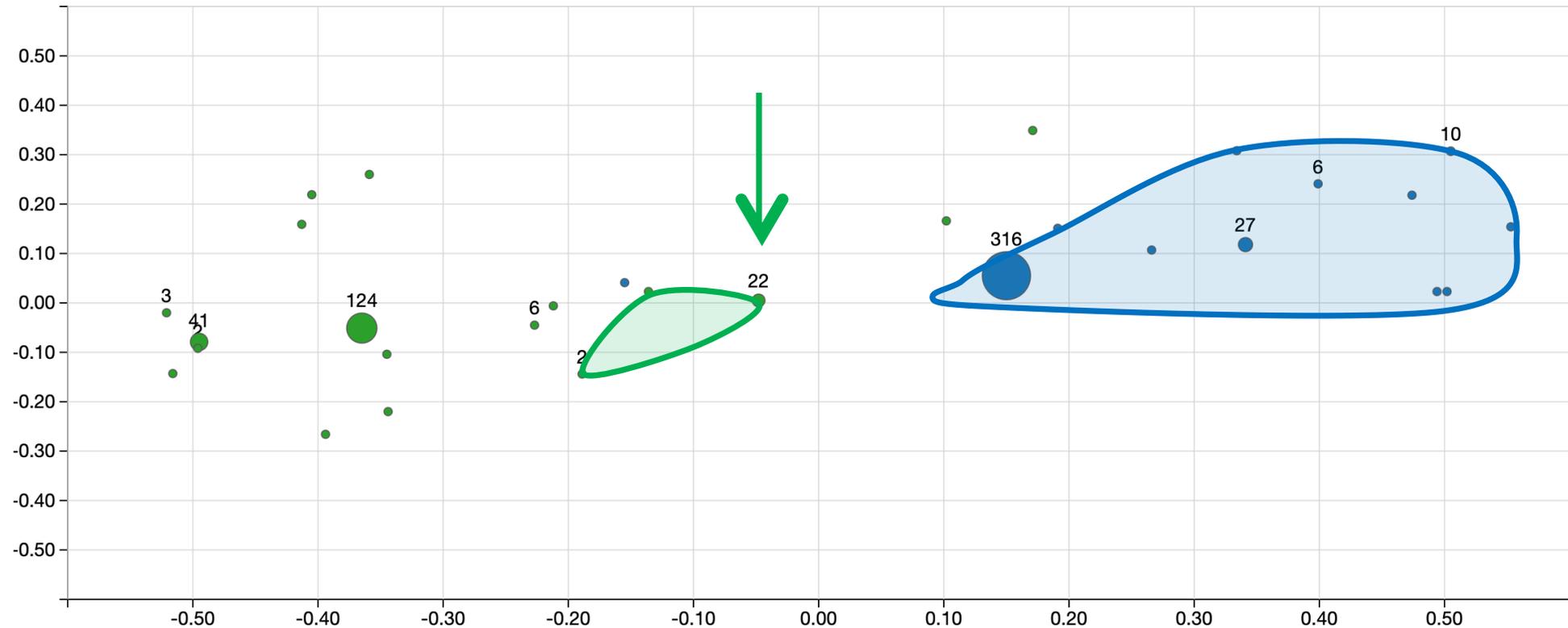
Go!

# German - PERFECT and PAST

## MDS visualization (scenario *Camus - finite indic. forms - KU Leuven*)

Normalized stress: 0.027 (0 indicates perfect fit, 0.025 excellent, 0.05 good, 0.1 fair and 0.2 poor)

Perfekt  Präteritum  Präsens  Futur I  Futur II  Plusquamperfekt



### Filters

Language:  German  English  Spanish  French  Dutch  On  Show clusters

Dimension on x-axis:  1  2  3  4  5 Dimension on y-axis:  ∅  1  2  3  4  5

Go!

# French > German: state verbs (in a narrative sequence)

- Lexical semantics drive the distinction between French and German.
- All languages but French require a PAST tense form with states in a narration.

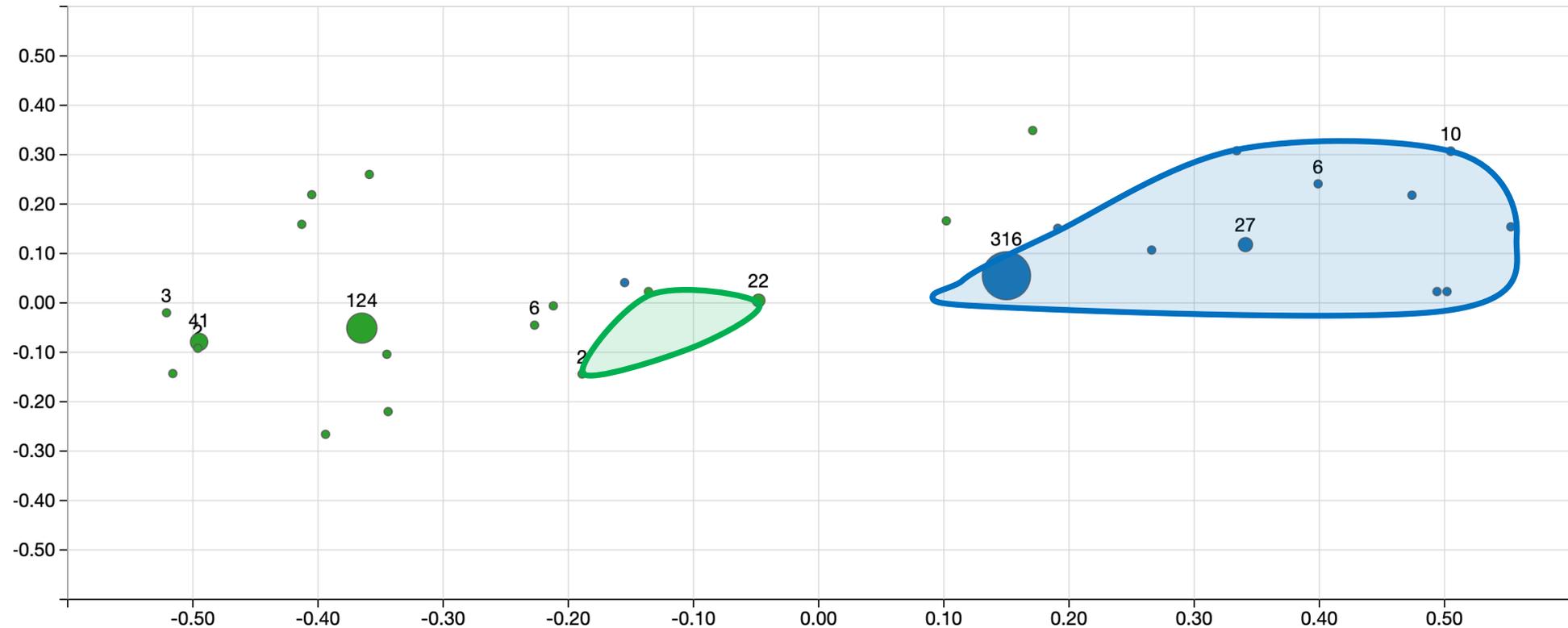
- |    |  |           |
|----|--|-----------|
| a. | J' <b>ai voulu</b> voir maman tout de suite. | [French]  |
| b. | Ich <b>wollte</b> sofort zu Mama.            | [German]  |
| c. | Ik <b>wilde</b> moeder meteen zien.          | [Dutch]   |
| d. | Yo <b>quería</b> ver a mamá inmediatamente.  | [Spanish] |
| e. | I <b>wanted</b> to see mother straight away. | [English] |

# German - PERFECT and PAST

## MDS visualization (scenario *Camus - finite indic. forms - KU Leuven*)

Normalized stress: 0.027 (0 indicates perfect fit, 0.025 excellent, 0.05 good, 0.1 fair and 0.2 poor)

Perfekt  Präteritum  Präsens  Futur I  Futur II  Plusquamperfekt



### Filters

Language:  German  English  Spanish  French  Dutch  On  Show clusters

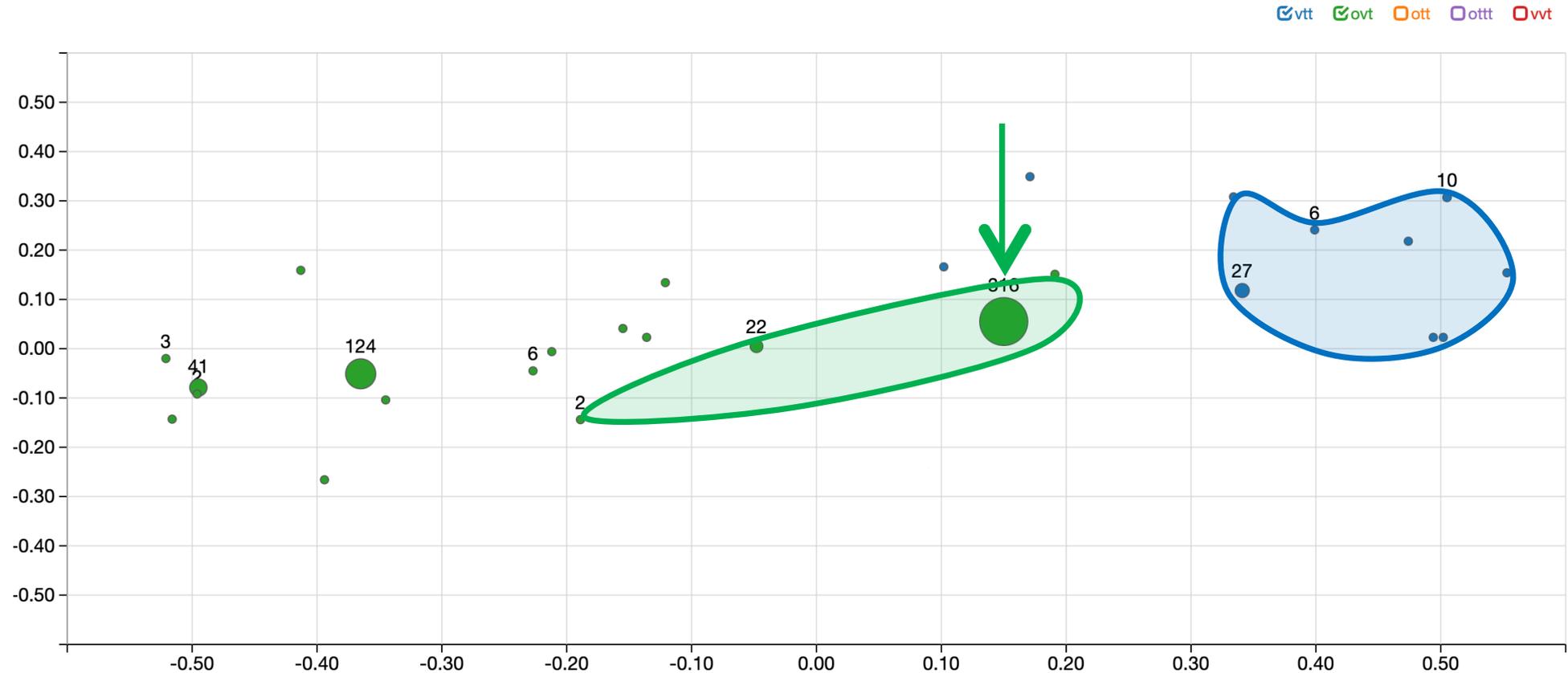
Dimension on x-axis:  1  2  3  4  5 Dimension on y-axis:  ∅  1  2  3  4  5

Go!

# Dutch - PERFECT and PAST

## MDS visualization (scenario *Camus - finite indic. forms - KU Leuven*)

Normalized stress: 0.027 (0 indicates perfect fit, 0.025 excellent, 0.05 good, 0.1 fair and 0.2 poor)



### Filters

Language: German English Spanish French Dutch  Show clusters

Dimension on x-axis: 1 2 3 4 5 Dimension on y-axis: ∅ 1 2 3 4 5

Go!

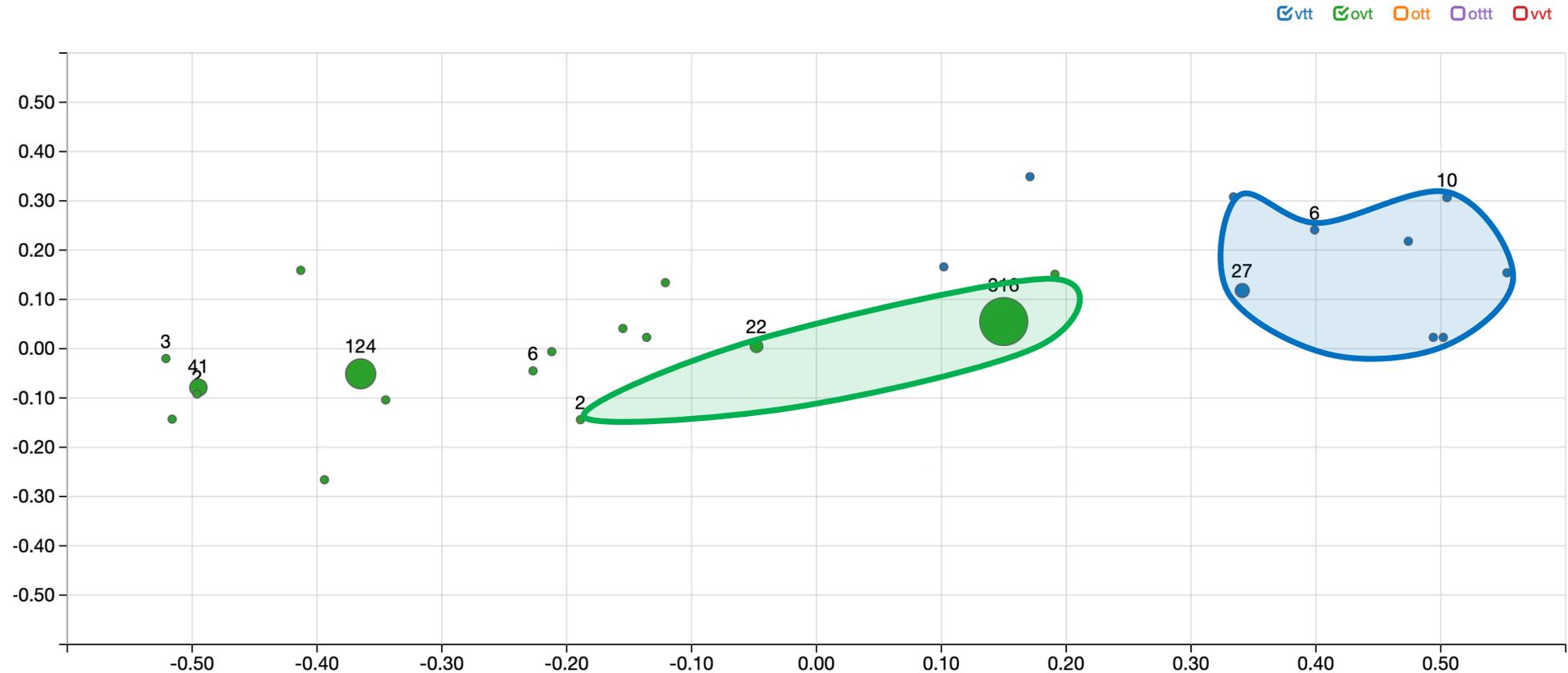
# German > Dutch: narration

- The German **Perfekt** can be used to tell a story (Löbner 2001, Schaden 2009).
- The Dutch **VTT** resists temporal progress in discourse (dynamic semantics).
  - a. Il **est sorti, est revenue, a disposé** des chaises. [French]
  - b. Er **ist hinaus gegangen, ist wieder gekommen, hat Stühle aufgestellt**.. [German]
  - c. Hij **ging** naar buiten, **kwam** weer terug en **zette** stoelen naar. [Dutch]
  - d. **Salió, volvió y colocó** las sillas. [Spanish]
  - e. He **went** in and out, arranging chairs. [English]

# Dutch - PERFECT and PAST

## MDS visualization (scenario *Camus - finite indic. forms - KU Leuven*)

Normalized stress: 0.027 (0 indicates perfect fit, 0.025 excellent, 0.05 good, 0.1 fair and 0.2 poor)



### Filters

Language: German English Spanish French Dutch  Show clusters

Dimension on x-axis: 1 2 3 4 5 Dimension on y-axis: ∅ 1 2 3 4 5

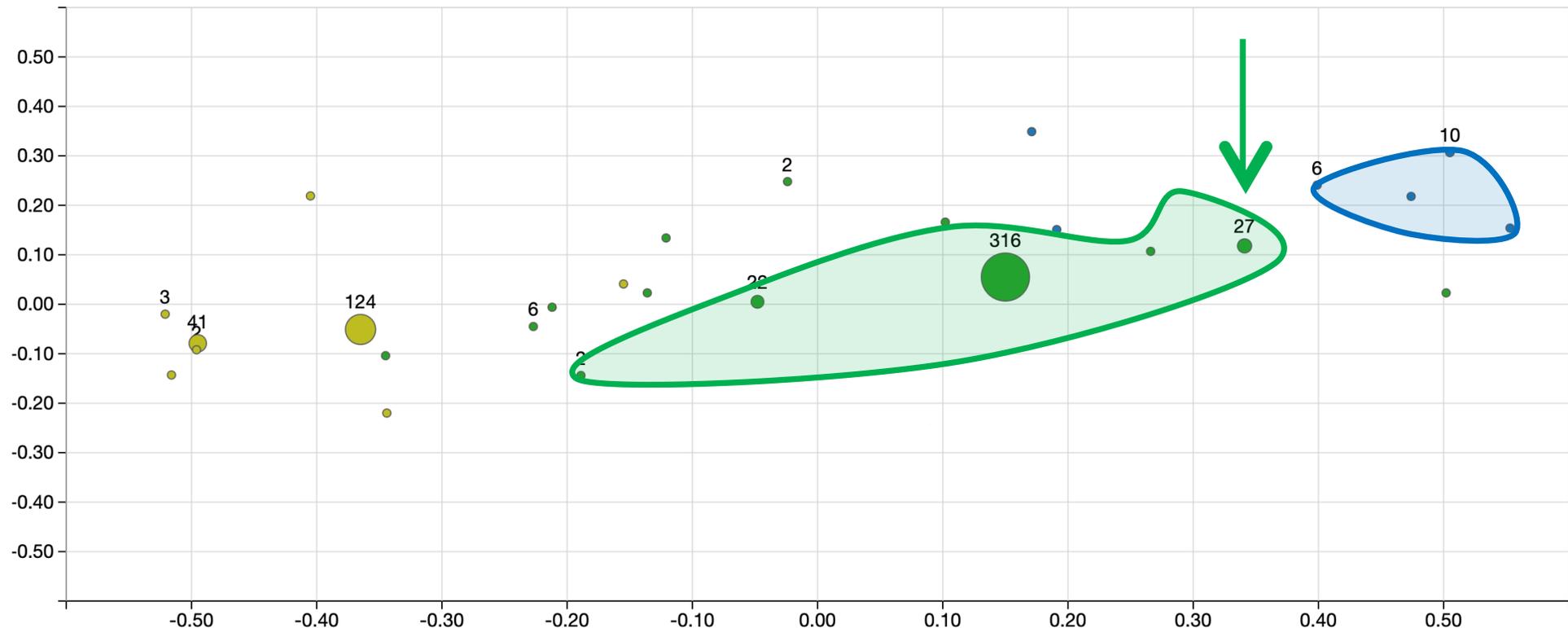
Go!

# Spanish - PERFECT and PAST

## MDS visualization (scenario *Camus - finite indic. forms - KU Leuven*)

Normalized stress: 0.027 (0 indicates perfect fit, 0.025 excellent, 0.05 good, 0.1 fair and 0.2 poor)

pretérito perfecto compuesto  pretérito indefinido  presente  futuro imperfecto  pretérito imperfecto  futuro perfecto  pretérito pluscuamperfecto  futuro próximo  condicional simple



### Filters

Language:  German  English  Spanish  French  Dutch  On  Show clusters

Dimension on x-axis:  1  2  3  4  5 Dimension on y-axis:  0  1  2  3  4  5

Go!

# Dutch > Spanish: past events

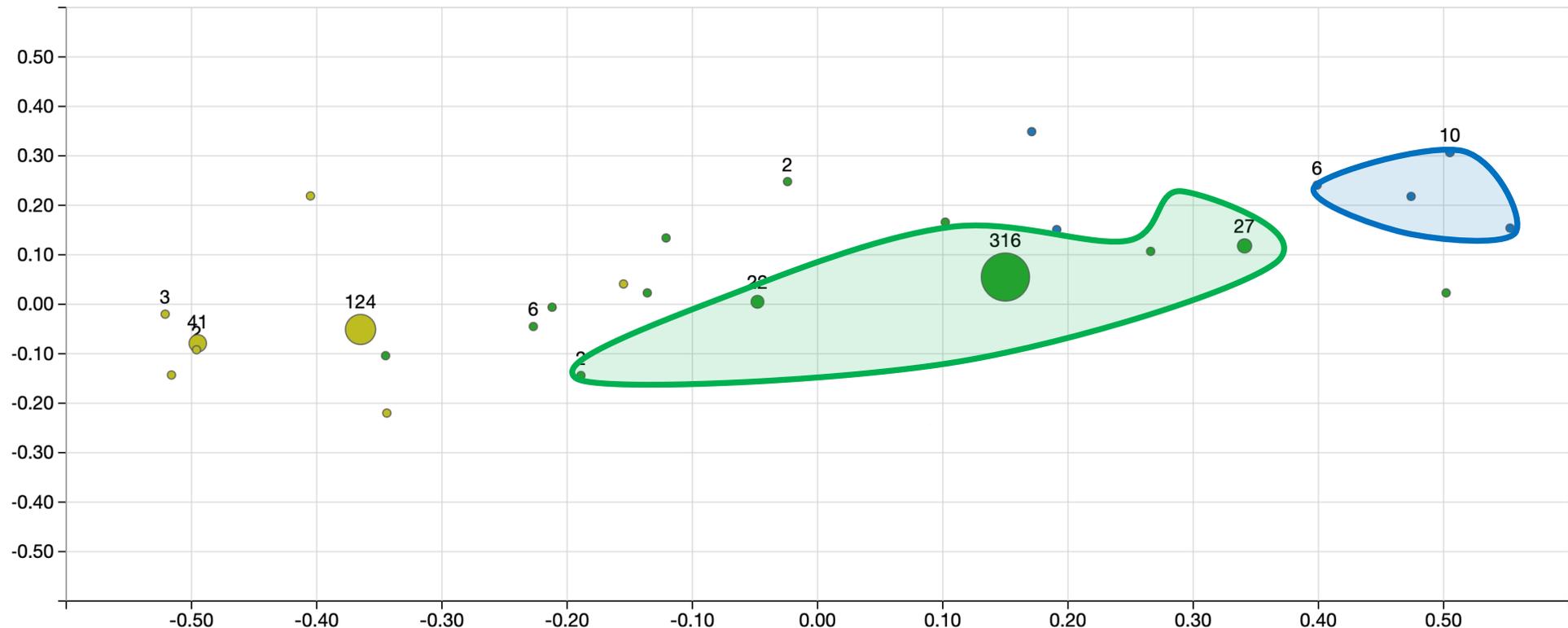
- Events delimited by a (prehodiernal) temporal adverbial are expressed by the PERFECT in Dutch/German/French, but require the *Préterito Indefinido* in Spanish. (compositional semantics).
  - a. Il **a perdu** son oncle, il y a quelque mois. [French]
  - b. Er **hat** voor ein paar Monaten seinen Onkel **verloren**, [German]
  - c. Hij **heeft** zijn oom een paar manden geleden **verloren**. [Dutch]
  - d. **Perdió** a su tío hace algunos meses. [Spanish]
  - e. He **lost** his uncle some months ago. [English]

# Spanish - PERFECT and PAST

## MDS visualization (scenario *Camus - finite indic. forms - KU Leuven*)

Normalized stress: 0.027 (0 indicates perfect fit, 0.025 excellent, 0.05 good, 0.1 fair and 0.2 poor)

pretérito perfecto compuesto  pretérito indefinido  presente  futuro imperfecto  pretérito imperfecto  futuro perfecto  pretérito pluscuamperfecto  futuro próximo  condicional simple



### Filters

Language:  German  English  Spanish  French  Dutch  On  Show clusters

Dimension on x-axis:  1  2  3  4  5 Dimension on y-axis:  0  1  2  3  4  5

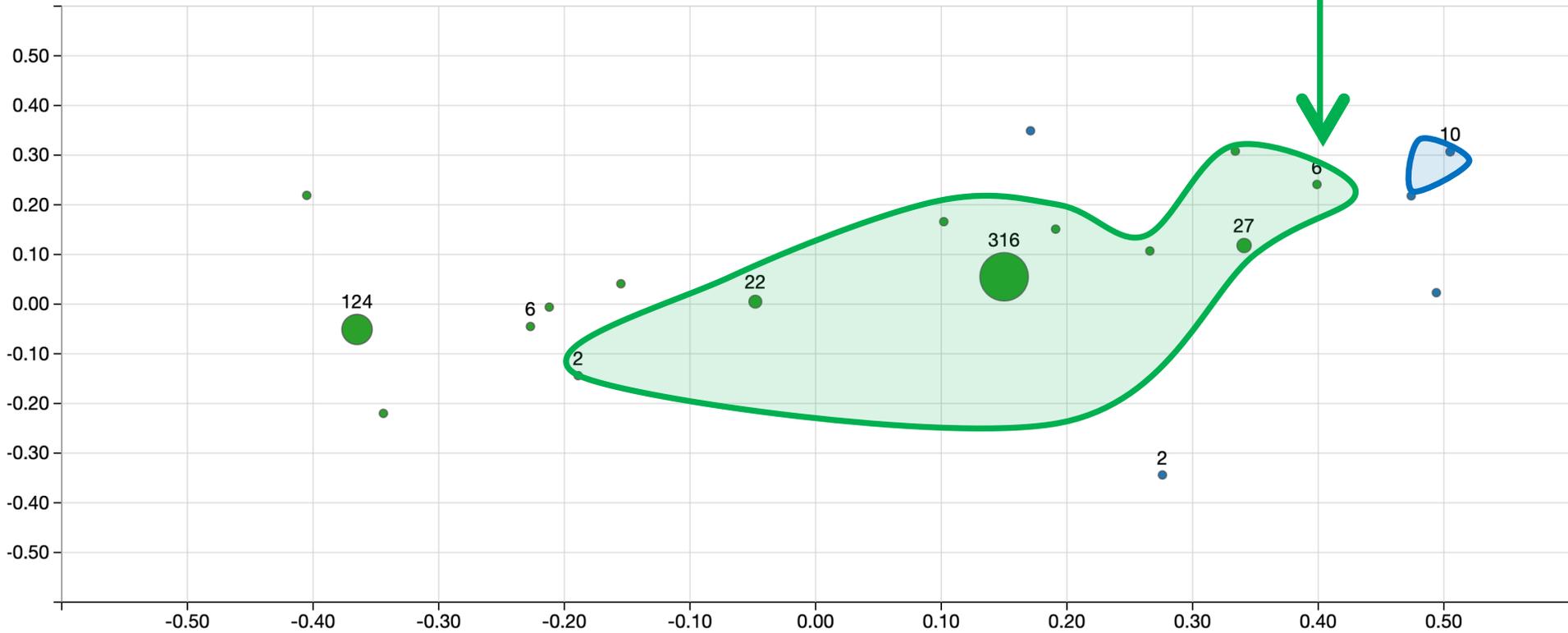
Go!

# English - PERFECT and PAST

## MDS visualization (scenario *Camus - finite indic. forms - KU Leuven*)

Normalized stress: 0.027 (0 indicates perfect fit, 0.025 excellent, 0.05 good, 0.1 fair and 0.2 poor)

simple past    present perfect    simple present    simple future    future perfect    past continuous    future in the past continuous    past perfect    future in the past



### Filters

Language:  German    English    Spanish    French    Dutch    On   Show clusters

Dimension on x-axis:  1    2    3    4    5   Dimension on y-axis:  0    1    2    3    4    5

Go!

# Spanish > English: pragmatics?

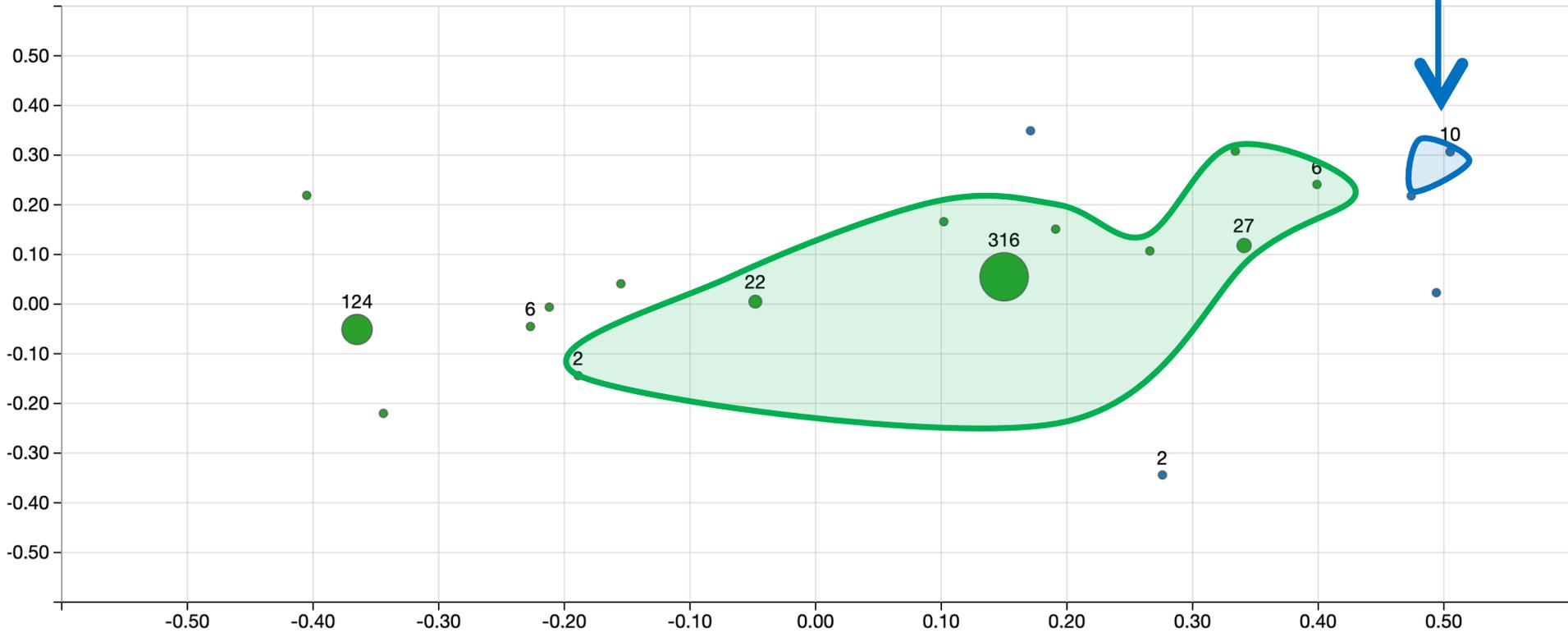
- *Pretérito Perfecto Compuesto* seems to be compatible with deictic adverb referring to ‘today’ in Spanish, but *Simple Past* in English.
  - a. Aujourd’hui maman *est morte*. [French]
  - b. Hoy mamá *ha muerto*. [Spanish]
  - c. Mother *died* today. [English]
  
- ‘Novelty’ of the PERFECT state and current relevance seem to play a role too. English does not allow to the *Present Perfect* to express pragmatically presupposed events (Michaelis 1994). The event needs to be hearer new
  - a. Il a begayé un peu: “On l’*a couverte*”. [French]
  - b. Tartamudeó un poco: “La *hemos cubierto*”. [Spanish]
  - c. He stuttered a bit: “We *covered* her up”. [English]

# English - PERFECT and PAST

## MDS visualization (scenario *Camus - finite indic. forms - KU Leuven*)

Normalized stress: 0.027 (0 indicates perfect fit, 0.025 excellent, 0.05 good, 0.1 fair and 0.2 poor)

simple past    present perfect    simple present    simple future    future perfect    past continuous    future in the past continuous    past perfect    future in the past



### Filters

Language:  German    English    Spanish    French    Dutch    On   Show clusters

Dimension on x-axis:  1    2    3    4    5   Dimension on y-axis:  ∅    1    2    3    4    5

Go!

# English: classical resultatives

- Results with current relevance allow a **PERFECT** in all languages.

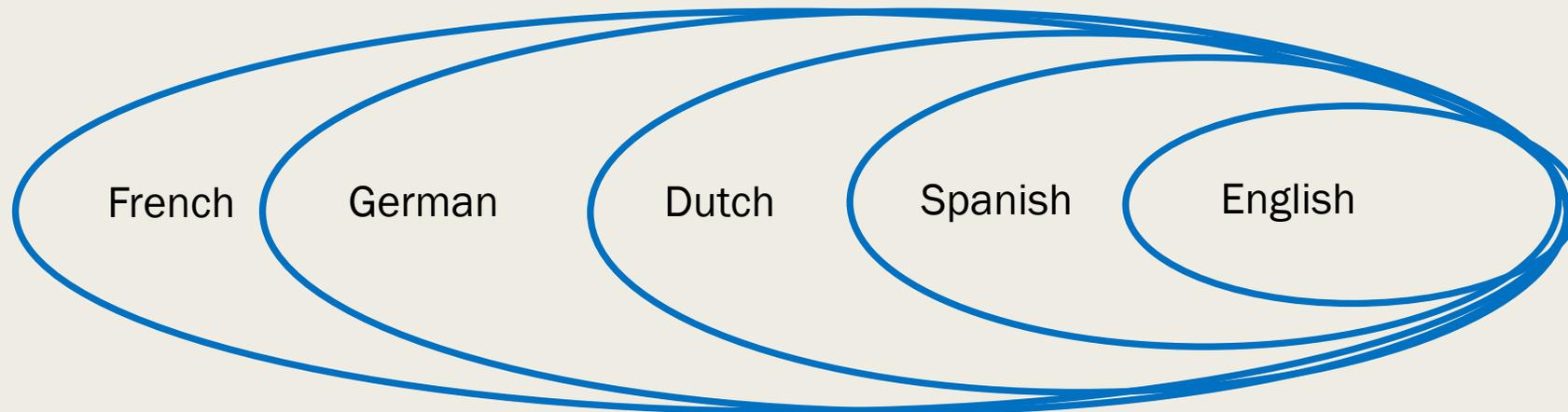
- a. Nous l'**avons transporté** dans notre petite morgue. [French]
- b. Wir **haben** sie in unsere kleine Leichenhalle **gebracht**, [German]
- c. Wij **hebben** haar naar ons lijkenhuisje **gebracht**. [Dutch]
- d. La **hemos transportado** a nuestro pequeño depósito. [Spanish]
- e. We've **transferred** her to our little mortuary. [English]

# English: classical existentials

- Existentials readings also allow a **PERFECT** in all languages.
  - a. Depuis huit ans, ils n'ont pas **changé** leur itinéraire. [French]
  - b. Seit acht Jahren **haben** sie ihre Route nicht **geändert**, [German]
  - c. In al die acht jaar **hebben** zij hun wandelroute niet **veranderd**. [Dutch]
  - d. Al cabo de ocho años, no **han cambiado** de itinerario. [Spanish]
  - e. In eight years they **haven't changed** their route. [English]

# Implicational hierarchy of PERFECT use

- ❑ Once a context switches from PERFECT to PAST in a particular language,
  - *the verb form remains a PAST tense in the next language: creates a*
  - *subset relation.*
- ❑ Not a dichotomy (Schaden 2009), but a sliding scale from more liberal to more restricted PERFECT languages (Van der Klis et al. 2020, 2021/22).



# The cross-linguistic distribution of the PERFECT

- Descriptive statistics show the global tendencies in the grammar.
- The temporal maps show that the distribution of tense forms is organized as a subset relation: the reduction in the domain of the PERFECT gives rise to a wider use of the PAST.
- Investigation of the individual data points provides the demarcation lines between each pair of languages.
- The linguistic principles governing the variation between languages imply (i) lexical semantics (stative verbs), (ii) compositional semantics (boundedness, continuity), (iii) dynamic semantics (narration) and (iv) pragmatics (information structure).

# Reflections on parallel corpus research

- Data driven approach, so results might be skewed due to the specific corpus.
- Camus is said to make a ‘special’ use of the *passé composé* in *L’Étranger* ~ potential translation bias towards extended PERFECT use in target languages.
- Can we reproduce the insights in a translation corpus with a different source language, preferably one that has a ‘classical’ PERFECT?
- 2<sup>nd</sup> TinT translation corpus: J.K. Rowling (2012) *Harry Potter and the Philosopher’s Stone* and its translations.
- Replicate the findings from Camus in a different corpus with a different source language to validate the *Translation Mining* methodology.

The image features two large, black, L-shaped brackets. One is positioned on the left side, with its vertical bar extending downwards and its horizontal bar extending to the right. The other is on the right side, with its vertical bar extending upwards and its horizontal bar extending to the left. These brackets frame the central text.

# RELIABILITY AND GRANULARITY

# Reliability issues

- Can we rely on the data? How generalizable is it?
- Are translations good representations of the language under study?
- Is the idiolect of one translator representative?
- *Grammar induced variation vs. ‘idiosyncratic’ variation: can we tell them apart?*
  - We rely on quantitative data (if there is a **systematic** change in usage, it’s probably grammar)
  - We don’t use our data blindly, but we **interpret** it and integrate it with previous findings in the literature.

# Granularity issues

- (Monolingual) Corpus work **only** shows attested forms ('best' options for each context): its dependent variable is necessarily **binary** (attested vs. non attested)
- Corpora provide data about the well-formedness of the **best structures only**, because alternatives that are still viable might not appear (especially in small-size corpora; Featherston 2005)
- **Parallel corpora** is better at picking up on **negative evidence**, since by keeping the context/meaning constant, we can see whether some forms do not show up in a given language.
- Grammaticality judgments tend to be **gradient**; and it has been repeatedly observed that **middle-rated constructions do not appear in corpora** (i.e., there is **discrepancy between frequency counts and grammaticality**; Kempen & Harbusch 2005, Gries 2005, Goldberg et al. 2004).

*We might be missing out on relevant constraints to build a valid cross-linguistic semantics.*

The image features two thick black L-shaped brackets. One is positioned in the top-left corner, and the other is in the bottom-right corner. They are oriented towards each other, framing the central text.

A CASE STUDY: THE PRESENT  
PERFECT PUZZLE IN SPANISH,  
ENGLISH AND DUTCH

# Introduction

The *present perfect puzzle*: “the **Present Perfect** does not go with an adverbial referring to the past” (Klein 1992: 526), so the **Simple Past** has to be used instead:

1) Chris *#has left* / *left* York today at six. (Klein 1992: 546)

But other languages in our cline (French, German, Dutch):

2) Chris *a quitté* (*quitta*) York aujourd’hui à 6 heures.

3) Chris *hat* York heute um 6 *verlassen* (*verließ*).

4) Chris *heeft* York vandaag om zes uur *verlaten*. (*verliet*)

# Introduction: hodiernality

Does Spanish pattern with English or with Dutch?

5) Chris se *ha ido* / *#fue* de York hoy a las seis.

6) Chris se *#ha ido* / *fue* de York ayer.

- Spanish seems to allow temporal adverbial that create the relation  $E=R \subseteq \text{day}(S)$  (e.g., Schwenter 1994)

# Introduction: deixis

Does English never allow temporal adverbials?

7) Chris *#has left* / *left* York today at six.

8) Chris *has left* / *left* York this morning

- English seems to allow *deictic* temporal adverbials (i.e., adverbials whose reference is calculated with respect to the speaker's time/space center of reference; e.g., Hitzeman 1995)

# Research questions

- Can we assess the role of both these constraints in PERFECT use across languages?
- Can we test for the strength of our corpora-based generalizations (reliability) and improving their level of detail (granularity)?
- What do we win by relying on two methodologies driven by empirical data? What is the link between corpora and experimental research in crosslinguistic semantics?

# Method: acceptability judgments

- 8 contexts that conveyed bounded **events** (lexical aspect: achievements) with different past-referring temporal adverbials.
- Three languages: Dutch > Spanish > English
- Three independent variables with two levels each (2x2x2):
  - Grammatical Marker: *PERFECT* / *PAST*
  - Temporal proximity:
    - +T : adverbials relate to day (S) by being included in it (e.g., *this morning*) overlapping (*today*) or including it (e.g., *this month*).
    - -T: adverbs do not include or are included in day (S) (e.g., *last month*).
  - Deixis:
    - +D: temporal reference of adverb is deictic in nature (e.g., *yesterday*)
    - -D: adverbs that can be placed on a timeline independently from speakers' center of reference (e.g., *in November*)

# Method: acceptability judgments

- Total of 64 stimuli (+96 fillers: PAST bad across languages, e.g., continuatives: Ella *ha vivido*/ \**vivió* en Berlín desde el 2010)
- Latin Square design.
- 1/2 of stimuli followed by comprehension questions.
- 160 subjects per language (recruited through MTurk and Neerlandistiek).
  - Dutch: The Netherlands - all provinces (mostly Utrecht)
  - Spanish: Spain- all regions (no differences, mostly Madrid)
  - English: UK- all regions (no differences, mostly greater London)
- Rating on a 5-point Likert scale

# Sample stimulus

Peter and Theresa are planning to go to a concert next weekend. Peter offers to go get the tickets later today, but Theresa tells him:

- +T, +D: “I *purchased* / *have purchased* mine **this morning**
- +T, -D: “I *purchased* / *have purchased* mine **at midnight**
- T, + D: “I *purchased* / *have purchased* mine **last month**
- T, -D: “I *purchased* / *have purchased* mine **in November**

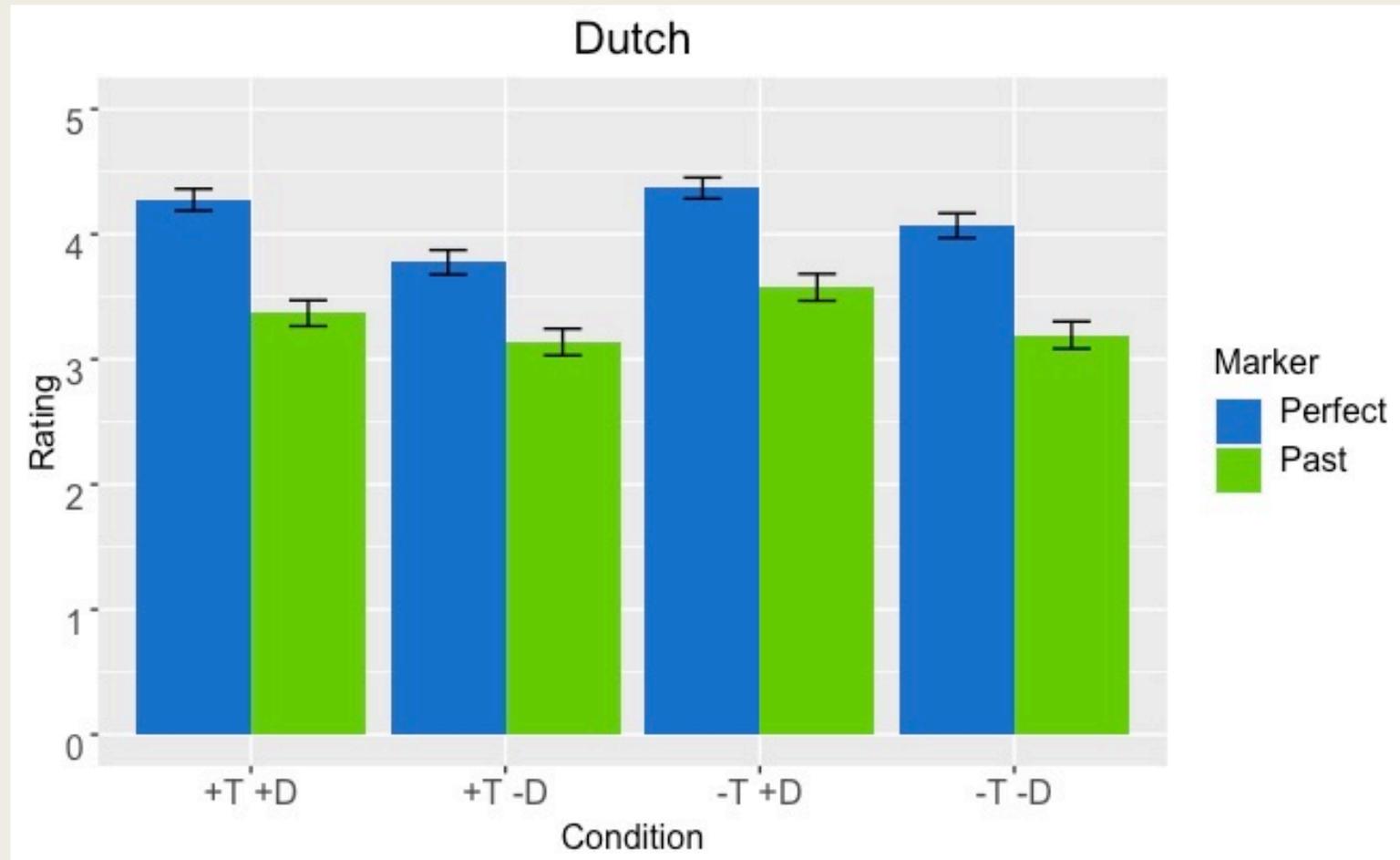
It was cheaper that way”.

# Data analysis & Results

- All participants performed at above 75% in comprehension questions.
- Linear mixed model analysis; *lme4* package (Bates et al. 2015), in R.
  - *Model selection by Likelihood Ratio Tests.*
  - *Post hoc tests in multcomp package (Hothorn et al. 2008).*
  - *p-values corrected by Tukey.*
- **Fixed effects:** interaction of Temporal Proximity\*Deixis\*Marker.
- **Random effects:** random intercepts for subject and item.

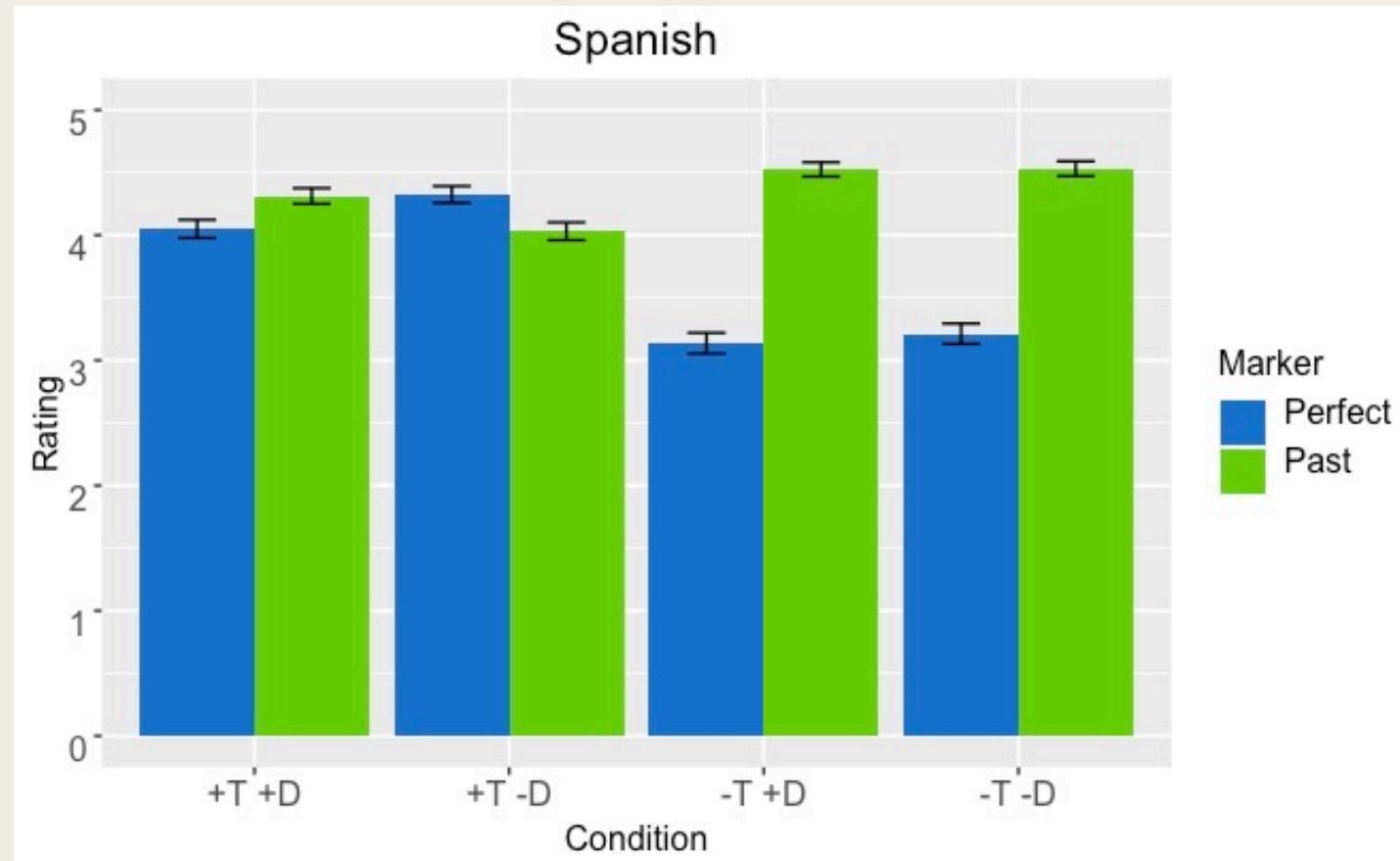
# Dutch

- Main effect of marker: **PERFECT** > **PAST** in all conditions ( $\chi^2(2) = 32.117$ ;  $p < .001$ ).



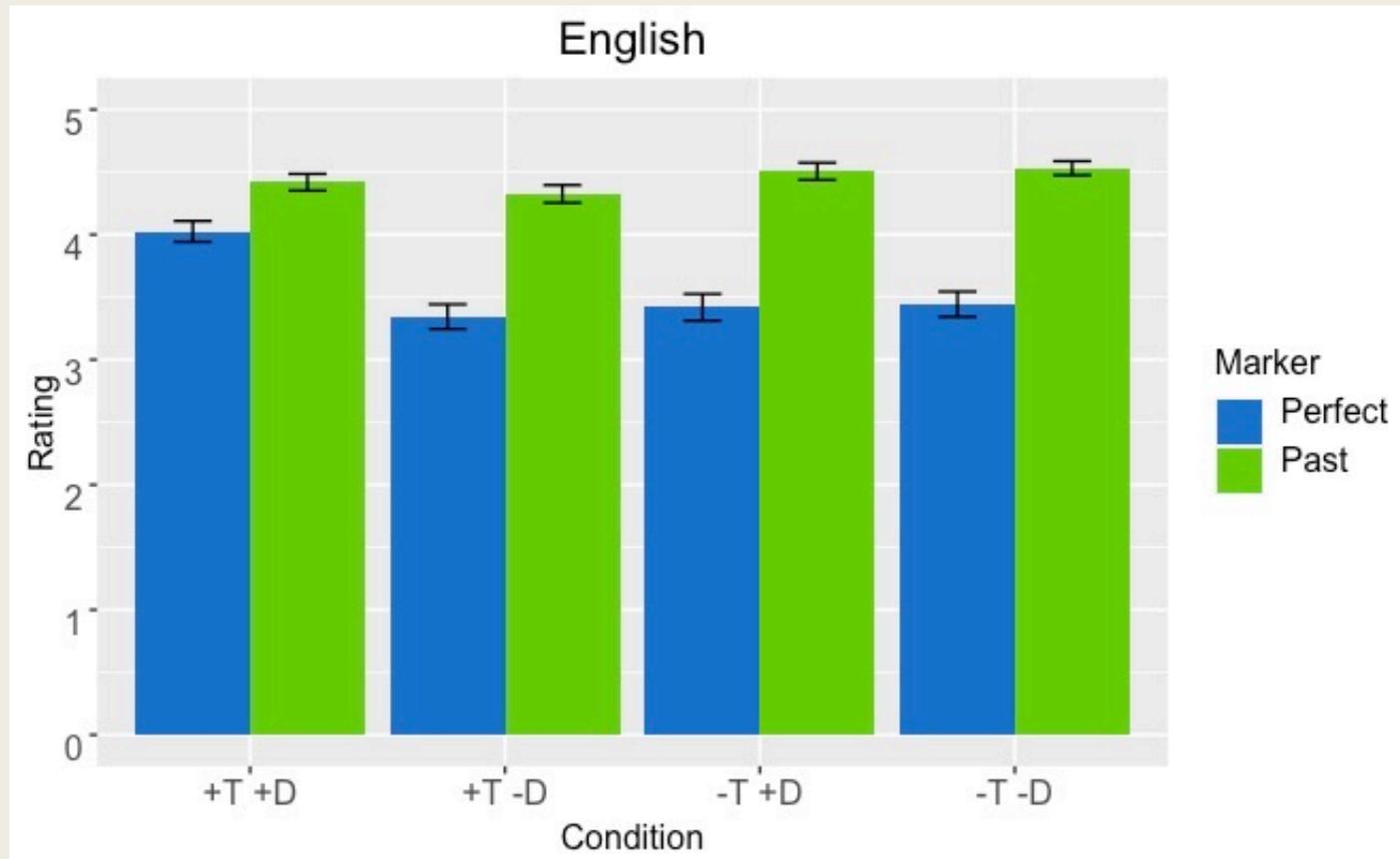
# Spanish

- Significant T\*Marker interaction: PAST is better than PERFECT only in -T ( $\chi^2(1) = 57.07$ ;  $p < .001$ ), but they are not different in +T ( $\chi^2(1) = 0.016$ ;  $p = .90$ ). No effect of deixis.



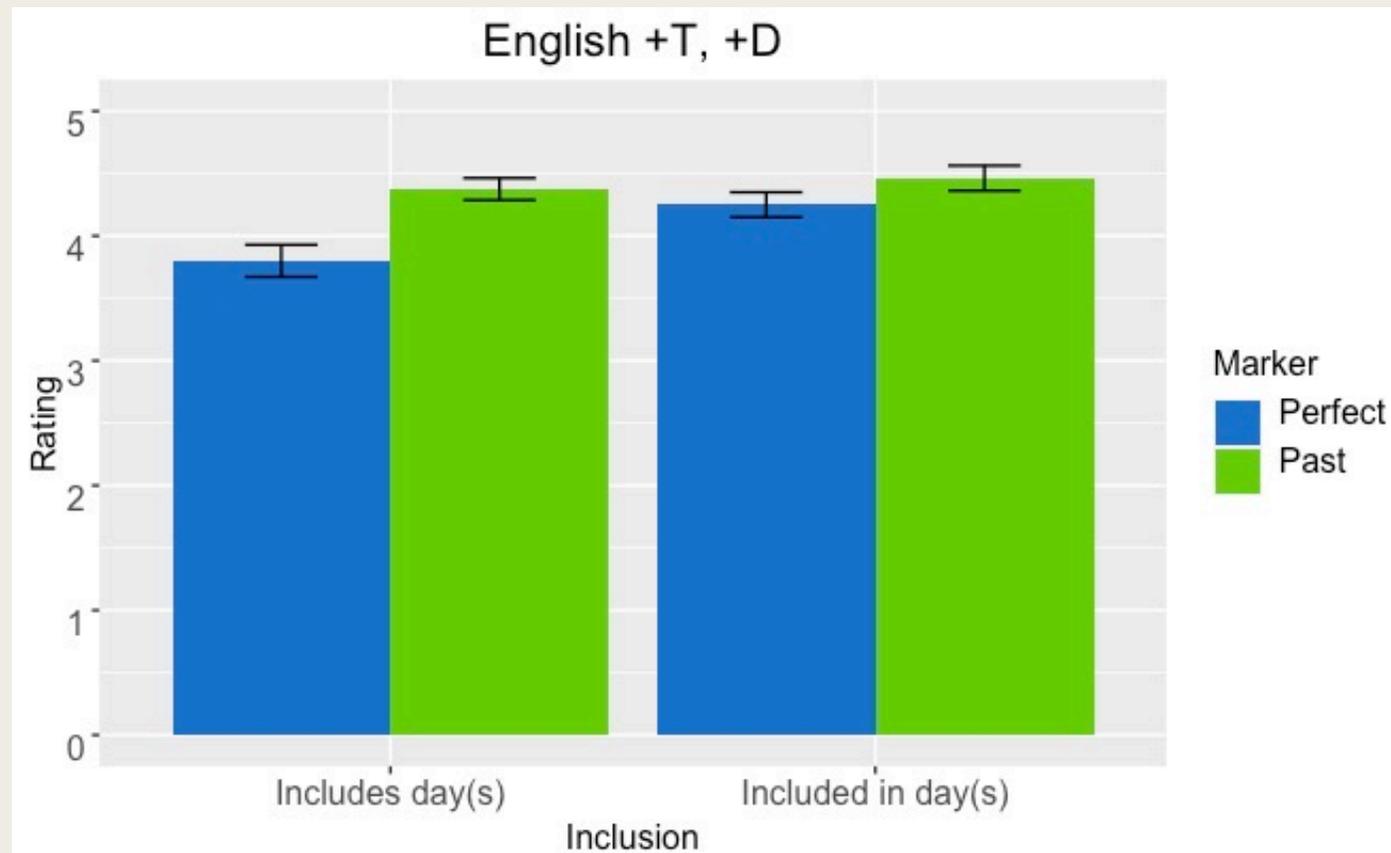
# English

- Significant T\*D\*Marker interaction: PAST is better than PERFECT in all conditions, but there is a less categorical difference in +T, +D. The difference is still significant in a posthoc test ( $\beta = 0.394$ ;  $p = .035$ ).



# English: within +T +D

- if we subdivide +T,+D adverbials when the adverb includes the day (S) (e.g., *this month*) the difference across markers is still significant ( $\chi^2(1) = 6.7711$ ;  $p < .01$ ), but when it is included in the day (S) (e.g., *this morning*), the difference across markers disappears ( $\chi^2(1) = 0.5942$ ;  $p = .4408$ ).



# Summary

- Dutch: **PERFECT** is allowed to refer to past events unconstrainedly. **PAST**, in turn, is dispreferred for this purpose.
- Spanish: **PERFECT** is accepted when the **event is clearly linked to the day of utterance**, but there is no preference for this marker (contra Schwenter 1994, a.o.), since the **PAST** can also be used.
- English: **PAST** is preferred across conditions over the **PERFECT** but this marker is more accepted with deictic hodiernal adverbials, especially when the adverb is **included in the day (S)** (e.g., *this morning*).

# Conclusion

- Both hodiernality and deixis (and considerations about ‘proper’ hodiernality vs. ‘extended’ hodiernality) will have to be included in a crosslinguistic semantic analysis of the PERFECT...
- We are working on it!



GENERAL  
CONCLUSIONS

# Take-home message

- **Translation Mining** is a new, data-driven methodology that is able to pick up nuanced constraints at play in crosslinguistic variation.
- **Experimental work** is also necessary to check the reliability of parallel-corpora generalizations, but also to refine them.
- The constraints at play in selecting each Perfect/Past form are diverse across different languages.
- **Addressing crosslinguistic variation from a variety of data sources is crucial for advancing semantic generalizations about tense-aspect categories.**

Time in Translation project: <http://time-in-translation.hum.uu.nl/>

# Selected references

- Condoravdi, C. y Deo, A. 2015. Aspect Shifts in Indo-Aryan and Trajectories of Semantic Change. En C. Gianollo, A. Jäger y D. Penka (Eds.), *Language Change at the Syntax-Semantics Interface* (pp. 261-292). Nueva York: De Gruyter Mouton.
- Iatridou, S., Anagnostopoulou, E., y Izvorski, R. 2003. Observations about the Form and Meaning of the Perfect. En A. Alexiadou, M. Rathert y A. von Stechow (Eds.), *Perfect Explorations* (pp. 153-202). Nueva York: De Gruyter Mouton.
- Klein, W. 1992. The Present Perfect puzzle, *Language* 68, 525-552.
- Le Bruyn, B.; M. van der Klis & H. de Swart. 2019. The Perfect in dialogue. *Linguistics in the Netherlands* 36(1), 162-175.
- Löbner, S. 2002. Is the German Perfekt a Perfect Perfect? In I. Kaufmann & B. Stiebels (eds.), *More than Words: A Festschrift for Dieter Wunderlich*, 369–391. Berlin: Akademie Verlag.
- Michaelis, L. 1994. The ambiguity of the English Present Perfect, *J. of Linguistics* 30, 111-157.
- Portner, P. 2003. The (temporal) semantics and (modal) pragmatics of the perfect. *Linguistics and Philosophy*, 26, 459–510.
- Schaden, G. 2009. Present Perfects compete. *Linguistics and Philosophy* 32(2), 115-141.
- Schwenter, S. 1994. The grammaticalization of an anterior in progress: evidence from a peninsular Spanish dialect. *Studies in Language* 18: 71–111.
- de Swart, H. 2007. A cross-linguistic discourse analysis of the perfect. *Journal of Pragmatics* 39, 2273-2307.

Merci de votre attention!  
Danke für Ihre Aufmerksamkeit!  
Bedankt voor jullie aandacht!  
Gracias por su atención!  
Thank you for your attention!