

Time in Translation

Henriëtte de Swart
Utrecht University

Martijn van der Klis

Computational linguist, PhD
in NWO-funded TinT.



Bert Le Bruyn

Semantics/corpus linguistics/ L2
acquisition, co-PI NWO-funded TinT.



Time and Tense

- Human life is marked by time passing by.
- Language has ways to express temporal reference ~ window on human cognition.
- **Lexical means**, words that:
 - refer to a time span (*yesterday*)
 - refer to an event that took place in time (*World War II*),
 - connect events in time (*before, since*).
- **Tense**: grammatical expression of temporal reference.

Philosophers' interest in time

- Debates on temporal reference, temporal reasoning and relation of time with truth, knowledge go back to ancient philosophers.
- Aristotle: 'There will be a sea battle tomorrow.'
- Tense Logics: Prior (1957).
- <https://plato.stanford.edu/entries/logic-temporal/>

Linguists' interest in time and tense

- Grammar : present/past/future **tense** (**deictic**).
- Reichenbach (1947): speech time (S), event time (E), reference time (R).
- Interaction with **aspect** (**non-deictic**): 'shape' of situation referred to.
 - Lexical aspect (state/event verbs)
 - Grammatical aspect(perfective/imperfective, Progressive).

Computational linguists' interest in time

- Tenses are anaphoric (~ pronouns): temporal structure supports discourse coherence.
- Large datasets: information extraction, text comprehension, summarizing.
- E.g. European Newsreader project: what happened to whom, when and where?

Cross-linguistic variation

- Comrie (1985): all cultures have a conceptualization of time, not all languages have tense.
- Wo shuaiduan-**le** tui [Mandarin Chinese]
I break.LE leg
'I broke my leg.' (still in plaster)
- Wo shuaiduan-**guo** tui
I break.GUO leg
'I broke my leg.' (once, fully healed now)



Cross-linguistic variation

- **Philosophical** questions: conceptualization of time and events, human cognition.
- **Linguistic** questions: different grammars, language change.
- **Computational** questions: language-dependent applications, machine translation, L2 tools.



Machine Translation of time: how hard can it be?

- ➡ Loáciga (2017): translation of the English *Simple Past*/*Present Perfect* into French.
- ➡ Eurparl dataset of >300.000 English verb phrases.
- ➡ About 50% of English *Simple Past* translated into French *Passé Composé*. Other verb forms found are: *Imparfait*, *Passé Simple*, *Présent*.
- ➡ About 60% of English *Present Perfect* translated into French *Passé Composé*. Other tense forms found are: *Présent*, *Subjonctif*.

Significant translation ambiguity

- Purely statistical machine translation models insufficient: default strategies do not take into account context/meaning.
- Native speakers perceive 'wrong' tense choice in context as grammatical error.
- Improve MT output by using annotation, e.g. **narrativity** (Meyer, Grisot, and Popescu-Belis 2013), **boundedness** (Loáciga 2017).



New linguistic questions..

- ▶ Can we predict which features will be most successful (lexical/sentence-level/discourse-level)?
- ▶ Are the same annotation features relevant in all translation pairs? (cross-linguistic variation!)



Usage-based approach to linguistics

- Can we use large data-sets to improve linguistic theory?
- That would place us in a better position to answer questions about temporal reference.
- Support for foundational research as well as applications.

Time in Translation: *have* PERFECT

- Morpho-syntax: auxiliary (*have/be*) + participle.
- Core of perfect meaning: past event + present state.
- 1) *Mary has visited Paris.* [experiential perfect]
(her past visit to Paris is relevant now)
- 2) *Mary has moved to Paris.* [resultative perfect]
(she currently lives in Paris)
- 3) *Mary has been living in Paris since 2010.*
(she currently lives in Paris) [continuative perfect]
- 4) *Malcolm X has just been assassinated.*
[‘hot news’ perfect]

Micro-variation: PERFECT, PRESENT

- Linguistic theory claims: continuative PERFECT meaning conveyed by PRESENT in Dutch, French.

a. Rien ne bouge. [PRESENT] [French]

b. Nothing has moved. [PERFECT] [English]

c. Niets beweegt. [PRESENT] [Dutch]

(No et moi, Delphine de Vigan, 2007)

- Manual research on literary translations confirms that not all languages have a continuative PERFECT.

Micro-variation: PERFECT, PAST, PRESENT

- Multilingual conversation data: subtitles of tv series.
 - a. In case you hadn't noticed, we just got a confession. [Eng]
 - b. Falls es ihnen entging, er hat gestanden. [German]
 - c. Si vous ne l'avez pas remarqué, on a des aveux. [French]
- Result meaning not always conveyed by PERFECT: PAST tense implies result, PRESENT tense implies preceding event.
- Manual research confirms that PERFECT use in Dutch, German, French is neither a subset, nor a superset of PERFECT use in English. **Many-to-many mapping.**

The PERFECT – the TinT approach

- **Semantic maps** (Haspelmath 1997)
 - Enables us to see variation within and between languages.
 - translation corpora: same meaning, different forms
 - Language-specific distribution of meanings over forms.
- **No preconceptualized meanings**, generate maps directly from the data.
- **Goal**: lexical+compositional+dynamic semantics of the PERFECT across languages.

PERFECT extraction

- Extraction of PERFECTS from multilingual corpora (Europarl, Subtitle corpus, literary translations)
- Word-level alignment of verb phrases in 5 languages (English, Dutch, German, French, Spanish).
- Tense attribution for translations
 - Discard 'other' translations (nominalizations, 'free' translations)
 - Manual/automatic tagging (language-specific)

TimeAlign

- Web application
- Manual alignment of translations of PERFECTS.

Annotation

Dutch (original)

We zijn allemaal blij dat de rechter hem heeft vraggesproken 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

✓ Submit

➔ Go to another fragment



Dissimilarity matrix and multidimensional scaling

- Create a **dissimilarity matrix**. Input to
- **Multidimensional scaling**: creates a low-dimensional representation of the data, while respecting distances in the original high-dimensional space.
- Output is **temporal map**.
- Methodology inspired by Wälchli & Cysouw (2012) (lexical variation) and data-driven dialectology (phonetic variation, Nerbonne 2009).

Distance function: a pair of source and translations is 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

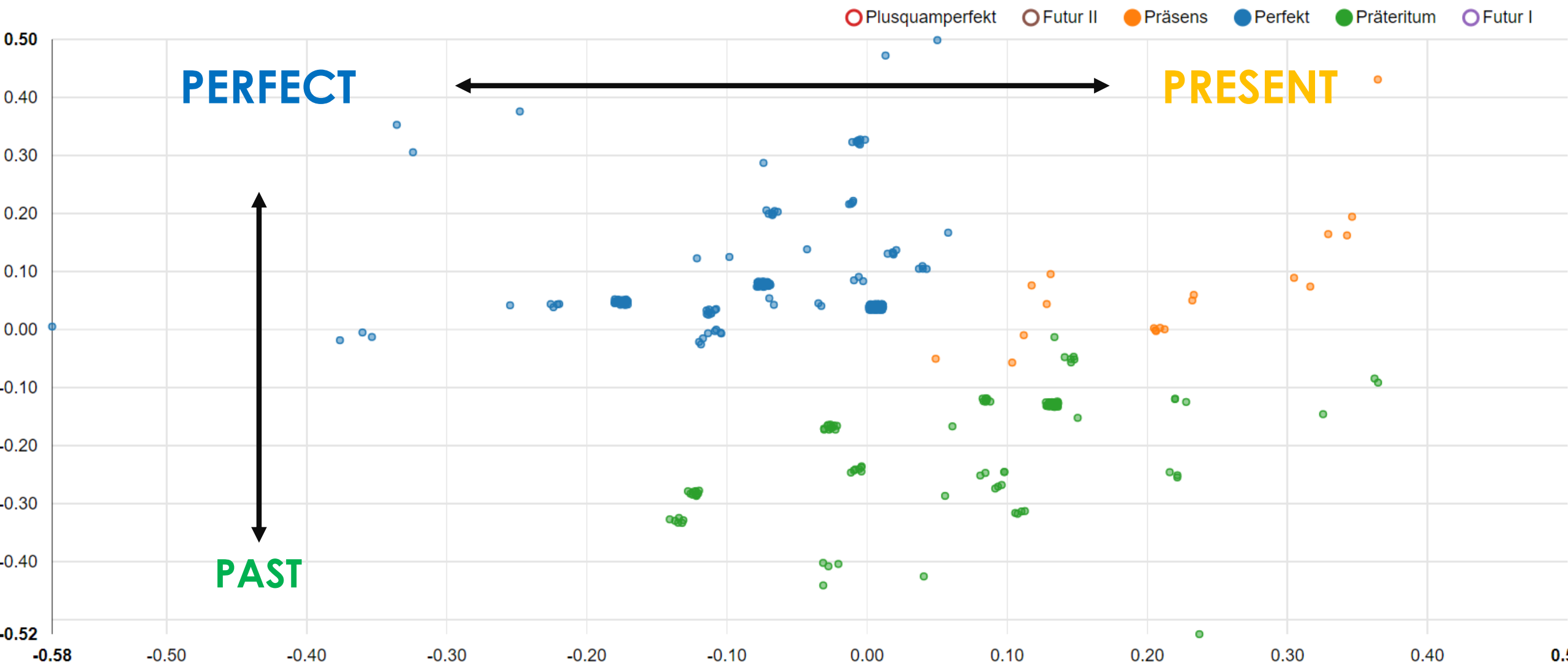
➡ $d(1, 2) = 2/5$, $d(1, 3) = 2/5$, $d(2, 3) = 4/5$

Classification of tense forms into PERFECT, PRESENT, PAST

	DE	EN	ES	FR	NL
PERFECT	360	347	371	481	438
PRESENT	19	18	47	20	20
PAST	124	146	89	8	36
PLUPERFECT	4	1	3	2	18
other	5	-	2	1	

Distribution of tense forms across languages in Europarl dataset

PERFECT map: German



Filters

Time in Translation



The project ▾

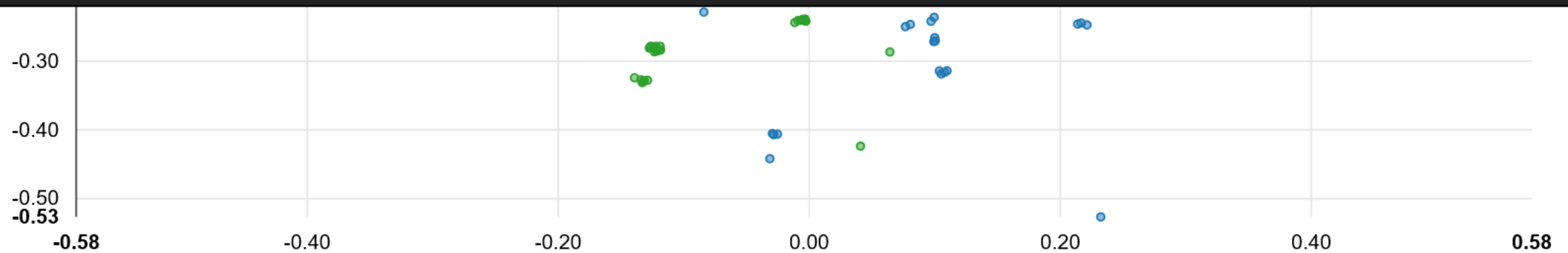
Student Research

Translation Mining ▾

Contact

Signed in as *henriette*

Log out



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.

Language:

German

English

Spanish

French

Dutch

Dimension on x-axis:

1

2

3

4

5

Dimension on y-axis:

1

2

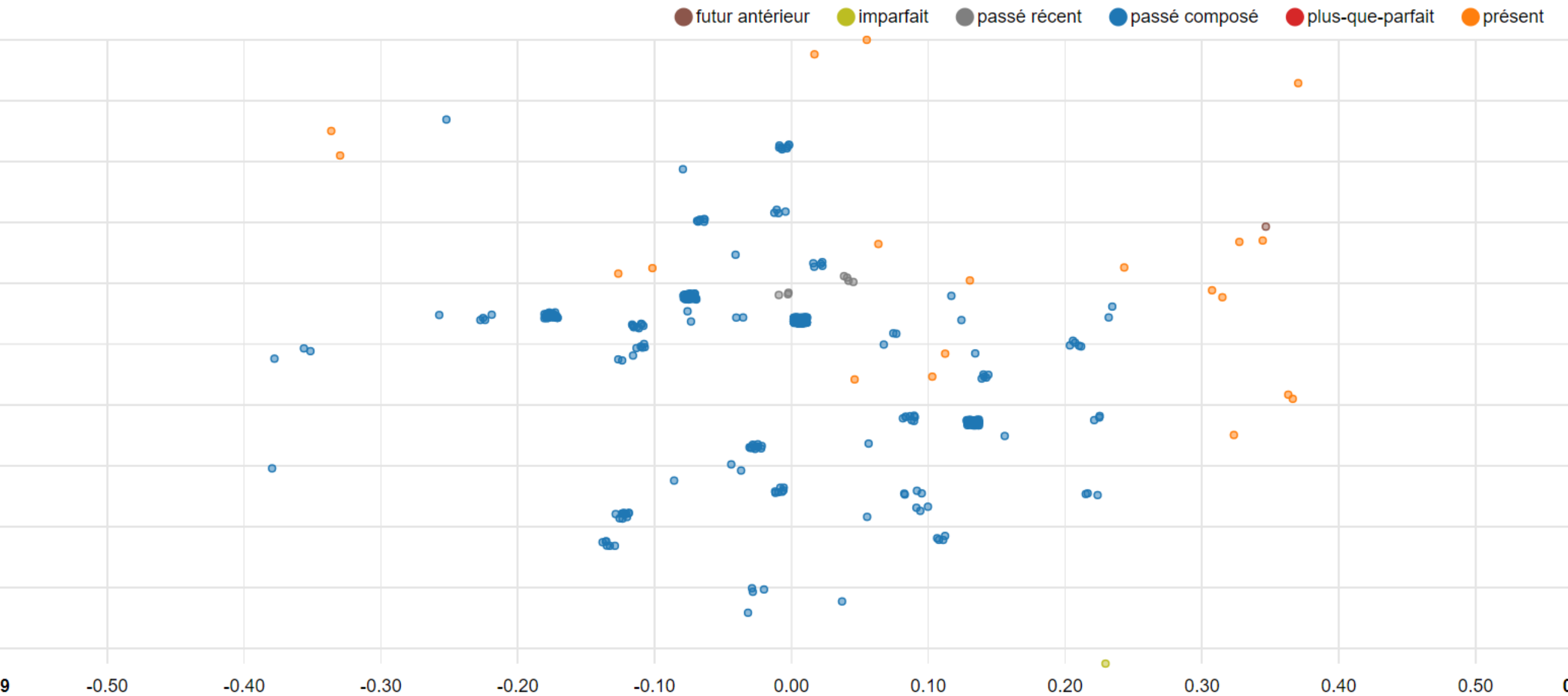
3

4

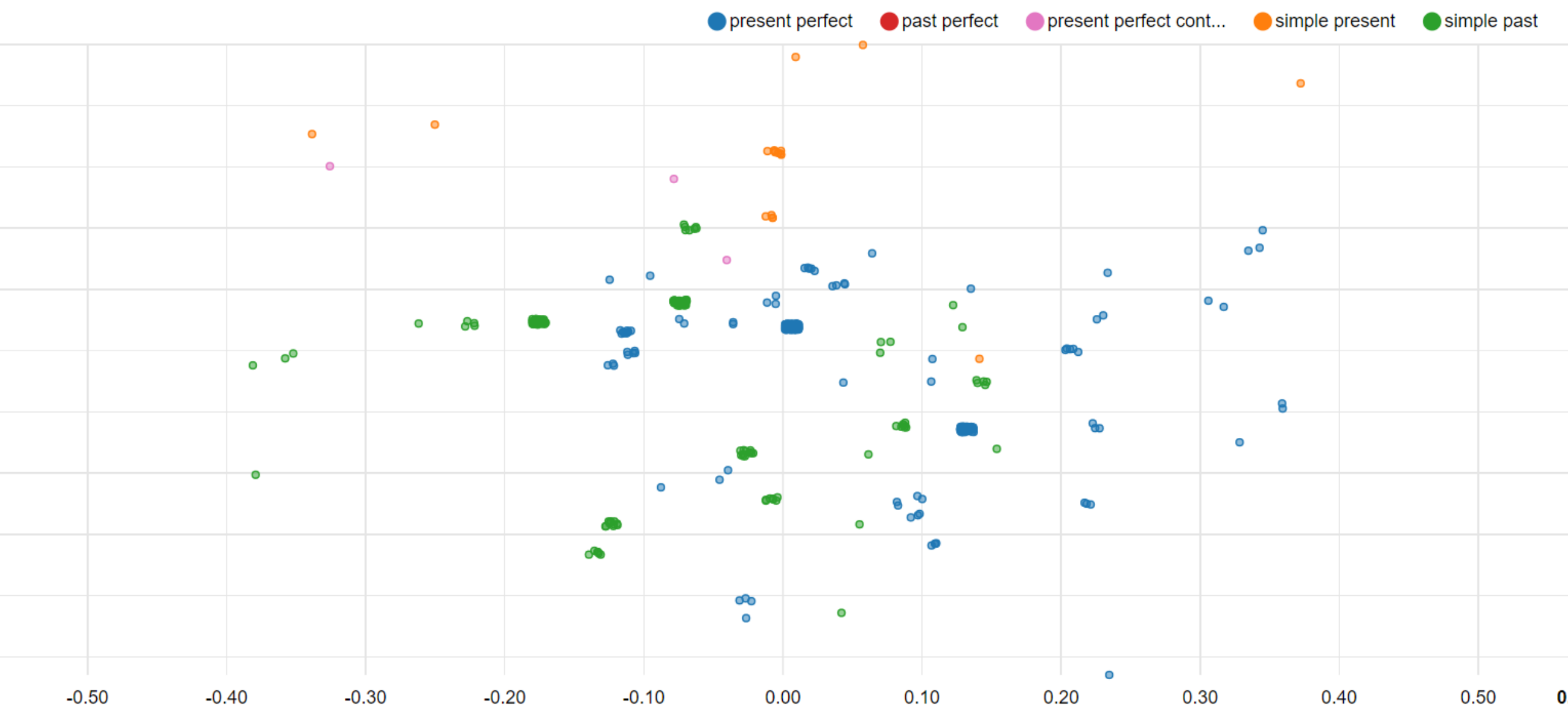
5

✓ Go!

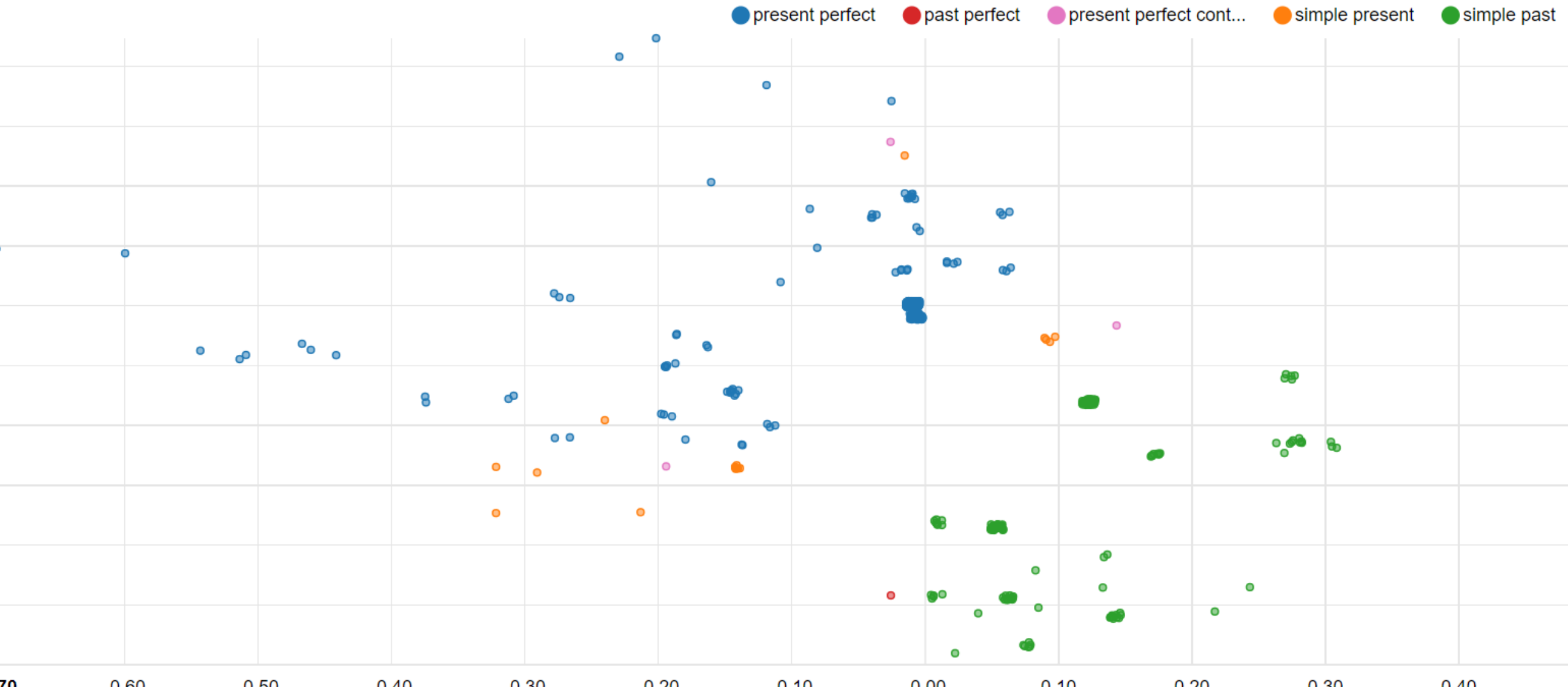
PERFECT map: French



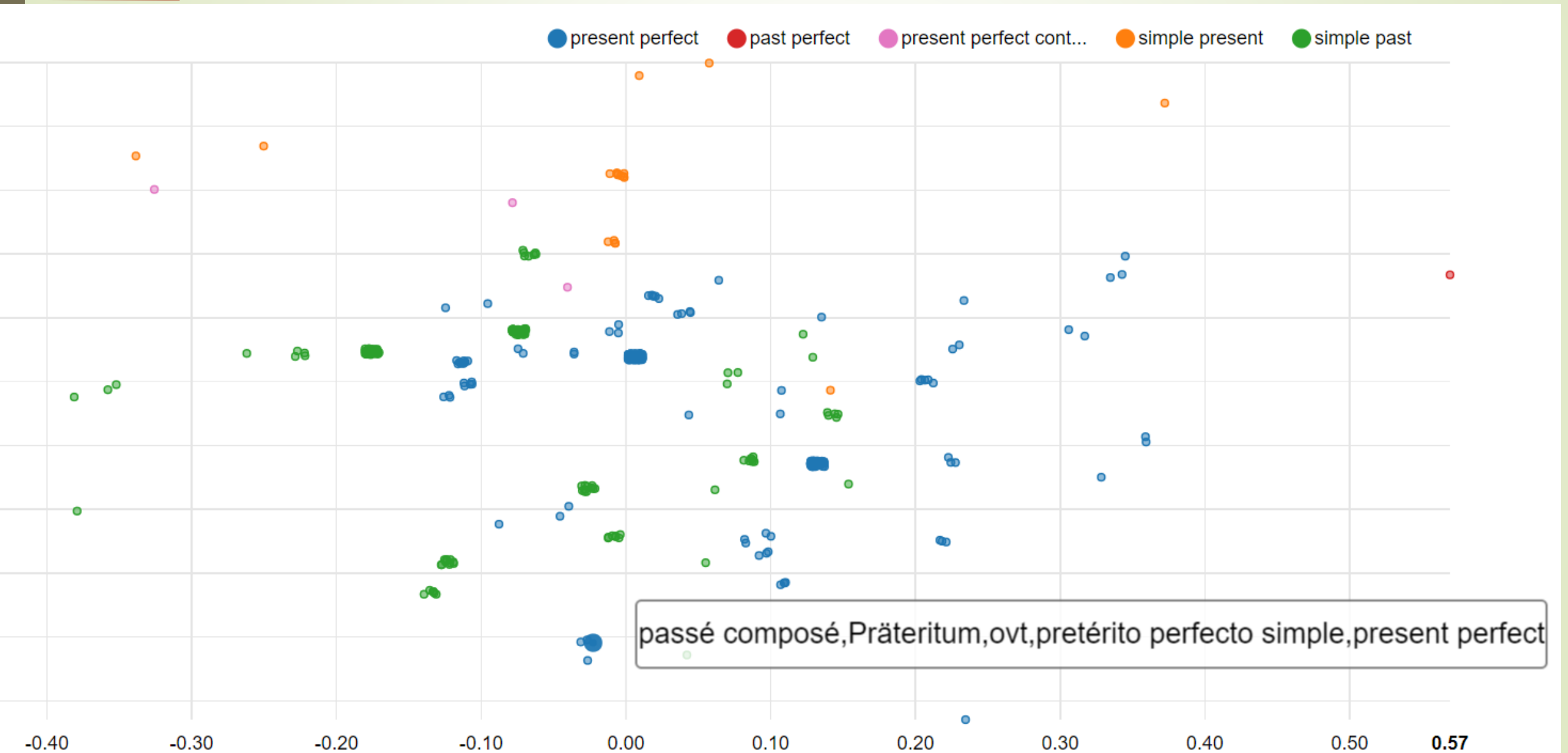
PERFECT map: English (plot 1/2)



PERFECT map: English (plot 5/3)



PERFECT map: English



Europarl 5-tuple: PERFECT extracted from English, matched with translations in context

English

Perfect

ep-00-12-14.xml - 11952

In the former Yugoslavia and Timor , the UNHCR **has decided** to protect all displaced persons and provide humanitarian aid for them , irrespective of whether or not they are refugees .

Translations

German

Präteritum

Im ehemaligen Jugoslawien und in Timor **beschloss** das UNHCR , allen Entwurzelten , unabhängig davon , ob es sich um Flüchtlinge handelt oder nicht , Schutz und humanitäre Hilfe zu gewähren .

Spanish

pretérito perfecto simple

En la antigua Yugoslavia y en Timor , el ACR **decidió** proteger y proporcionar ayuda humanitaria a todas las personas desterradas , trátase o no de refugiados .

French

passé composé

En ex-Yougoslavie et au Timor , le HCR **a décidé** de protéger et de fournir une aide humanitaire à toutes les personnes déracinées , qu' elles soient ou non des réfugiés .

Dutch

ovt

In het voormalige Joegoslavië en Timor **besloot** het UNHCR bescherming en humanitaire hulp te bieden aan alle mensen die uit hun huizen verdreven waren , of dat nu vluchtelingen waren of niet .



Temporal maps

- We visualize the results of MDS on a **temporal map**.
- We can choose which language to show.
- We can click/unclick tenses to filter them.
- We can hover over the map to visualize the tuples.
- We can click on a tuple to recover the underlying Europarl datapoint.
- We use the tenses per language as labels; color coding helps to see patterns.
- Dimensions reveal competition between forms, cross-linguistic variation.

Using the maps to test linguistic hypotheses

- We can test hypotheses from the literature:
 - French language change – replacement of *Passé Simple* by *Passé Composé*).
 - Confirm that English/Spanish requires a PAST with a locating time adverbial, whereas German, Dutch, French tolerate a PERFECT in this configuration.

Time in Translation

Time in Translation

×

+

←

→

↺

timealign.pythonanywhere.com/timealign/show/3552

📖

☆

|

☰

🔗

👤

Time in Translation

🏠

The project ▾

Student Research

Translation Mining ▾

Contact

Signed in as henriette

Log out

Source

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 .

Both English and Spanish should allow time adverbials that refer to 'now'; confirmed in Europarl.

Time in Translation [Home](#) [The project](#) [Student Research](#) [Translation Mining](#) [Contact](#) Signed in as *henriette* [Log out](#)

French **ep-00-12-14.xml - 14906**

Les députés sociaux-démocrates danois **ont voté** aujourd'hui contre la proposition visant à porter organisation commune des marchés dans le secteur de la banane .

Translations

German **Perfekt**

Die dänischen Sozialdemokraten **haben** heute gegen den Vorschlag betreffend die gemeinsame Marktorganisation für Bananen **gestimmt** .

English **present perfect**

The Danish Social Democrats **have** today **voted** against the proposal for a common market arrangement for bananas .

Spanish **pretérito perfecto compuesto**

.
(DA) Los socialdemócratas daneses **hemos votado** hoy en contra de la propuesta relativa a la organización común de mercados en el sector del plátano .

Dutch **ott**

De Deense sociaal-democraten **stemmen** tegen het voorstel betreffende een gemeenschappelijke marktordening voor bananen .

Subtle differences in what counts as 'now'

The screenshot shows a web browser window with the URL `http://timealign.pythonanywhere.com/timealign/st`. The browser's address bar and tabs are visible at the top. Below the browser window is a dark navigation bar with the following items: "Time in Translation" (with a home icon), "The project" (with a dropdown arrow), "Student Research", "Translation Mining" (with a dropdown arrow), "Contact", "Signed in as *henriette*", and "Log out".

The main content area is titled "German" on the left and "ep-00-12-14.xml - 8260" on the right. The German source text is: "Letzte Nacht habe ich davon geträumt, ein Rentner zu sein, denn obwohl ich die italienische Rentnerpartei vertrete, beziehe ich keine Rente." The words "Letzte Nacht" and "habe" are circled, and "geträumt" is highlighted in green.

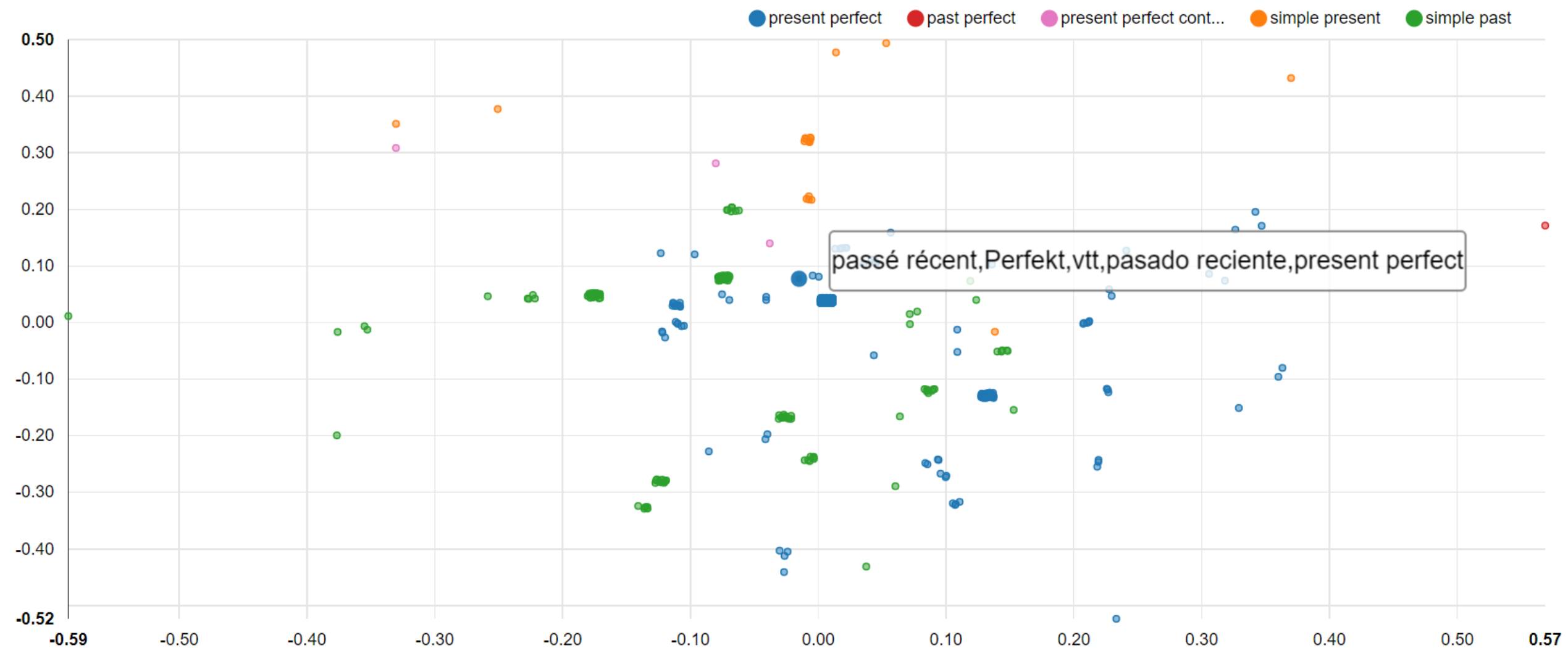
Below the German text is a section titled "Translations". It contains four language-specific translation blocks:

- English** (simple past):
I must tell you that last night I dreamt I was a pensioner .
I do not actually receive a pension although I am the representative of the Pensioners ' Party in Italy .
- Spanish** (pretérito perfecto compuesto):
Debo decir que esta noche he soñado con que era un pensionista .
En efecto , aunque soy el representante del Partido de los Pensionistas en Italia , no disfruto de pensión alguna .
- French** (passé composé):
Je dois dire que cette nuit, j' ai rêvé que j' étais retraité .
- Dutch** (overtone/overtone):
Ik moet u vertellen dat ik afgelopen nacht droomde dat ik met pensioen was

Perfect meanings

- ➡ 'Hot news' PERFECT in English (McCawley 1971): reports a noteworthy event, something that just happened and the listener has no knowledge of yet.
- ➡ Under debate in the literature whether this qualifies as a separate meaning of the PERFECT, or whether this is a recent PAST.
- ➡ Use the Europarl dataset to answer the question: is 'hot news' a dimension of the PERFECT or of the PAST?

Explore the range of 'other' verb forms besides PAST/PRESENT/PERFECT



Hot news PERFECT + 'just' in English/German/Dutch; recent past in French/Spanish

English

Perfect

ep-00-12-11.xml - 1851

The fact is that the Court of Justice **has** just **repealed** the 1998 Directive banning advertising and sponsorship of tobacco products .

Translations

German

Perfekt

Der Gerichtshof **hat** nämlich gerade die Richtlinie aus dem Jahr 1998 , die Werbung und Sponsoring für Tabakerzeugnisse verbietet , **aufgehoben** .

Spanish

pasado reciente

El Tribunal de Justicia , efectivamente , **acaba de anular** la directiva de 1998 que prohibía la publicidad y el patrocinio de los productos del tabaco .

French

passé récent

La Cour de justice , en effet , **vient d' annuler** la directive de 1998 interdisant la publicité et le parrainage en faveur des produits du tabac .

Dutch

vtt

Het Hof van Justitie **heeft** kortgeleden de richtlijn van 1998 betreffende het verbod op reclame en sponsoring in de tabakssector **geannuleerd** .

Limits of the temporal maps

- Not every language shows easily interpretable clusters.
- We built a model of 5 dimensions, but we don't fully understand the role of the dimensions.
- We can test hypotheses about known cross-linguistic variation, but this quickly becomes **qualitative** research on individual examples.

Beyond the original dataset

- **Intermediate conclusion:** the Europarl dataset is too complex: too many meaning distinctions, too many variables, not enough datapoints.
- **Next step:** zoom in on more homogeneous datasets.
- Datasets that focus on one particular competition, e.g.
 - PERFECT vs. PAST in narrative discourse (Camus dataset),
 - PERFECT vs. PRESENT in continuative readings (Europarl dataset of *has been V-ing* and its translations).

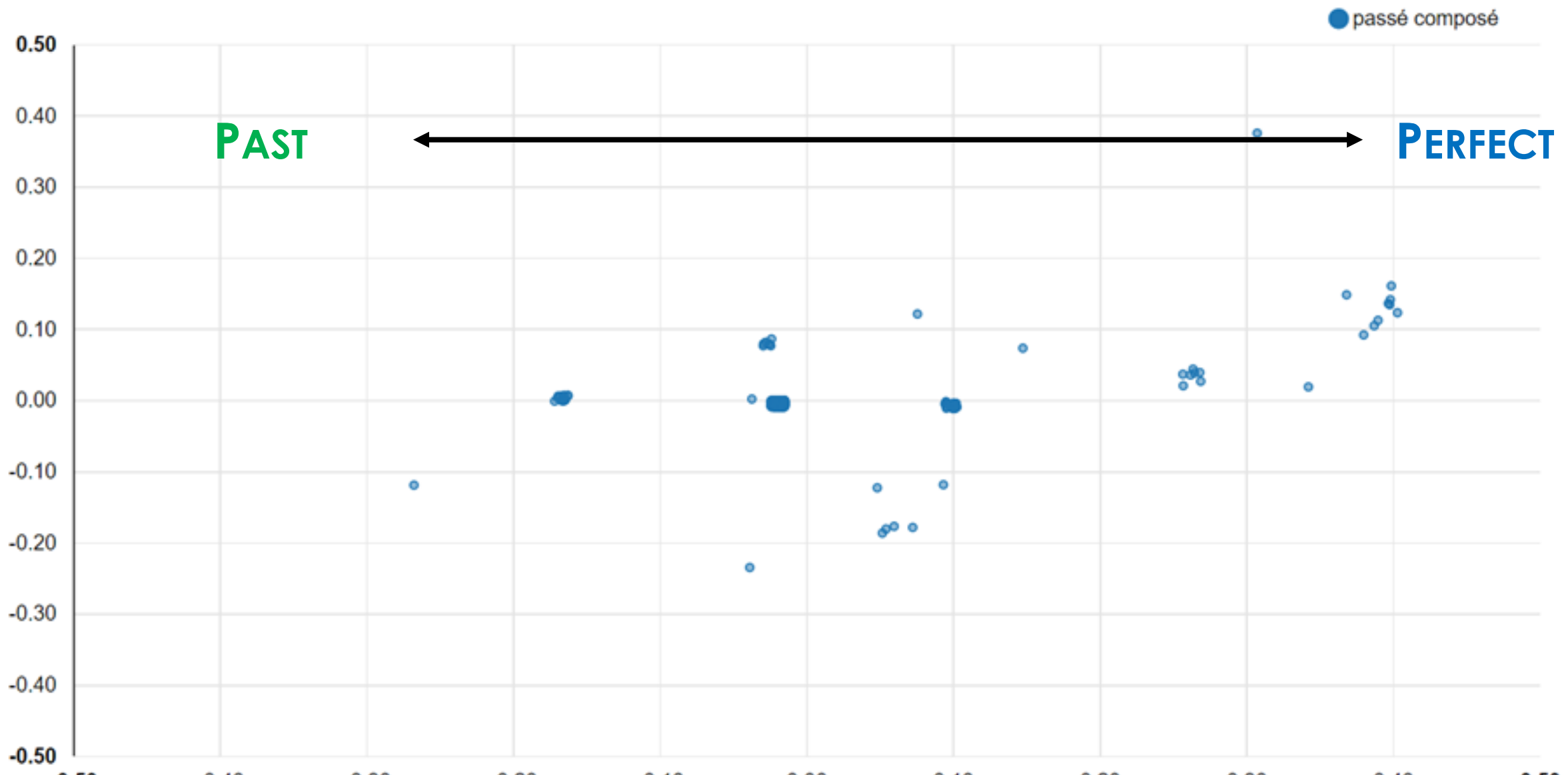
PERFECT vs. PAST in narrative discourse

- Starting point: French is losing perfective simple PAST (*Passé Simple*), and the *Passé Composé* replaces it.
- Lindstedt (2000): French *Passé Composé* is not a PERFECT anymore, but has become a PAST, because it allows narrative use (story telling).
- Löbner (2000): German also allows narrative PERFECT.
De Swart (2007): Dutch does not allow narrative PERFECT.
Schaden (2009): Spanish patterns like English.
Dahl (2015): Swedish patterns like English.

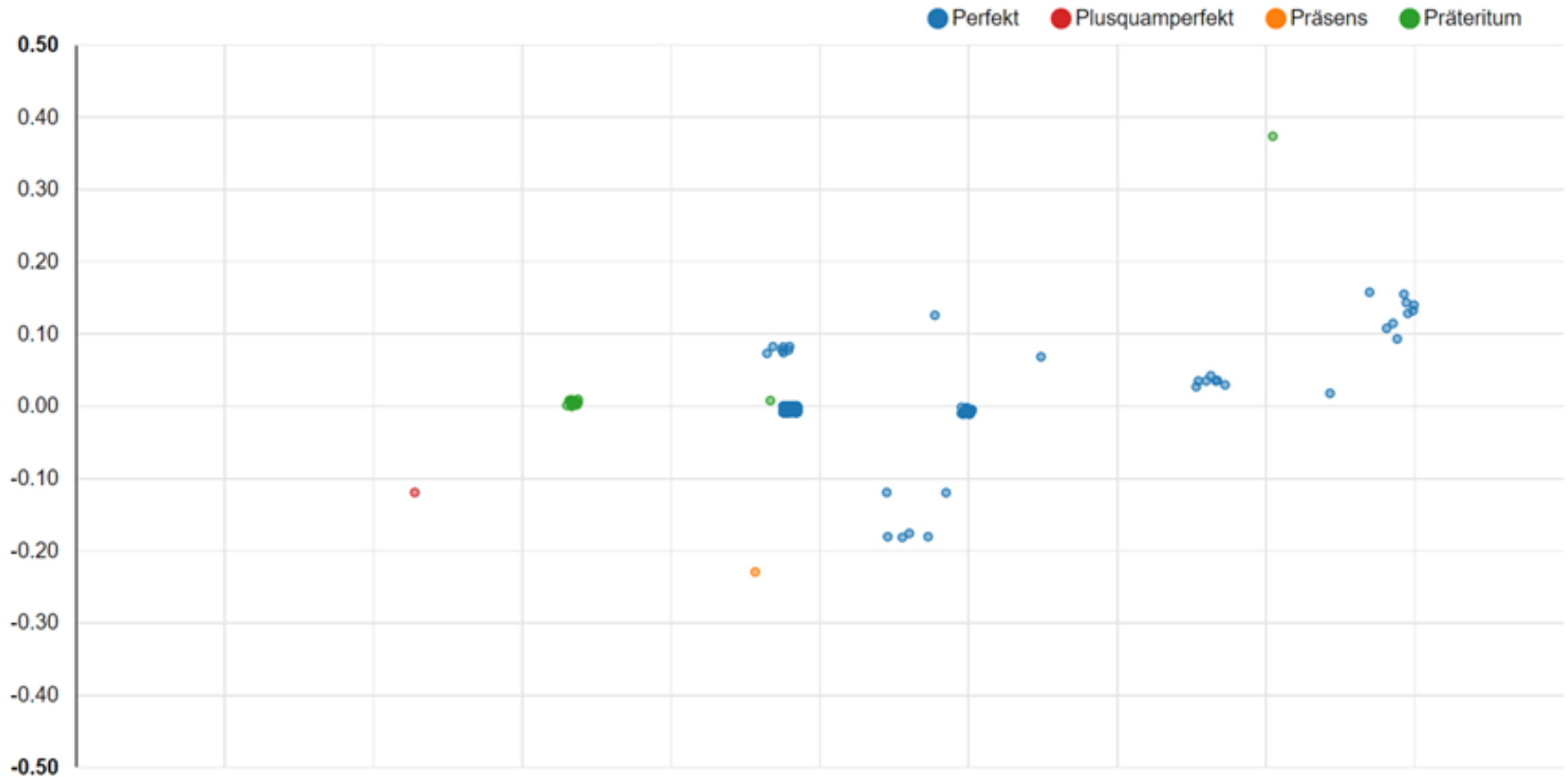
How to translate the narrative *Passé Composé*?

- A. Camus (1942). *L'étranger*: tells the story in an alternation of *Passé Composé* and *Imparfait*.
- Extract all the *Passé Composés* from the first three chapters of *L'étranger*, and match them with their German, Dutch, Spanish, English translations.
- Dataset: tense forms used to move the storyline forward ('and then and then').
- Build a temporal map of the translations of this novel.

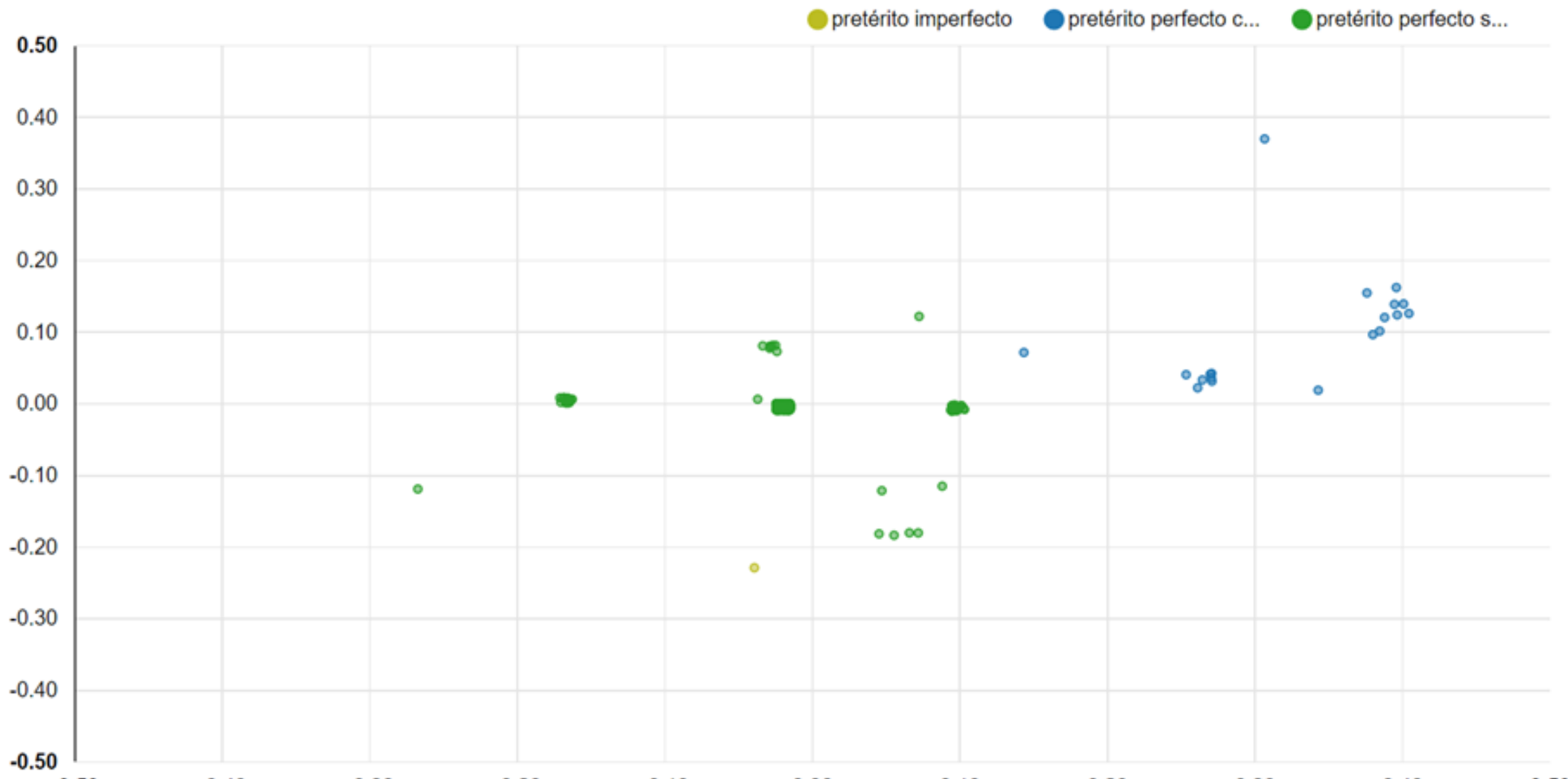
French source: all *Passé Composé*



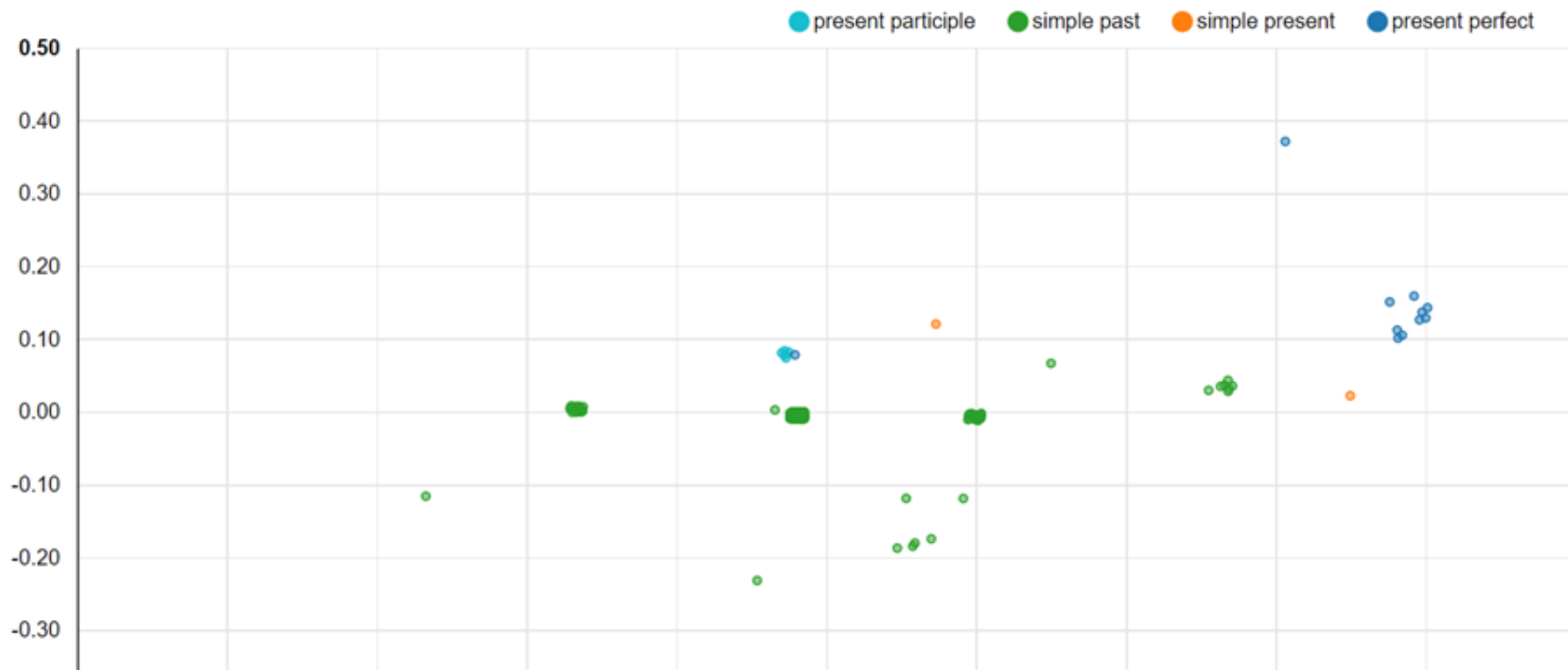
German translation: green cluster to the left



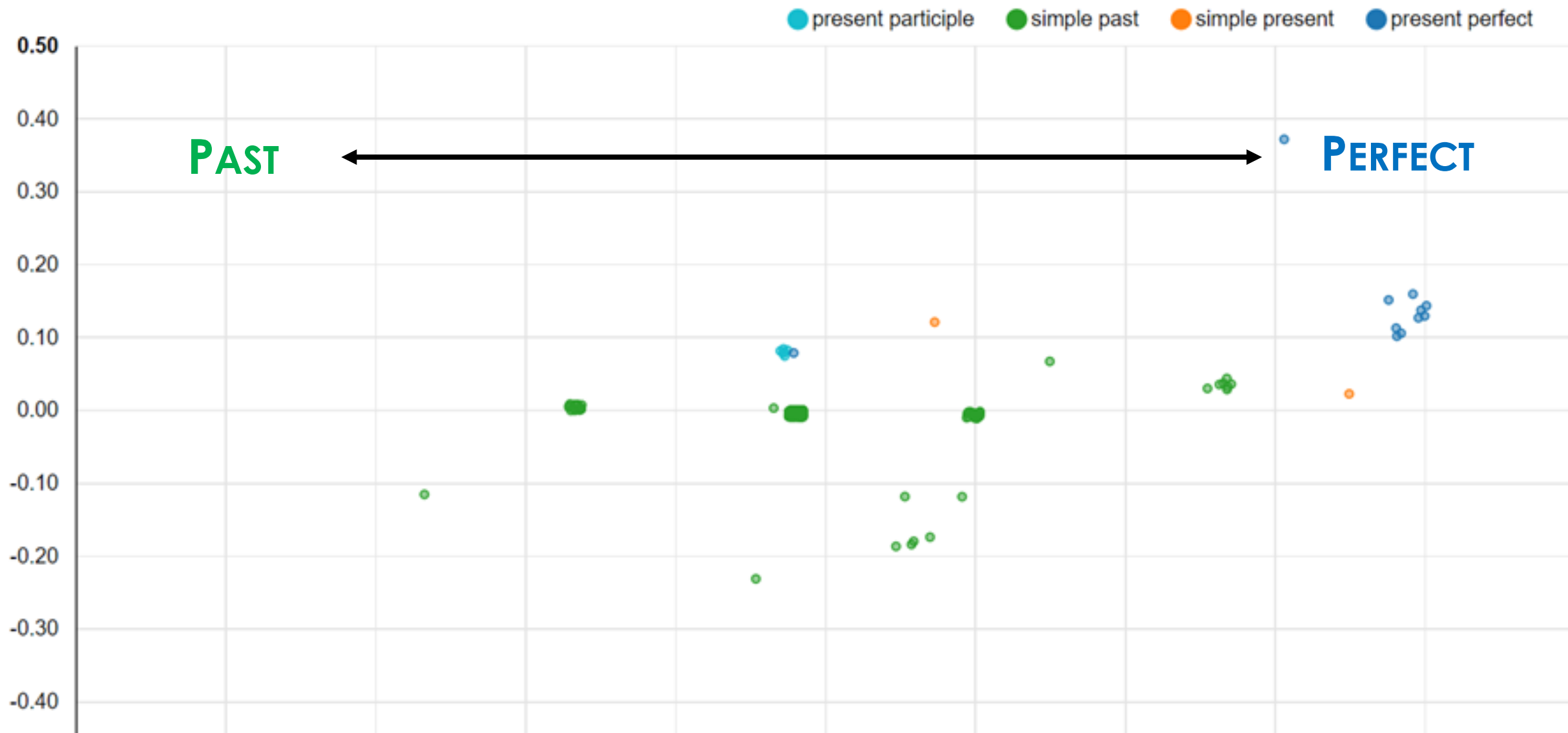
Spanish translation



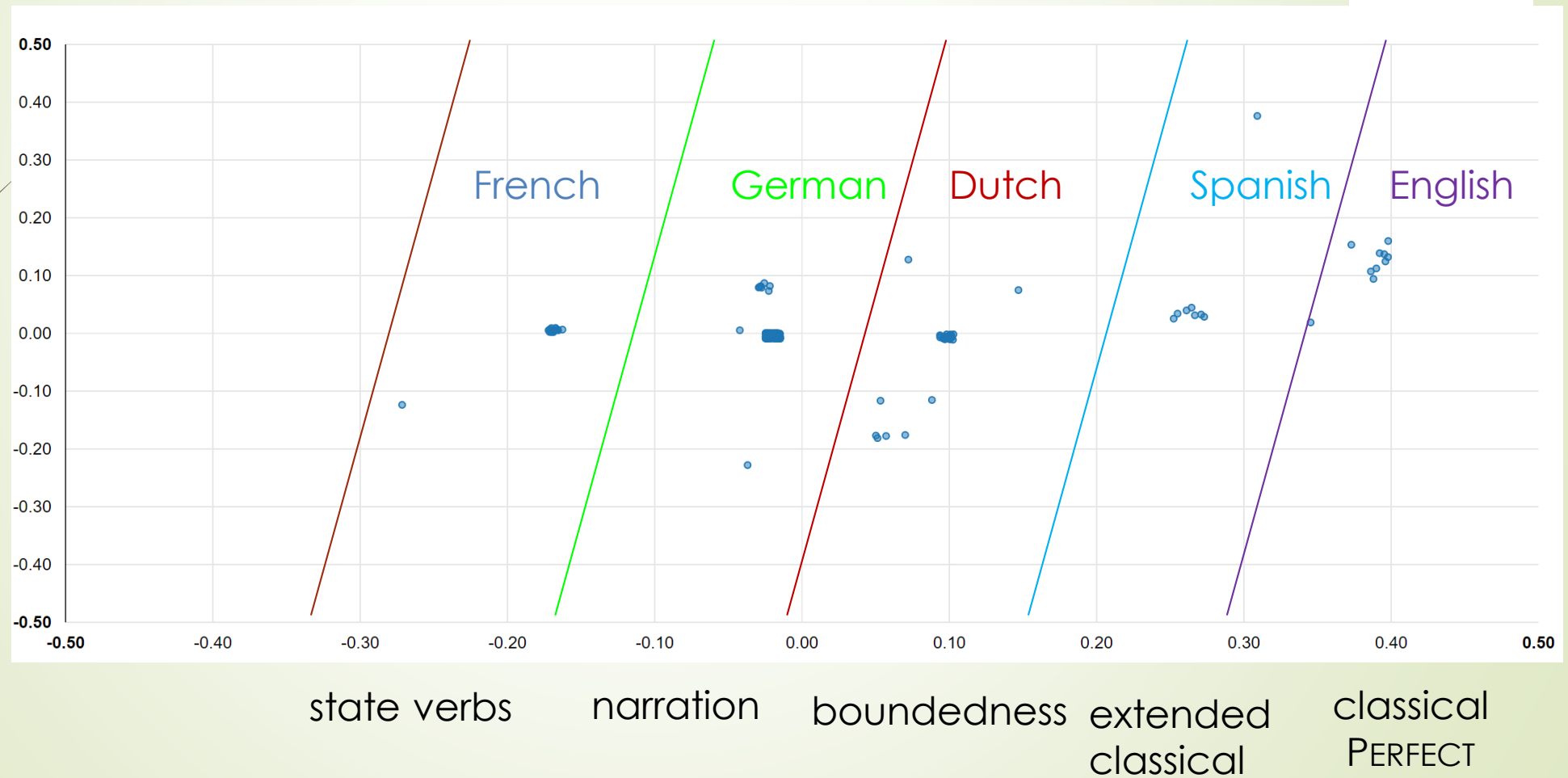
English translation: blue cluster to the right



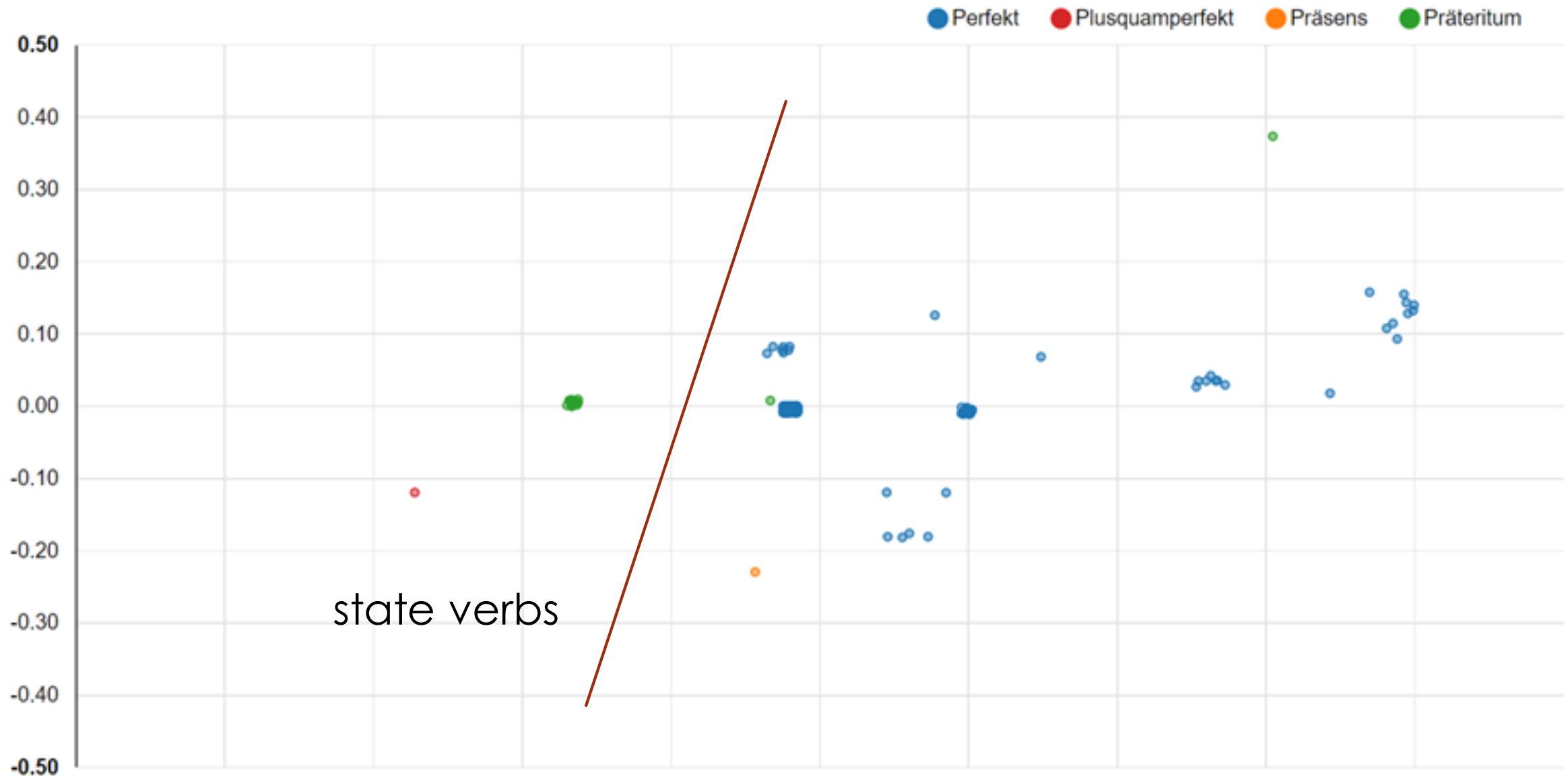
Competition PAST and PERFECT



Drawing the line between PERFECT and PAST



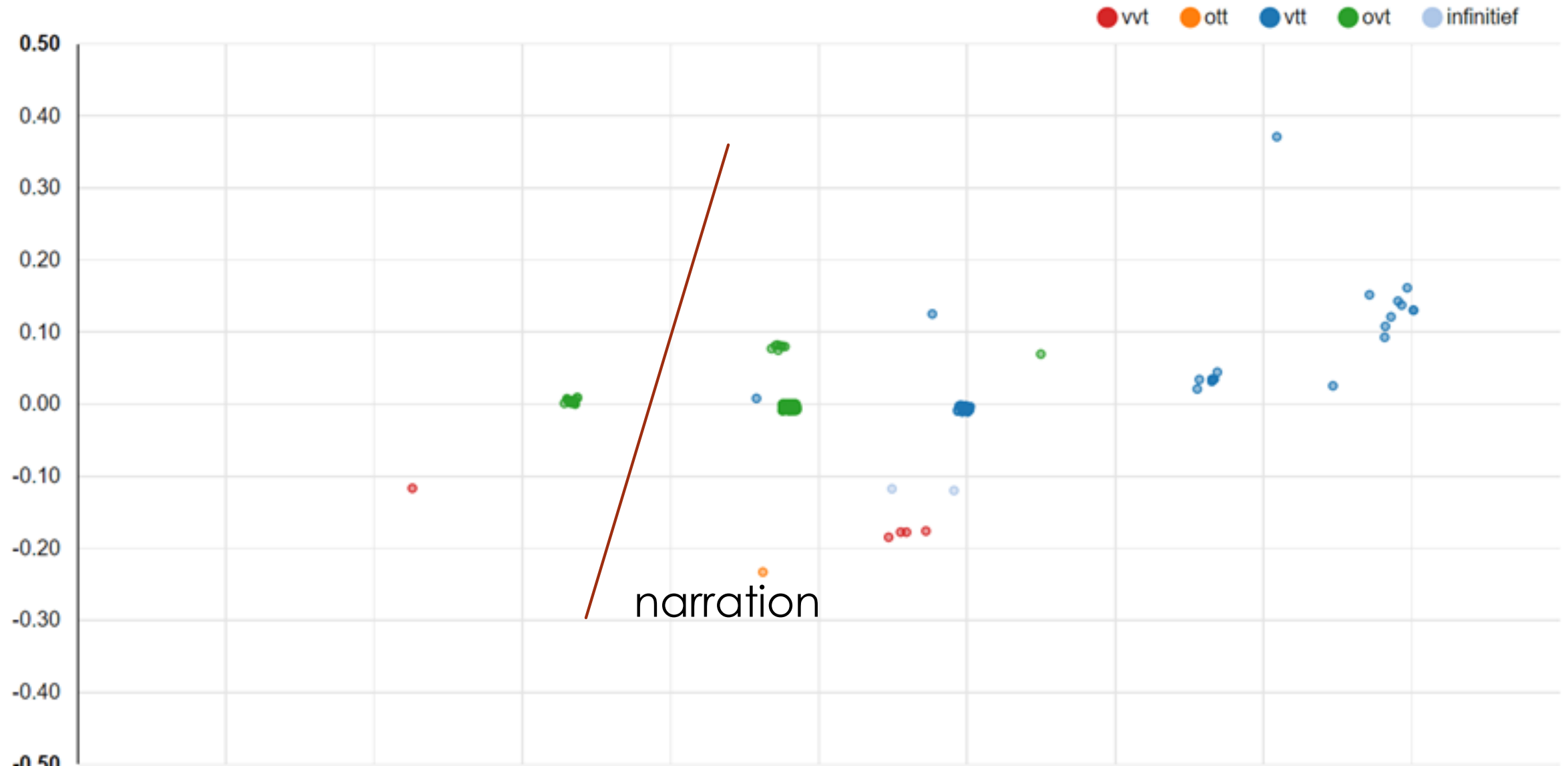
German translation: green cluster to the left



Cognitive state verbs: French/German

- *J'ai trouvé qu'en effet, de cette façon, elle serait punie.* [Fr]
I agreed that that would punish her all right. [En]
Ich fand dass sie auf diese Weise tatsächlich bestraft wäre. [De]
- *Je n'ai pas cru devoir lui refuser l'autorisation.* [Fr]
I didn't think I could refuse him permission. [En]
Ich dachte ich dürfte ihm die Erlaubnis nicht verwehren. [De]

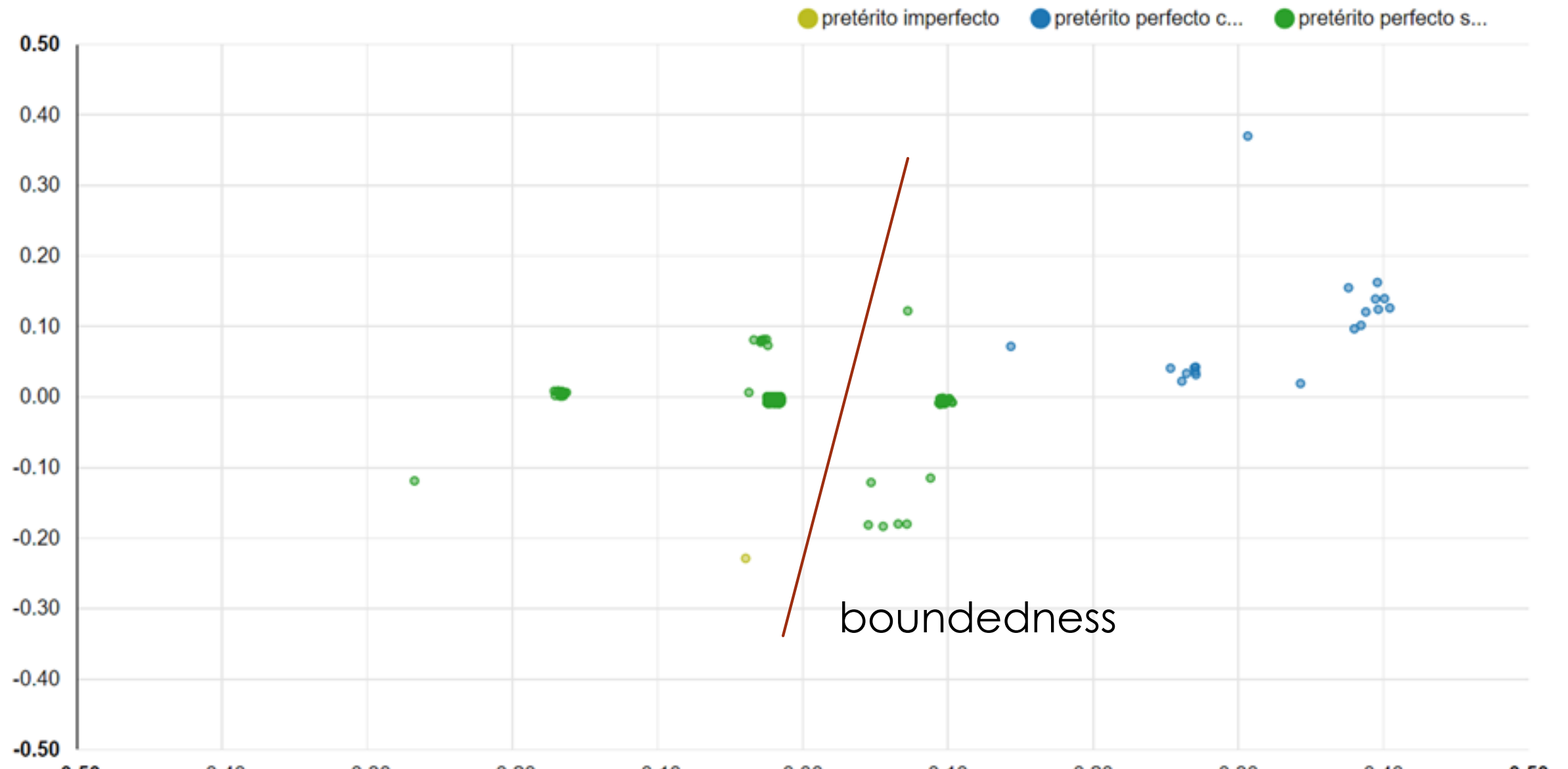
Dutch map



Narration: German/Dutch

- Il a *cacheté* la lettre, et nous *avons fini* le vin. [Fr]
- He *put* the letter in the envelope and we *finished* off the wine. [En]
- Er *hat* den Brief *zugemacht*, und wir *haben* den Wein ausgetrunken. [De]
- Hij *plakte* de brief dicht en wij *dronken* de rest van de wijn op. [NI]

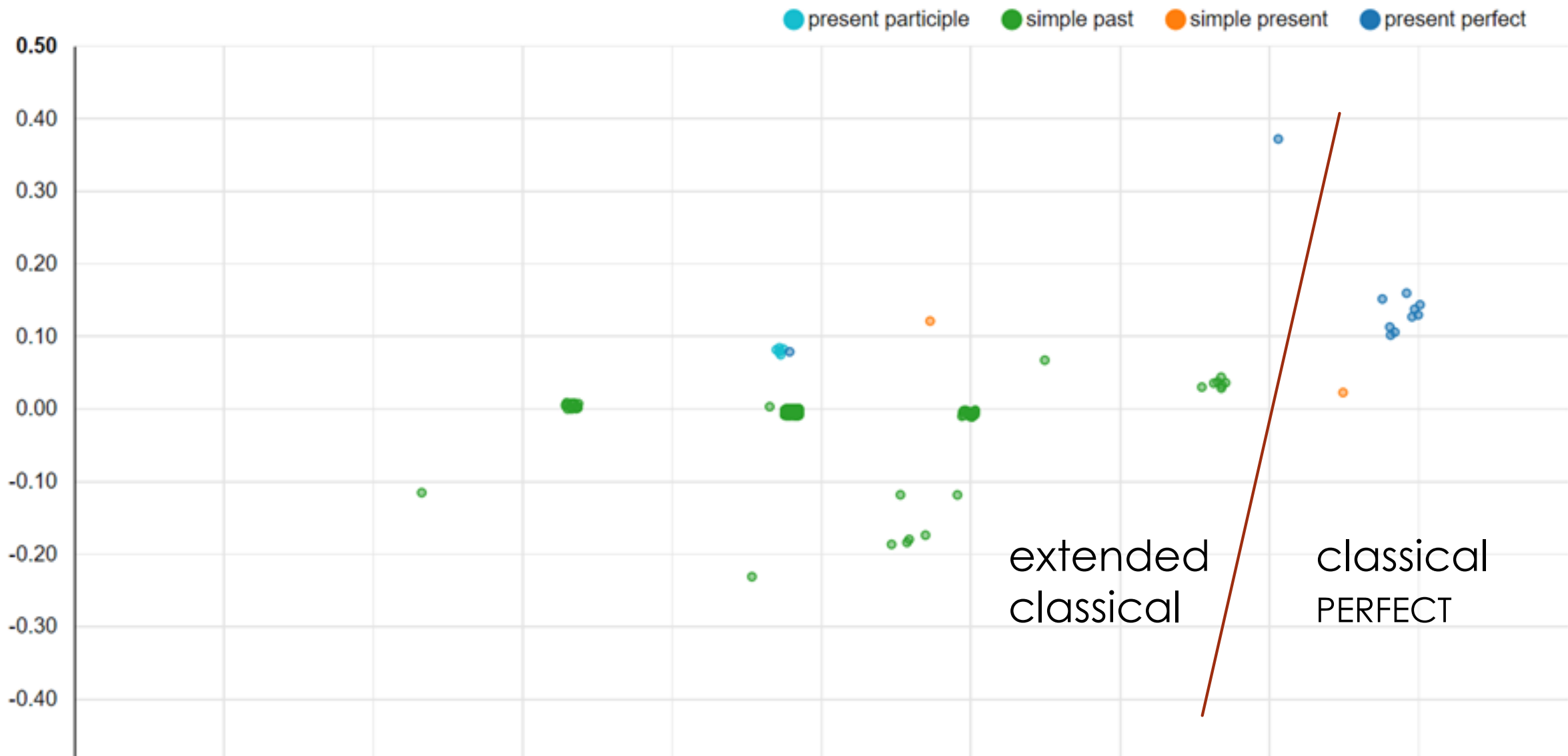
Spanish map



Boundedness: Dutch/Spanish

- Je crois que j'*ai somnolé* un peu. [Fr]
- I think I *dozed* off for a while. [En]
- Ik geloof dat ik een beetje *heb gedoezeld*. [NI]
- Creo que *dormité* un poco. [Es]

English map



Extended classical PERFECT use: Spanish/English

- Aujourd' hui , *maman est morte.* [Fr]
- Mother *died* today. [En]
- Hay, *mamá ha muerto.* [Es]

- L'un m'a même crié: "On les *a eus!*" [Fr]
- One of them even shouted to me, "We *thrashed* them." [En]
- Uno gritó incluso : "Les *hemos ganado.*" [Es]

'Classical' Perfect

- Je les voyais comme je n'*ai* jamais *vu* personne. [Fr]
- I saw them more clearly than I 've ever *seen* anyone. [En]
- Maintenant il est trop grand pour moi et j'*ai dû* transporter dans ma chambre la table de la salle à manger. [Fr]
- But now it 's too big for me and I've *had* to move the dining-room table into my bedroom. [En]
- Nishyama & Koenig (2010): reverse causality.

French \supseteq German \supseteq Dutch \supseteq Spanish \supseteq English

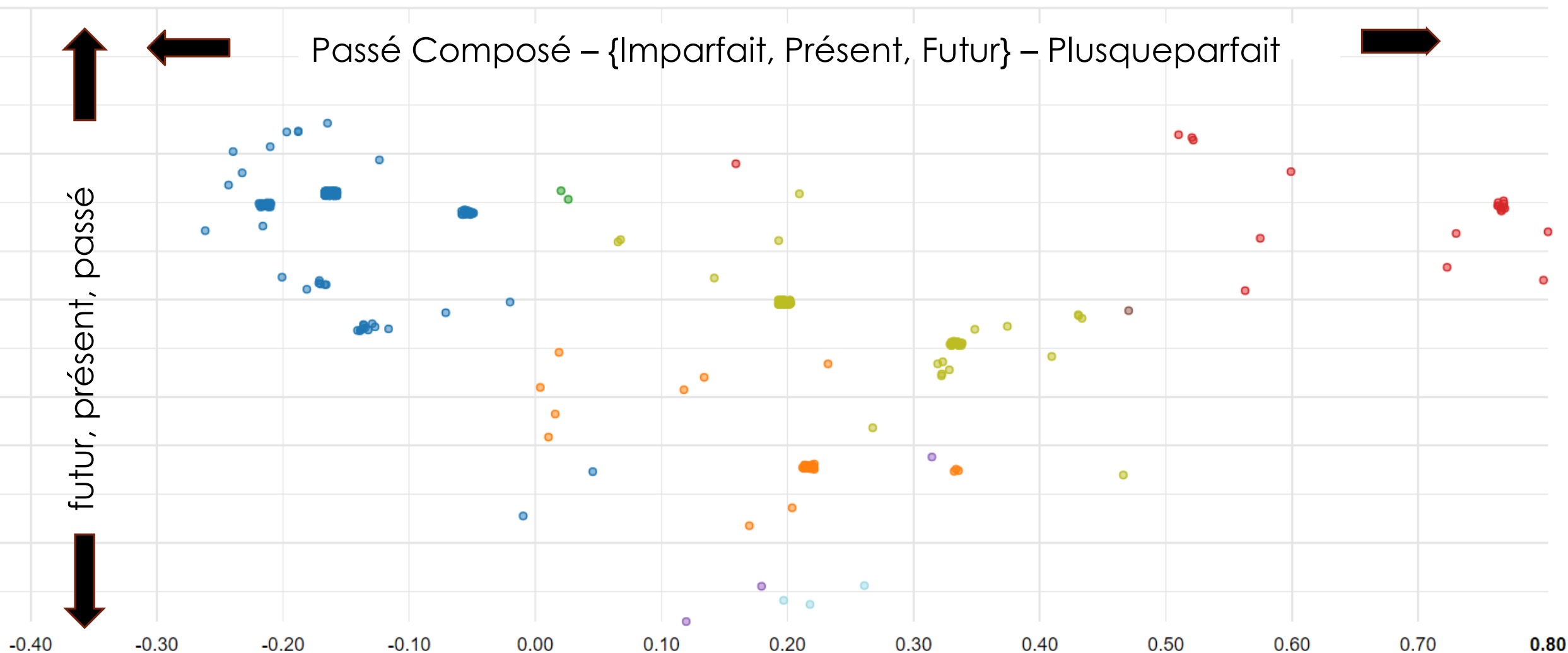
- Grønn & von Stechow (2016): French Passé Composé = English Present Perfect + Simple Past.
- How to test this?
- Add other verb forms from *L'étranger* to the competition.
- Strategy: manually extract all verb forms from the 1st Chapter of *L'étranger*, match them with their translations, label the tense forms, and create a temporal map of the entire temporal structure.

French map Camus (all)

○ conditionnel passé ○ futur proche du pass... ● présent ○ conditionnel ● imparfait ● passé simple ● passé composé ● plus-que-parfait ● futur proche
● futur antérieur ● futur

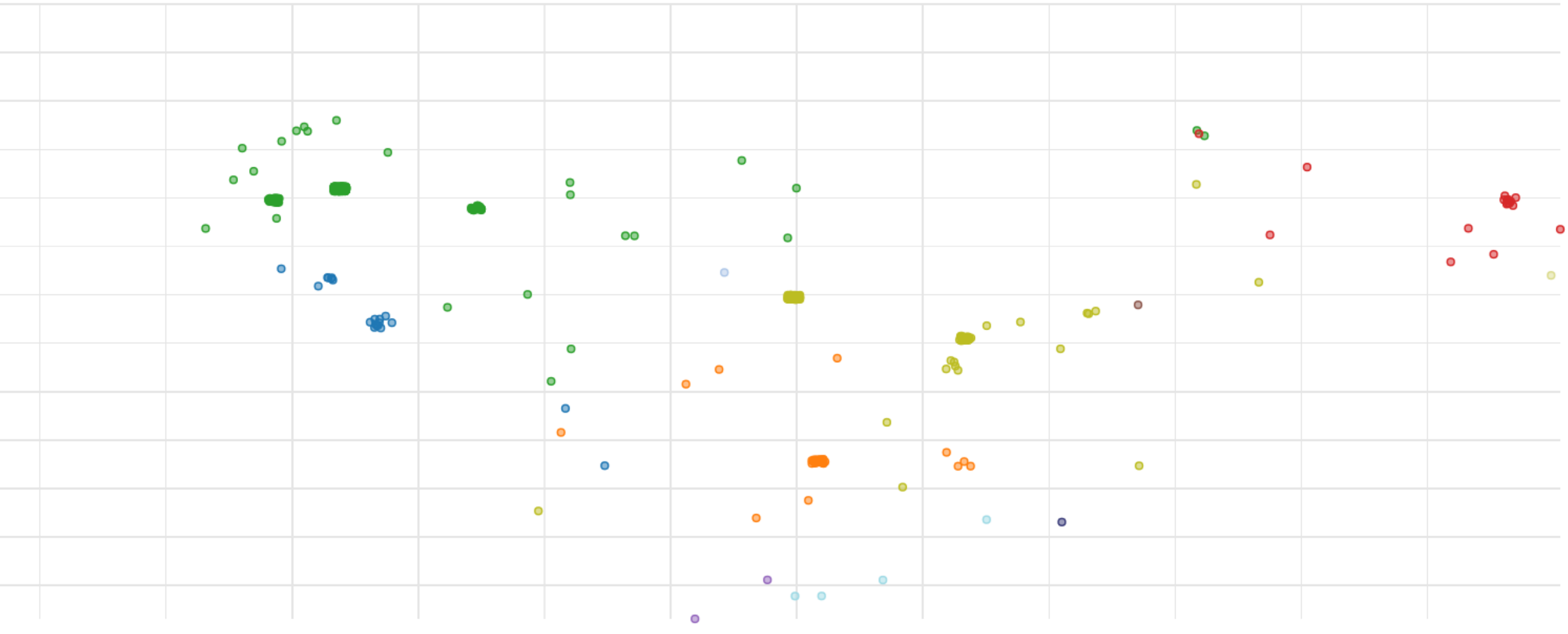
Passé Composé – {Imparfait, Présent, Futur} – Plusqueparfait

futur, présent, passé

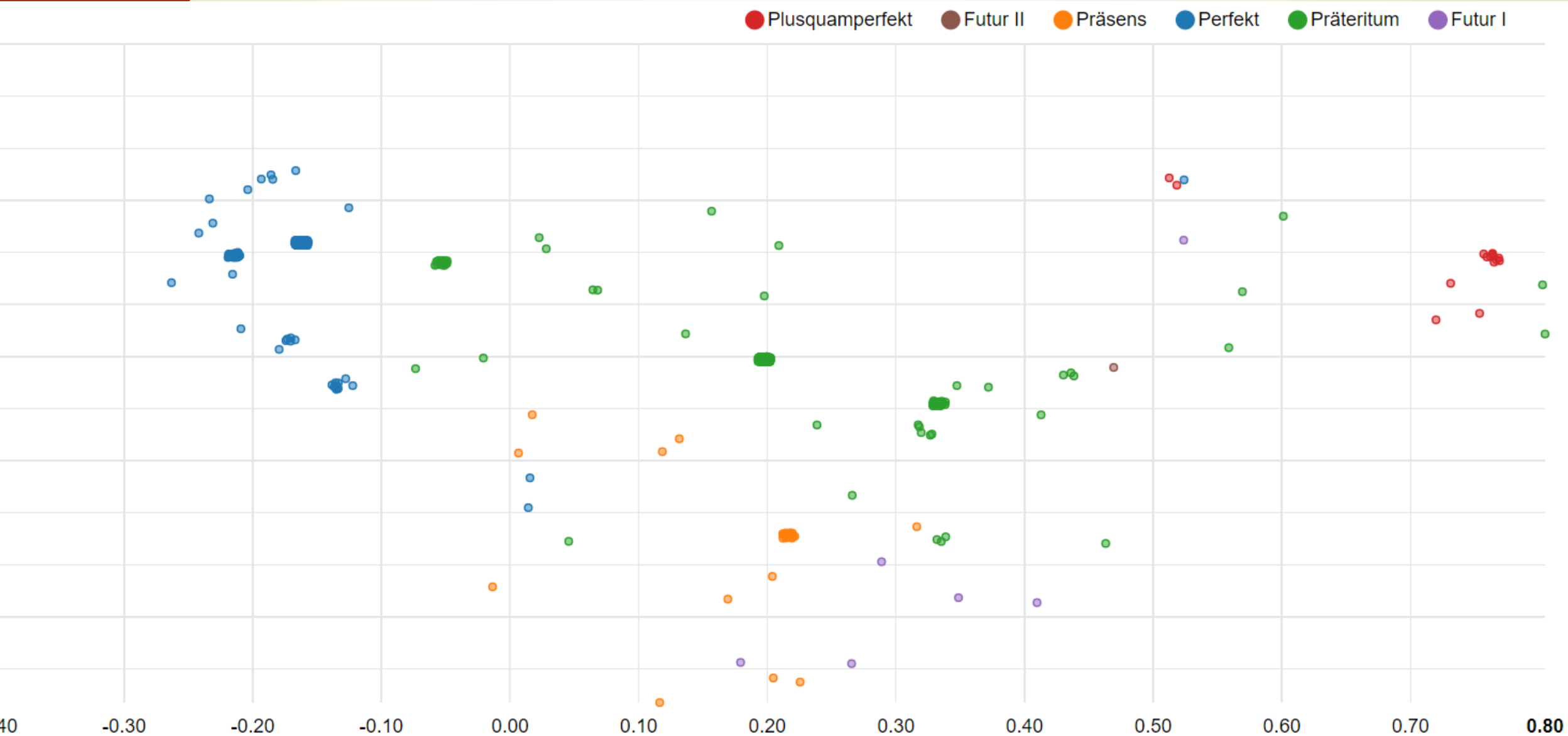


Spanish map Camus (all)

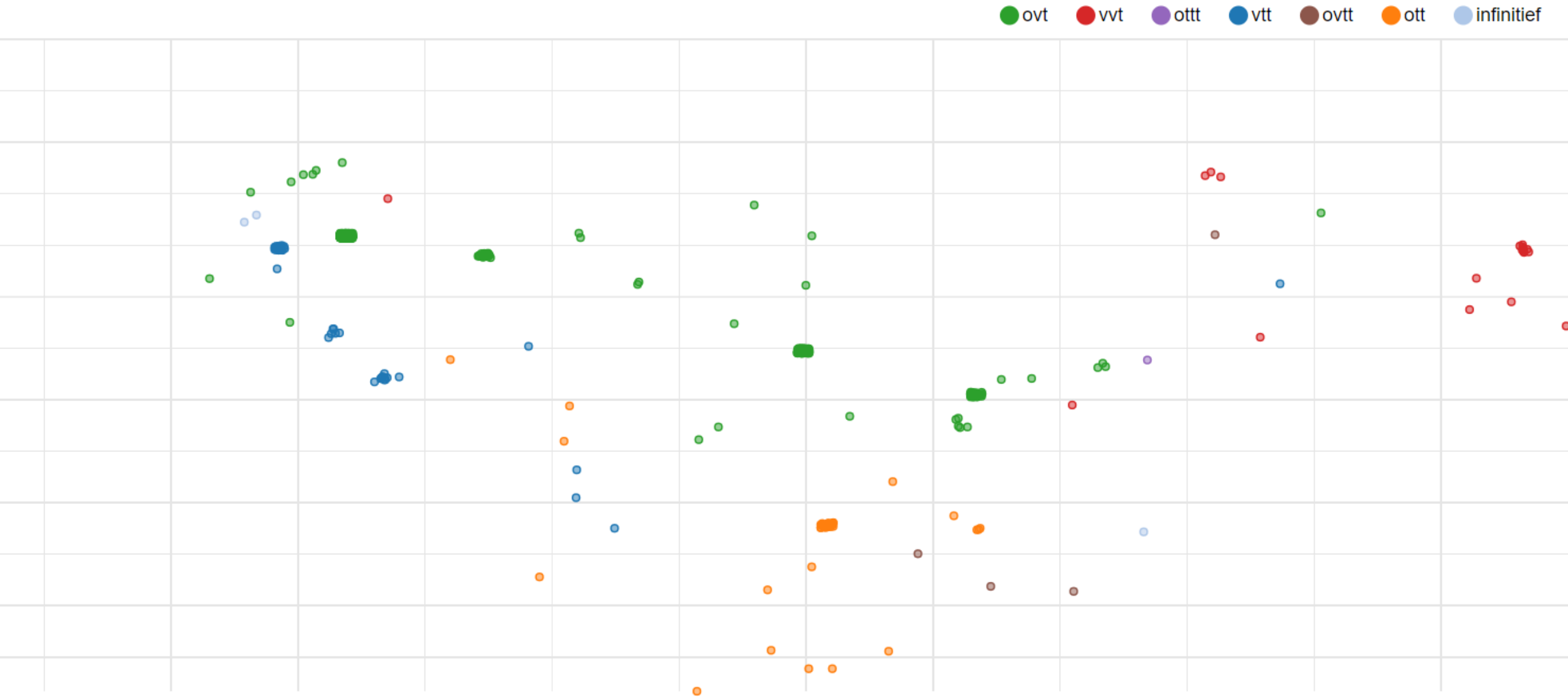
● infinitivo ● futuro próximo en im... ● presente ● pretérito perfecto s... ● infinitivo perfecto ● futuro perfecto ● pretérito imperfecto
● futuro próximo ● pretérito pluscuampe... ● futuro ● pretérito perfecto c...



German map Camus (all)

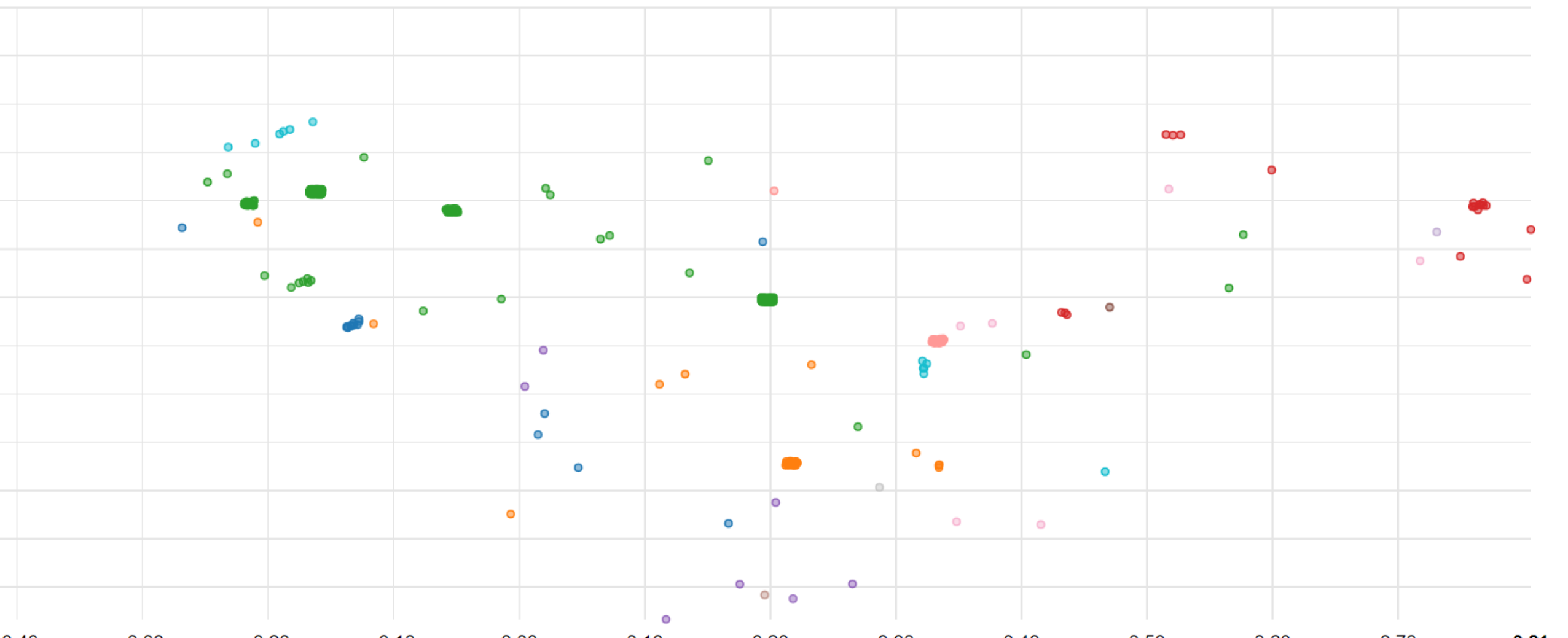


Dutch map of Camus (all)



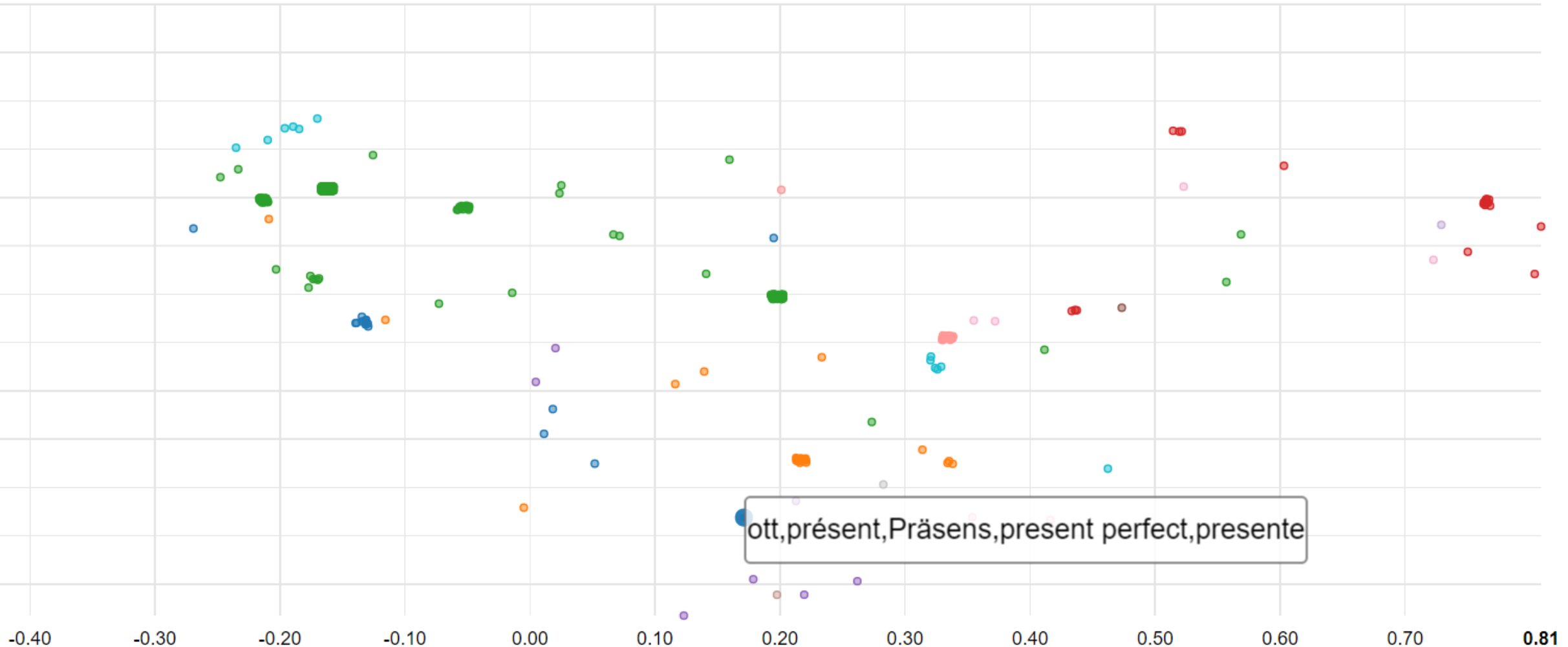
English map Camus (all)

future in the past c... past perfect simple past past continuous future in the past future continuous present participle simple present
present perfect simple future future perfect past perfect continu...



Generating PERFECTS in translation

● future in the past c... ● past perfect ● simple past ● past continuous ● future in the past ● future continuous ● present participle ● simple present
● present perfect ● simple future ● future perfect ● past perfect continu...



Continuative English Present Perfect

- Il y a longtemps que vous êtes là? [Fr]
- *Have you been* here long? [En]
- *Sind Sie schon lange* hier? [De]
- *Hoe lang bent u al* hier? [NI]
- Hace mucho tiempo que *está* usted aquí? [Es]

Work in progress..

- More on our research programme on the website:
<http://timealign.pythonanywhere.com/>
- We are just getting started, so feedback is appreciated!

