

Constraint-based Syntax and Semantics:
journée d'étude in honor of Danièle Godard
IEA, Paris, March 28 2017

Time in Translation

Henriëtte de Swart

Utrecht Institute of Linguistics (UiL OTS)



Joint work with
Martijn van der Klis and
Bert Le Bruyn




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*)
 - an event that took place in time (*World War II*),
 - that connect events in time (*before, since*).
- **Grammatical means:** tense.




Cross-linguistic stability and variation

- Tense is always **deictic**: present (*now*), past, future.
- **Aspect**: non-deictic, situation type (state, activity event); lexical and grammatical aspect.
- Cross-linguistic variation:
 - number and type of temporal distinctions established.
 - Number and type of aspectual distinctions established.



Two routes for the investigation of cross-linguistic variation

- Route 1: Macro-typological research: large set of unrelated languages (WALS)
 - Pre-conceptualized categories.
 - Data from different sources; difficult to compare.
 - Emphasis on description rather than constraints, limited prediction.
- 

Route 2: micro-typological research

- Small set of languages:
 - Closely related languages (similar grammar).
 - Parallel corpora: comparable data.
- New descriptions and new insights:
 - Generalizations emerge from the data.
 - Use results to test existing theories, and build new ones.
 - Empirical predictions can be tested against new corpora and native speaker intuitions.

Empirical domain:

past, present, *have* PERFECT

- Morpho-syntax PERFECT: 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]

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]
- Multilingual translation research confirms that PERFECT use in Dutch/German/ French is neither a subset, nor a superset of PERFECT use in English.
- This is **not** in line with common assumptions in the linguistics literature (e.g. Grønn & von Stechow 2016).

Time in Translation: a new project

- Existing approaches look at the PERFECT in isolation, we aim to look at the PERFECT in context.
- Language-internal variation:
 - Competition of PERFECT with PAST and PRESENT.
- Cross-linguistic variation:
 - Same meaning conveyed by other tense-aspect forms.
- From prototypical examples to corpus data:
 - Include discourse/conversational context.
 - Sensitivity to register.

Scaling up the translation approach

- Translation equivalents: **form variation** across languages in contexts where the **meaning is stable**.
- 5 languages (English, Dutch, German, French, Spanish).
- 3 main tense-aspect forms (PAST, PRESENT, PRESENT PERFECT).
- Digital resources (large datasets): Europarl (formal spoken language), Subtitle corpus (informal conversation), literary corpora (story telling).
- Infer constraints on tense use at the sentential and discourse/conversational level.

Methodology

- Extraction of PERFECTS from multilingual corpora
- Word-level alignment of verb phrases
- Tense attribution for translations (language specific)
 - Discard 'other' translations (nominalizations, 'free' translations)
 - Manual/automatic tagging (language-specific)
- Create a dissimilarity matrix
- Multidimensional scaling: project variation onto two-dimensional space (a semantic map).
 - Methodology inspired by Wälchli & Cysouw 2012

Translation

Time in Translation

timealign.pythonanywhere.com/timealign/show/2382

Time in Translation

The project

Student Research

Translation Mining

Contact

Signed in as henriette

Log out

Source

English

ep-00-12-11.xml - 2382

Everybody , including the nationals of third countries who have been victims of a crime in any State of the European Union , must receive the same minimum level of treatment .

Translations

German

Perfekt

Jede Person , die im Hoheitsgebiet irgendeines Mitgliedstaats der Europäischen Union Opfer einer Straftat geworden ist , muss eine vergleichbare Mindestbehandlung erhalten .

Das gilt auch für Angehörige von Drittstaaten .

Spanish

pretérito perfecto compuesto

Es necesario que toda persona , incluidos los nacionales de terceros países que han sido víctimas de un delito en el territorio de cualquier Estado de la Unión Europea , reciba un trato mínimo similar .

French

passé composé

Toutes les personnes qui ont été victimes d'une infraction sur le territoire de l'un des États membres de l'Union européenne doivent faire l'objet d'un même traitement minimal .

Dutch

vtt

Iedereen , met inbegrip van de onderdanen van derde landen die het slachtoffer zijn geworden van een misdrijf in een lidstaat van de Europese Unie , moet gelijk behandeld worden .

Different tenses in 5-tuple

Time in Translation

×

+

←

→

↻

timealign.pythonanywhere.com/timealign/show/3523

📖

☆

|

☰

🔍

🔔

⋮

Time in Translation

🏠

The project ▾

Student Research

Translation Mining ▾

Contact

Signed in as *henriette*

Log out

Dutch

ep-00-12-11.xml - 3523

Mevrouw de Voorzitter , niemand onderschat het belang en de reikwijdte van deze Top , die bovendien de langste in de geschiedenis van de Europese Unie **is geworden** .

Translations

German

Präteritum

Frau Präsidentin , niemand unterschätzt die Bedeutung und Tragweite dieses Gipfels , der noch dazu der längste in der Geschichte der Europäischen Union **war** .

English

present perfect

Madam President , no one underestimates the significance and scope of this summit , which **has** also **been** the lengthiest in the history of the European Union .

Spanish

presente

Señora Presidenta , nadie subestima la importancia y el alcance de esta Cumbre que , además , **es** la más larga de la historia de la Unión Europea .

French

présent

Madame la Présidente , personne ne sous-estime l' importance et la portée de ce Sommet , qui , en plus , **est** le plus long de l' histoire de l' Union européenne .

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	

Number of tense forms across languages (Europarl)

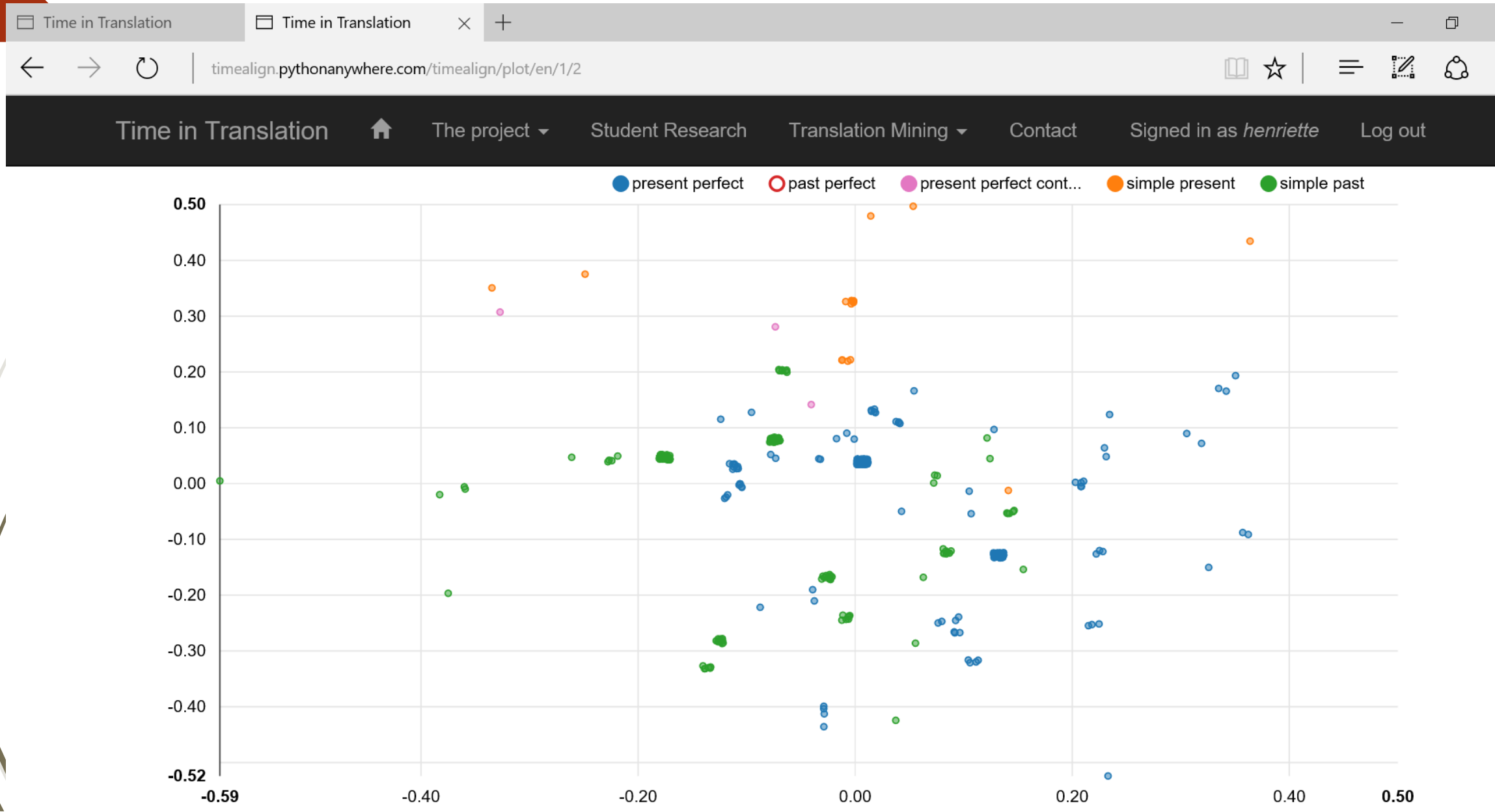
Methodology II

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

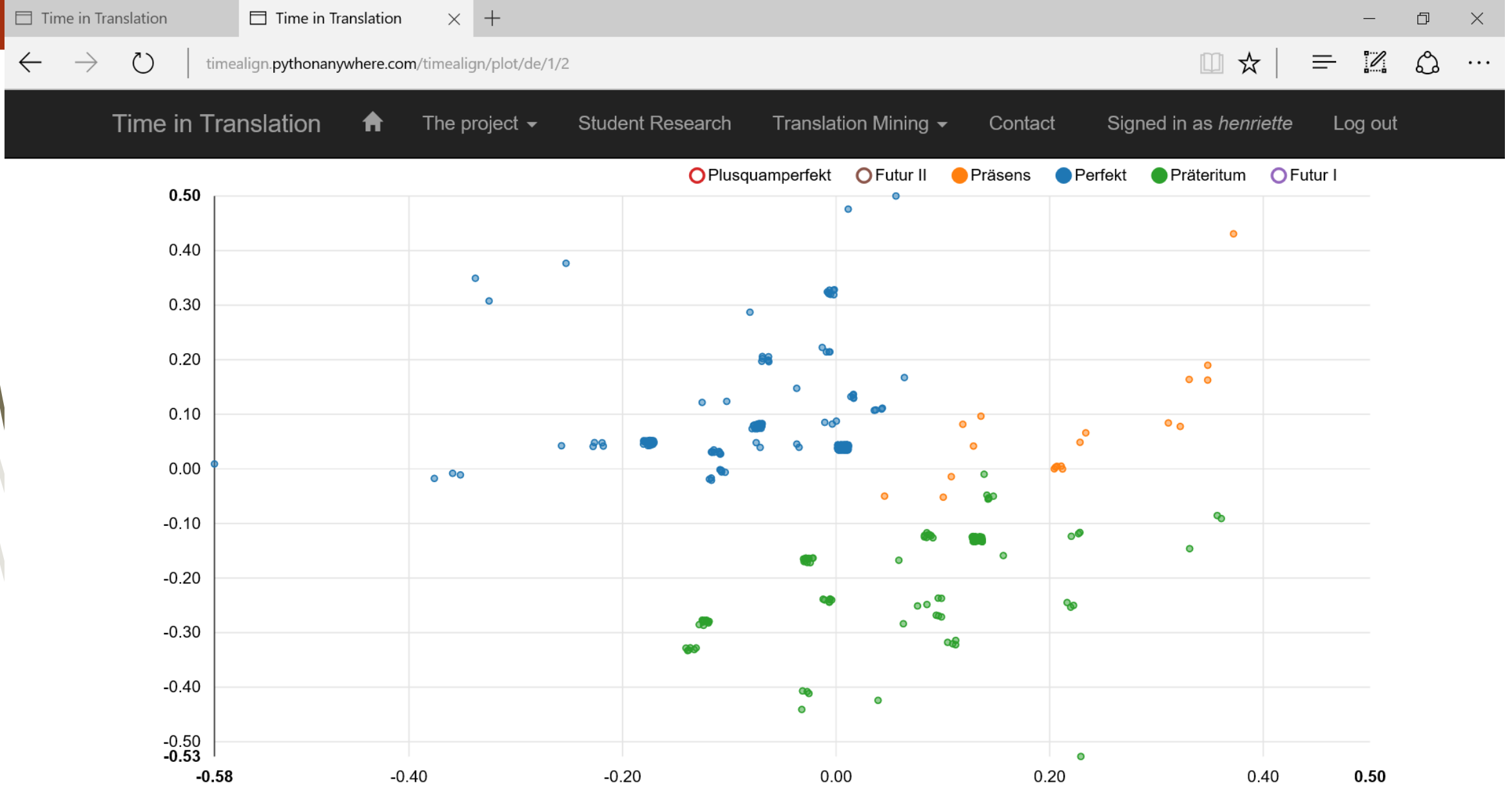
#	German	English	French	Spanish	Dutch
1	Perfekt	Present perfect	Passé composé	Pretérito perfecto conpuesto	Voltooid tegenwoordige tijd
2	Präterium	Simple past	Passé composé	Pretérito perfecto conpuesto	Voltooid tegenwoordige tijd
3	Perfekt	Present perfect	Passé récent	Pasado receinte	Voltooid tegenwoordige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

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

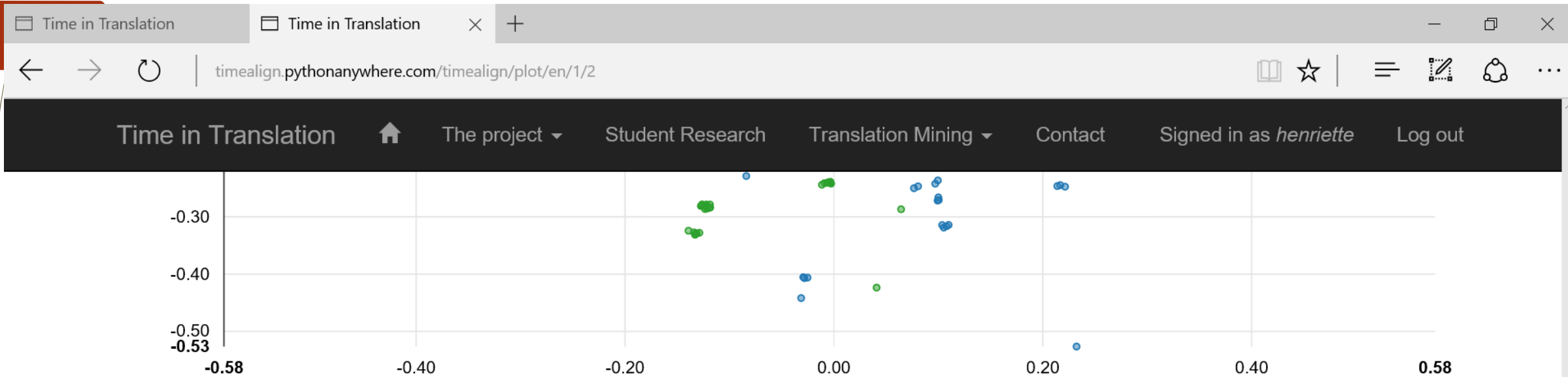
PERFECT map: English plot1/2



Perfect map: German plot 1/2



Filters



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!

English plot 2/3

Time in Translation

+



timealign.pythonanywhere.com/timealign/plot/1/en/2/3



Time in Translation



The project ▾

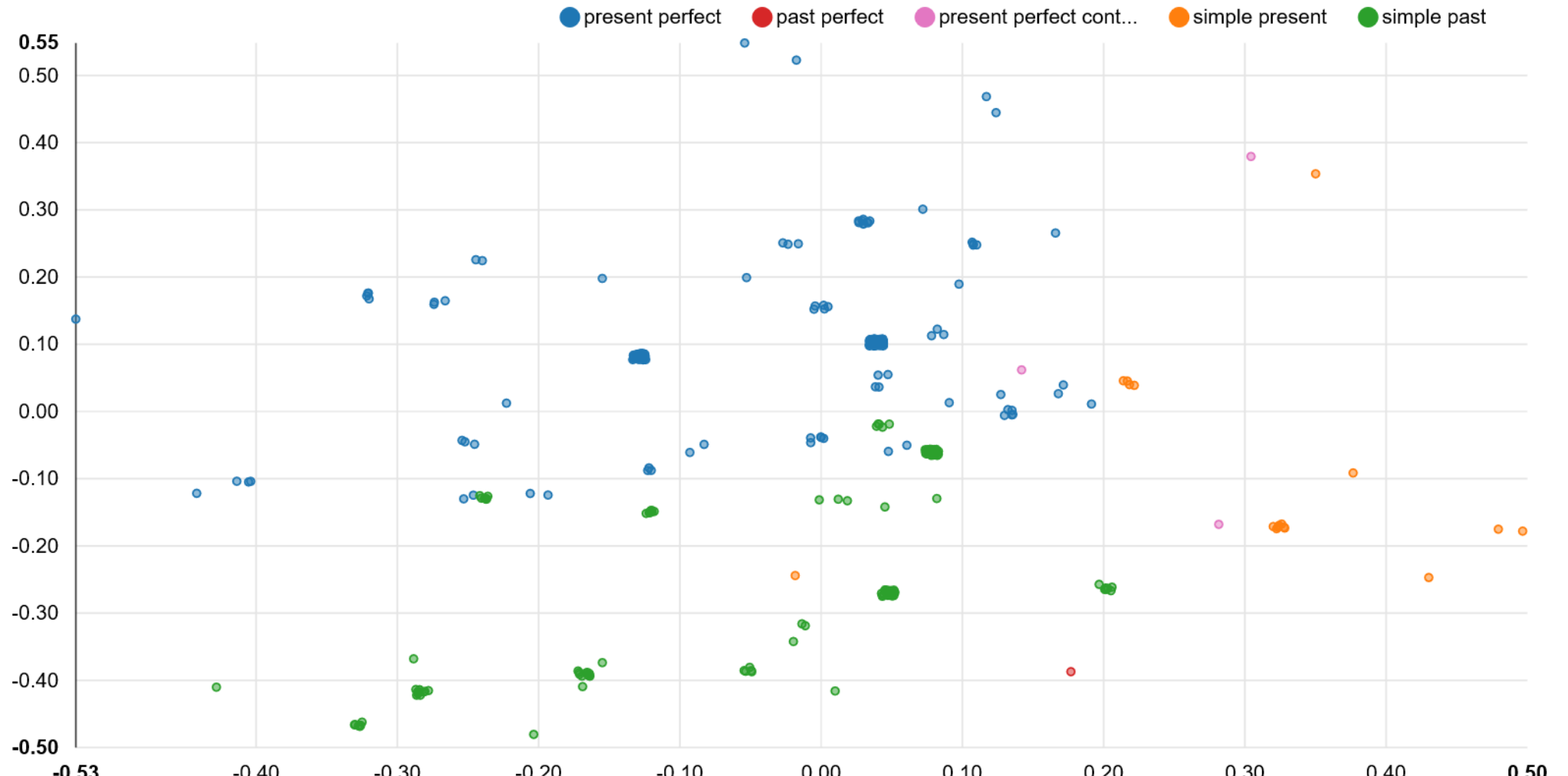
Student Research

Translation Mining ▾

Contact

Signed in as *henriette*

Log out

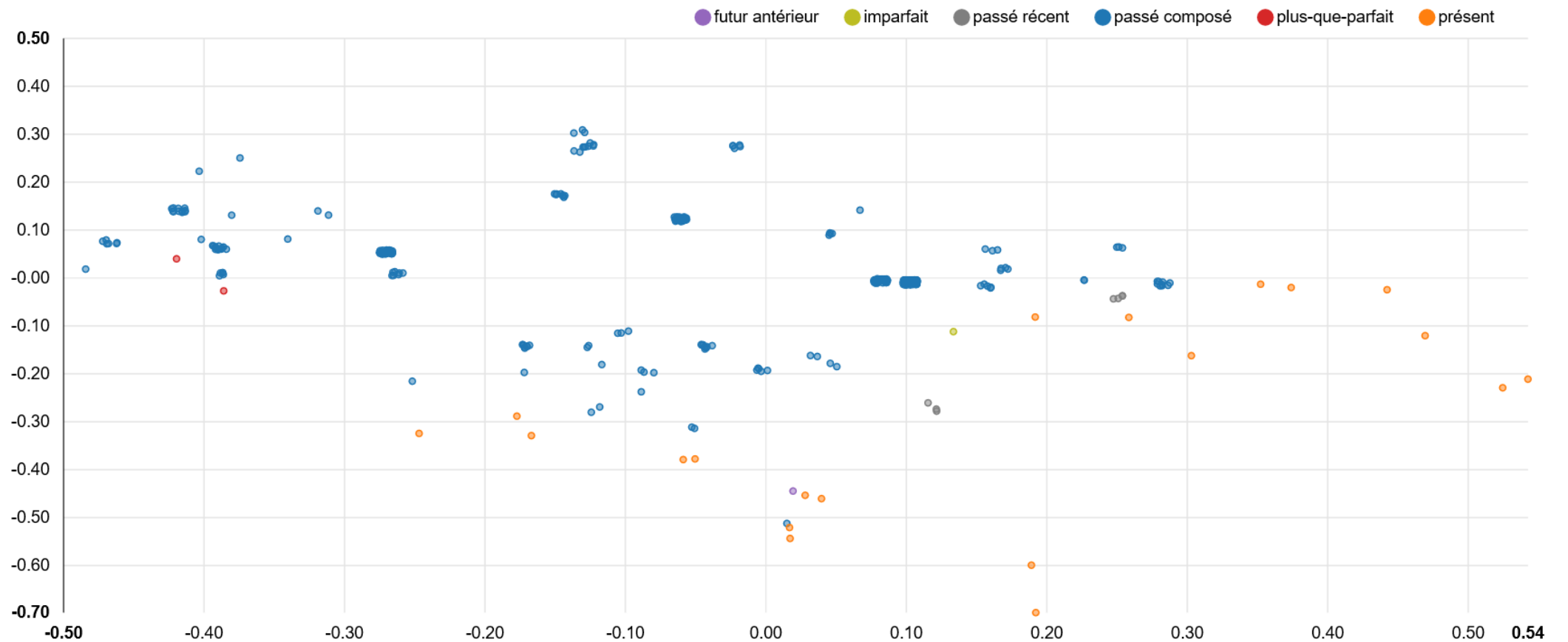


How to interpret the dimensions?

- MDS plots dissimilarity along (statistical) dimensions – no interpretation given.
- We ran MDS for 5 dimensions.
- Plot 1/2 carves out the opposition between PAST, PRESENT and PERFECT in German.
- Other dimensions are relevant in other languages.
- Question for linguistics: how to link dimensions to lexical and grammatical constraints?

French plot 3/5

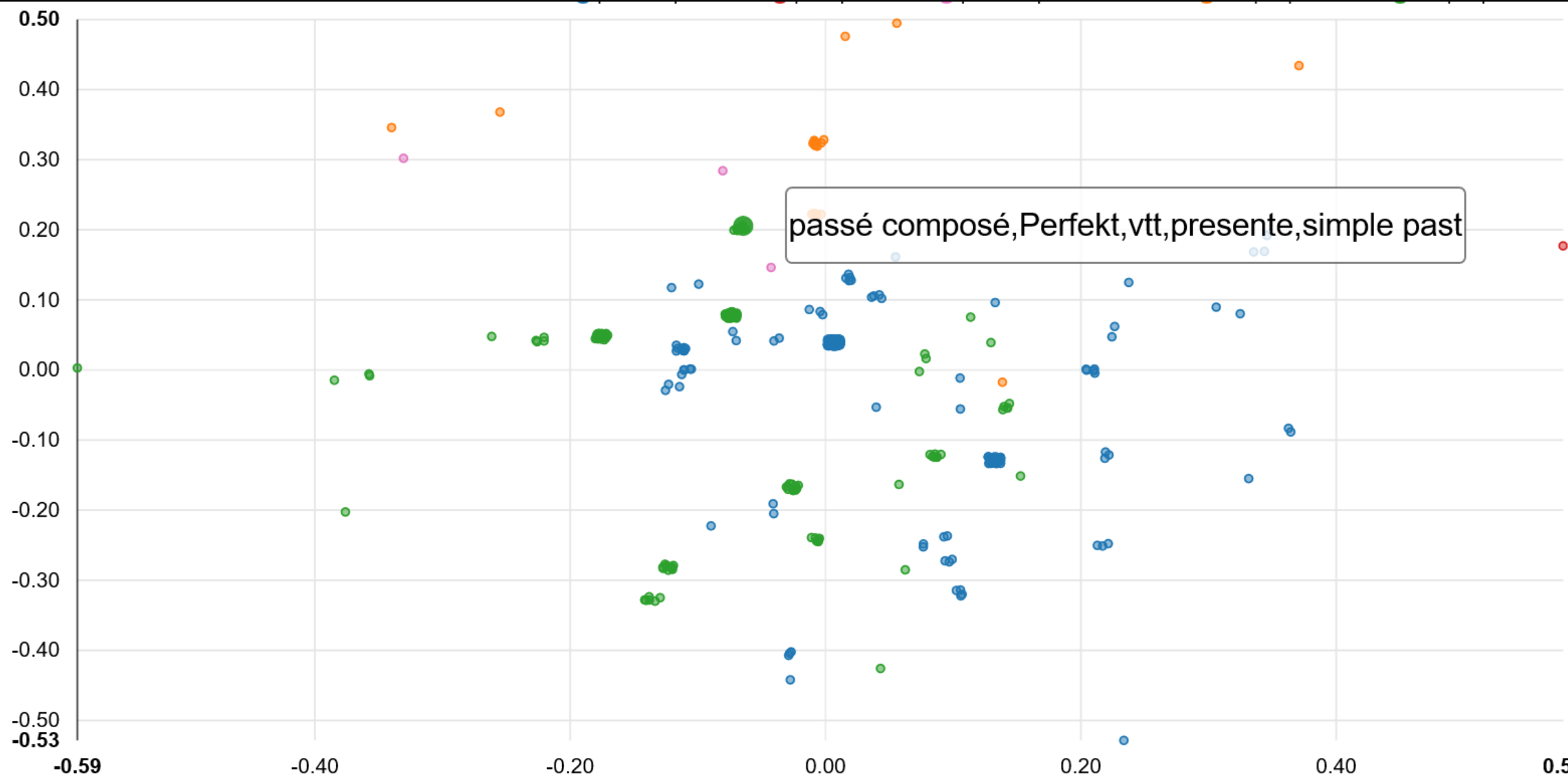
Time in Translation [The project](#) [Student Research](#) [Translation Mining](#) [Contact](#) Signed in as *henriette* [Log out](#)





Diving into the data

- The more the points are distant from each other, the more their tenses are different.
- We can investigate 'outliers' to find cases where one language is different from the other languages.
- Test linguistic theory: 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 (Schaden 2009).



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.

The screenshot shows the 'Time in Translation' web application. The browser address bar displays the URL `http://timealign.pythonanywhere.com/timealign/st`. The application's navigation bar includes links for 'Time in Translation', 'The project', 'Student Research', 'Translation Mining', 'Contact', and a user status 'Signed in as henriette' with a 'Log out' option. The main content area is titled 'French' and shows a text snippet: '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.' The words 'ont voté' are highlighted in green, and 'aujourd'hui' is circled. A file identifier 'ep-00-12-14.xml - 14906' is visible in the top right. Below this, the 'Translations' section is divided into five language-specific boxes:

- German** (Perfekt): 'Die dänischen Sozialdemokraten haben heute gegen den Vorschlag betreffend die gemeinsame Marktorganisation für Bananen gestimmt.' The words 'haben' and 'gestimmt' are highlighted in green, and 'heute' is circled.
- English** (present perfect): 'The Danish Social Democrats have today voted against the proposal for a common market arrangement for bananas.' The words 'have' and 'voted' are highlighted in green, and 'today' is circled.
- Spanish** (pretérito perfecto compuesto): The text is partially obscured but includes '(DA) Los socialdemócratas daneses hemos votado hoy en contra de la propuesta relativa a la organización común de mercados en el sector del plátano.' The words 'hemos' and 'votado' are highlighted in green, and 'hoy' is circled.
- Dutch** (ott): 'De Deense sociaal-democraten stemmen tegen het voorstel betreffende een gemeenschappelijke marktordening voor bananen.' The word 'stemmen' is highlighted in green.

Subtle differences in what counts as 'now'



The screenshot shows the 'Time in Translation' web application. The browser's address bar displays the URL `http://timealign.pythonanywhere.com/timealign/sl`. The application's navigation bar includes links for 'Time in Translation', 'The project', 'Student Research', 'Translation Mining', 'Contact', and a user status 'Signed in as henriette'. The main content area is titled 'German' and shows a source sentence: 'Letzte Nacht habe ich davon geträumt, ein Rentner zu sein, denn obwohl ich die italienische Rentnerpartei vertere, beziehe ich keine Rente.' The words 'habe' and 'geträumt' are highlighted in green. A file identifier 'ep-00-12-14.xml - 8260' is visible in the top right. Below the German text, the 'Translations' section is divided into four language pairs, each with a tense label in a blue box:

- English** (simple past): 'I must tell you that last night I dreamt I was a pensioner.' The words 'last night' and 'dreamt' are highlighted in green.
- Spanish** (pretérito perfecto compuesto): 'Debo decir que esta noche he soñado con que era un pensionista.' The words 'esta noche' and 'he soñado' are highlighted in green.
- French** (passé composé): 'Je dois dire que cette nuit j'ai rêvé que j'étais retraité.' The words 'cette nuit' and 'ai rêvé' are highlighted in green.
- Dutch** (ovt): 'Ik moet u vertellen dat ik afgelopen nacht droomde dat ik met pensioen was.' The words 'afgelopen nacht' and 'droomde' are highlighted in green.

Each language translation section also contains a second sentence in the same language, providing further context or a different perspective on the original German sentence.

Work in progress..

- More on our research programme on the website:
- <http://timealign.pythonanywhere.com/>
- We are just getting started, so feedback is appreciated!
- The project has an opening for a PhD student (fall 2017) and a post-doc (fall 2018).
- Jop openings posted on:
<https://www.uu.nl/en/organisation/working-at-utrecht-university/jobs>

