

VARIATION IN PERFECT/PAST COMPATIBILITY WITH TEMPORAL ADVERBIALS IN ENGLISH AND SPANISH

Martín Fuchs & Martijn van der Klis
Utrecht University

Linguistic Symposium on Romance Languages (LSRL) 52
University of Wisconsin Madison
23 April 2022



Universiteit Utrecht



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 method: **Translation Mining** (parallel-corpora: compare translations of the **same text** into different languages).
- Corpora: Camus (1942) *L'Etranger*, and J.K. Rowling's *Harry Potter and the Philosopher's stone*.

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.

Algorithms created by the Digital Humanities Lab
of Utrecht University



Reflections on parallel corpus research

- Data driven approach where 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 translations.
- Can we reproduce the insights from one corpus in a different corpus with a different source language, preferably one that has a ‘classical’ PERFECT?
- 2nd TinT translation corpus: J.K. Rowling’s *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.

Converging methodologies

- Since the object of linguistic inquiry cannot be directly accessed, we should use **several manifestations of linguistic behavior to reconstruct it** (Kepser & Reis 2005).
- Data types: introspection, experiments, corpora, typology, etc.
- **The more converging the data, the more support for hypotheses** (Adli 2005, Hoffmann 2006, Francis 2022)

Our methods

- Research question:
 - *Which features differentiate the **PAST** and the **PERFECT** in languages that have two markers?*
- How to tackle the problem:
 - *Parallel corpus : translation (same context/meaning, different forms).*
 - *Experimentally check the reliability (and the granularity) of the results.*

Roadmap

1. Introduction & Research Questions: The **PERFECT**, its competition with the **PAST**, and its crosslinguistic variation.
2. A case study: the *Present Perfect puzzle* in English and Spanish. through an **acceptability judgment task**.
3. General conclusion: **triangulate your data for better generalizations and wider empirical coverage**.



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 *ate* breakfast at that café. (Perfective) **Past** **E-R < S**
- 2) Mary *has eaten* breakfast at that café. (Present) **Perfect** **E < R,S**

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:

- 3) Mary (#*has*) *left* work at 5pm. **R=S restricts past-time adverbs**

The **Present Perfect** is also not available in narrative contexts:

- 4) When Mary (#*has*) *left*, she (#*has*) *waved* at me. **R=S blocks anaphors /narration**

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.
- Different languages use their corresponding **PERFECT** and (Perfective) **PAST** markers to different extents.
 - Marie **a pris** / (**#prit**) le petit-déjeuner dans ce café
 - Maria **hat** / (**#frühstücke**) in diesem Café **gefrühstückt**
 - Maria **heeft** (in dat café) / (**ontbeet**) in dat café **ontbeten**
- There are some **language-specific analysis** other than English (Bertinetto 1986, Vet 1992, Boogaart 1999, Löbner 2002, Howe 2006) and also, more recently, some **awareness of cross-linguistic variation** (e.g., de Swart 2007, Schaden 2009, Dahl & Velupillai 2013).

Constraints in other languages

- French, Dutch, German **PERFECT**: compatible with past time adverbials.
 - Marie *est sortie* du travail à 17 heures. [French]
 - 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]
 - Toen Marie *#is weggegaan* / *ging weg*,
#heeft / *zwaaide* 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.

Parallel corpus work (van der Klis et al. 2021)

Spanish: constrain on preodiernal past

- Events delimited by a (preodiernal) temporal adverbial are expressed by the **PERFECT** in Dutch/German/French, but require the *Préterito Indefinido* in Spanish.

- 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: constrain on prehodieral past

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 $E=R \subseteq \text{day}(S)$ to combine with the **PERFECT**, becoming a **hodiernal past marker** (e.g., Schwenter 1994).

Parallel corpus work (van der Klis et al. 2021)

English: constrain on past events

- *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]

English: deixis

Does English never allow temporal adverbials with the **Present Perfect**?

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) to combine with **Present Perfect**.
- Adverbs that include the speech time **S**, like *this morning*, target **R=S** (de Swart 2007) and do not create an incompatibility with this marker.

A thick black L-shaped frame is positioned on the left and bottom edges of the slide, framing the central text.

THE PRESENT PERFECT PUZZLE IN
SPANISH AND ENGLISH:
AN ACCEPTABILITY JUDGMENT TASK

Research questions

- Can we assess the role of both these constraints in **PERFECT** use in Spanish and English?
- Can we test for the strength of our corpora-based generalizations (reliability) and improve their level of detail (granularity)?

Method: acceptability judgments

- 8 contexts that conveyed bounded **events** (lexical aspect: achievements) with different past-referring temporal adverbials.
- **Two languages:** Peninsular Spanish / UK 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: She **has lived**/ ***lived** in Berlin since 2010)
- Latin Square design.
- 75% of stimuli followed by comprehension questions.
- 160 subjects per language (recruited through MTurk).
 - 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:

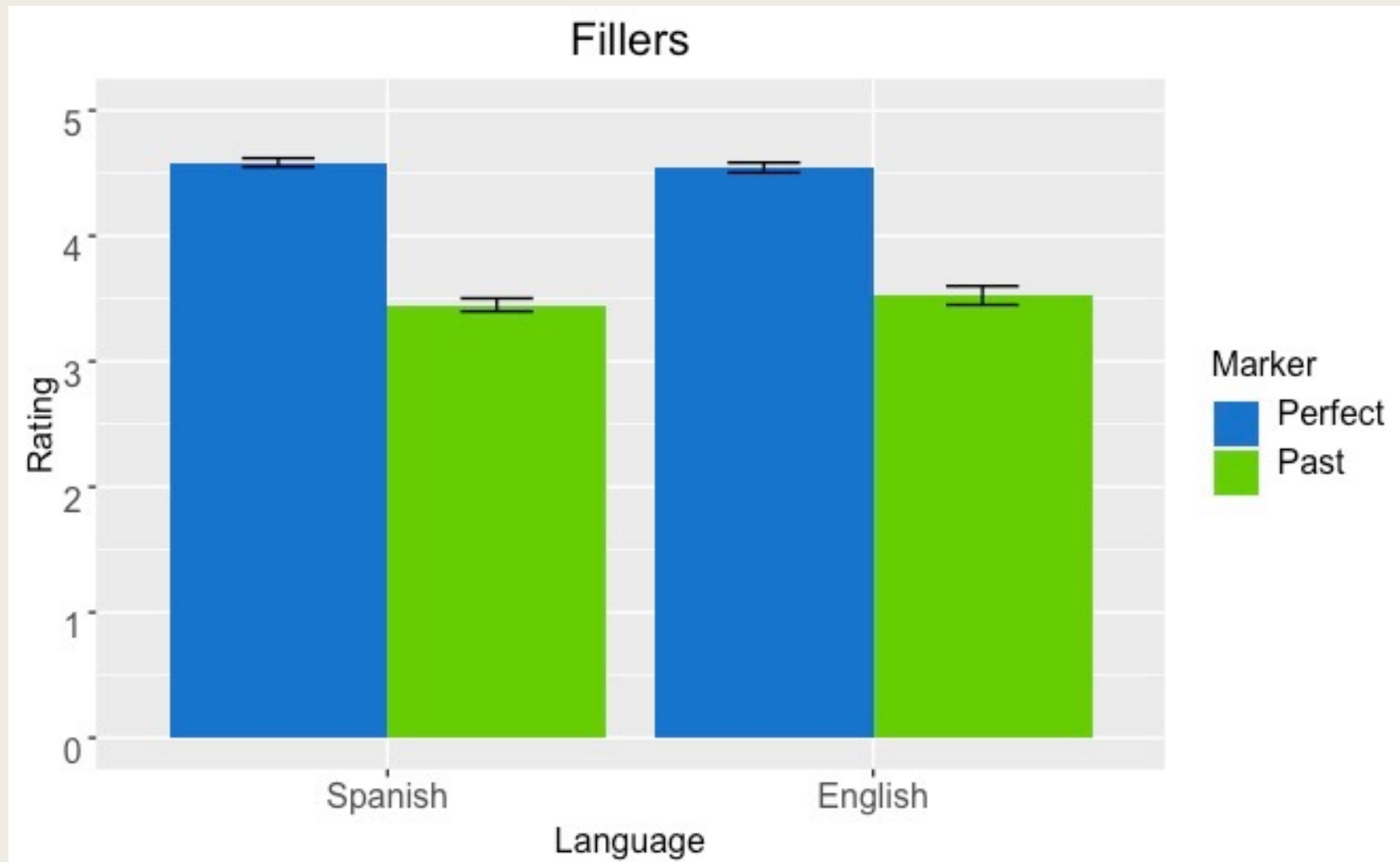
- +T, +D: “I *purchased* / *have purchased* mine this morning
- +T, -D: “I *purchased* / *have purchased* mine at midnight
- T, + D: “I *purchased* / *have purchased* mine last month
- T, -D: “I *purchased* / *have purchased* mine in November

It was cheaper that way”.

Data analysis & Results

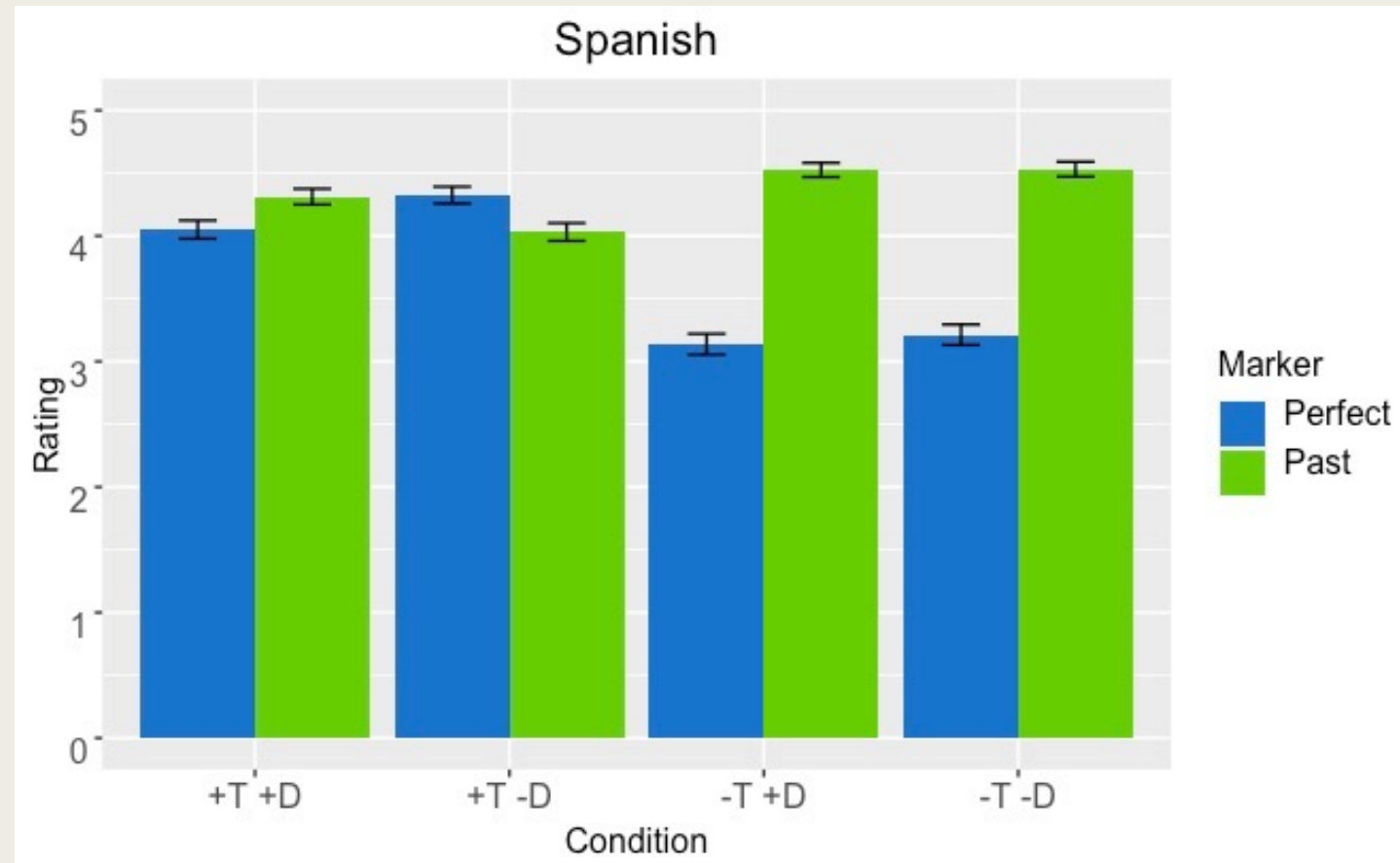
- All participants performed at above 75% in comprehension questions.
- Linear mixed model analysis; *lme4* package (Bates et al. 2015), in R.
 - *Model selection by Likelihood Ratio Tests.*
 - *Post hoc tests in multcomp package (Hothorn et al. 2008).*
 - *p-values corrected by Tukey.*
- **Fixed effects:**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

Fillers (e.g., continuatives)



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.

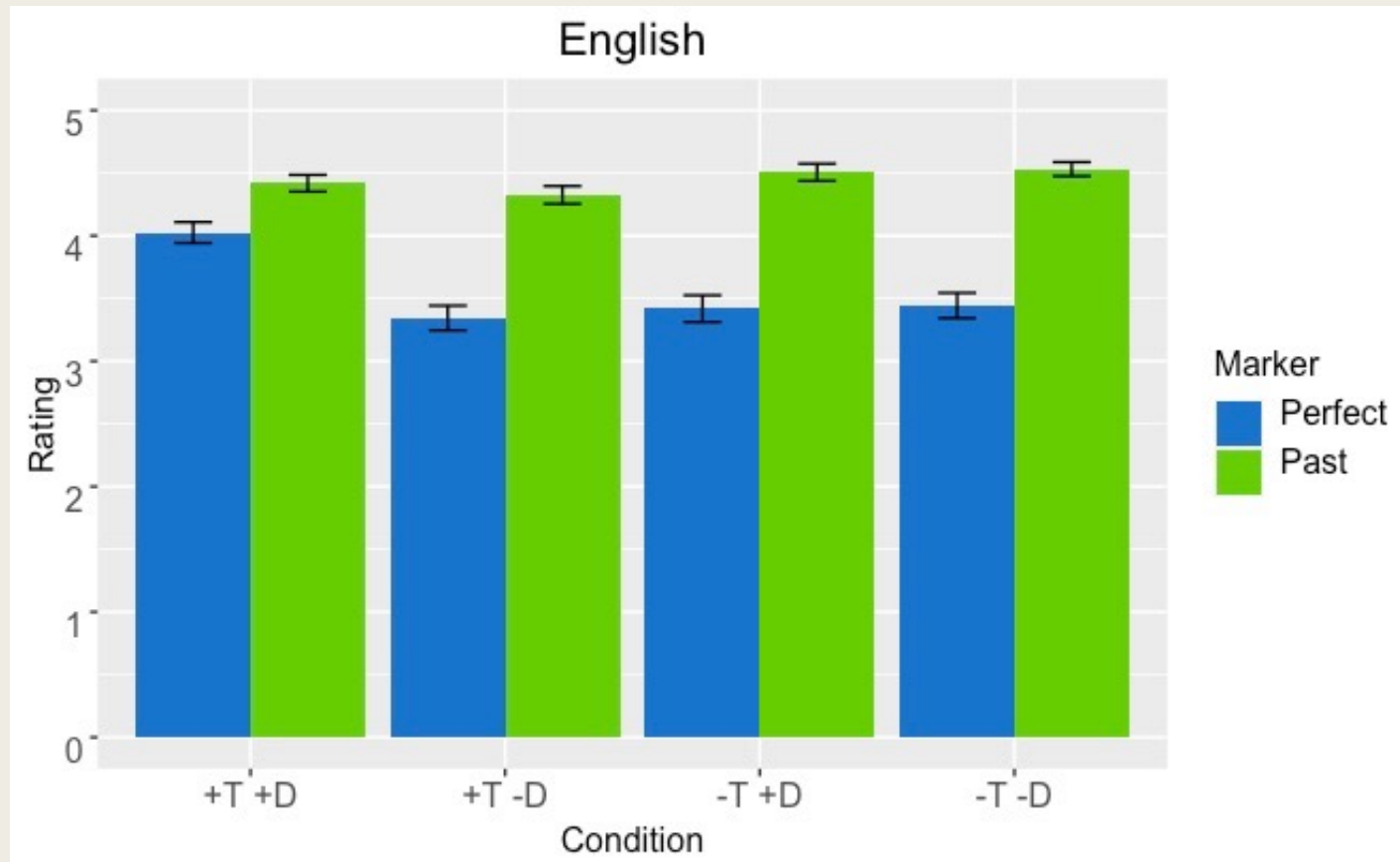


Spanish: No effect of regional variation

- Use of *Pretérito Perfecto Compuesto* had reported in Central/Castilian Spanish, but northwest Spanish (by influence of Gallego) and Andalusian Spanish should have a preference for *Pretérito Indefinido* use.
- We found **no effect of region (or interaction)** in our data.
- Maybe regional variation is limited to **production** tasks (e.g., Schwenter 1994, Azpiazu 2014, 2015, etc.)

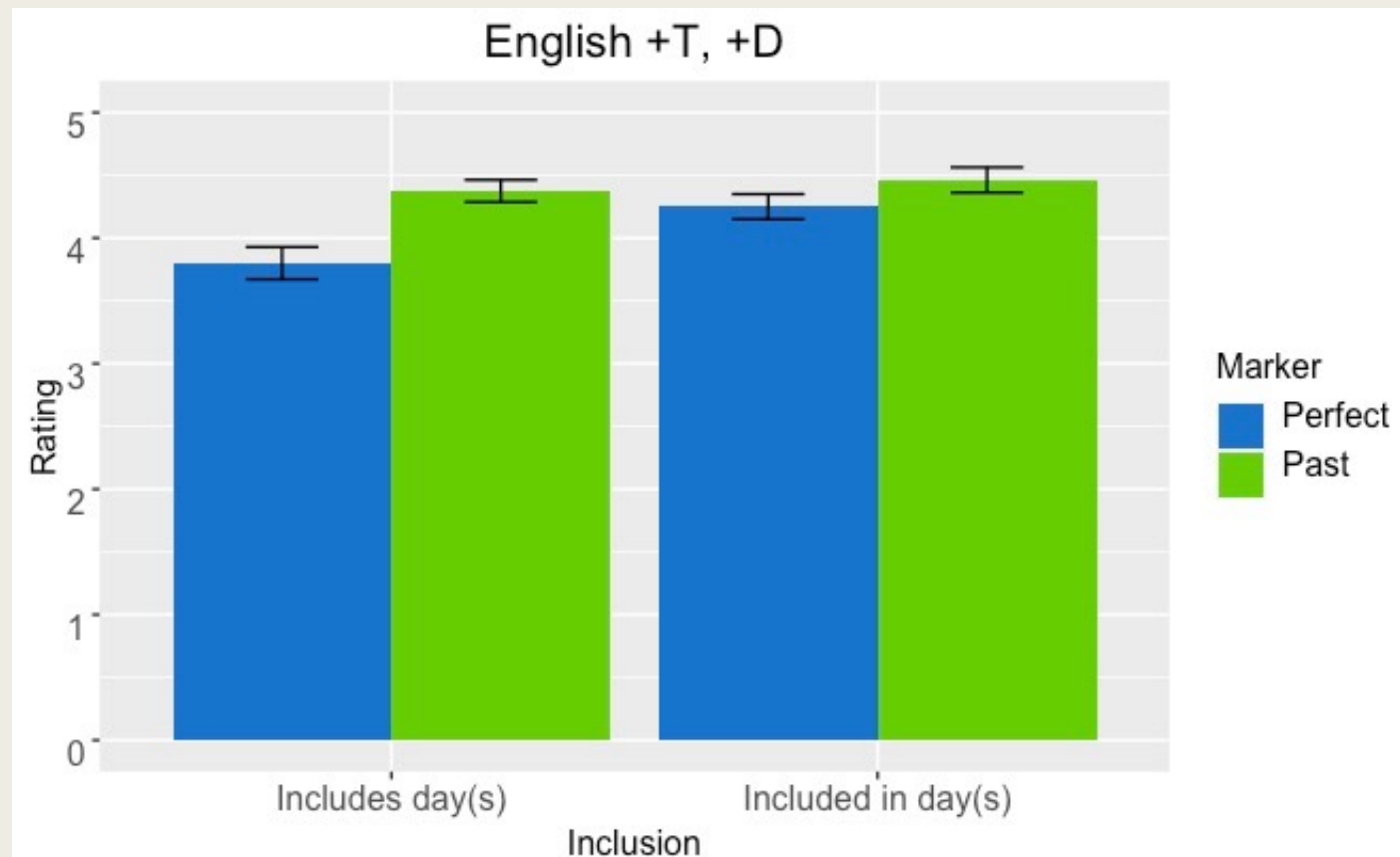
English

- 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

- **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 et seq), since the PAST can also be used and receives similar ratings.
- **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*).

Conclusions

- We produced a more **fine-grained typology** of the elements that are part and parcel of a crosslinguistically valid semantics of the **PERFECT** (that needs to be able to account for its **adverbial compatibility** across languages).
- Both hodiernality and deixis (and considerations about ‘proper’ hodiernality vs. ‘extended’ hodiernality) will have to be included in that crosslinguistic semantic analysis of the **PERFECT** ...
- We are working on it!

Take-home message

- Translation Mining/Parallel corpora research 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, and **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

- Adli, A. 2005. Gradedness and consistency in grammaticality judgments. In S. Kesper & M. Reis (eds.) *Linguistic Evidence: Empirical, Theoretical and Computational Perspectives*. New York: De Gruyter.
- Francis, E. J. 2022. *Gradient Acceptability and Linguistic Theory*. Oxford: OUP //
- Hitzeman, J. 1995. A Reichenbachian Account of the Interaction of the Present Perfect with Temporal Adverbials. *NELS* 25, 17.
- Hoffman, T. 2006. Corpora and Introspection as Corroborating Evidence: The case of preposition placement in English relative clauses. *Corpus Linguistics and Linguistic Theory* 2(2), 165-195
- Klein, W. 1992. The Present Perfect Puzzle. *Language* 68, 525-552.
- van der Klis, M., Le Bruyn, B. & de Swart, H. 2021. A multilingual corpus study of the competition between past and perfect in narrative discourse. *Journal of Linguistics*. 1-35
- Portner, P. 2003. The (Temporal) Semantics and (Modal) Pragmatics of the Perfect. *Linguistics and Philosophy* 26, 459-510.
- Reichenbach, H. 1947. *Elements of Symbolic Logic*. London: McMillan.
- Schwenter, S. 1994. The Grammaticalization of an Anterior. *Studies in Language* 18, 71-111.
- Squartini, M. & Bertinetto, P. 2000. The Simple and Compound Past in Romance languages. In Ö. Dahl, *Tense and Aspect in the Languages of Europe*. 403-440.
- de Swart, H. 2007. A cross-linguistic discourse analysis of the perfect. *Journal of Pragmatics* 39, 2273-2307.
- Vet, C. 1992. Le passé composé: contextes d'emploi et interprétation. *Cahiers de Praxématique* 19, 37-59.

Gracias por su atención!
Thank you for your attention!

Research funded by [NWO](#) under grant 360-80-070.