FACULTY OF PSYCHOLOGY
AND EDUCATIONAL SCIENCES

DEPARTMENT OF EXPERIMENTAL PSYCHOLOGY

PRODUCTIVITY OF THE SPANISH INCHOATIVE CONSTRUCTION: FROM CORPUS TO ACCEPTABILITY JUDGMENTS

Mariia Baltais (Ghent University) & Robert J. Hartsuiker (Ghent University)
Spruik presentation at AMLaP 2022, University of York

GHENT
UNIVERSITY

BACKGROUND

People use language “in creative but constrained ways”
(Goldberg, 2019)

- Syntactic productivity = a construction’s ability to attract new or existing lexical items
- Traditional corpus measures: token frequency of (co-)occurrence, type/token ratio, hapax/token ratio, etc.

How is productivity attested in corpora related to productivity “at work” in the mind of language users?

- Constructions are ***extensible*** beyond closed-ended corpora
  - ➢ Are corpus measures of productivity predictive of acceptability ratings?

- Grammaticality-frequency discrepancy: ratings tend to be more lenient than corpus data
  - ➢ Do ratings reflect a construction’s ***extensibility***?

➢ Influence of participants’ individual characteristics?

GHENT UNIVERSITY

MATERIALS

- Spanish inchoative construction [V + Prep ”a” + INF]: the onset of an event
- Two sources of productivity: inchoative verb slot, infinitive slot

[Subj]	**[V]**	[Prep]	**[INF]**	Translation
Pedro	**empieza / rompe / …**	a	**reír**	'Pedro begins / breaks / … to laugh'
Pedro	**empieza / ?rompe**	a	**entender**	'Pedro begins / ?breaks to understand'

- Van Hulle & Enghels, in press: data from the Spanish Web corpus
  - 6 inchoatives with different degrees of productivity of the INF slot
  - Each combined with 10 infinitives of different token frequency of co-occurrence (incl. hapaxes and non-attested infinitives)

GHENT UNIVERSITY

PARTICIPANTS & PROCEDURE

- 96 native speakers of European Spanish via https://prolific.co/
  - 37 women, 59 men
  - Mean age: 29 y, SD: 10.4

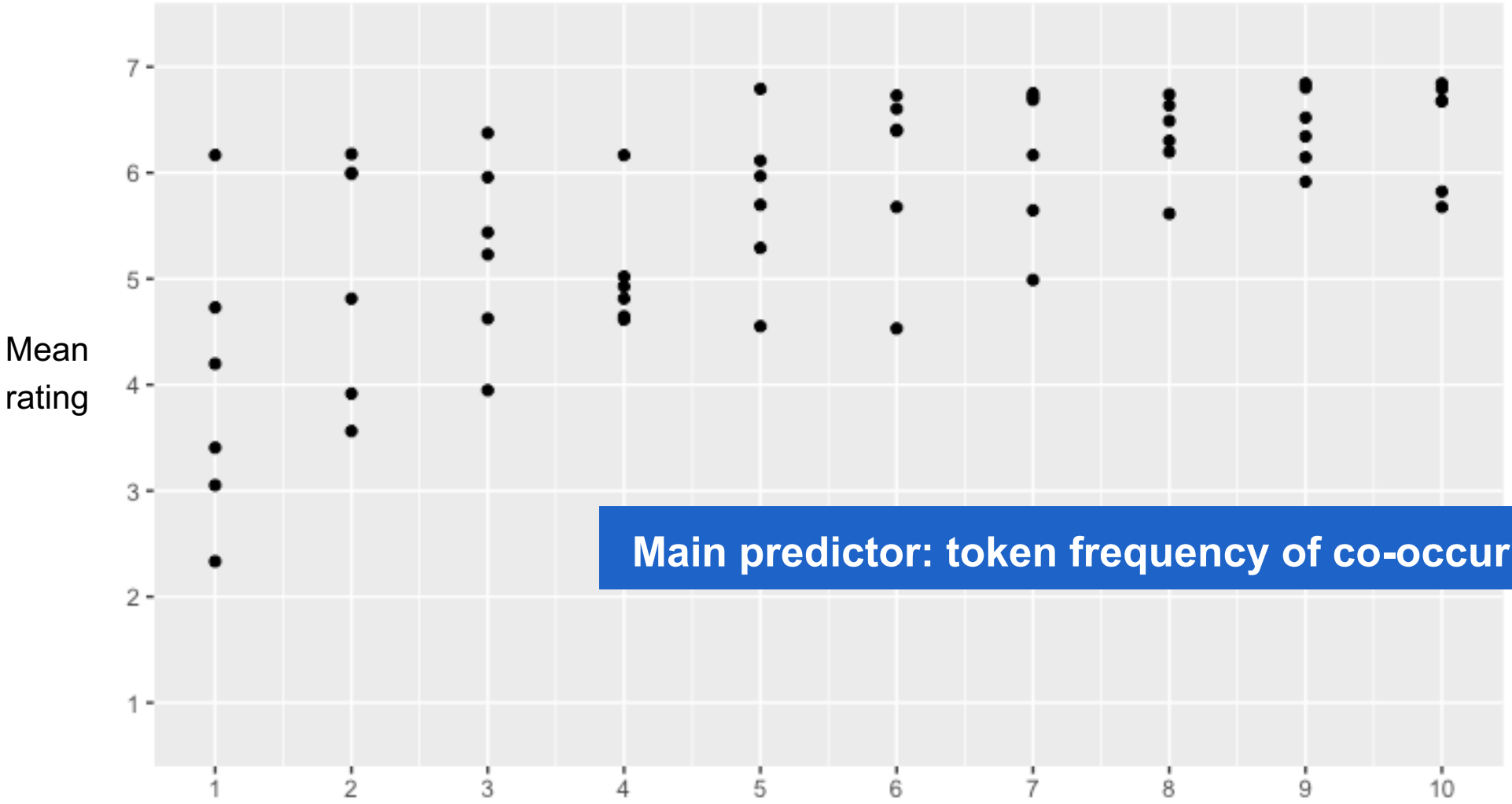
- 60 critical sentences + 140 filler sentences = 200 in total
  - Authentic (simplified) corpus sentences
  - 21 ‘yes/no’ comprehension questions

- 7-point Likert scale

- Sociobiographic questionnaire
- Big Five Inventory (BFI-2) for personality traits

GHENT UNIVERSITY

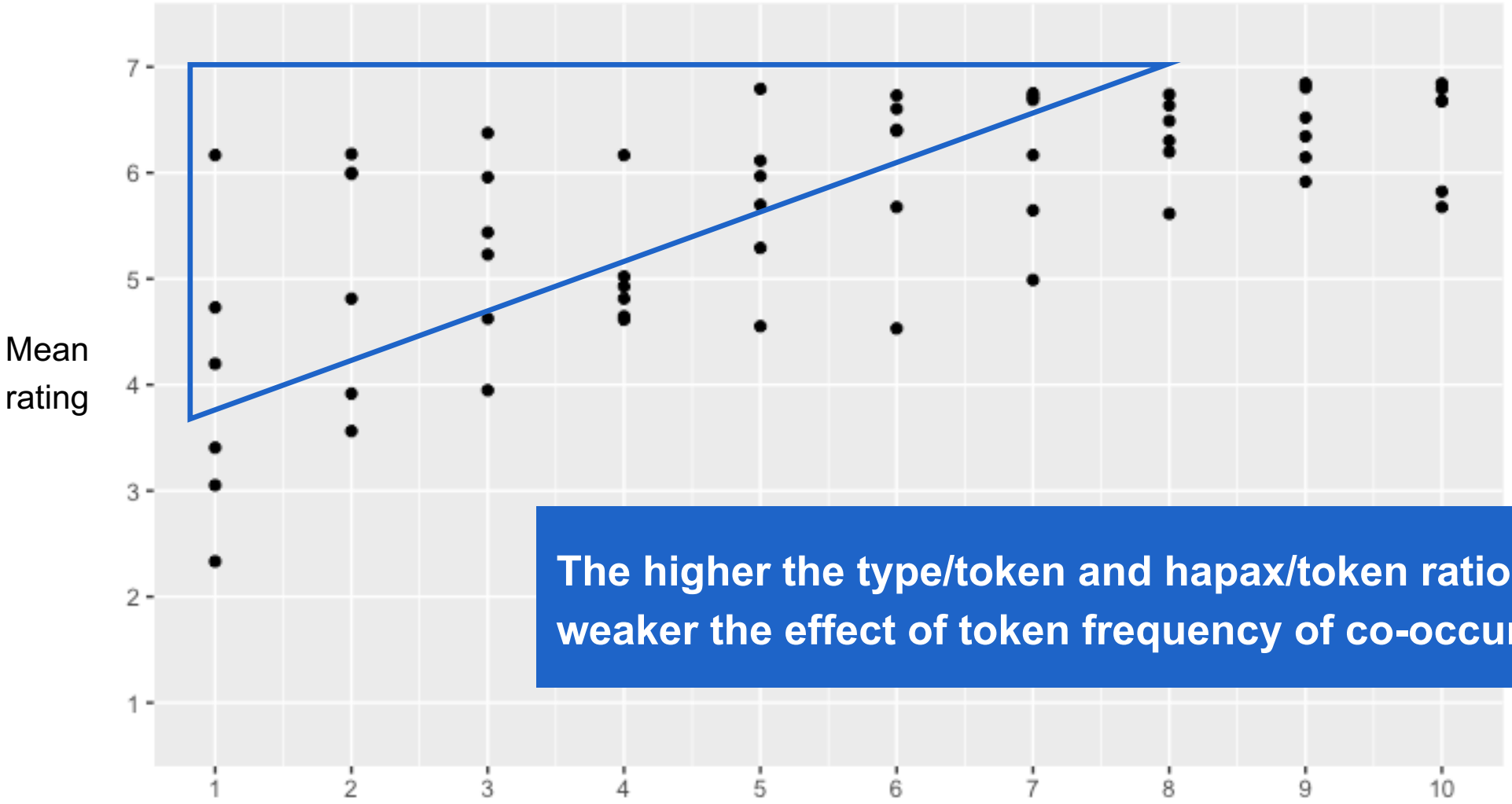
HIGHER FREQUENCY CORRESPONDS TO HIGHER RATINGS



Main predictor: token frequency of co-occurrence

GHENT UNIVERSITY

ACCEPTABILITY-FREQUENCY DISCREPANCY REFLECTS EXTENSIBILITY



The higher the type/token and hapax/token ratios, the weaker the effect of token frequency of co-occurrence

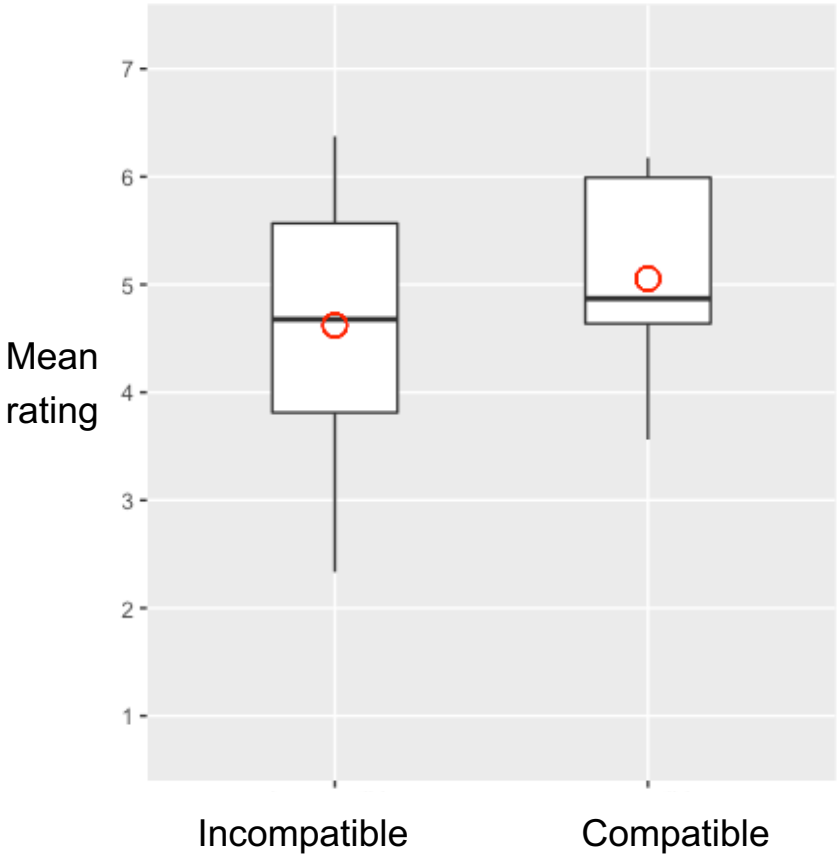
GHENT UNIVERSITY

SEMANTIC COMPATIBILITY ENHANCES EXTENSIBILITY TO NON-ATTESTED ITEMS

- Subset hapaxes and non-attested INFs
- Inchoatives with high type/token and hapax/token ratios received higher ratings
- **Semantically "compatible"*** combinations received higher ratings

*An infinitive is semantically compatible with the inchoative verb if it belongs to a semantic class (http://adesse.uvigo.es/) that is frequently attested with this inchoative in the dataset

Example compatible: class "physiology"
echarse a respirar *'throw oneself to breath'*
Example incompatible: class "phase"
echarse a desarrollar *'throw oneself to develop smth'*



GHENT UNIVERSITY

PARTICIPANT VARIABLES

Interaction with corpus measures of productivity:

- Multilingualism (more than one L2)
- Reading experience (books per year)
  - ➢ Conservative attitude towards less conventional extensions of the inchoative construction?

- No significant effects of personality traits (BFI-2)
  - ➢ Future research with more sensitive measures (e.g., eye-tracking during reading)

DISCUSSION & CONCLUSION

How is productivity attested in corpora related to productivity "at work" in the mind of language users?

- Corpus measures are predictive of acceptability ratings
  - When measures reflecting different aspects of productivity are taken into account all together

- Experimental data can provide insight on ***extensibility*** of inchoative verbs to low-frequent and non-attested infinitives
  - Inchoatives with higher type/token and hapax/token ratios get higher ratings
  - Low-frequent infinitives are more acceptable if they are semantically "compatible" with the inchoative

- Experimental data can provide insight on the role of individual differences regarding syntactic productivity

Mariia Baltais & Robert J. Hartsuiker
Department of Experimental Psychology

Corresponding author:
mariia.baltais@ugent.be

“Language Productivity @ Work” project:
https://www.languageproductivity.ugent.be/

Thank you!

REFERENCES

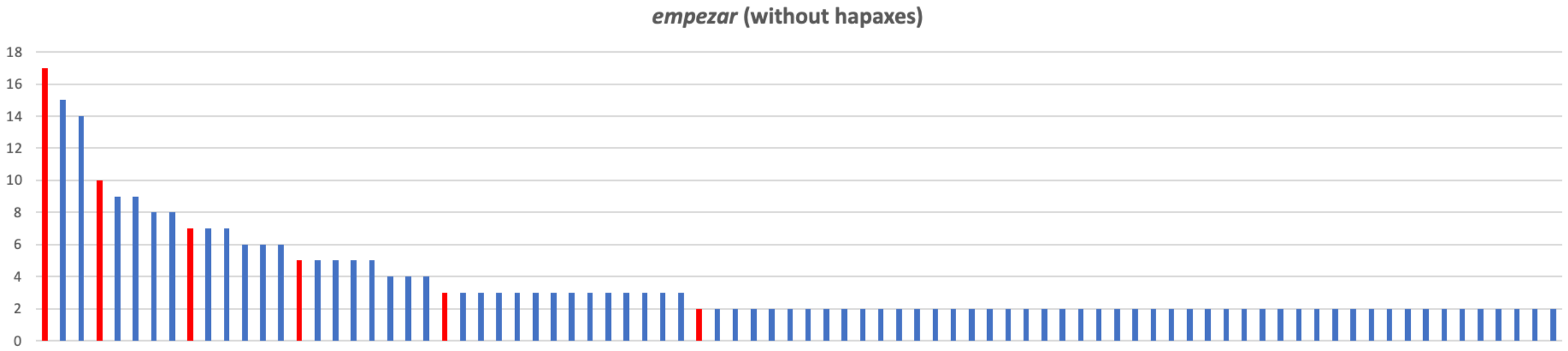
- Baayen, R. H. (2009). 43. Corpus linguistics in morphology: morphological productivity. *Corpus linguistics. An international handbook*, 900–919.
- Barðdal, J. (2008). *Productivity: Evidence from case and argument structure in Icelandic.* Amsterdam: John Benjamins Publishing Company.
- Divjak, D. (2017). The role of lexical frequency in the acceptability of syntactic variants: Evidence from *that*-clauses in Polish. *Cognitive Science*, 41, 35–82.
- Enghels, R., & Van Hulle, S. (2018). El desarrollo de perífrasis incoativas cuasi-sinónimas: entre construccionalización y lexicalización. *ESTUDIOS DE LINGUISTICA-UNIVERSIDAD DE ALICANTE-ELUA*, 32, 91–110.
- García Fernández, L. (2012). *Las perífrasis verbales en español.* Barcelona: Castalia.
- Goldberg, A. (2019). *Explain me this: Creativity, competition, and the partial productivity of constructions.* Princeton: Princeton University Press.
- Kempen, G., & Harbusch, K. (2008). Comparing linguistic judgments and corpus frequencies as windows on grammatical competence: A study of argument linearization in German clauses. In A. Steube (Ed.), *The discourse potential of underspecified structures*, 179–192. Berlin: Walter de Gruyter.
- Van Hulle, S., & Enghels, R. (in press). De Spaanse inchoatiefconstructie in beeld. Clusteranalyse als antwoord op het quasi-synonymie vraagstuk. *Handelingen – Koninklijke Zuid-Nederlandse maatschappij voor taal-en letterkunde en geschiedenis.*

ANNEX: CHOICE OF INCHOATIVES

- 6 inchoatives with different productivity characteristics
- Equal samples of 500 sentences each

Inchoative	Translation	Estimated token frequency in the subcorpus	Type/token ratio	Hapax/token ratio
empezar	'to begin, to start'	429583	0.56	0.39
ponerse	'to put oneself'	60728	0.36	0.24
lanzarse	'to throw oneself'	7476	0.43	0.27
meterse	'to put oneself'	1648	0.42	0.28
romper	'to break'	1976	0.06	0.03
echarse	'to throw oneself'	4242	0.03	0.01

ANNEX: CHOICE OF INFINITIVES



10 infinitives per inchoative

- 6 from high to intermediate to low token frequency of co-occurrence
- 2 hapaxes: frequent and infrequent semantic class*
- 2 non-attested: frequent and infrequent semantic class*

*http://adesse.uvigo.es/: creation, perception, displacement, physiology…

GHENT UNIVERSITY

ANNEX: ANALYSIS FULL DATA SET

Fixed effects	Estimate	SE	*t-value*
(Intercept)	5.64774	0.10989	51.395
Ranked token fq INCH x INF	**0.70366**	0.08626	**8.158***
Hapax/token ratio INCH	0.13266	0.11114	1.194
Estimated token fq INCH	**0.28833**	0.11300	**2.551***
Ranked token fq INCH x INF: Hapax/token ratio INCH	**-0.31554**	0.08483	**-3.720***
*AR ~ Ranked token fq INCH x INF * Hapax/token ratio + Estimated token fq INCH + (1 | item) + (1 + Ranked token fq INCH x INF * Hapax/token ratio + Estimated token fq INCH | participant)*			

- Main effect *ranked token frequency of co-occurrence*
- Main effect *estimated token fq of the inchoative in the corpus*
- Significant interaction *ranked token frequency of co-occurrence* and *hapax/token ratio of the inchoative*

GHENT UNIVERSITY

ANNEX: ANALYSIS HAPAXES & NON-ATTESTED

Fixed effects	Estimate	SE	*t-value*
(Intercept)	5.2037	0.2476	21.018
Hapax/token ratio INCH	**0.5909**	0.1950	**3.031***
Semantic class (infrequent)	**-0.7320**	0.3433	**-2.132***
Lemma fq INF	**-0.4823**	0.1683	**-2.866***
Nu. words in sentence	-0.3783	0.1962	-1.928
AR ~ Hapax/token ratio INCH + Semantic class + Lemma fq INF + Nu. words in sentence + (1 | item) + (1 + Hapax/token ratio INCH + Semantic class + Lemma fq INF + Nu. words in sentence | participant)			

- Main effect *hapax/token ratio of the inchoative*
- Main effect *semantic class* (incompatible vs. compatible combinations)
- Main effect *lemma frequency INF* (due to some infinitives with high lemma frequency and low ratings for the combination)

GHENT UNIVERSITY

ANNEX: ANALYSIS INDIVIDUAL VARIABLES

Fixed effects	Estimate	SE	*t-value*
(Intercept)	5.68461	0.14031	40.514
Token fq INCH x INF	**0.48396**	0.12726	**3.803***
Hapax/token ratio INCH	**0.46705**	0.12758	**3.661***
Many L2s	-0.35391	0.23977	-1.476
Books per year	0.01997	0.05581	0.358
Token fq INCH x INF: Many L2s	**0.23901**	0.08659	**2.760***
Many L2s: Hapax/token ratio INCH	**0.18722**	0.09117	**2.054***
Hapax/token ratio INCH: Books per year	**-0.05545**	0.02158	**-2.569***
*AR ~ Token fq INCH x INF * Many L2s + Hapax/token ratio * Many L2s + Hapax/token ratio * Books per year + (1 | item) + (1 + Token fq INCH x INF + Hapax/token ratio | participant)*			

GHENT UNIVERSITY