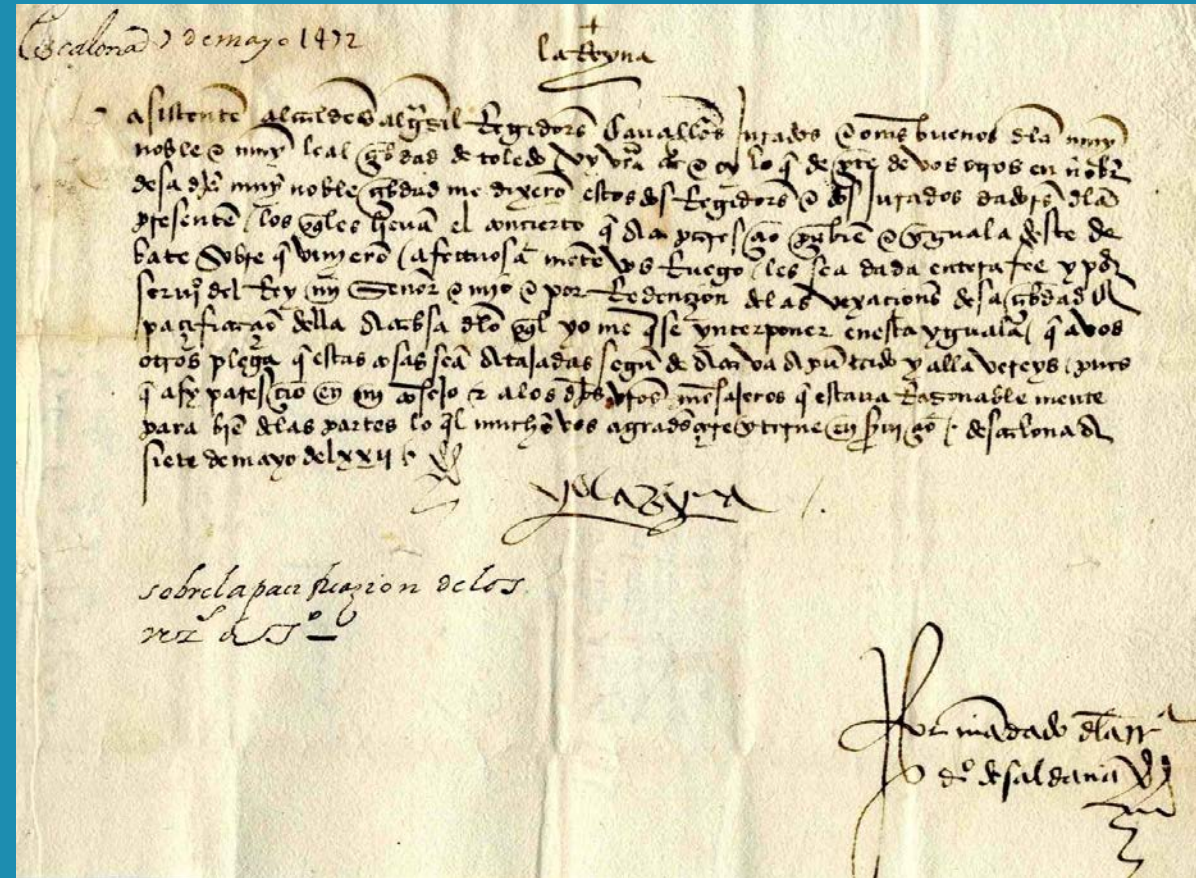# Socio-stylistic aspects of syntactic variation:
the case of Spanish asyndetic complementation (15th-18th c.)

Giulia Mazzola (KU Leuven)

Bert Cornillie (KU Leuven)

Malte Rosemeyer (Universität Freiburg / KU Leuven)

Stefano De Pascale (KU Leuven)

Codea-0377, 1522. (http://www.corpuscodea.es/)

53rd Annual Meeting of the

Societas Linguistica Europaea

26 August – 1 September 2020

(A)syndetic alternation in Spanish

(Pre-)classical Spanish (15th-18th century)

(1) […] *le* *dijo*$_{V1}$ *que* *ella* *sabía*$_{V2}$ *un* *medio* *para* *sacar*

DAT.3SG say.IND.PFV.3SG COMP she know.IND.IPFV.3SG INDF.M way to take.out.INF

a *su* *hermano* *de* *la* *prisión*

ACC POSS.M.SG brother of DET.F.SG prison

'[…] she said to him that she knew a way to take his brother out of prison' (Valencia, 1731, CODEA-2119)

(2) […] *don Francisco Gris* *dijo*$_{V1}$ *al* *testigo* Ø *tenía*$_{V2}$ *que* *delatar*

Mr. Francisco Gris say.IND.PFV.3SG to.DET.M.SG witness have.IND.IPFV.3SG that denounce.INF

al *reo* *por* *cosas* *de* *molinista*

ACC.DET.M.SG defendant for thing.PL of molinist

' […] don Francisco Gris said to the witness he had to denounce the defendant for molinist issues' (Cuenca, 1729, CODEA-2140)

KU LEUVEN

(A)syndetic alternation in Spanish

(Pre-)classical Spanish (15th-18th century)

SYNDETIC

(1) […]

le	dijo_{V1}	que	ella	sabía_{V2}	un	medio	para	sacar
DAT.3SG	say.IND.PFV.3SG	COMP	she	know.IND.IPFV.3SG	INDF.M	way	to	take.out.INF

a	su	hermano	de	la	prisión
ACC	POSS.M.SG	brother	of	DET.F.SG	prison

'[…] she said to him that she knew a way to take his brother out of prison' (Valencia, 1731, CODEA-2119)

ASYNDETIC

(2) […]

don Francisco Gris	dijo_{V1}	al	testigo	Ø	tenía_{V2}	que	delatar
Mr. Francisco Gris	say.IND.PFV.3SG	to.DET.M.SG	witness		have.IND.IPFV.3SG	that	denounce.INF

al	reo	por	cosas	de	molinista
ACC.DET.M.SG	defendant	for	thing.PL	of	molinist

' […] don Francisco Gris said to the witness he had to denounce the defendant for molinist issues' (Cuenca, 1729, CODEA-2140)

KU LEUVEN

Asyndetic construction in Spanish

- It started spreading in the 15th century (Pountain 2015)
- Heydays: 16th and 17th century
- It is still popular at the end of the 18th century

→ closely competing with the syndetic variant

(Mazzola, Cornillie & Rosemeyer *submitted*)

KU LEUVEN
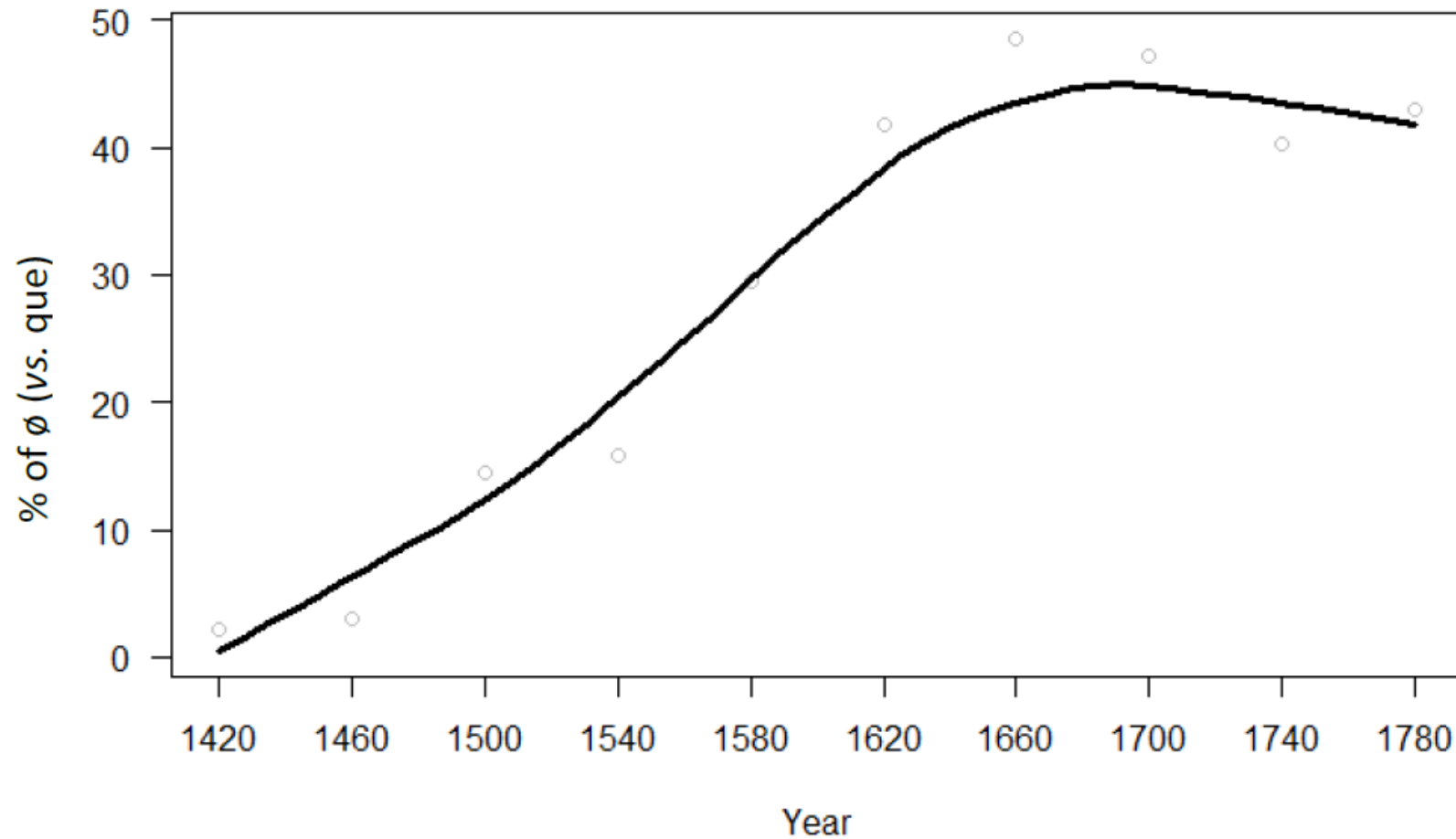
Asyndetic complements in Spanish

(Pre-)classical Spanish (15th-18th century)

- Distinctive syntactic feature of the period
  - Frequent with indicative and subjunctive complement clauses
  - With a wide range of predicates

Present-day Spanish

- Restricted to subjunctive-selecting verbs
- Stylistically marked → Formal genres

KU LEUVEN

Stylistic path of the construction

classical Spanish **?** >

> modern Spanish (formal)
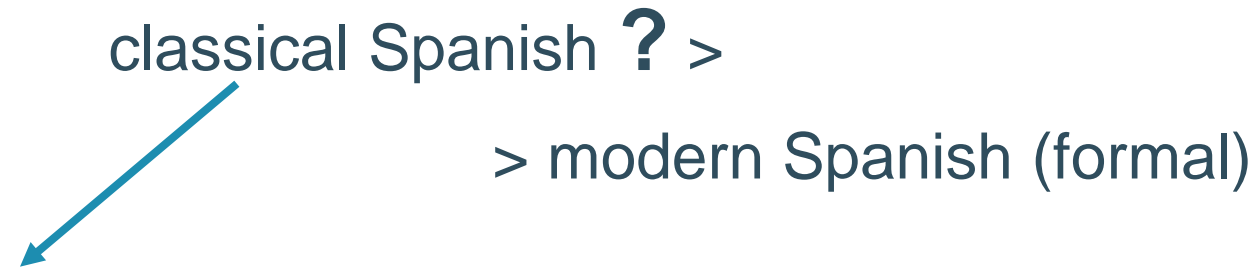
- Language of **distance** (literary, administrative language)

(i.a. Serradilla 1997; Girón Alconchel 2005; Espinosa Elorza 2014; Barraza Carbajal 2014)

- Language of **immediacy** (epistolary and informal language)

(Fernández Alcaide 2009; Blas Arroyo & Porcar Miralles 2016)

- **Idiolect** of some authors

(Keniston 1937; Octavio de Toledo y Huerta 2011)

KU LEUVEN

Document Type

From the corpus
CODEA+2015

- Ø is more frequent than *que* in Private Letters and Notes

- Ø is ~25% in administrative/bureaucratic documents

- Insufficient variable to explain the variation
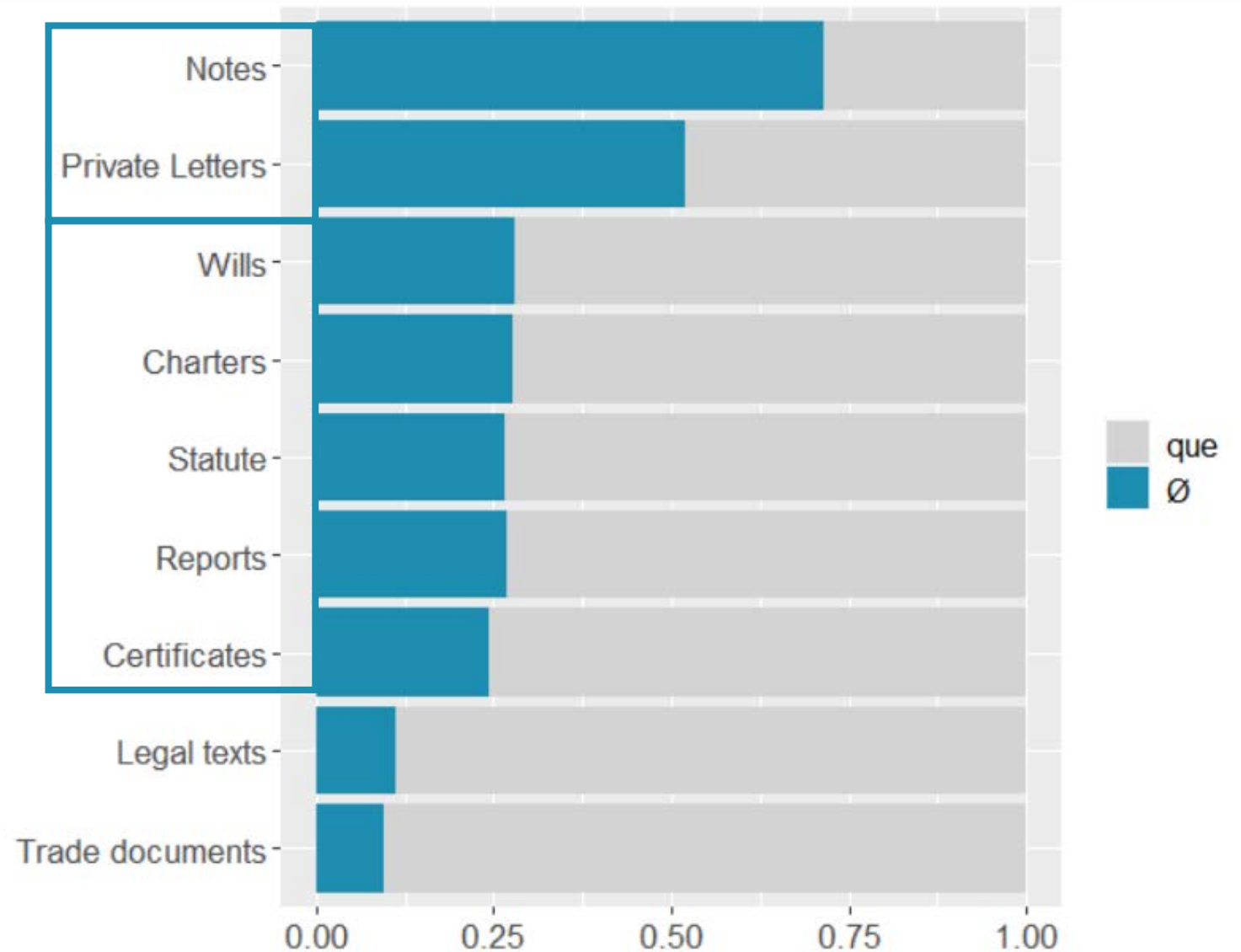
Mazzola Cornillie & Rosemeyer (2019)

KU LEUVEN

Research Questions

- Was the asyndetic construction a stylistically marked variant?
  - Is it bound to some specific textual genres or Discourse Traditions (distance vs. immediacy)? → the role of *Junktion* (Kabatek et al. 2010)
  - Is the alternation between *que* and Ø influenced by the intended audience of a text? (Audience Design, cf. Bell 2001)

- Which VARIABLES explain better the observed variation?
  - ❑ Main verb (language-internal predictor)
  - ❑ Relationship of social power/solidarity between writer and addressee (manual annotation)
  - ❑ **New documents classification**: clusters of documents based on word frequency (automatic annotation)

KU LEUVEN

Data & Methods

Corpus CODEA+2015

- Spanish documents written until 1799

- High level of philological accuracy

- Cornillie & Rosemeyer (2018): extraction and manual selection of the variable context:

main verb/construction + que/ø + finite complement clause

Today's study includes

Tot. 3943 tokens

- **Mixed effects logistic regression**

KU LEUVEN

Variables

VERB CLASS (verbs with more than 30 occurrences in the dataset)

Verb class	Verbs	Meanings	Syndetic tokens	Asyndetic tokens
Communication	*Decir*	'to say'	832	138
Volition	*Querer*	'to want'	245	39
Order	*Mandar, Ordenar*	'to order'	996	212
Request	*Pedir, Rogar, Suplicar*	'to ask, to beg, to plead'	375	576
Others	*Entender*	'to understand'	447	83
	Parecer	'to seem'		
	Saber	'to know'		
	Creer	'to believe'		
	Ver	'to see'		

KU LEUVEN

Variables

ADDRESSEE + STATUS: - is the text addressed to someone? - relationship between issuer and addressee

Superior: the addressee has a higher social status than the issuer

(a citizen to an institution, a nun to an aristocrat…)

Equal: addressee and issuer have the same status

(family members…)

Inferior: the addressee has a lower social status than the issuer

(the king to the citizens, the bishop to some monks…)

No addressee: document not addressed to someone in particular

KU LEUVEN

Variables

DOCUMENT CLUSTERS – Bottom-up classification of the documents by words and word frequencies

Vector Space Model (see De Pascale 2019)

- main idea =
  - Characterize every text ('document') by means of the frequency of the words ('terms') used in that text
  - **Discourse Traditions** are defined by the relative frequency and distribution of lexical, grammatical and formulaic elements (cf. Kabatek et al. 2010: 7)
  - Texts that share similar words with similar frequency profiles belong to the same Discourse Traditions
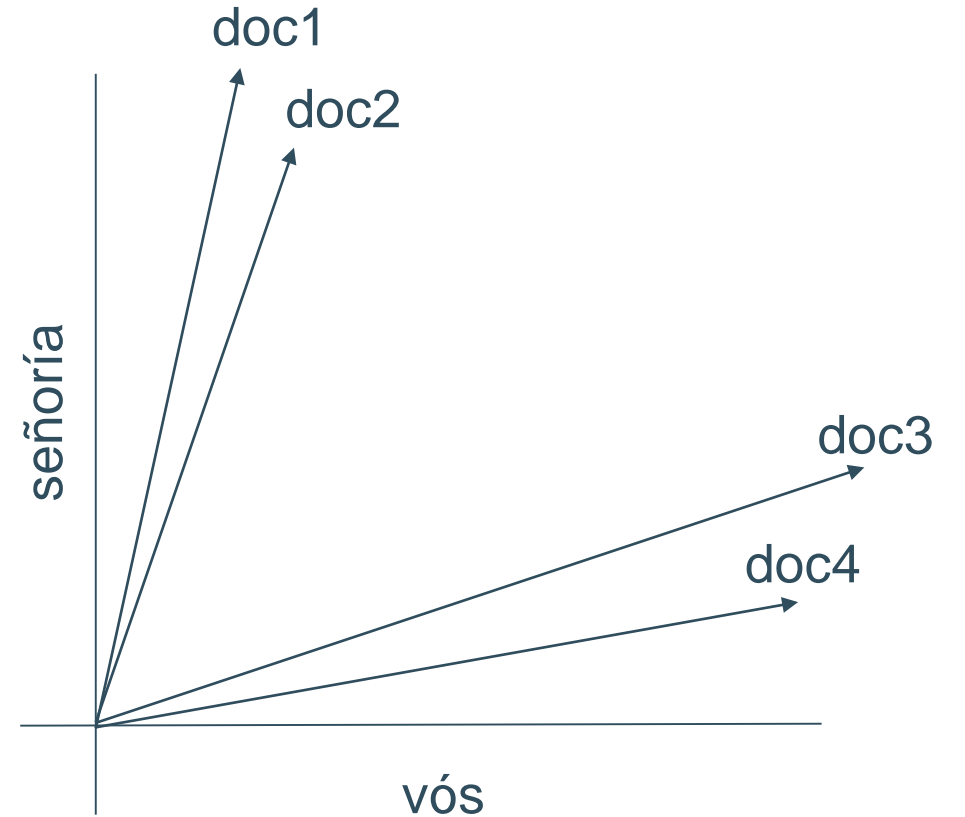
KU LEUVEN

Variables

Vector space model

Document-by-term matrix

each document is represented as a vector of (weighted) counts of the (most significant) terms

in that document

|	señoría	merced	señor	convento	vós	ciudad
doc1	284	79	125	134	0	3
doc2	163	178	114	52	0	0
doc3	8	4	5	7	123	184
doc4	0	2	0	1	360	354

Document vector

Numeric vectors (the arrows, i.e. the documents) are compared for similarity

KU LEUVEN

Variables

DOCUMENT CLUSTERS

two-dimensional visualization of document-based vector space model (technique: t-sne)

- 1 point in scatterplot = 1 text from corpus

- points close to each other = text with similar words and words' frequencies

- Each document clustered in 5 groups based on their vector similarity (technique: hierarchical cluster analysis)
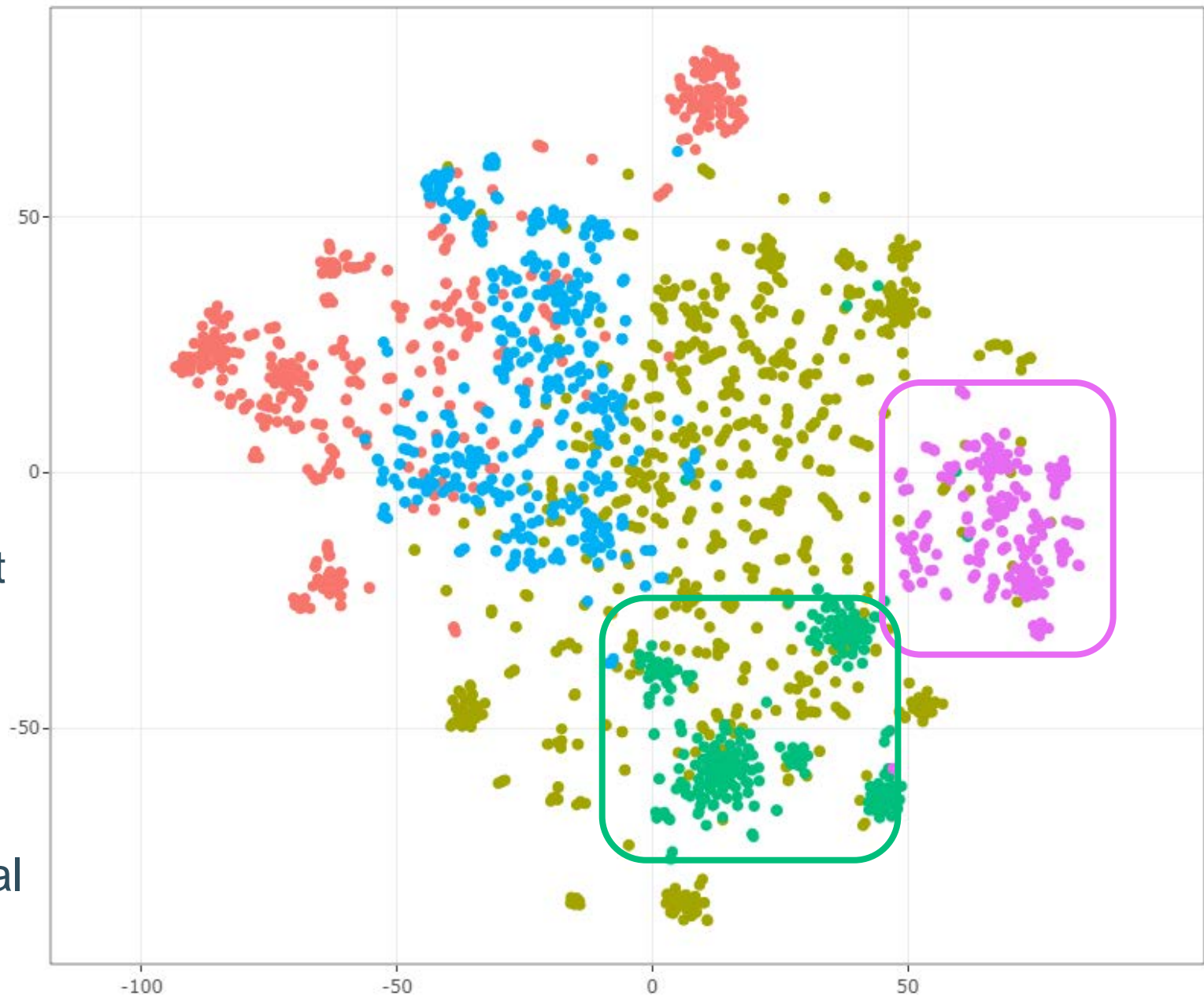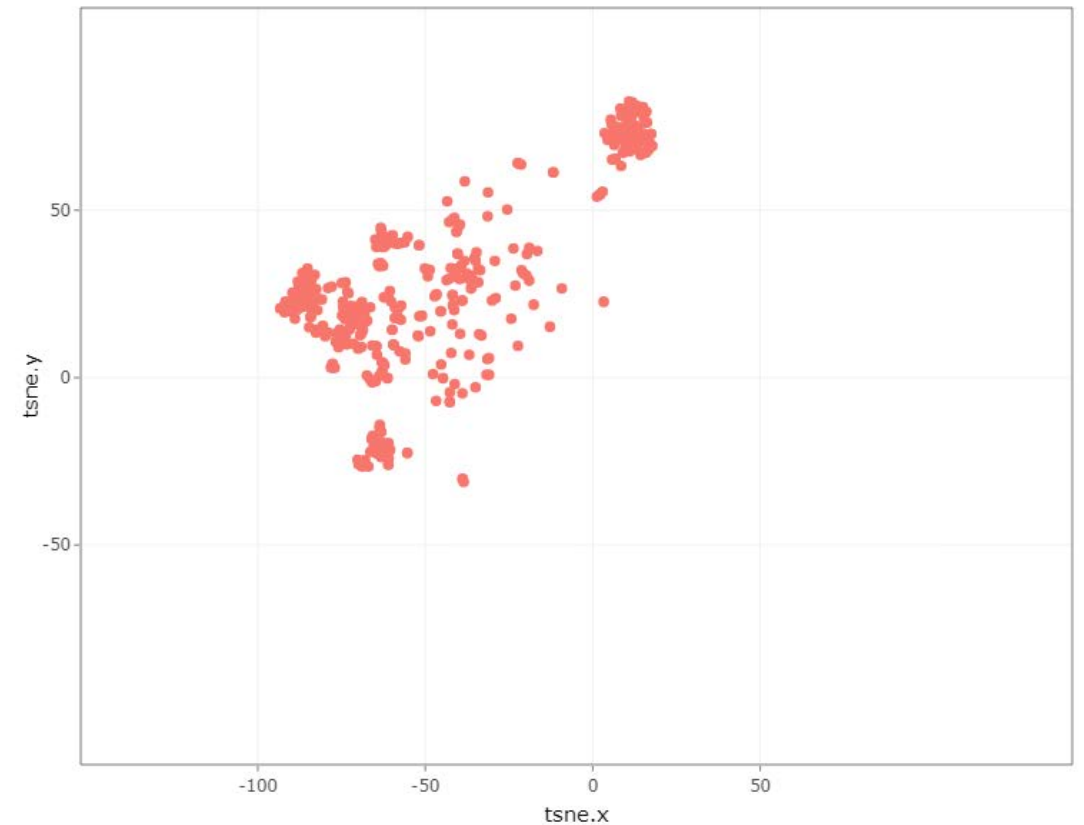
KU LEUVEN

Variables

- **DOCUMENT CLUSTERS**

Cluster 1

Typical words: *et, testist, confirmat, cum, ciudat, testament, dito*

Cluster	Label	N. Tokens	N. Documents
1	Latin and Oriental	175	48

KU LEUVEN

Variables

- **DOCUMENT CLUSTERS**

Cluster 2

Typical words: *reales, villa, ciudad, señorías*

Cluster	Label	N. Tokens	N. Documents
1	Latin and Oriental	175	48
2	Bureaucratic	2048	508

KU LEUVEN
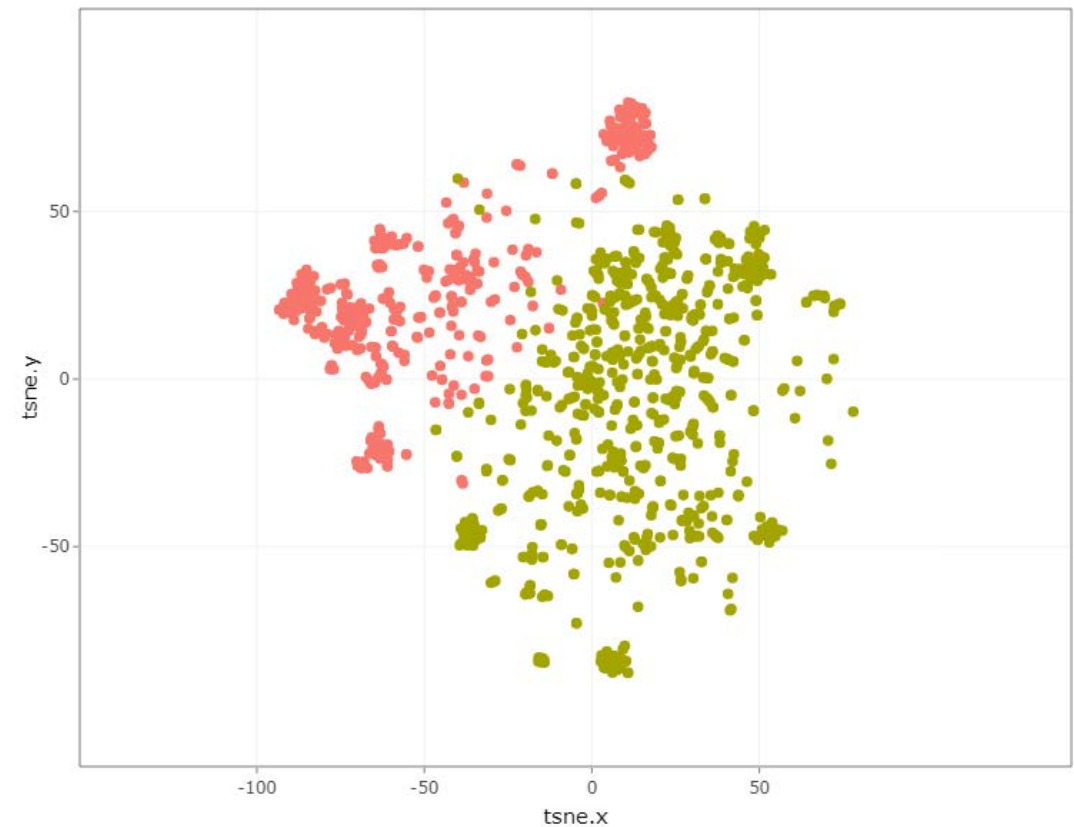
Variables

- **DOCUMENT CLUSTERS**

Cluster 3

Cluster	Label	N. Tokens	N. Documents
1	Latin and Oriental	175	48
2	Bureaucratic	2048	508
3	Letters / Polite Requests	480	187

Typical words: *vuestra, señoría, merced, excelencia, magestad, ilustrísima*
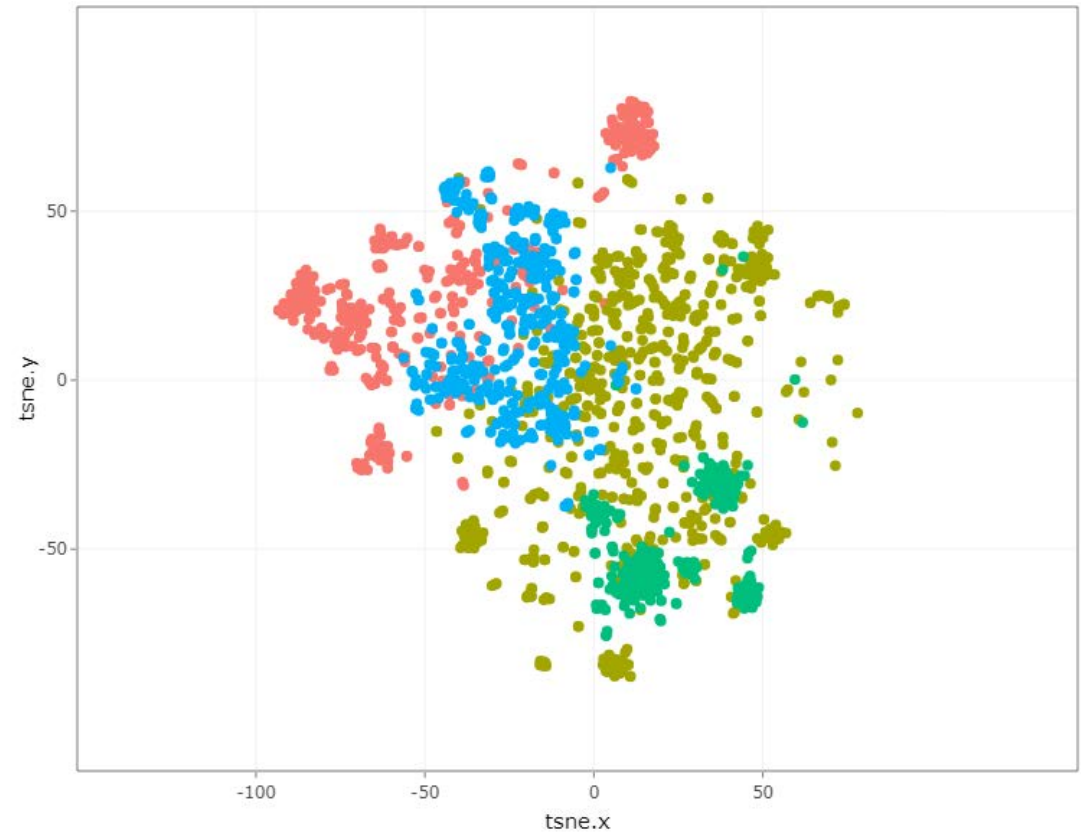
KU LEUVEN
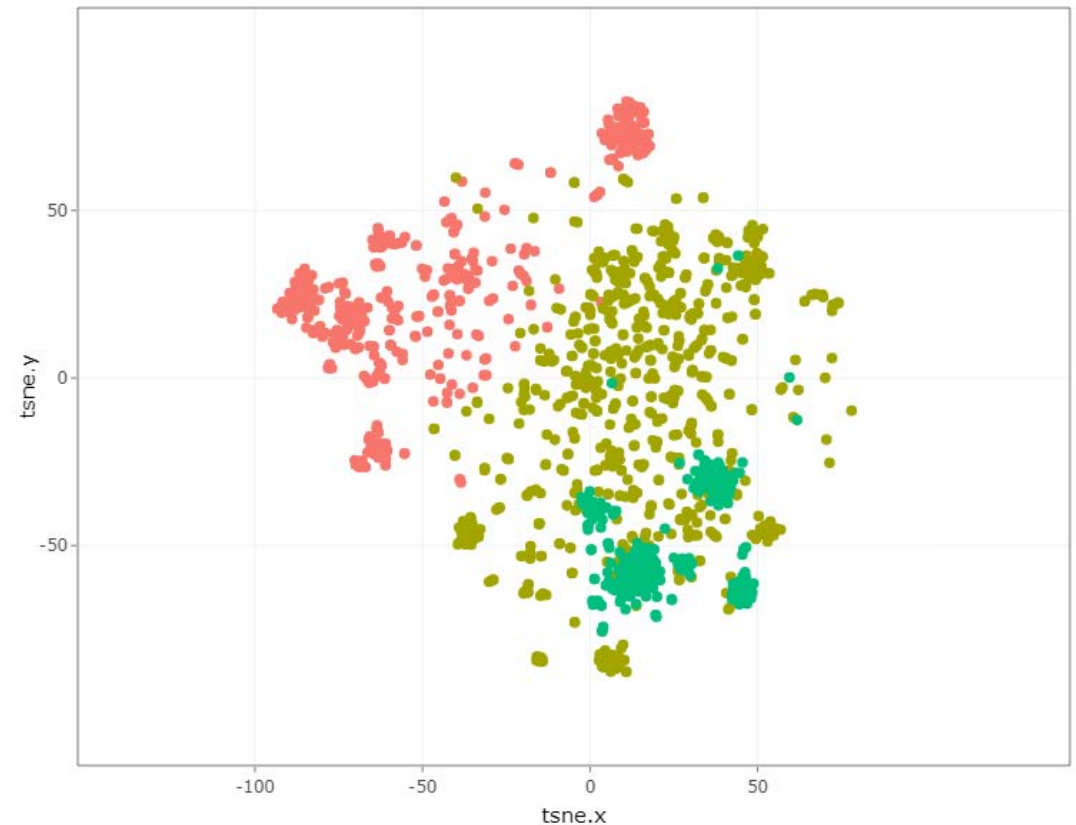
Variables

- **DOCUMENT CLUSTERS**

Cluster 4

Cluster	Label	N. Tokens	N. Documents
1	Latin and Oriental	175	48
2	Bureaucratic	2048	508
3	Letters / Polite Requests	480	187
4	Laws	1197	214

Typical words: *vós, nós, monasterio, ciudad, iglesia, cabildo*

KU LEUVEN

Variables

- **DOCUMENT CLUSTERS**

Cluster 5

Cluster	Label	N. Tokens	N. Documents
1	Latin and Oriental	175	48
2	Bureaucratic	2048	508
3	Letters / Polite Requests	480	187
4	Laws	1197	214
5	Notes	43	35

Typical words: *niño/a, cristiano/a, llamase, agua, criatura*

KU LEUVEN
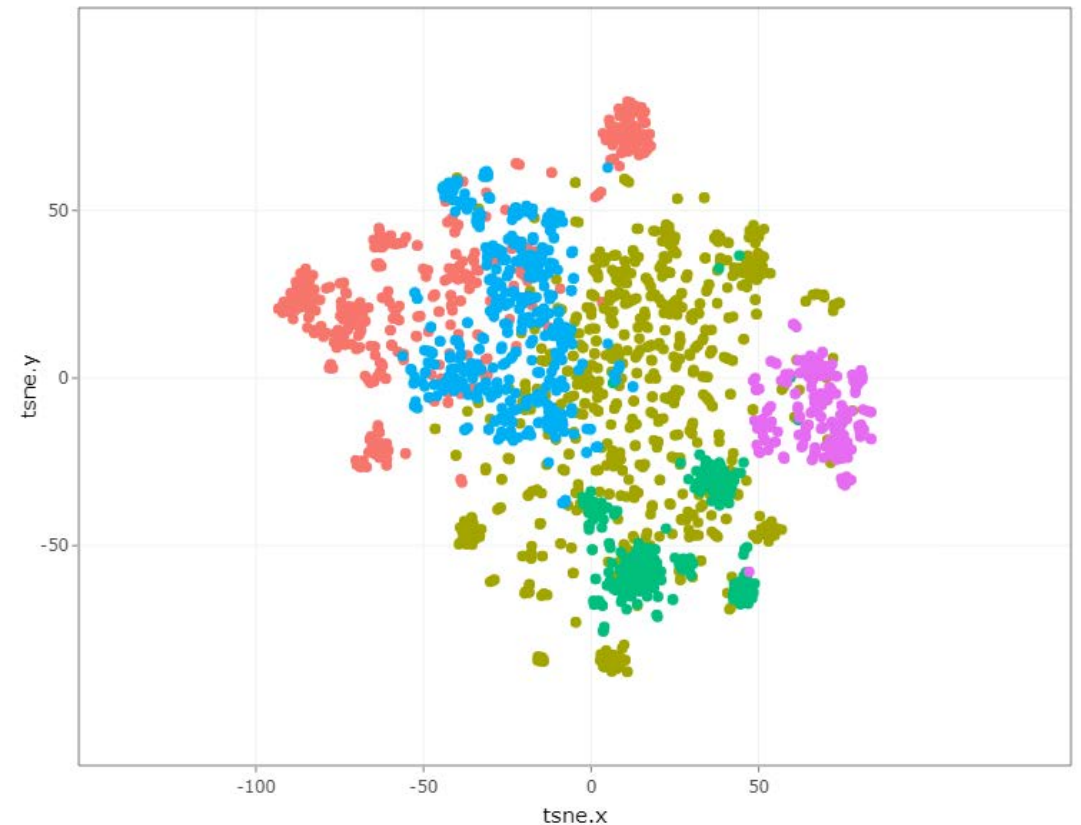
Mixed effects logistic regression

R package: lme4
(Bates et al. 2015)

Predictor variables

Addressee + status

Verb Class
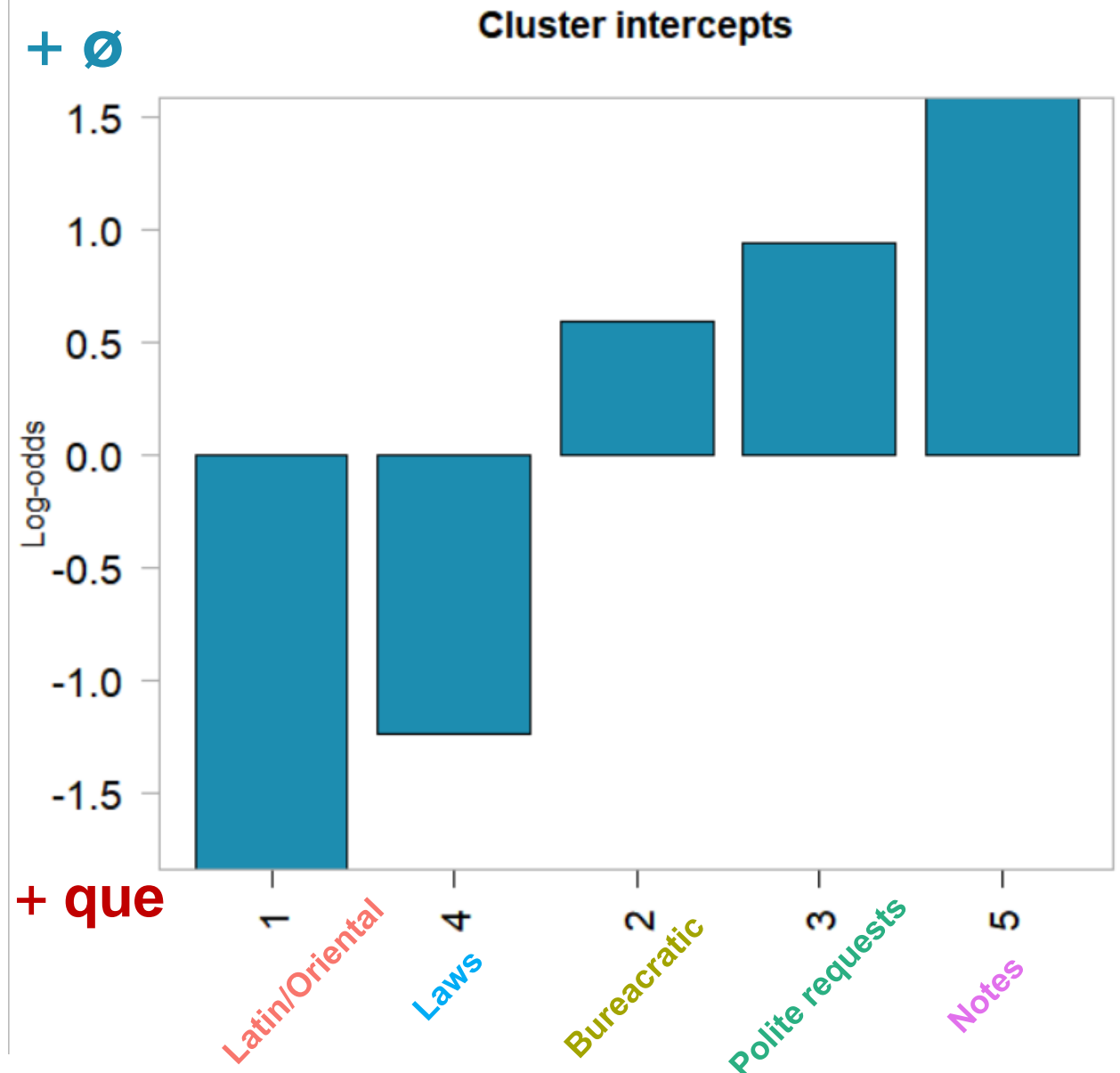
Random effect:

Document clusters

que/ø

Dependent variable
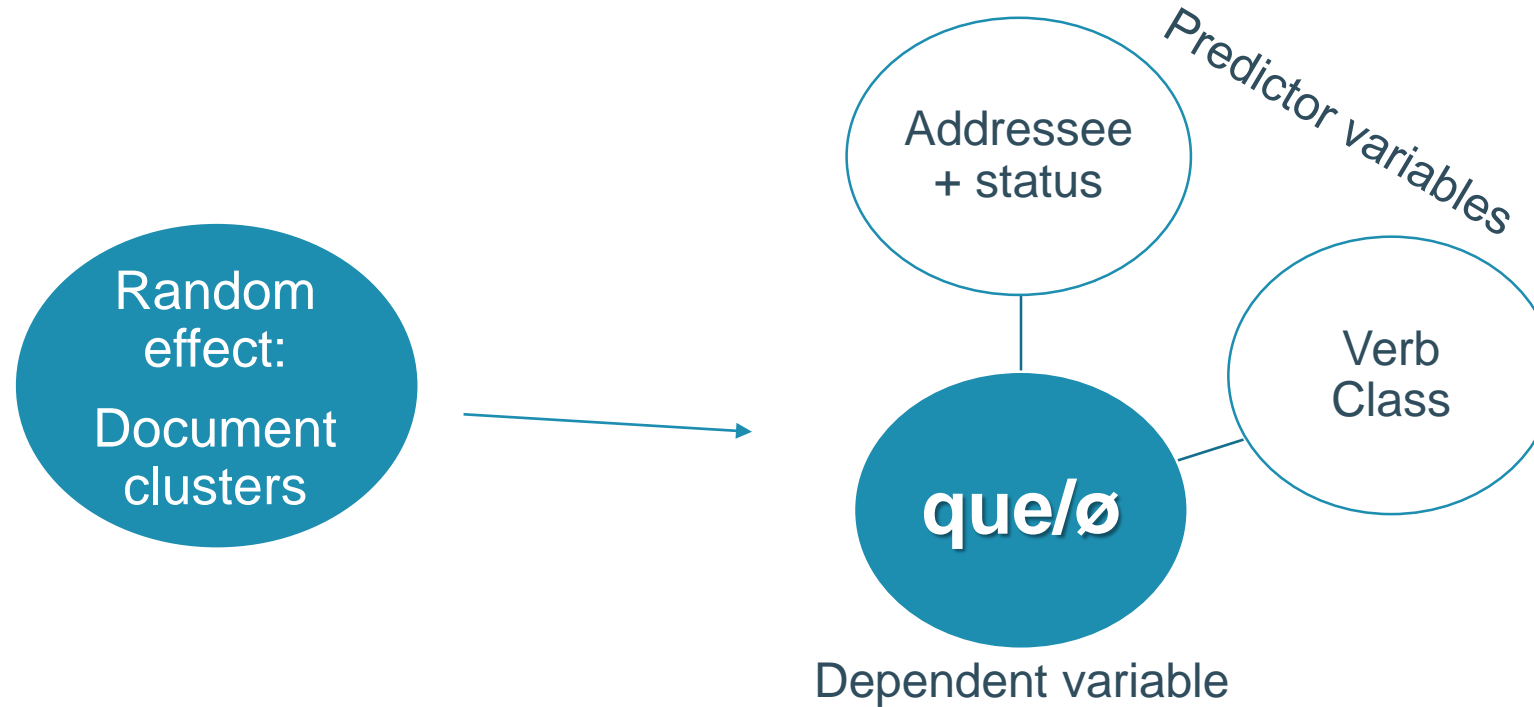
- A sample of 5 document clusters out of an indefinite number of topics and possible clusters

AIC= 3420.8
C index= 0.81

KU LEUVEN

Document Clusters

- Asyndetic complements are mostly predicted by cluster **3**-Polite requests and **5**-Notes

- To a minor extent, asyndetic complementation is favored by group **2**-Bureaucratic

- Group **1**-Latin/Oriental and **4**-Laws favor the presence of *que* in the complement clause

Cluster intercepts

+ ø

+ que

Y-axis: Log-odds (1.5, 1.0, 0.5, 0.0, -0.5, -1.0, -1.5)

X-axis: 1 Latin/Oriental, 4 Laws, 2 Bureaucratic, 3 Polite requests, 5 Notes

KU LEUVEN

Addressee + Status

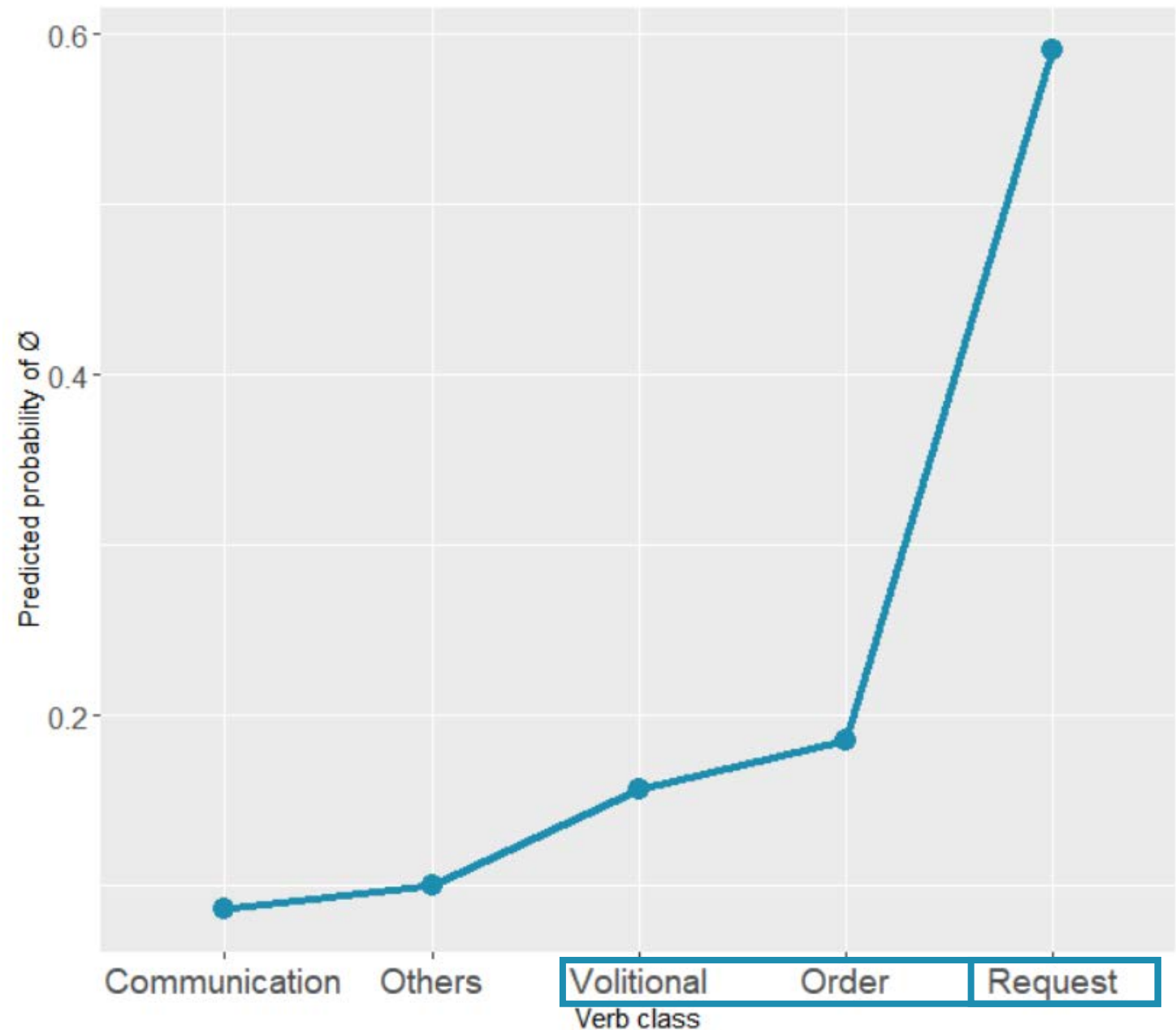- Asyndetic complementation is
  - mostly favored when addressing someone with a higher social status
  - Avoided when addressing subordinated people

- Difference between *addressee vs. no addressee* = statistically significant
  χ^2=81.572, df = 1, p-value < 2.2e-16

KU LEUVEN
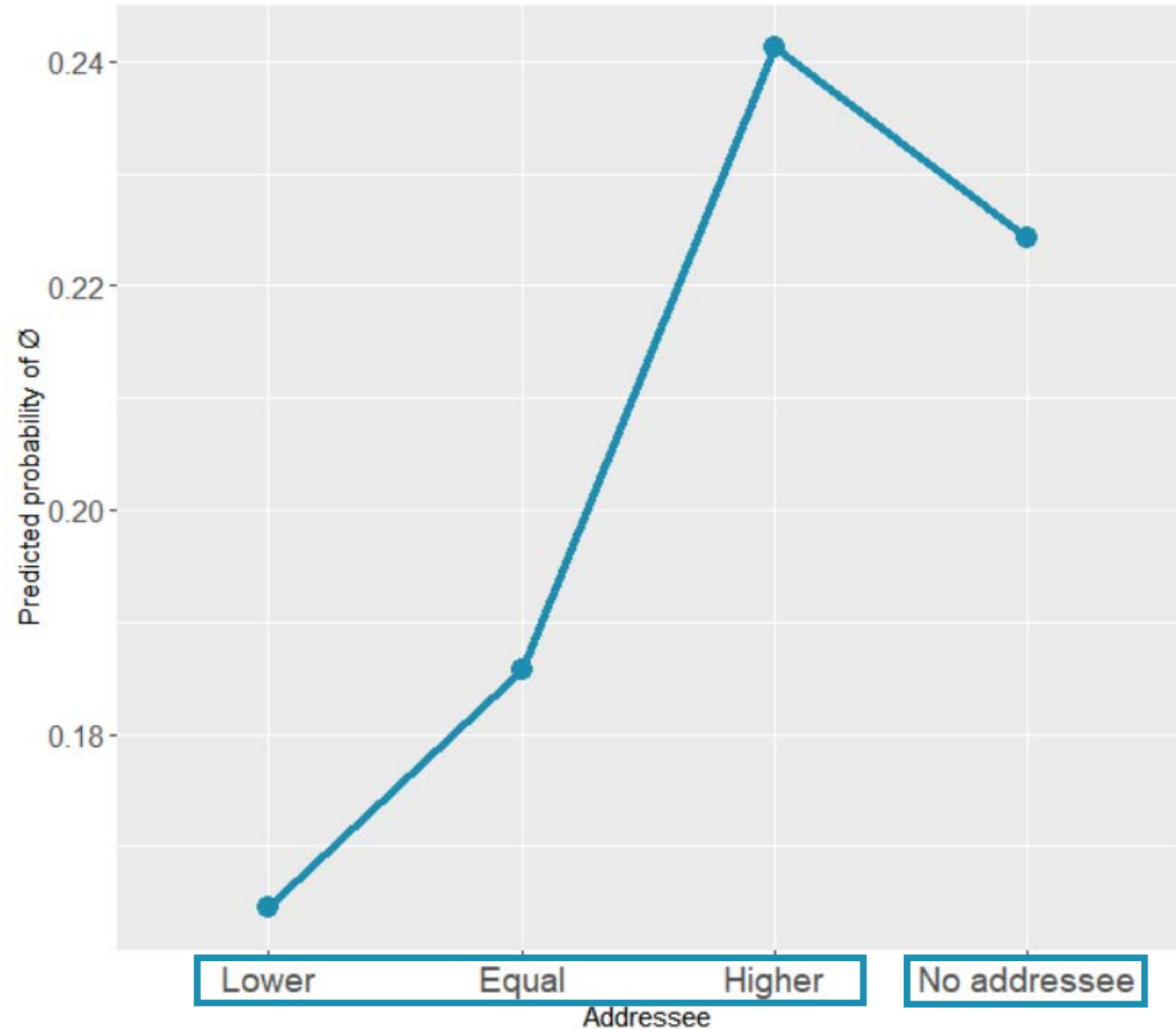
Verb Class

- Best predictor: verbs of request (neutral verb *pedir* 'to ask', deferential verbs *rogar* and *suplicar* 'to beg')

- Surprising difference between **request** and **order/volitional verbs**
→ very often they share semantic and syntactic features

KU LEUVEN

Summary of the results

Predicted probabilities of asyndetic complementation by context:

CODEA-1757
→ The King to the deputy mayor of Alcalá de Henares: <u>order to inferior addressee</u>, cluster **4-Laws**

5) E ***mandamos***$_{V1}$ *a vós las nuestras justicias* **que** ***tengáis***$_{V2}$ *especial cuidado de saber si se cumple esto ansí, y executéis las dichas penas en los que no la cumplieren.*

'And we order to you, our authorities, to be especially careful in monitoring whether this [rule] is respected, and to execute the mentioned punishments to those who do not obey'
(CODEA-1757, Madrid, 1535, Textos Legislativos)

Low probability of asyndetic complementation: 5%

KU LEUVEN

Summary of the results

Predicted probabilities of asyndetic complementation by context:

CODEA-1705
→ A man to the director of an orphanage: <u>polite request to superior addressee</u>, cluster **5-Notes**

6) **Suplico**$_{V1}$ *a vuestra merced* Ø **mande**$_{V2}$ *recibir essa criatura.*

'I beg Your Mercy to order to welcome this child' (CODEA-1705, Madrid, 1596, Notas y Breves)

Highest probability of asyndetic complementation: 90%

KU LEUVEN

Discussion: Document clusters

- Cluster **3-polite requests** and **5-notes** mostly favor ø

- They represent Discourse Traditions that
  - are characterized by a high share of polite forms of address
  - contain a lot of requests
  - are addressed to superior addressees

KU LEUVEN
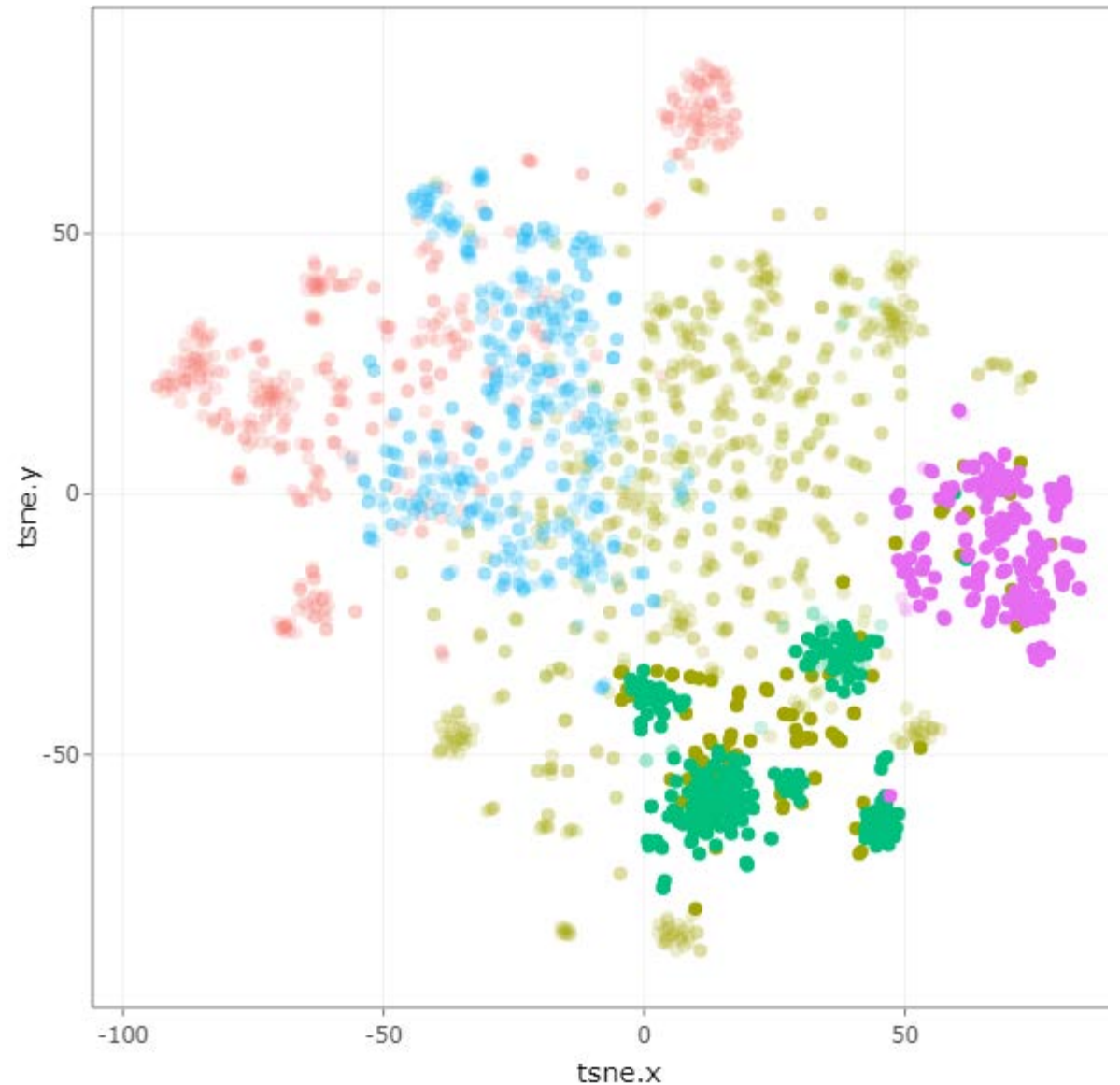
Discussion: Addressee's status

- Asyndetic complementation appropriate especially to address people with a higher social status→ elegant feature

- Second version of *Camino de Perfección* by Teresa of Ávila (Octavio de Toledo y Huerta 2011)
    - She modifies the second version to address it to a less intimate, more prestigious public
    - Among other stylistic corrections, she eliminates a lot of *que* linkers

    "the omission, at least with some main verbs and with contiguous predicates, was considered as an elegant feature" (2011: 13, translation mine)

KU LEUVEN

Discussion: Verb class

- Mazzola, Cornillie & Rosemeyer (*submitted*) on language-internal factors

asyndetic construction favored by the typical features of a **manipulation schema** → syntactic configuration typical of verbs of requests AND command (like *pedir, mandar, ordenar, suplicar etc…*)

- Verbs of request like *suplicar, rogar* ('to beg')→ mitigation of a directive speech act (cf. Iglesias Recuero 2016): speakers in a position of inferiority compared to the addressee

- Mitigated directive speech acts → suitable to address people with a higher social status and perform "polite" requests

- New pragmatic dimension: not ANY kind of manipulation verbs, but especially those which express mitigated requests

KU LEUVEN

Conclusions

- Semantic factor of verb class is a powerful predictor of the alternation between *que* and *zero*

- Asyndetic complementation is a stylistically marked variant
  - ✓ marker of the language of communicative distance (formal)
  - ✓ typical of Discourse Traditions used to perform polite requests addressed to prestigious addresses

KU LEUVEN

Conclusions

- Vector-based model → suitable to reveal Discourse Traditions in large corpora?

- Multidimensional categorization: Word frequency and cosine similarities (document clusters), types of verbs, textual genre

  - The variable of document clusters has proved to be an excellent predictor of the alternation between syndetic and asyndetic complement clauses.

→Future research: explore document clusters and their syntactic markers diachronically

KU LEUVEN

- Barraza Carbajal, Georgina. 2014. Oraciones subordinadas sustantivas de objeto directo. In Concepción Company Company (ed.), *Sintaxis histórica de la lengua española* (Lengua y estudios literarios), vol. 3, 2971–3106. Mexico: UNAM: Fondo de cultura económica.

- Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software* 67(1). https://doi.org/10.18637/jss.v067.i01. http://www.jstatsoft.org/v67/i01/ (22 January, 2020).

- Bell, Allan. 2001. Back in style: Reworking audience design. In Penelope Eckert & John R. Rickford (eds.), *Style and sociolinguistic variation*, 139–169. Cambridge: Cambridge University Press.

- Blas Arroyo, José Luis & Margarita Porcar Miralles. 2016. Un marcador sociolingüístico en la sintaxis del Siglo de Oro: patrones de variación y cambio lingüístico en completivas dependientes de predicados doxásticos. *Revista internacional de lingüística iberoamericana* (28). 157–185.

- Cornillie, Bert & Malte Rosemeyer. 2018. La elisión de la conjunción completiva que en V1 V2 entre 1400 y 1800. Presented at the XI Congreso Internacional de Historia de la Lengua Española, Pontificia Universidad Católica del Perú, Lima.

- De Pascale, S. 2019. Token-based vector space models as semantic control in lexical lectometry. *Thesis dissertation.*

- Espinosa Elorza, Rosa María. 2014. Coordinación y Subordinación. Panorama general, relaciones diacrónicas básicas, y nexos. In Concepción Company Company (ed.), *Sintaxis histórica de la lengua española* (Lengua y estudios literarios), vol. 3, 2231–2337. Mexico: UNAM: Fondo de cultura económica.

- Fernández Alcaide, Marta. 2009. *Cartas de particulares en Indias del siglo XVI: edición y estudio discursivo* (Textos y Documentos Españoles y Americanos v. 6). Madrid : Frankfurt am Main: Iberoamericana ; Vervuert.

- Girón Alconchel, José Luis. 2005. Cambios gramaticales en los Siglos de Oro. In Rafael Cano (ed.), *Historia de la lengua española* (Ariel lingüística), 859–893. 2a ed. Barcelona: Ariel.

- GITHE. 2015. Codea+ 2015. *Corpus de documentos españoles anteriores a 1800*. http://www.corpuscodea.es/ (10 January, 2019).

- Iglesias Recuero, Silvia. 2016. Otra cara de la pragmática histórica: la historia de los actos de habla en español: peticiones y órdenes en las "novelas ejemplares" de Miguel de Cervantes. In Araceli López Serena, Antonio Narbona Jiménez, Santiago del Rey Quesada & Rafael Cano Aguilar (eds.), *El español a través del tiempo: estudios ofrecidos a Rafael Cano Aguilar. Volumen 2:* (Colección lingüística 50), vol. 2, 971–994. Sevilla: Editorial Universidad de Sevilla.

Faculty of Arts, Department of Linguistics

KU LEUVEN

- Kabatek, Johannes, Philipp Obrist & Valentina Vincis. 2010. Clause linkage techniques as a symptom of discourse traditions: Methodological issues and evidence from Romance languages. In Heidrun Dorgeloh & Anja Wanner (eds.), *Syntactic Variation and Genre* (Topics in English Linguistics), vol. 70, 247–276. Berlin, New York: De Gruyter Mouton. Keniston, Hayward M. 1937. *Spanish Syntax List.* new York: Holt.

- Mazzola, Giulia, Malte Rosemeyer & Bert Cornillie. 2019. La complementación asindética entre el siglo XV y el siglo XVIII: Variación socioestilística. Presented at the V Congreso Internacional de Corpus Diacrónicos en Lenguas Iberorrománicas, Université libre de Bruxelles / Ghent University.

- Mazzola, Giulia, Malte Rosemeyer & Bert Cornillie. 2020. Asyndetic complementation and referential integration in Spanish: a probabilistic diachronic grammar account. Manuscript submitted for peer-review.

- Octavio de Toledo y Huerta, Alvaro. 2011. Santa Teresa y la mano visible. Sobre las variantes sintácticas del Camino de perfección. In Mónica Castillo Lluch & Lola Pons Rodríguez (eds.), *Así se van las lenguas variando: nuevas tendencias en la investigacíon del cambio lingüistico en español* (Fondo Hispánico de Lingüística y Filología v. 5), 241–304. Bern ; New York: Peter Lang.

- Pountain, Christopher. 2015. Que-deletion: the rise and fall of a syntactic fashion. In *Estudios dedicados a David Mackenzie*, 17. Universidad de Santiago de Compostela.

- Serradilla Castaño, Ana. 2010. La subordinación completiva en español antiguo: continuidad y ruptura de los modelos latinos. In Mónica Castillo Lluch & Marta López Izquierdo (eds.), *Modelos latinos en la Castilla medieval* (Medievalia Hispánica 14), 145–158. Madrid : Frankfurt am Main: Iberoamericana ; Vervuert.

Faculty of Arts, Department of Linguistics

KU LEUVEN

Thank you!

Live Discussion:
Thursday 27th August 2020, CET 9.20-9.30

Find extra contents in the pdf presentation on our OSF Project page

KU LEUVEN