WHAT CAN YOU DO WITH 200 STUDIES OF LANGUAGE CHANGE?

Terttu Nevalainen, Tanja Säily, Turo Vartiainen and Joonas Kesäniemi

with Jefrey Lijffijt, Agata Dominowska and Aatu Liimatta

REASSESSING LANGUAGE CHANGE PROJECT

- Lack of access to data & previous corpus-based research → research on real-time change largely not cumulative
- **Language Change Database**: Comparative, real-time baseline data for modelling change in progress
  - Open-access **linked data** web application, easy to **edit & query**
  - Currently in beta stage, c. 200 entries
  - Could be used to carry out **meta-analyses** of a large number of linguistic changes

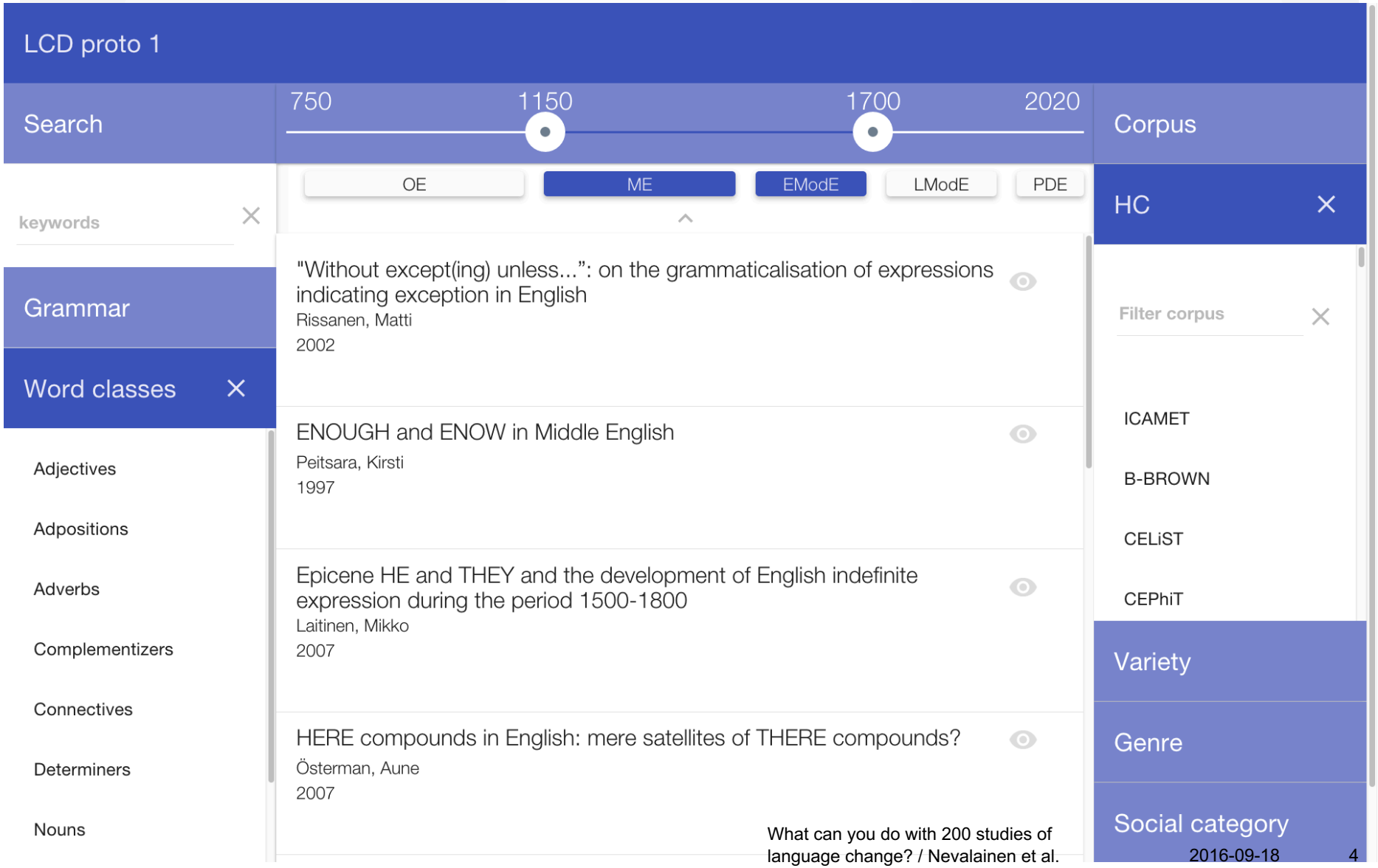
WHAT THE DATABASE WILL CONTAIN

- Metadata, e.g.
  - Bibliographical details
  - Linguistic items & topics studied
  - Corpora & time periods studied

- **Data**
  - Tables
  - Summary of results

More information:
helsinki.fi/lcd

SEARCH INTERFACE



Publication

"Without except(ing) unless...": on the grammaticalisation of expressions indicating exception in English

Topic

EXCEPT

Time periods:

Modern English
Present-Day English
Middle English
Old English

Custom time periods:

(-)

Keywords

EXCEPT; preposition; subordinator;

Corpora and databases

BROWN Corpus
Freiburg–LOB Corpus of British English
Freiburg-Brown corpus of American English
Helsinki Corpus
Lancaster-Oslo/Bergen Corpus
Middle English Compendium

Other sources

Genre

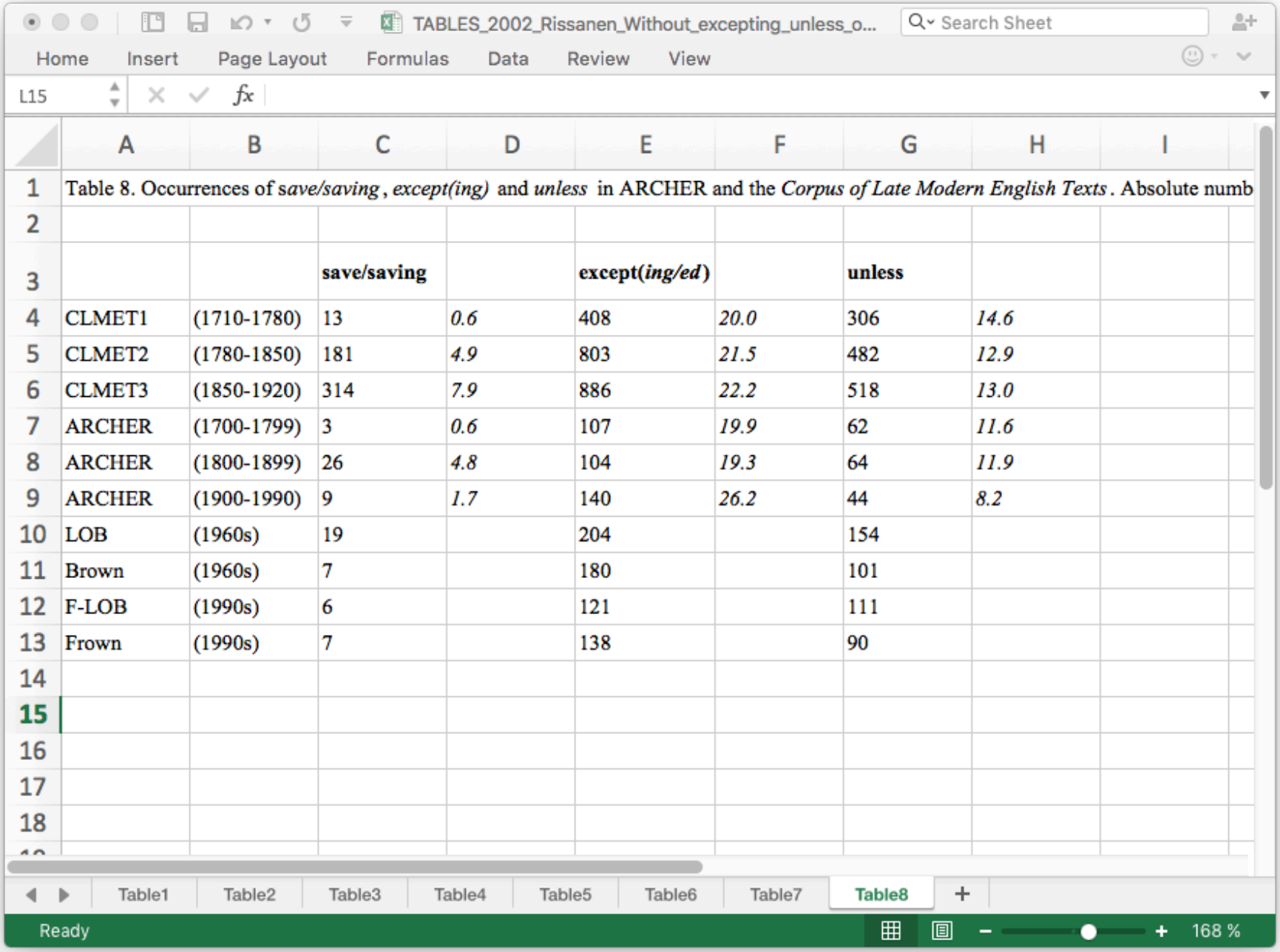
Variety

Summary of results

- The subordinator and preposition EXCEPT first appears in 15th-century texts.
- The connective was obviously borrowed from Latin; the verb EXCEPT appears roughly at the same time as the grammaticalized form.
- In the course of centuries, EXCEPT becomes more popular than SAVE but is superseded by UNLESS in subordinator use.
- It is used mostly in statutes and official correspondence.
- It's first appearance as a connective can be traced back to LME

Grammar

Preposition
Connectives

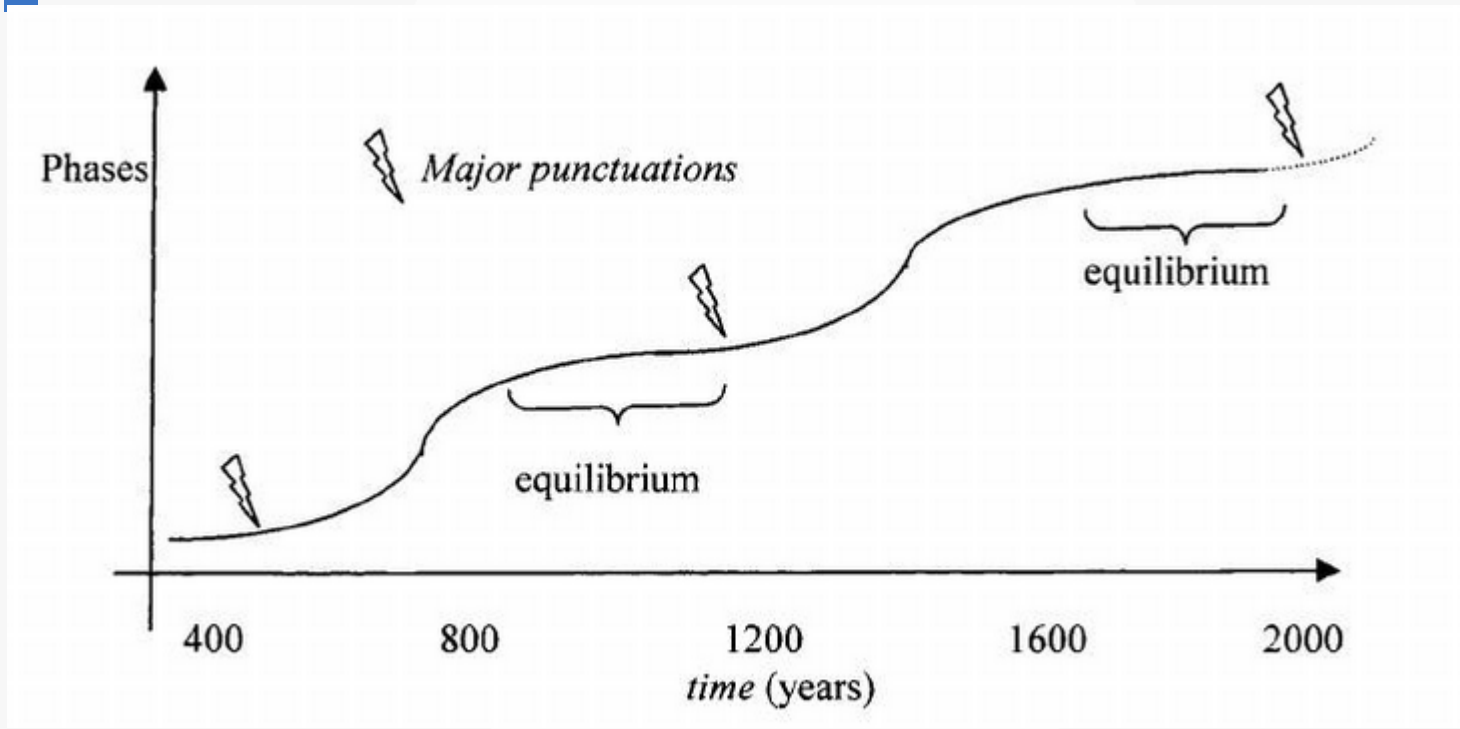


	A	B	C	D	E	F	G	H	I
1	Table 8. Occurrences of *save/saving*, *except(ing)* and *unless* in ARCHER and the *Corpus of Late Modern English Texts*. Absolute numb								
2									
3			**save/saving**		**except(*ing/ed*)**		**unless**		
4	CLMET1	(1710-1780)	13	*0.6*	408	*20.0*	306	*14.6*	
5	CLMET2	(1780-1850)	181	*4.9*	803	*21.5*	482	*12.9*	
6	CLMET3	(1850-1920)	314	*7.9*	886	*22.2*	518	*13.0*	
7	ARCHER	(1700-1799)	3	*0.6*	107	*19.9*	62	*11.6*	
8	ARCHER	(1800-1899)	26	*4.8*	104	*19.3*	64	*11.9*	
9	ARCHER	(1900-1990)	9	*1.7*	140	*26.2*	44	*8.2*	
10	LOB	(1960s)	19		204		154		
11	Brown	(1960s)	7		180		101		
12	F-LOB	(1990s)	6		121		111		
13	Frown	(1990s)	7		138		90		
14									
15									
16									
17									
18									

CASE STUDY: RISSANEN & CONNECTIVES

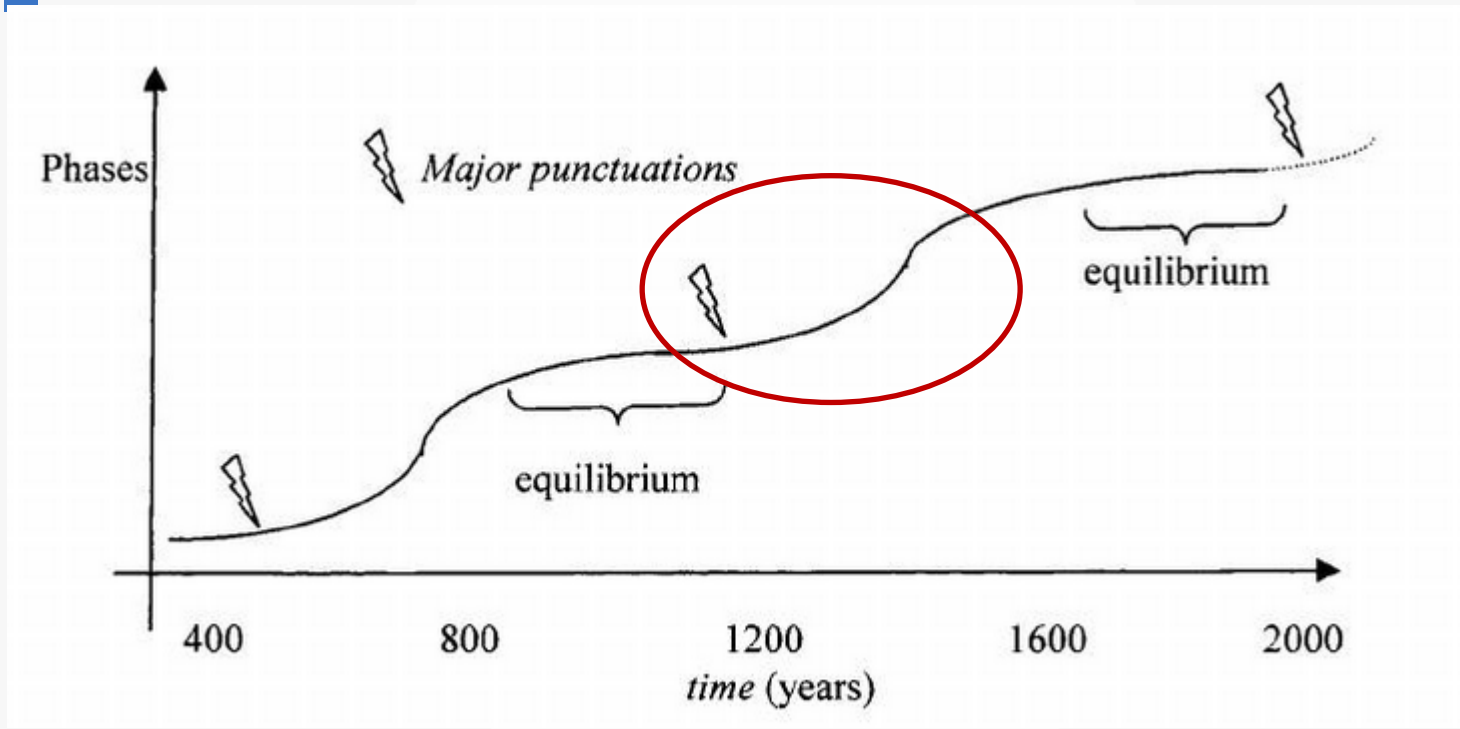
Punctuated equilibria?

HISTORY OF ENGLISH AS PUNCTUATED EQUILIBRIA



(Bergs 2005: 54)

HISTORY OF ENGLISH AS PUNCTUATED EQUILIBRIA

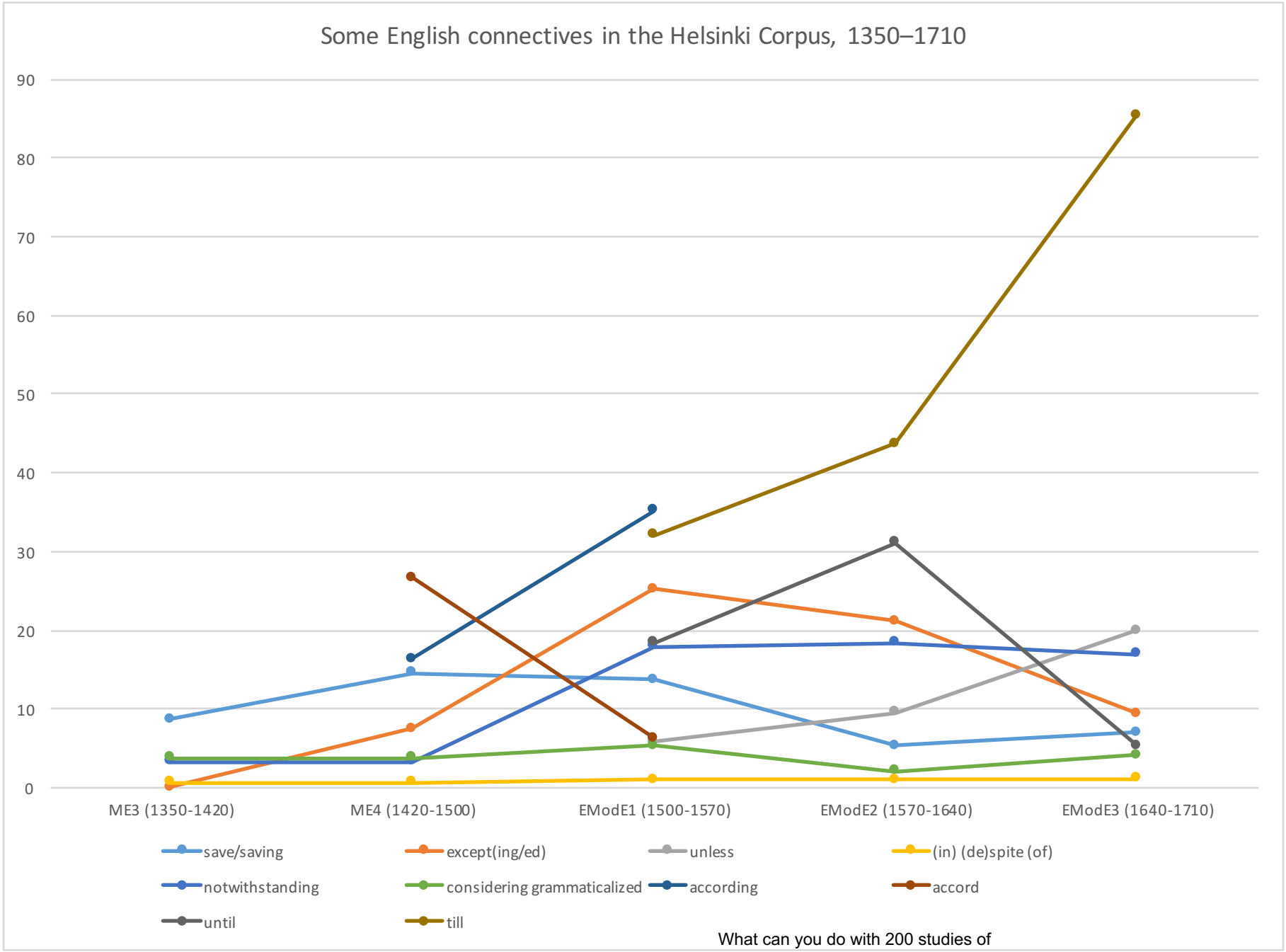


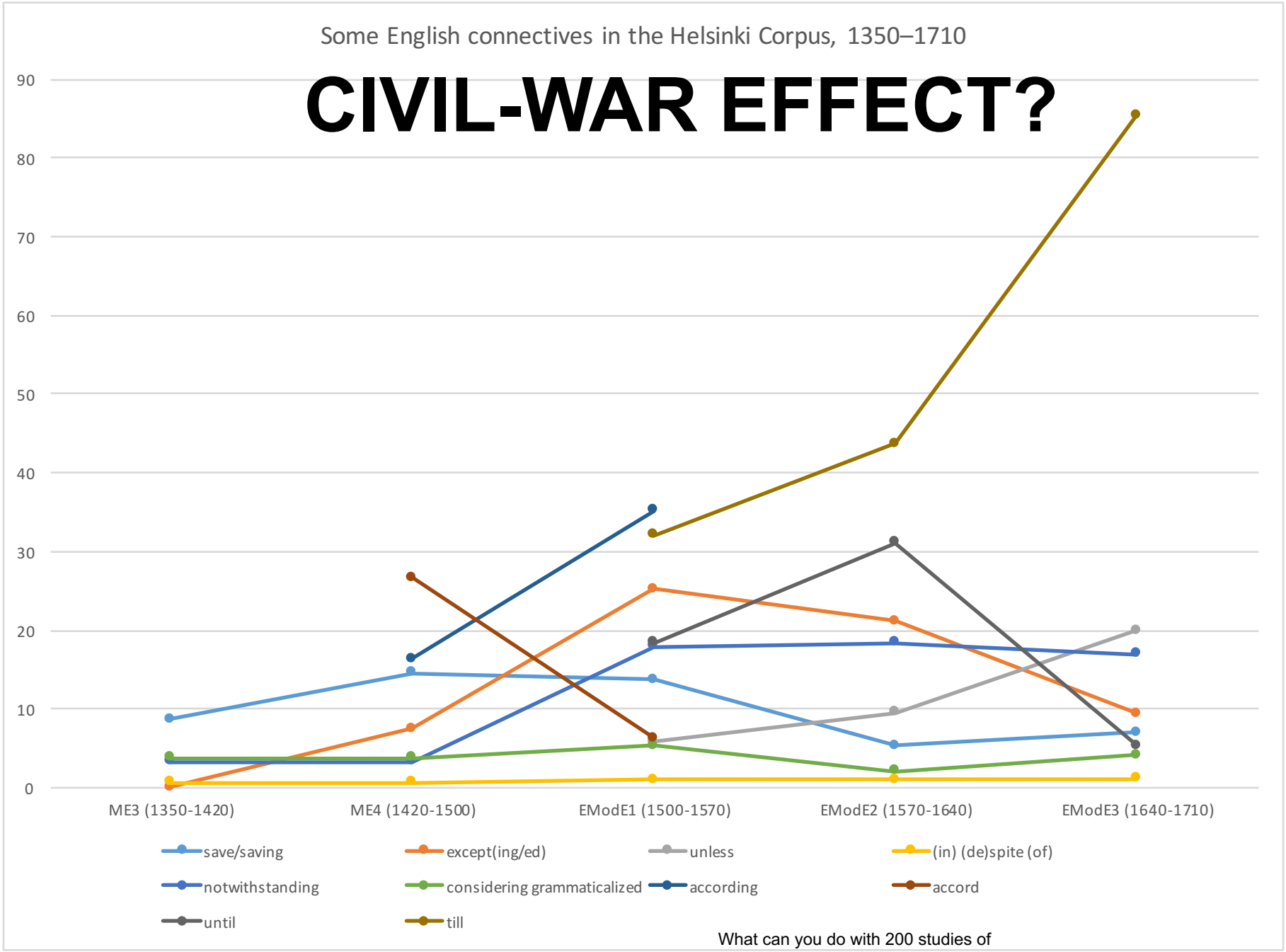
(Bergs 2005: 54)

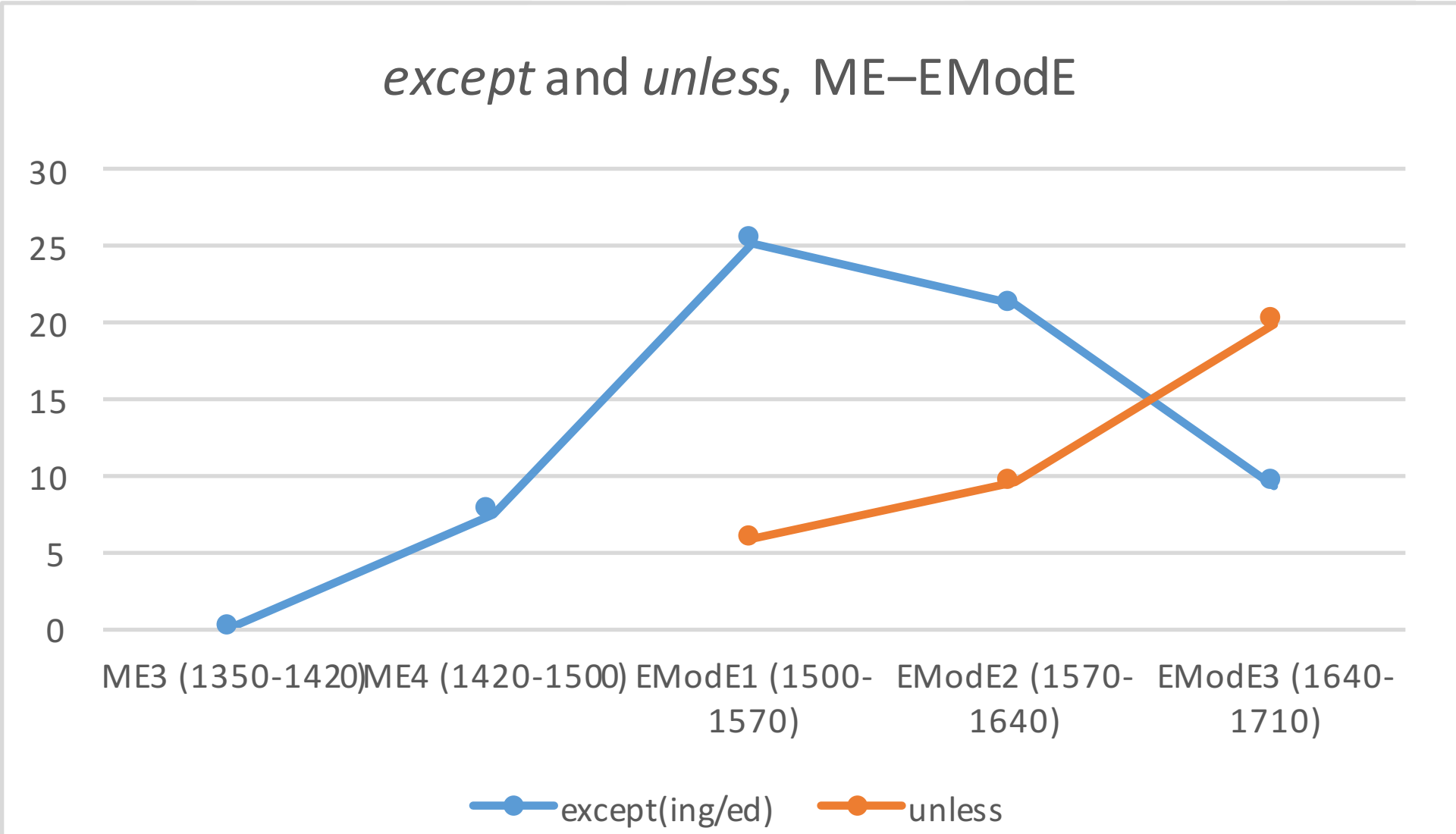
RISSANEN'S STUDIES ON CONNECTIVES

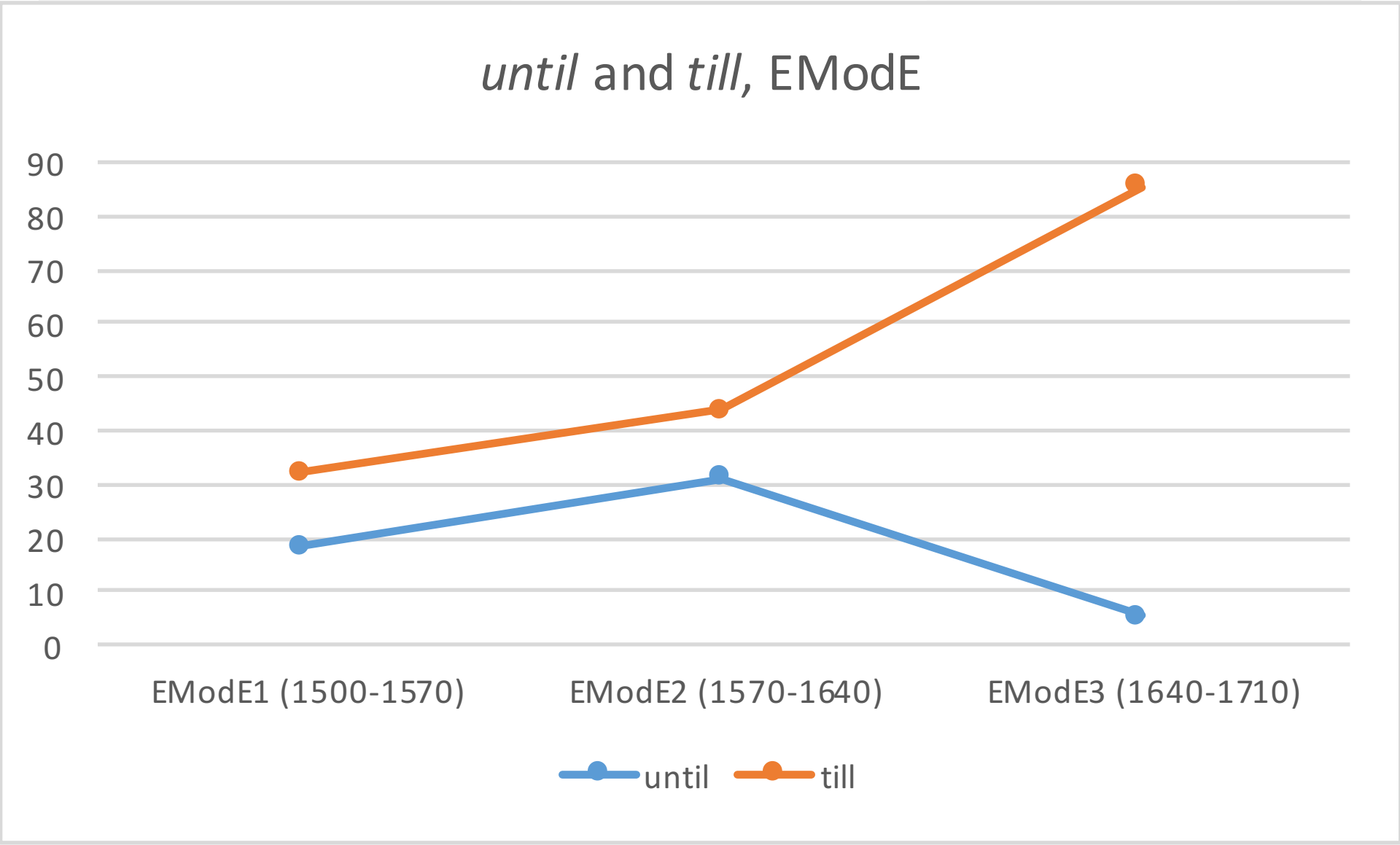
Grammaticalization / first appearance as a connective:

- Old English (or earlier)
  - *without, but, oþ, till, lest, þe/þæt*
- Middle English
  - *until, that*
- **Late Middle English** (C14–15) – punctuation?
  - *except, in case, because, save, unless, notwithstanding, outtaken, according, provided that, despite, seeing, considering*

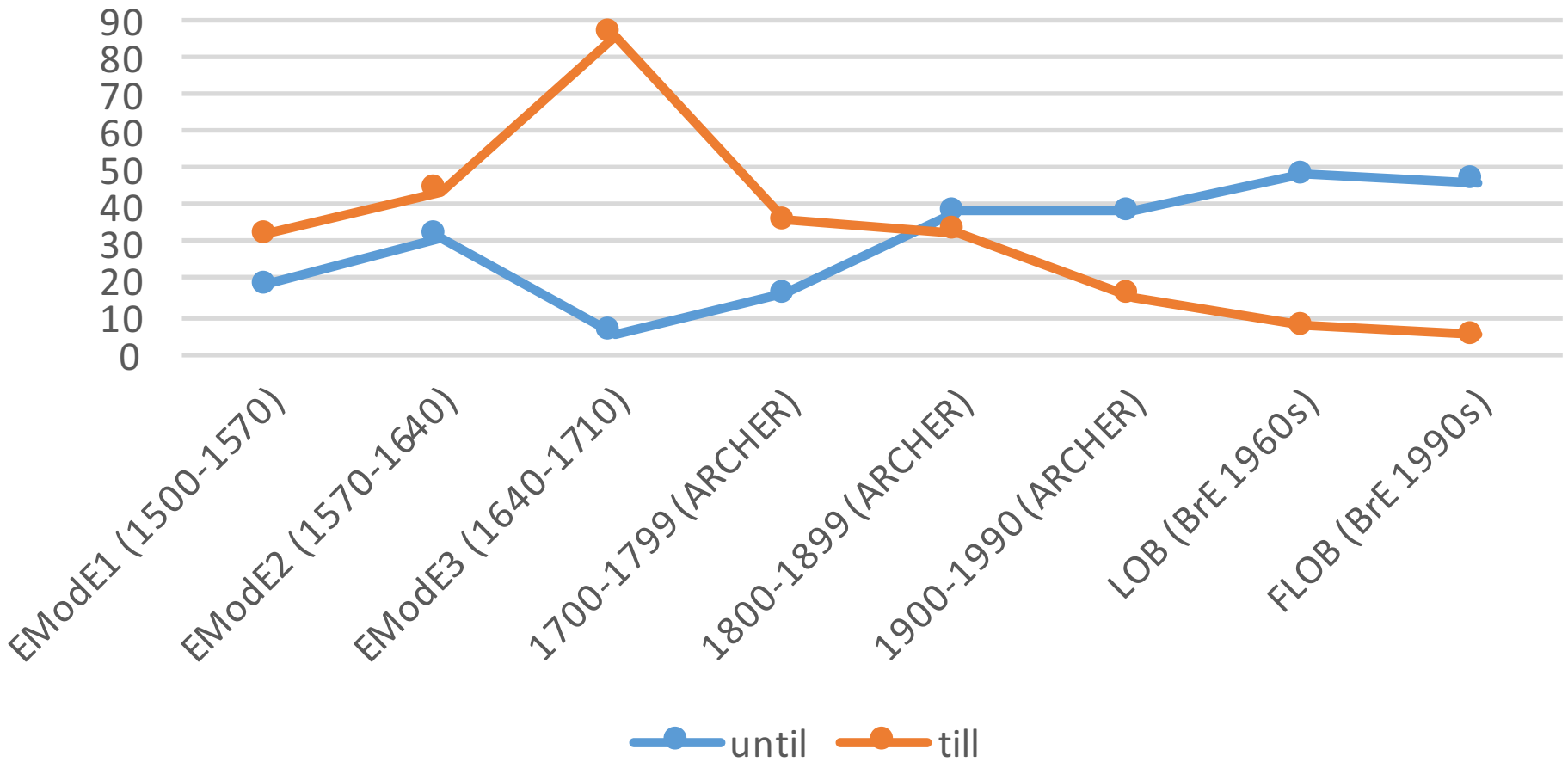

Some English connectives in the Helsinki Corpus, 1350–1710
90
80
70
60
50
40
30
20
10
0
ME3 (1350-1420)
ME4 (1420-1500)
EModE1 (1500-1570)
EModE2 (1570-1640)
EModE3 (1640-1710)
save/saving
except(ing/ed)
unless
(in) (de)spite (of)
notwithstanding
considering grammaticalized
according
accord
until
till


Some English connectives in the Helsinki Corpus, 1350–1710
CIVIL-WAR EFFECT?
90
80
70
60
50
40
30
20
10
0
ME3 (1350-1420)
ME4 (1420-1500)
EModE1 (1500-1570)
EModE2 (1570-1640)
EModE3 (1640-1710)
save/saving
except(ing/ed)
unless
(in) (de)spite (of)
notwithstanding
considering grammaticalized
according
accord
until
till

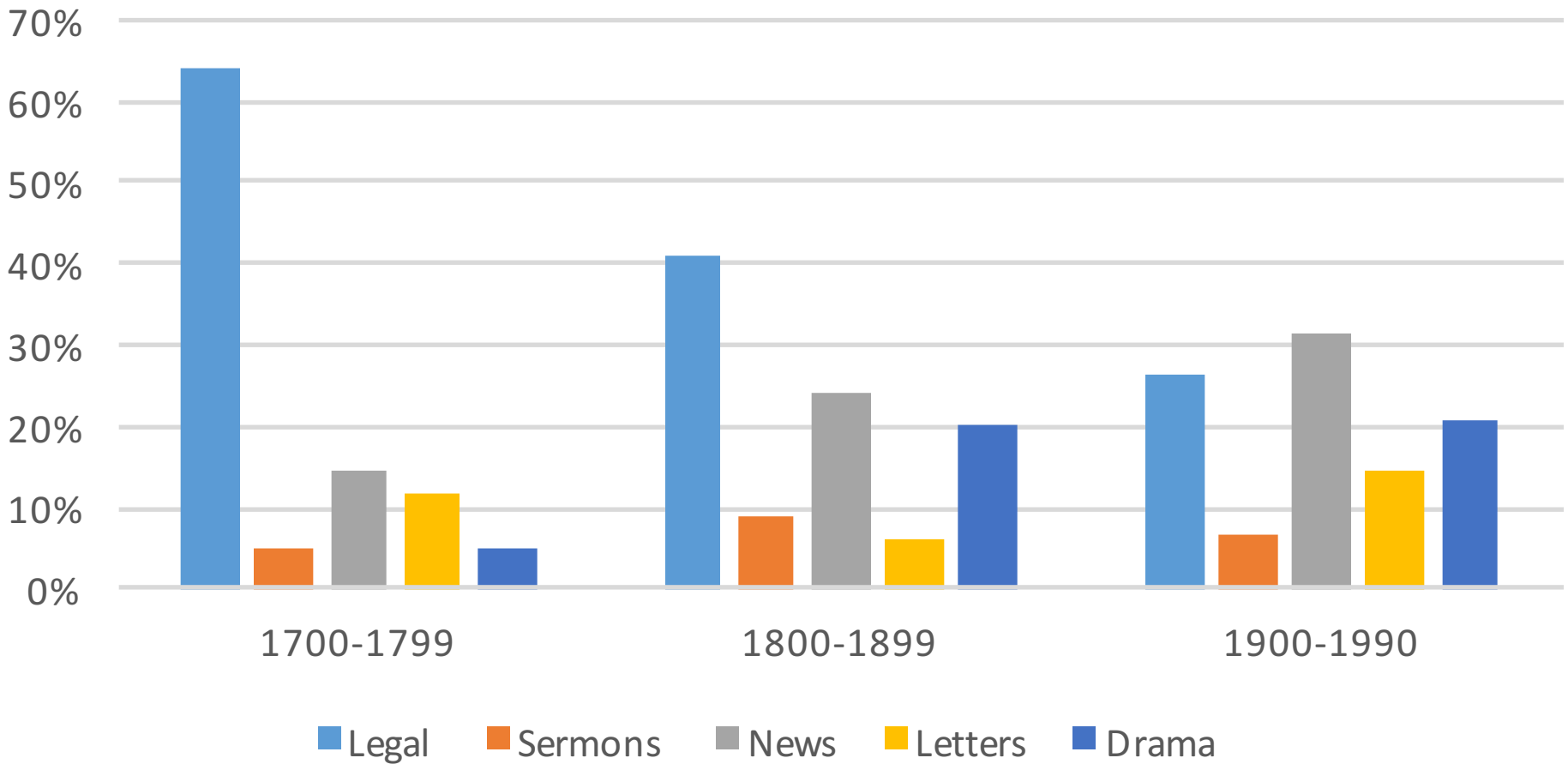

except and unless, ME–EModE
30
25
20
15
10
5
0
ME3 (1350-1420)
ME4 (1420-1500)
EModE1 (1500-1570)
EModE2 (1570-1640)
EModE3 (1640-1710)
except(ing/ed)
unless

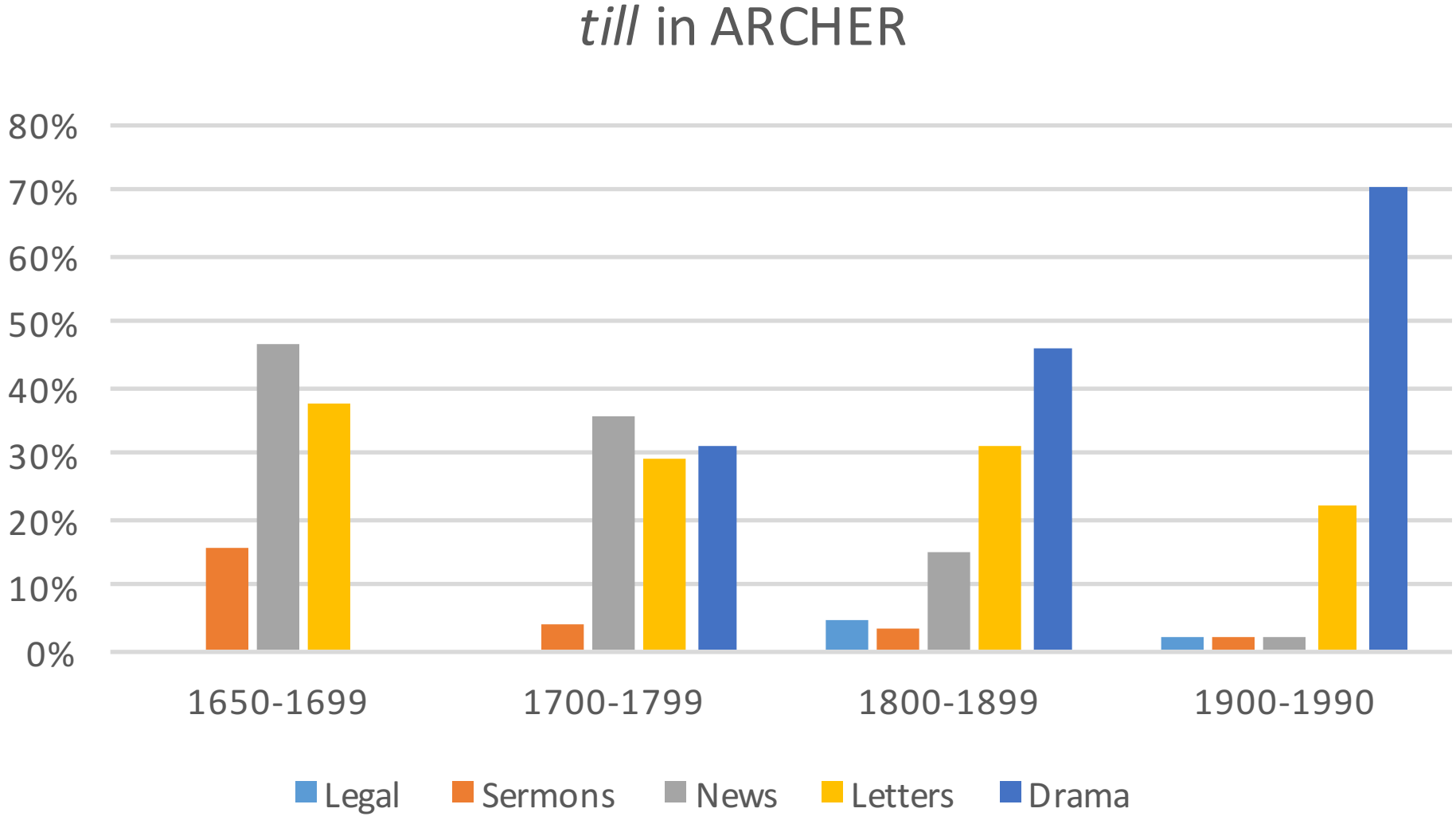

until and *till*, EModE
90
80
70
60
50
40
30
20
10
0
EModE1 (1500-1570)
EModE2 (1570-1640)
EModE3 (1640-1710)
until
till

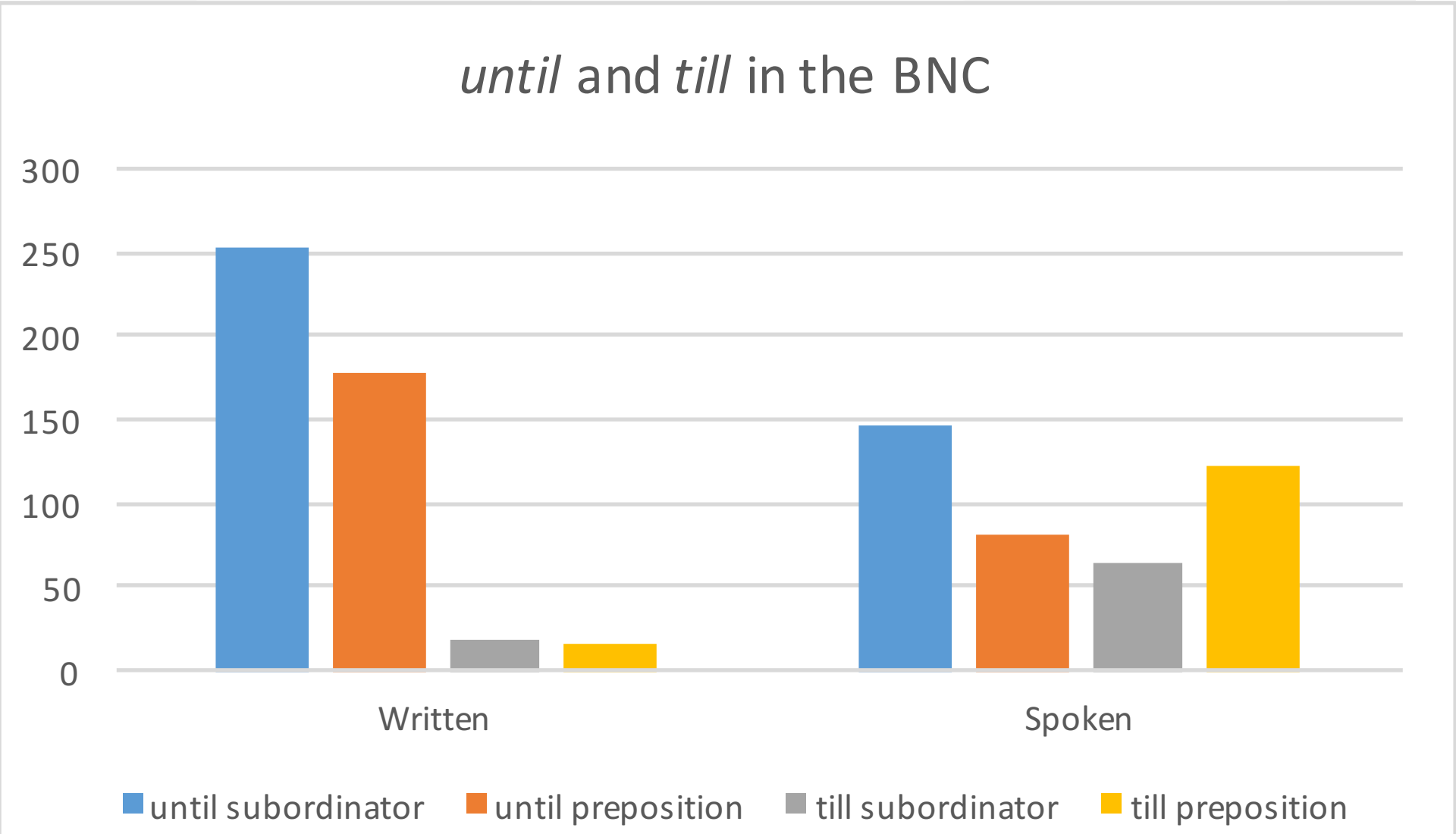
until and *till* in the long diachrony



until in ARCHER




till in ARCHER
80%
70%
60%
50%
40%
30%
20%
10%
0%
1650-1699
1700-1799
1800-1899
1900-1990
Legal
Sermons
News
Letters
Drama


until and *till* in the BNC
300
250
200
150
100
50
0
Written
Spoken
until subordinator
until preposition
till subordinator
till preposition

CASE REVISITED: FUTURE WORKFLOW

From Excel to **machine-readable tabular data** (RDF Data Cube)

- Filter datasets
  - Grammar = Connectives or any narrower grammar concept
  - Expression = contains “unless”
- Possibility to automatically calculate normalised frequencies
  - Validate normalised frequency values against corpus details
- Query filtered dataset
  - Sum of frequencies
  - Group by time period
- Visualize

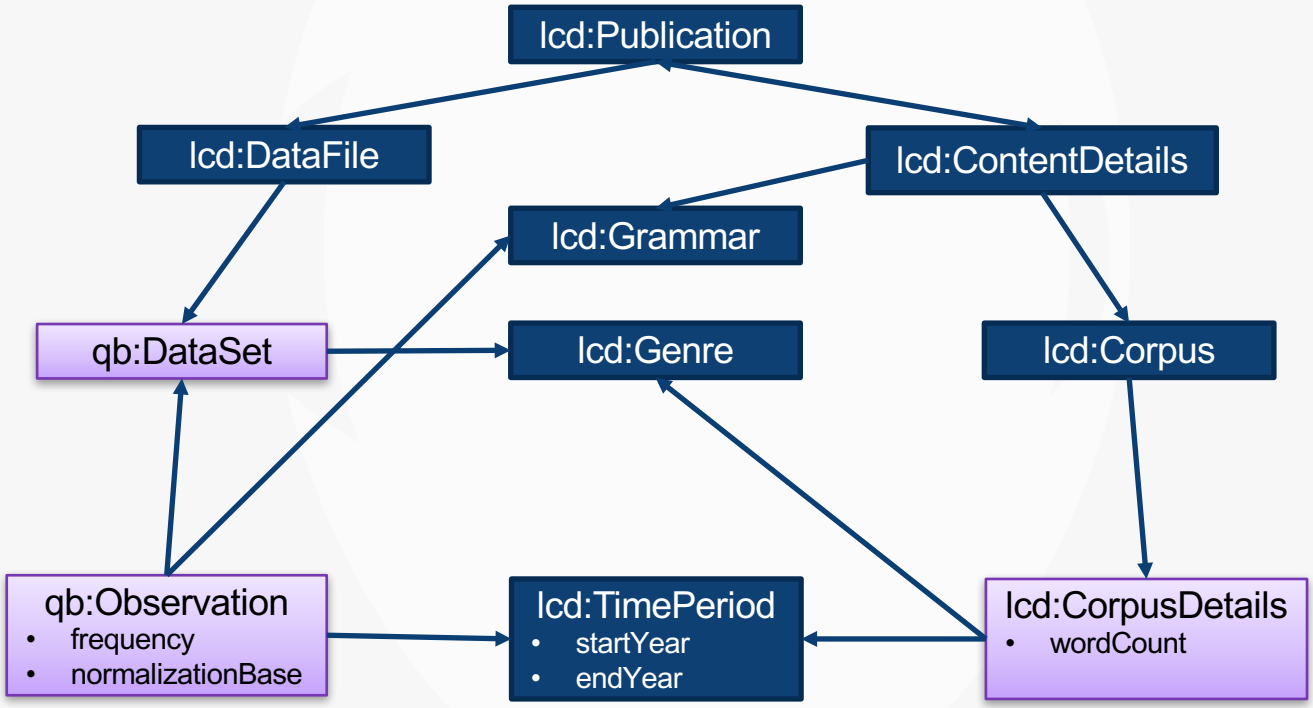
CONCLUSIONS (1)

- Workflow can be supported by one or more applications
- LCD works as common data source
- Proof-of-concept template for **storing and querying linguistic research data** for meta-analysis
  - Makes published research data ready for any kind of automated processing
  - Step towards re-usable research data

CONCLUSIONS (2)

- **Meta-analysis** possible at the level of:
  - Individual researchers
  - Linguistic topics
  - Time periods
- **LCD**: useful tool even at this stage
  - Will be published as an **open access** online resource by the end of the project (2018)
    - Wiki-style: researchers will be able to input their own studies
  - See Nevalainen et al. (2016)

DATA MODEL (SIMPLIFIED)



REFERENCES

- Bergs, Alexander. 2005. *Social networks and historical sociolinguistics: Studies in morphosyntactic variation in the Paston Letters (1421–1503)*. Berlin & New York: Mouton de Gruyter.
- Nevalainen, Terttu, Turo Vartiainen, Tanja Säily, Joonas Kesäniemi, Agata Dominowska & Emily Öhman. 2016. Language Change Database: A new online resource. *ICAME Journal* 40: 77–94.
- Rissanen, Matti. 22 studies on connectives. **Thank you, Matti!**