



FROM LINKED OPEN DATA TO META-ANALYSIS IN HISTORICAL LINGUISTICS

Joonas Kesäniemi, Turo Vartiainen, Tanja Säily, Agata Dominowska, Aatu Liimatta, and Terttu Nevalainen

HELDIG DIGITAL HUMANITIES SUMMIT 2017

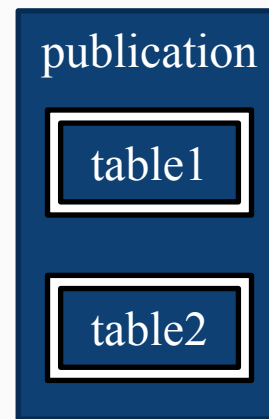
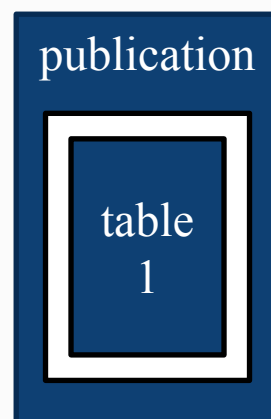
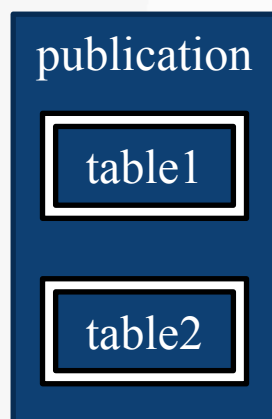


LANGUAGE CHANGE DATABASE (LCD)

- **LCD**: a research database that will provide comparative, real-time baseline data for modelling change in progress
- Summarises the results of research articles on variation and change in English
- Each entry annotated for linguistic and extra-linguistic features; **easy to edit & query**
- **Numerical data** in a computer-readable form
- Open-access **linked data** web application
- Developing **new tools** that can be used in **meta-analyses**

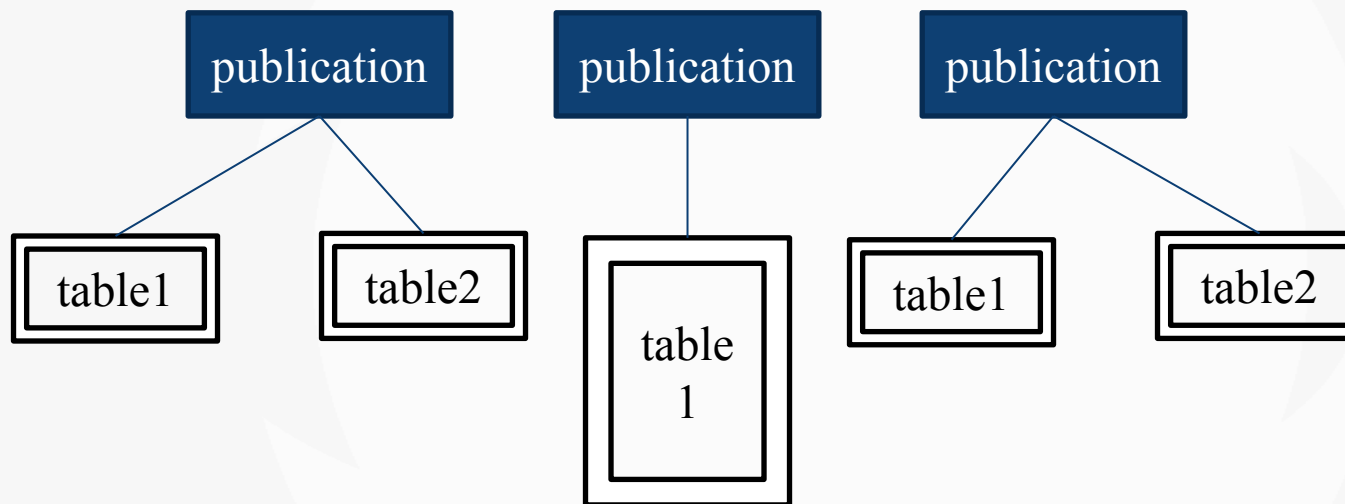


STARTING POINT



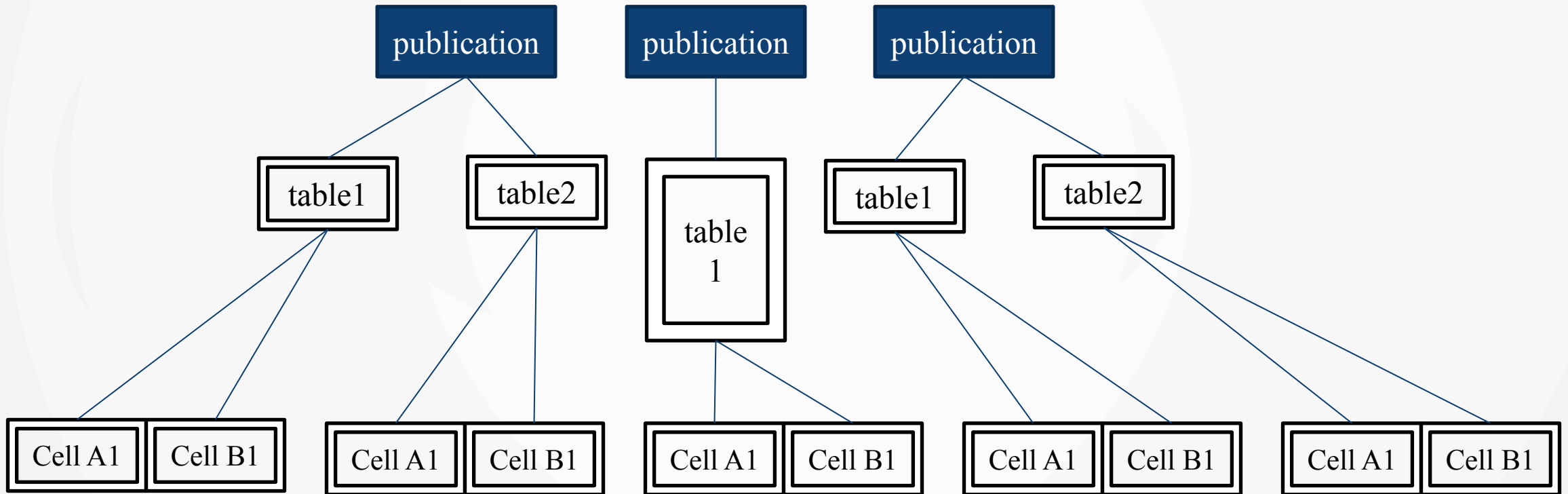


EXTRACTED DATA TABLES



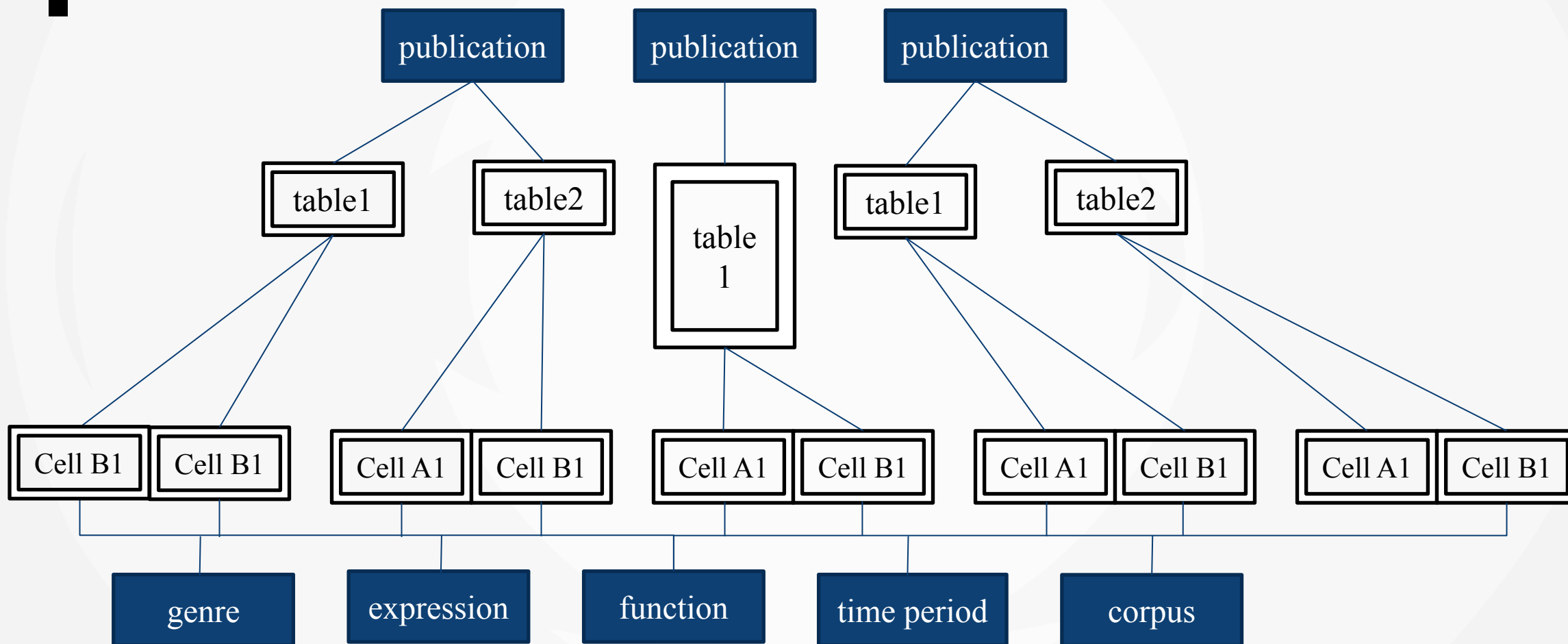


EXTRACTED DATA CELLS



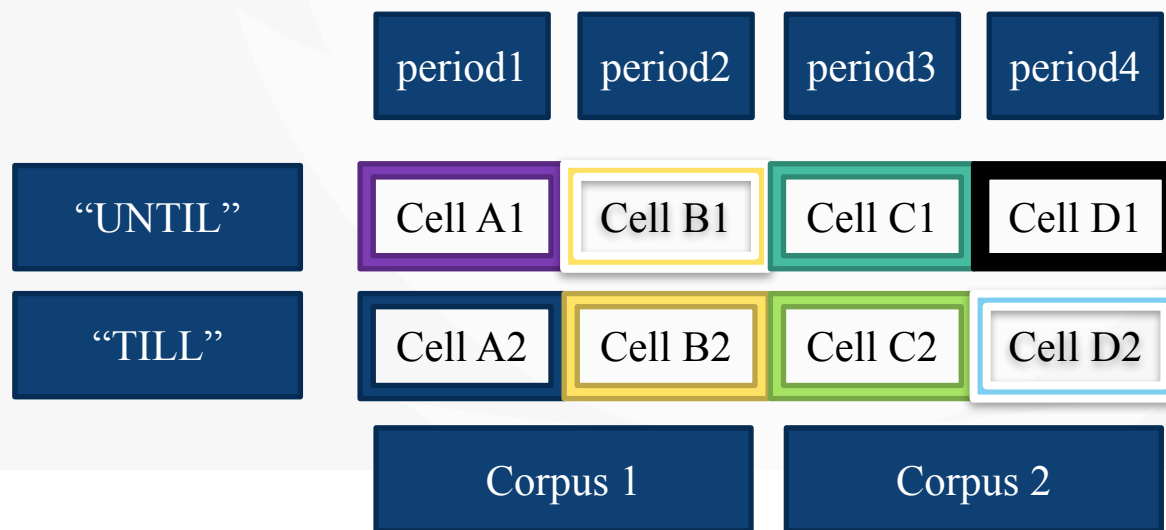
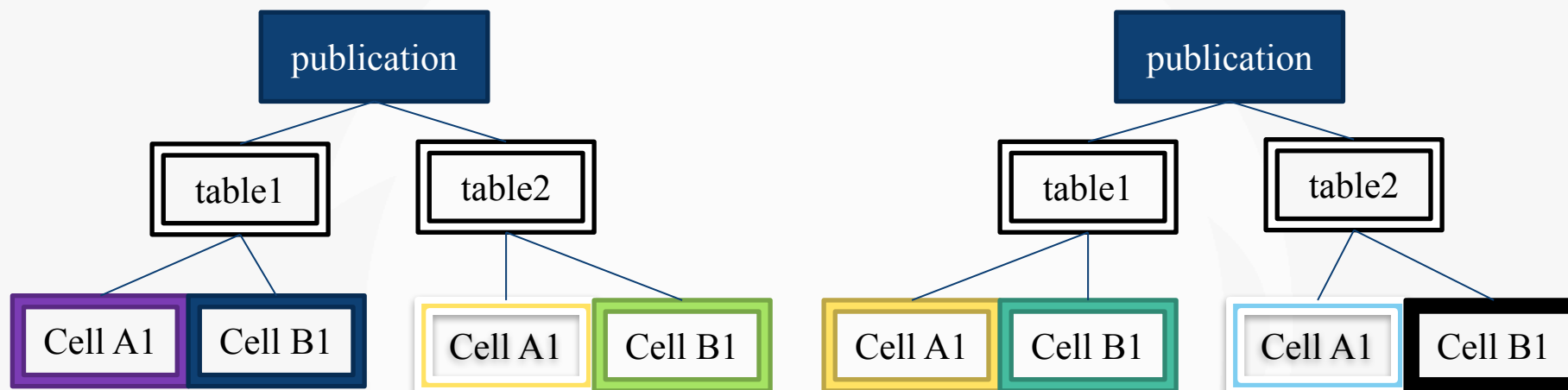


ANNOTATED DATA CELLS





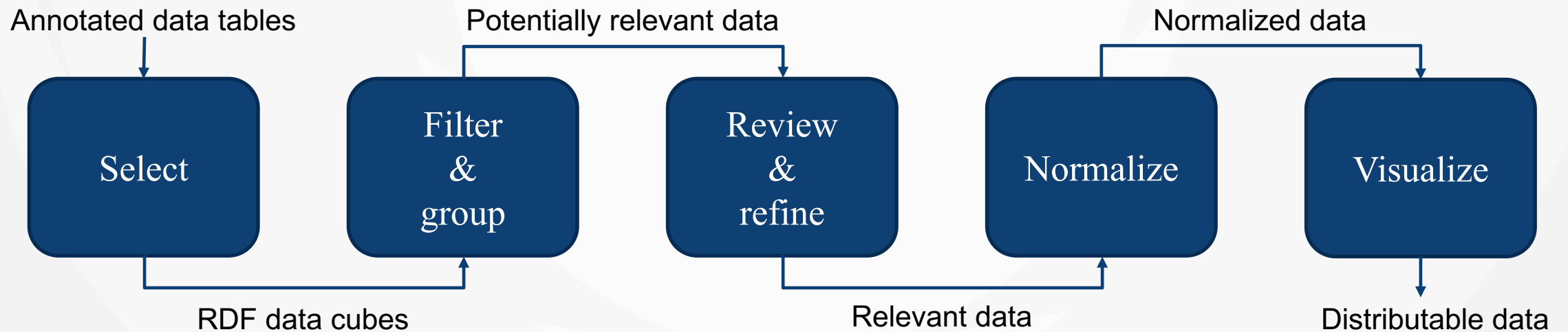
AGGREGATED DATASET





LADA

- LCD Aggregated Data Analysis workbench
- Generates RDF Data Cube representation of the annotated data
- Provides a workflow for creating new aggregated datasets





COME AND SEE THE DEMO!

TOC ▾
Select
Filter & Group
Review
Normalize
Visualize
LADA
LCD online
⚙️

Review

Filters

ARCHER
HC
LOB
F-LOB
TILL
UNTIL
Any or no value
Any or no value

Some time period

Groups

ARCHER
HC
LOB
F-LOB
TILL
UNTIL
Some time period

Source tables and filtered values

The development of TILL and UNTIL in English
Rissanen, 2005

Exclude publication

Table1

Table 1. Until and till in the Early Modern English sub-sections of the Helsinki Corpus. Figures per 100,000 words in brackets. (2005)

+
-

	until			till			HC
	subord.	prep.	total	subord.	prep.	total	
EModE1 (1500-1570)	24	11	35 (18.4)	45	16	61 (32.1)	
EModE2 (1570-1640)	43	16	59 (31.1)	57	26	83 (43.7)	
EModE3 (1640-1710)	6	3	9 (5.3)	87	59	146 (85.4)	

Selected and grouped values

Value(s)	Corpus	Expression	Period
0 0 12 15 5	ARCHER	TILL	1650-1699
0 0 1 0 0	ARCHER	UNTIL	1650-1699
30 0 0			