"History of English II: Focus on verbs"
LV 17 340, SoSe 2017
Tuesdays 8 a.m., room KL 32/102
Institut für Englische Philologie, Freie Universität Berlin

Kirsten Middeke, M.A.
kirsten.middeke@fu-berlin.de
office hour: Thursdays 1 p.m.
office: JK 29/223

The Penn Corpora

Alhadji Jallow, Jan Reimer, Georg Hartisch

13.06.2017

Penn Parsed Corpora of Historical English (http://www.ling.upenn.edu/hist-corpora/)

PPCEME (**P**enn-Helsinki **P**arsed **C**orpus of **E**arly **M**odern **E**nglish	Anthony Kroch, Beatrice Santorini, and Lauren Delfs. 2004. *The Penn-Helsinki Parsed Corpus of Early Modern English (PPCEME).* Department of Linguistics, University of Pennsylvania. CD-ROM, first edition, (http://www.ling.upenn.edu/hist-corpora/)
PPCME2 (**P**enn-Helsinki **P**arsed **C**orpus of **M**iddle **E**nglish **2**nd edition)	Anthony Kroch and Ann Taylor. 2000. *The Penn-Helsinki Parsed Corpus of Middle English (PPCME2*). Department of Linguistics, University of Pennsylvania. CD-ROM, second edition, (http://www.ling.upenn.edu/hist-corpora/).
PPCEEC (**P**arsed **C**orpus of **E**arly **E**nglish **C**orrespondence	*Parsed Corpus of Early English Correspondence, tagged version*. 2006. Annotated by Arja Nurmi, Ann Taylor, Anthony Warner, Susan Pintzuk, and Terttu Nevalainen. Compiled by the CEEC Project Team. York: University of York and Helsinki: University of Helsinki. Distributed through the Oxford Text Archive.
PPCMBE2 (**P**enn **P**arsed **C**orpus of **M**odern **B**ritish **E**nglish **2**nd edition)	Anthony Kroch, Beatrice Santorini, and Lauren Delfs. 2010.*The Penn-Helsinki Parsed Corpus of Modern British English* (PPCMBE). Department of Linguistics, University of Pennsylvania. CD-ROM, first edition. (http://www.ling.upenn.edu/hist-corpora/)

What is it?

PPCEME2: 56 text samples, 1.2 million words, 1150-1500
PPCEME: 448 text samples, 1.7 million words, 1500-1720
PPCEEC: 84 letter collections, 2.2 million words = 4970 letters, 1410-1695
PPCMBE2: 2.8 million words, 1st edition 101 text samples, 2nd edition 275 text samples

Difficulties:

- It is not lemmatized → very long queries

What do we use it for?

- To analyse how grammatical phenomena have changed over time (e.g. *going to*)

Kirsten Middeke, M.A.
kirsten.middeke@fu-berlin.de
office hour: Thursdays 1 p.m.
office: JK 29/223

Sub-Periods of the Helsinki-Corpus

Period designation	Composition date	Manuscript date
MX1	unknown	1150-1250
M1	1150-1250	1150-1250
M2	1250-1350	1250-1350
M23	1250-1350	1350-1420
M24	1250-1350	1420-1500
M3	1350-1420	1350-1420
M34	1350-1420	1420-1500
MX4	unknown	1420-1500
M4	1420-1500	1420-1500
E1	1500-1569	
E2	1570-1639	
E3	1640-1720	

PPCME2 set PrintStructures…

```
text_id
text_year
text_period
text_genre
text_dialect_reg      # region
text_dialect_dist     # district (where
      available)
text_dialect_hels     # code as documented
      in Helsinki corpus
```

PPCEME set PrintStructures…

```
text_id
text_year
text_period
text_genre
```

PPCEEC set PrintStructures…

```
text_id
letter_id
letter_auth
letter_auth_sex
letter_relToRec
letter_auth_dob
letter_auth_age
letter_rec
letter_rec_sex
letter_relToAuth
letter_rec_dob
letter_rec_age
letter_period
letter_date
letter_authenticity
```