

Digitizing historical print dictionaries I

General idea

- ▶ implicit structuring of data
- ▶ theory bound analysis/presentation
- ▶ inconsistencies
- ▶ scope ambiguities (e. g. unmarked glosses in discursive text)
- ▶ often discursive structure (beware of negations!)
- ▶ non-standard transliteration, normalization
- ▶ digitization is (re)modeling, e. g.:
 - typographic – 2-dimensional page
 - editorial – character and control “stream”
 - lexical – database representation (sets, graphs, . . .)in TEI terms: “views”
- ▶ modeling is abstraction and generalization

Digitizing historical print dictionaries II

Challenges

- ▶ IPR issues
- ▶ availability of copies
- ▶ image to text transformation (OCR, n-keying)
- ▶ we typically cannot keep all views (too complex, too expensive, probably unnecessary)
- ▶ reuseable and sustainable encoding (XML, Unicode)
- ▶ interoperabel and standard-conforming encoding (TEI, BCP 47, data category registries, linked open data)

Hands-on examples I

Mueller (1878)

- ▶ Eduard Mueller: Etymologisches Wörterbuch der englischen Sprache. Erster Teil, A–K. 2nd edition. Cöthen 1878.
- ▶ English headwords, implicit part-of-speech
- ▶ German glosses (mainly)
- ▶ German notes and commentary
- ▶ often contrasting diverging etymologies
- ▶ relations often left implicit
- ▶ related entries for derivatives
- ▶ (sometimes unattributed) citations
- ▶ bibliographical links

Hands-on examples II

Kluge/Lutz (1898)

- ▶ Friedrich Kluge, Frederick Lutz: Englisch Etymology: A Select Glossary. Boston 1898.
- ▶ English headwords, part-of-speech
- ▶ English glosses
- ▶ English notes and commentary
- ▶ occasionally provides reconstructions for PIE, Proto-Germanic
- ▶ rarely related entries, some derivatives as headwords
- ▶ very short entries (indeed more of a glossary)
- ▶ no citations
- ▶ no bibliographical links

Hands-on examples III

Uhlenbeck (1900)

- ▶ C. C. Uhlenbeck: Kurzgefasstes etymologisches Wörterbuch der gothischen Sprache. 2nd edition. Amsterdam 1900.
- ▶ Gothic headwords
- ▶ German glosses
- ▶ German notes and commentary
- ▶ relations often left implicit
- ▶ derivatives are (short) main entries
- ▶ few (underspecified) reconstructions
- ▶ attested forms in earlier and later languages
- ▶ bibliographical links

Hands-on examples IV

Falk/Torp (1910)

- ▶ H. S. Falk, Alf Torp: Norwegisch-dänisches etymologisches Wörterbuch. Erster Teil, A–O. Heidelberg 1910.
- ▶ Danish/Norwegian headwords
- ▶ German glosses
- ▶ German notes and commentary
- ▶ relations sometimes unspecified (“is related to”)
- ▶ derivatives, sometimes compounds and phrases as related entries
- ▶ no reconstructions
- ▶ no citations
- ▶ no bibliographical links

Hands-on examples V

Walde (1910)

- ▶ Alois Walde: Lateinisches etymologisches Wörterbuch. 2nd edition. Heidelberg 1910.
- ▶ Latin headwords
- ▶ German glosses (mostly)
- ▶ German notes and commentary
- ▶ many reconstructions and other hypothetical forms
- ▶ complex entries comprising certain and uncertain etymologies
- ▶ many related entries (many of which also appear as main entries)
- ▶ no citations
- ▶ extensive use of bibliographical links