



HeidelGram

A corpus-based network analysis of grammarians' references in 19th-century British grammars

Beatrix Busse

Kirsten Gather

Ingo Kleiber

Heidelberg University

1. Introduction



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

grammar books: indicators of changing and stable language norms

vast number of English grammar books in the 19th century

surprisingly few studies focussing on 19th-century grammar writing

"this area of investigation still constitutes a gap" (Anderwald 2016: 3)

1. Introduction



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

19th century: major changes in British grammar writing

- ♦ from the prescriptive tradition to the first modern descriptive grammars
- ♦ emergence of comparative and historical linguistics
- ♦ 'phonetic turn': phonetics/phonology becomes a separate field

2. The *HeidelGram* Project



HeidelGram Project

- ◆ Compilation and investigation of historical English grammar books
- ◆ Combination of (historical) corpus-linguistic and network-analytic methods

Pilot Study

Focus on a **corpus of 19th-century grammar books** with the aims

- ◆ to show connections between grammar books as manifested in grammarians' references to other grammars or grammarians
- ◆ to categorise the references in 19th-century grammars and examine distributions and developments
- ◆ to check the results against established assumptions

→ design, implementation, and analysis of a **network of grammars and grammarians**

2. The *HeidelGram* Project



Pilot Corpus

corpus of 19th-century grammars (40 texts, approx. 2.9 mio. words)

criteria for text selection

(a) popularity and distribution of grammar books

- ♦ bibliographic listings of grammar books (e.g. in Michael 1987, Görlach 1998)
- ♦ numbers of editions
- ♦ book catalogues, advertisements, etc.
- ♦ contemporaries' comments, e.g. in literary genres, private letters
- ♦ curricula of schools, colleges, etc.

(b) variety in function, audience, and text type



3. The Network

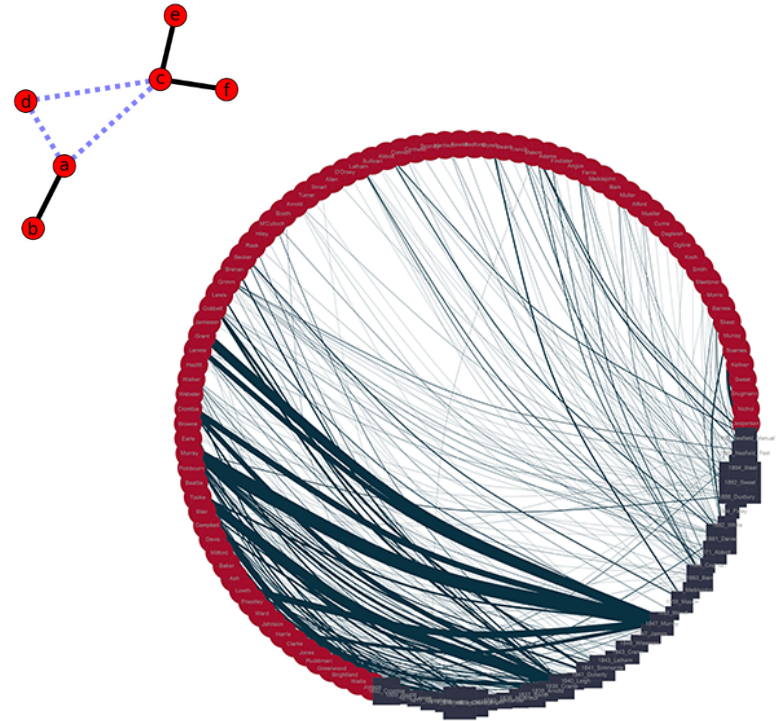
3. The Network



network approach (relationships as *nodes*, *edges/ties* and *attributes*)

White (2011):

- ♦ social vs. scholarly networks
- ♦ scholarly networks: cultural ties between authors irrespective of their possible acquaintance
- ♦ author intercitation: "the record of who has cited whom within a fixed set of authors" (2011: 275)
- ♦ a record (concordance) can carry multiple references



3. The Network



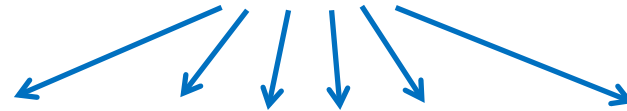
References

reference: the occurrence of another grammarian's name in a grammar book

example:

It is justly observed by Priestley, "that a term, which only implies the idea of persons, and expresses them by some circumstance or epithet, will hardly authorize the use of who." (Crombie 1802 on Priestley 1761)

quotation



opinion

3. The Network



Types of reference



3. The Network



type of reference: **quotation**

description: citation of text passages from other grammar books

subtypes: direct, indirect, direct/indirect, embedded, cross-reference

example: "We may consider," says Mr. C.P. **Mason**, "that the grammatical copula in every sentence consists of the personal inflections of the verb." (Bain 1863)
→ *quotation – direct*

3. The Network



type of reference: **opinion**

description: expression of positive or negative evaluation

subtypes: agreement, criticism, etc.

example: **Priestley**, in defending the other phraseology, appears to me to have *greatly erred* (Crombie 1802)
→ *opinion – criticism*

3. The Network



type of reference: **comparison/contrast**

description: comparison or grammarians' approaches, terminologies, etc.

subtypes: ---

example: "Mr. **Harris** and Mr. **Wallis** have each applied a partial principle in a universal sense, [...]" (Doherty 1841)

3. The Network



type of reference: **acknowledgment**

description: reference to grammarians or their works

subtypes: praise, approval, disapproval, neutral

example: "[...] the author of this Class Book has consulted the works of other grammarians, among whom may be mentioned Lindley **Murray**, Drs. **Latham**, **Beard**, **Angus**, **Morell**, [...]". (Duxbury 1886)
→ *acknowledgment – neutral*

3. The Network



type of reference: **mention**

description: simple mention of a grammarian without context

subtypes: ---

example: Give the substance of the quotation from Dr. **Angus**.
(Exercise in Duxbury 1886)

3. The Network



type of reference: **application**

description: application of a rule or concept introduced by and named after a grammarian

subtypes: ---

example: "The following table will be sufficient to illustrate **Grimm's** law: [...]" (Abbott 1871)

3. The Network



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Procedure

- ♦ manual compilation of list of search terms (last names of grammarians mentioned in the grammar books, supplemented by what secondary literature considers the most famous and influential grammarians)
- ♦ generating concordances (L/R± 20) based on the search terms
- ♦ problem: untrained OCR due to limited resources leads to corrupted data

3. The Network



Procedure

- ♦ our solution: automated search in digitised corpus texts with threshold-based pattern matching (gestalt pattern approach)
(*HGSimpleCorpusNetwork*, available via GitHub)
- ♦ matching threshold of 0.8 to cover all misreadings
e.g. the search for "Cobbett" produces results like *Cobbett* (1.0) and *Cohbett*, *Cobbeit*, *Gobbett*, *obbett* (0.82 to 0.92)
- ♦ manual correction by eliminating false positives

sweat / *Sweet*

Richard Johnson
(Latin Grammar)

vs.

Samuel Johnson
(English Dictionary)

3. The Network

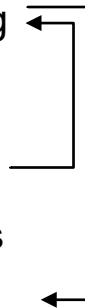


Procedure

- ♦ manual classification of all references (categories devised by substantive/ open coding influenced by *Grounded Theory*; two independent raters)

- ♦ coding process:

1. developing a robust coding scheme based on sample data
2. independent rating/coding
3. discussion of problematic cases
4. agreement on classification



independent rating/coding (multiclass, aggregated label \rightarrow *quotation_opinion*):

$$\kappa_{N=1963, c=3} = 0.5$$

$$\kappa_{N=1963, c=6} = 0.47$$

automated classification (experimental):

bag-of-words, multiclass (4) classification
unoptimized *TfidfVectorizer*, *KNeighbors*

$$\text{model acc.} = 0.38$$

$$F_1 = 0.62$$

3. The Network



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Procedure

Methodological pitfalls

- ♦ the search-term approach does not cover ambivalent names and pronouns
- ♦ choice of an optimal concordance length (including the actual reference/s while minimizing overlap)
- ♦ possibly some references have been lost due to the approximate string matching
- ♦ certain level of arbitrariness and subjectivity in coding the references



4. Main Results

4. Main Results



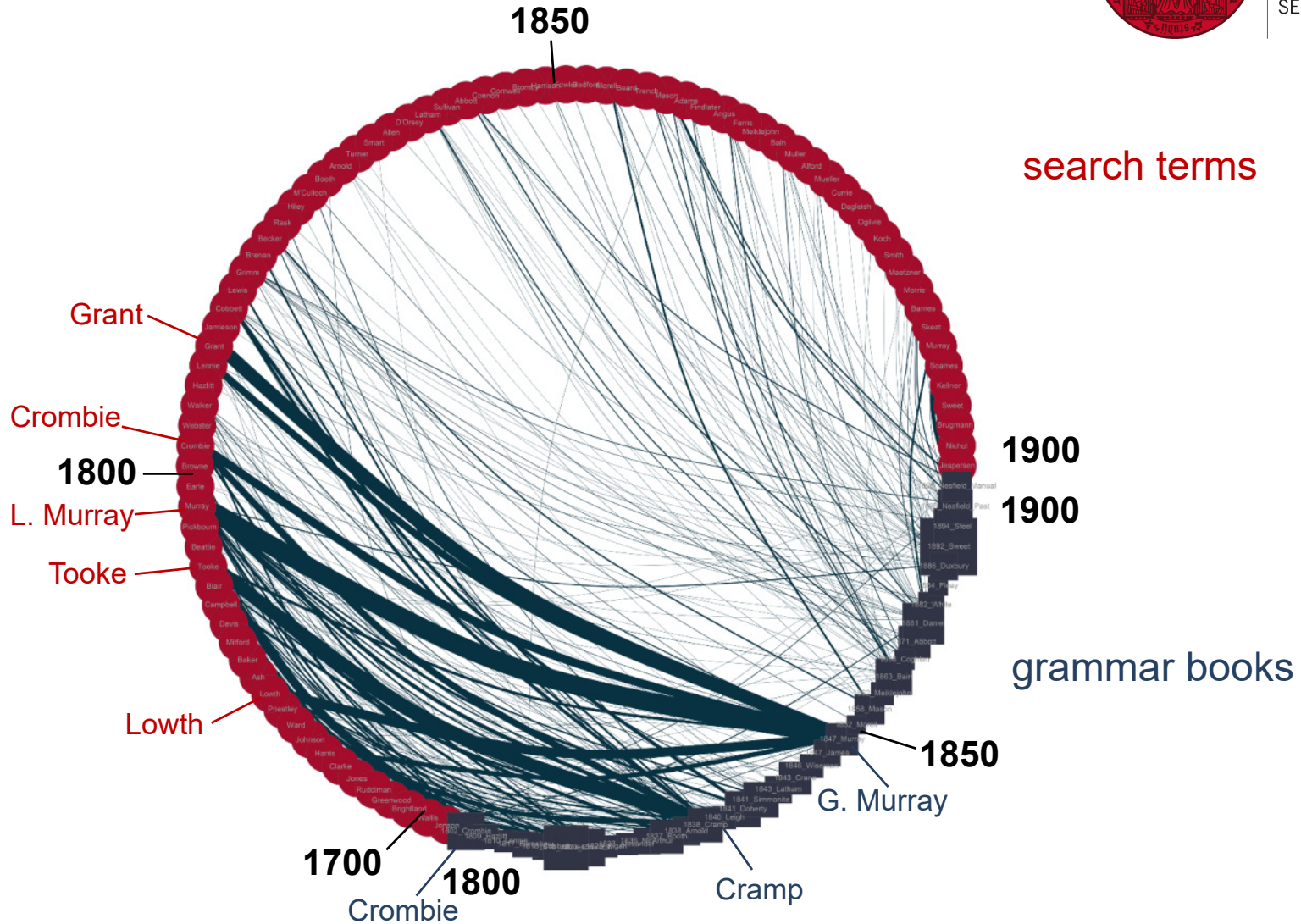
40 grammar books, list of 100 search terms (grammarians)

→ list of 2,003 concordances (after manual elimination of false positives),
containing 2,770 references

five out of 40 grammars do not contain any references to other grammarians

83 of the 100 search terms are referenced

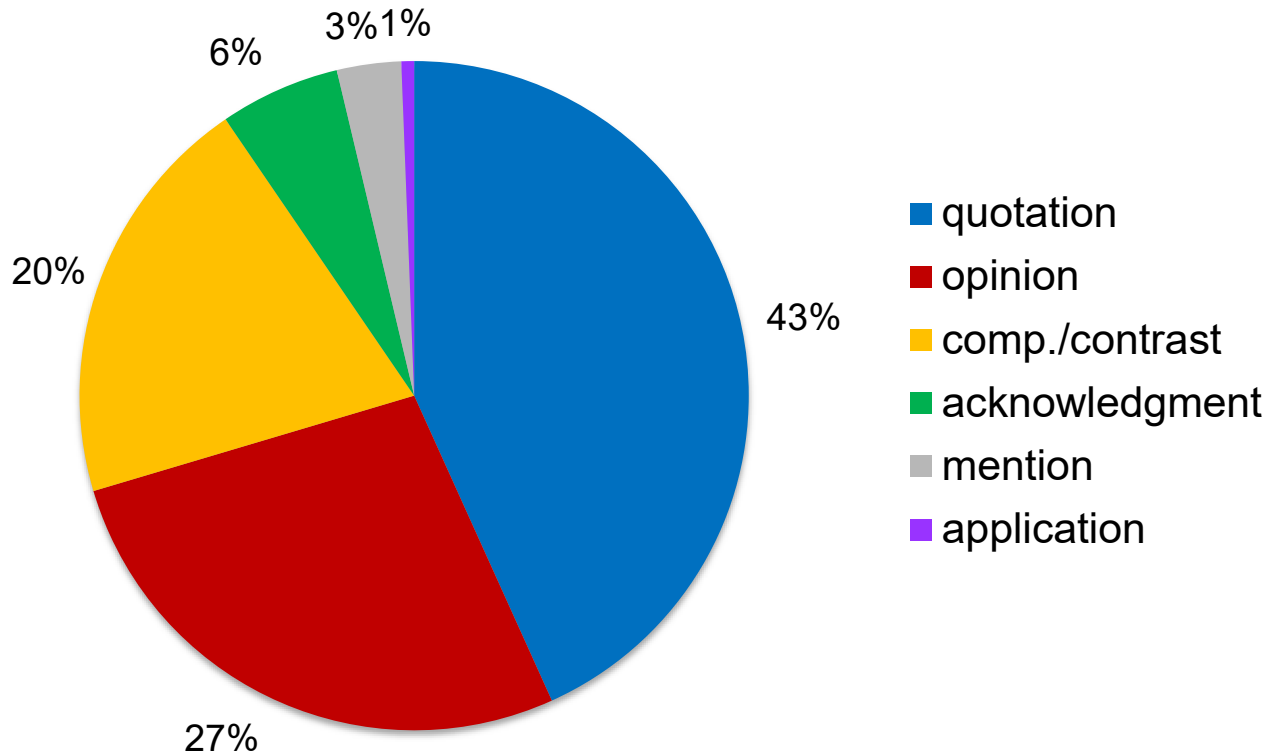
4. Main Results



4. Main Results



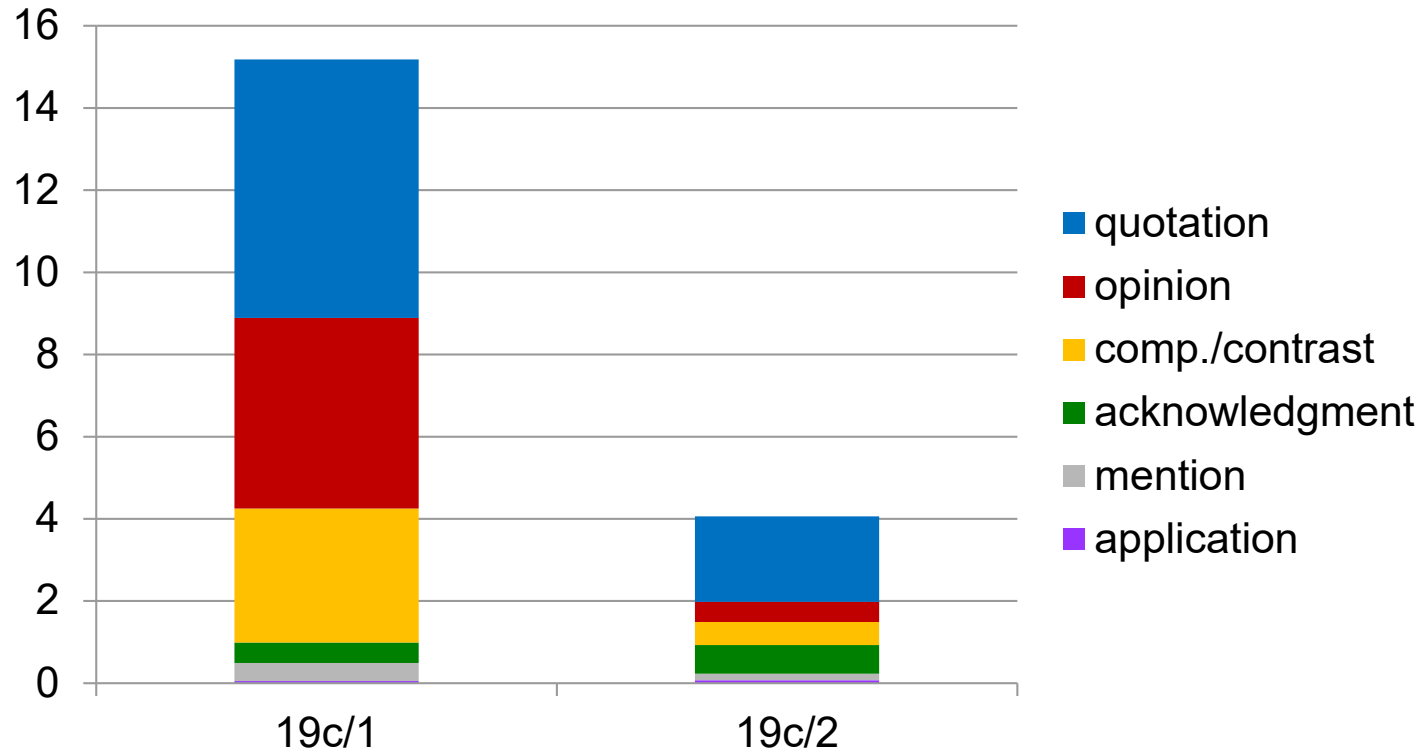
distribution of different **types of references** in 19th-century grammars
(in per cent of all references)



4. Main Results



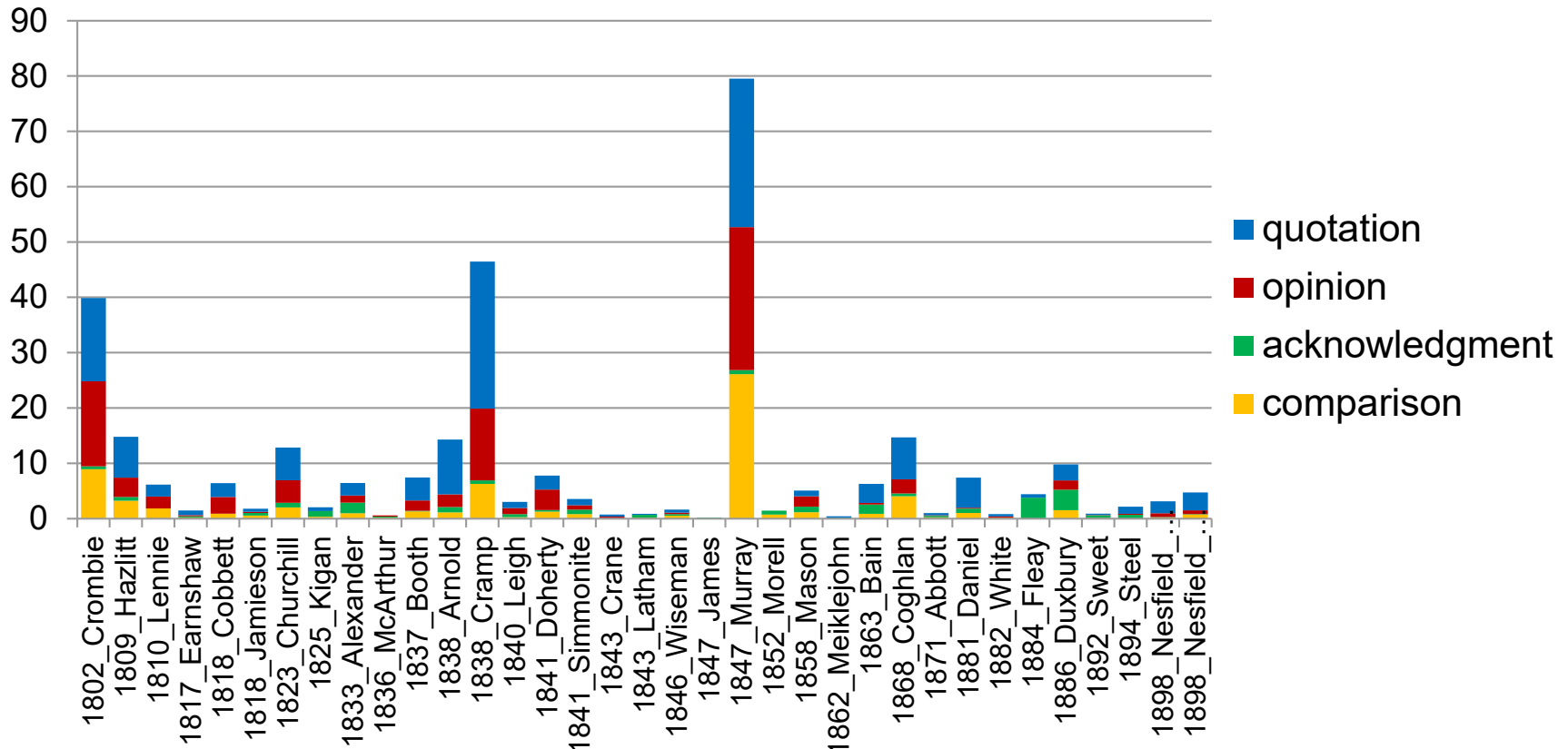
distribution of different types of references **in the first and second half of the 19th century** (frequencies per 10,000w)



4. Main Results



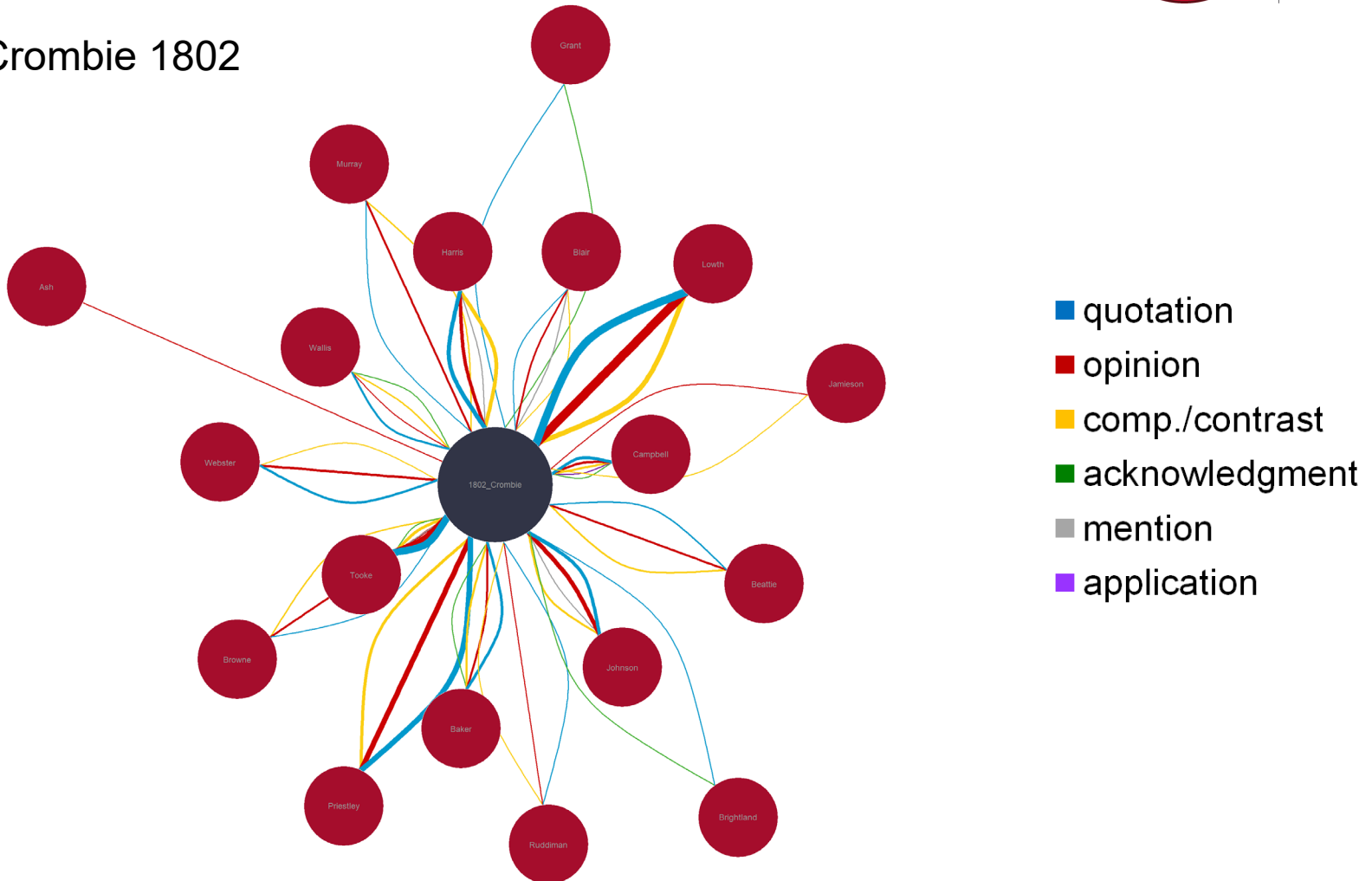
distribution of different types of references in 19th-century grammars
per author (normalised frequencies per 10,000w)



4. Main Results



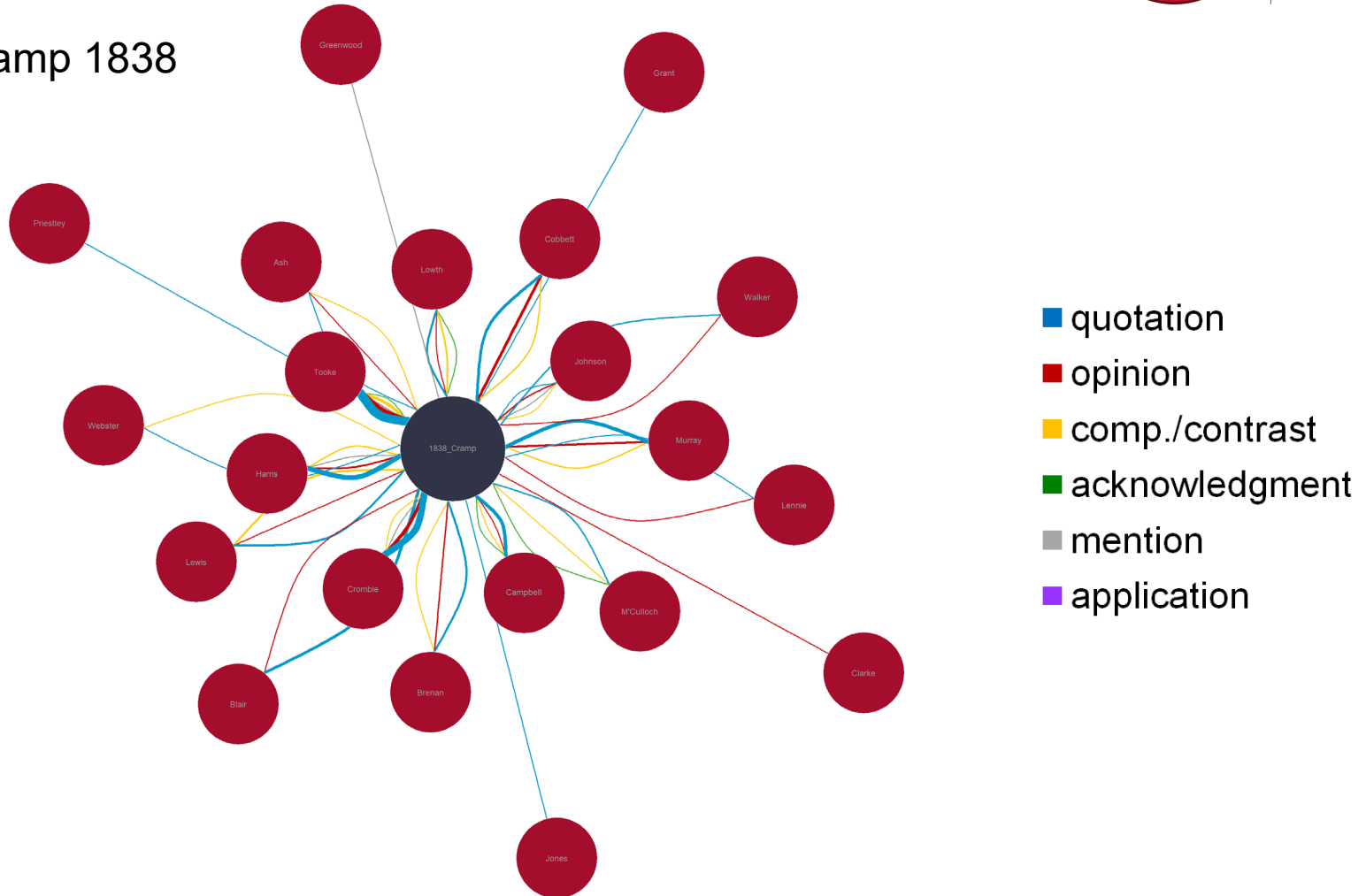
Crombie 1802



4. Main Results



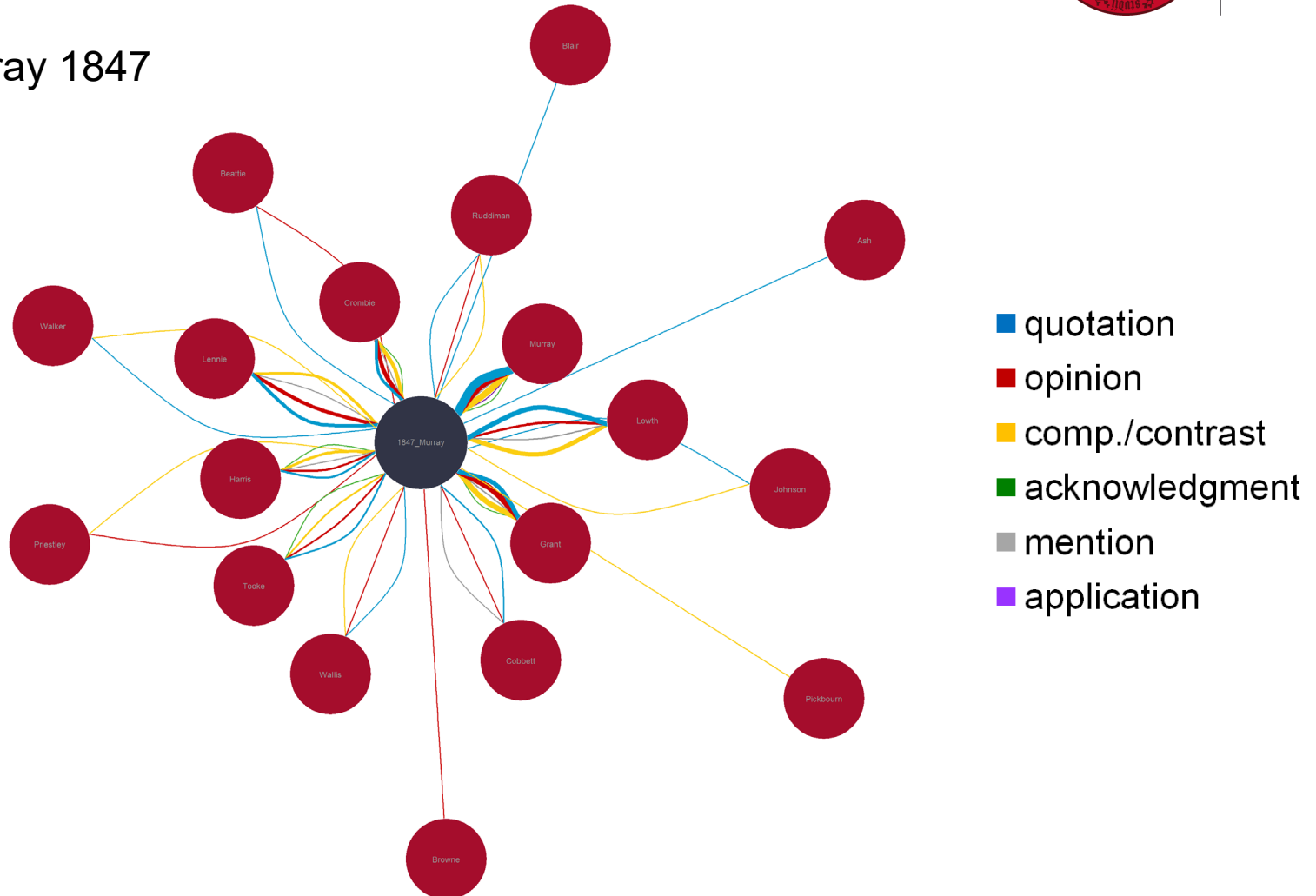
Cramp 1838



4. Main Results



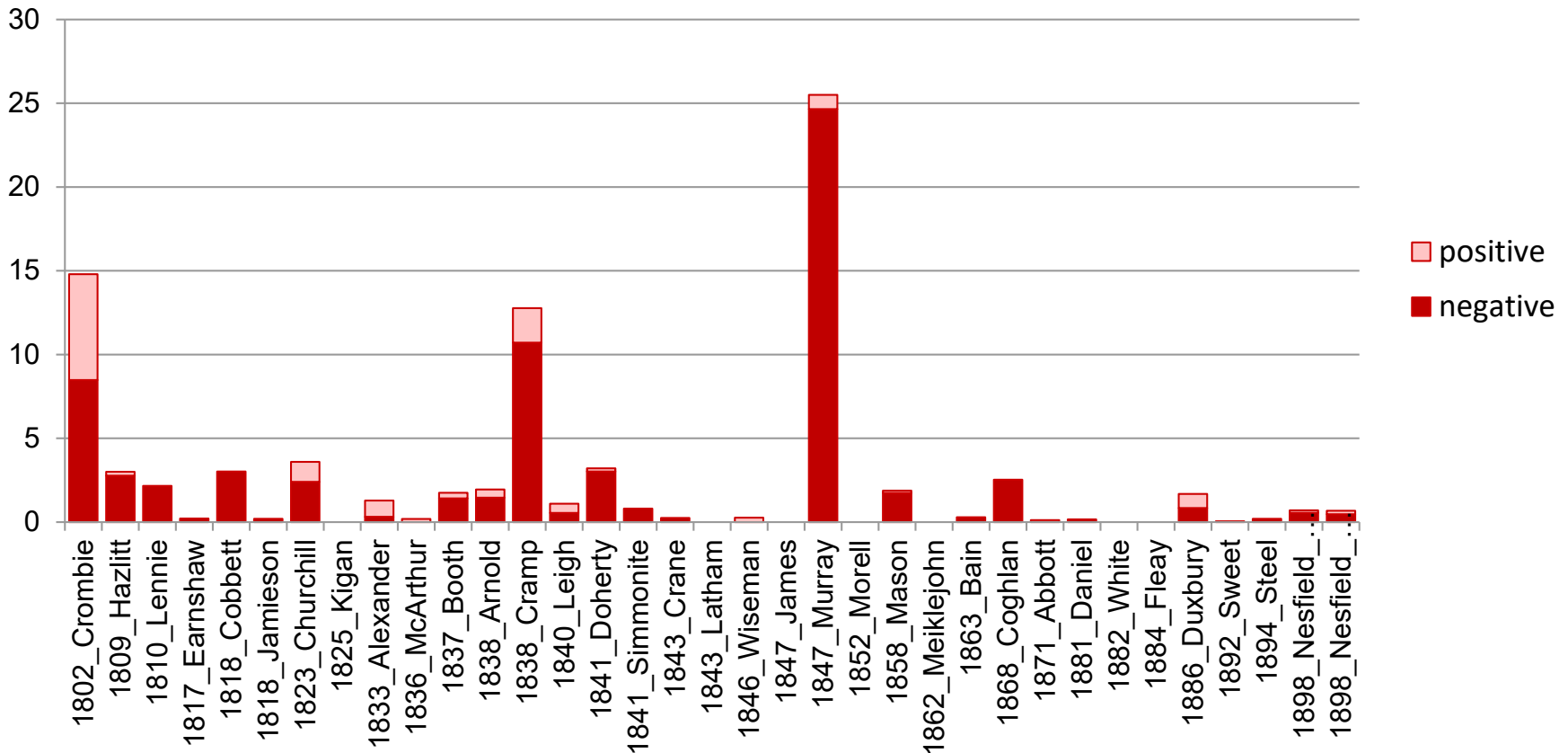
Murray 1847



4. Main Results



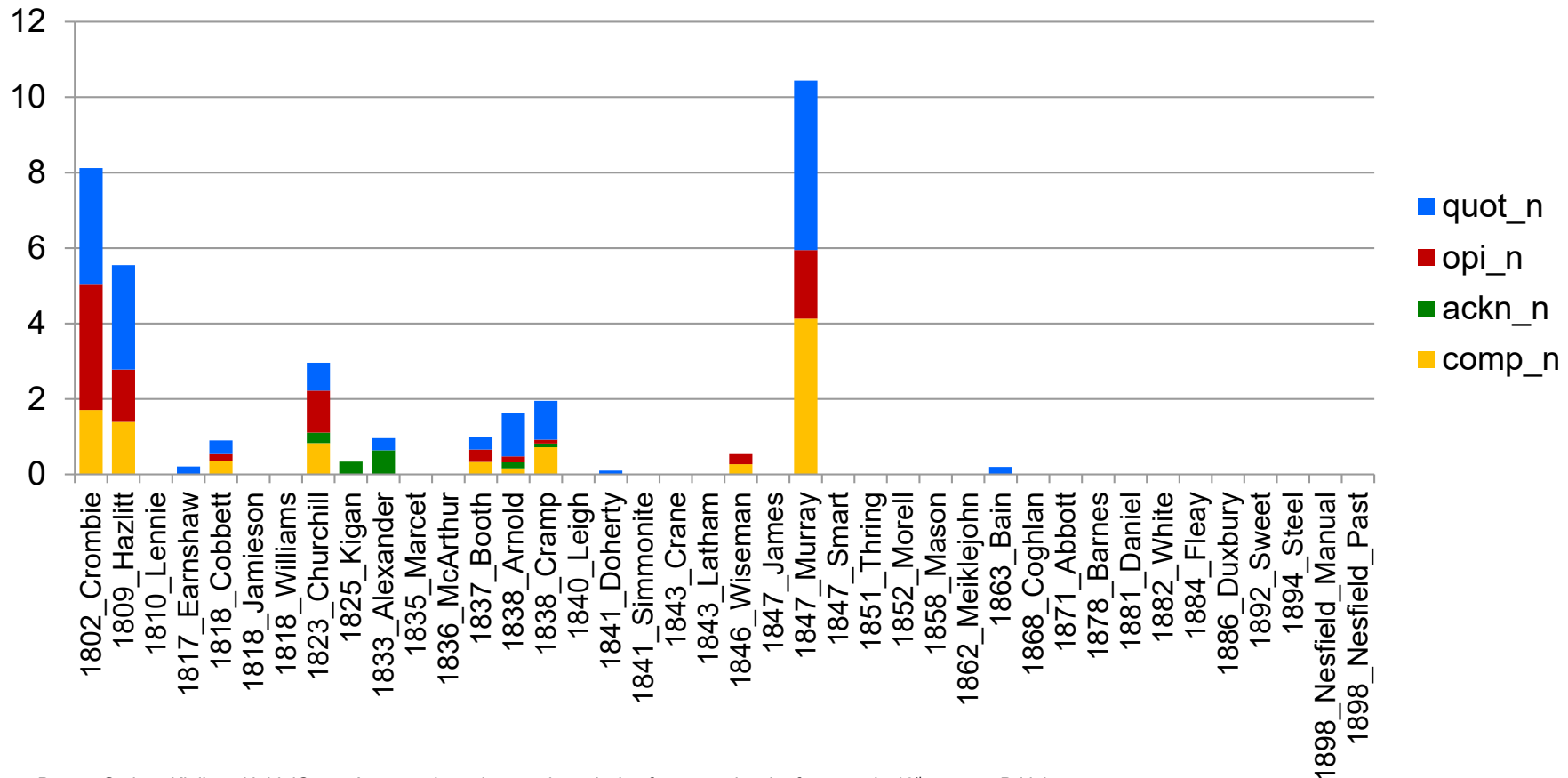
instances of positive and negative opinions (frequencies per 10,000w)



4. Main Results



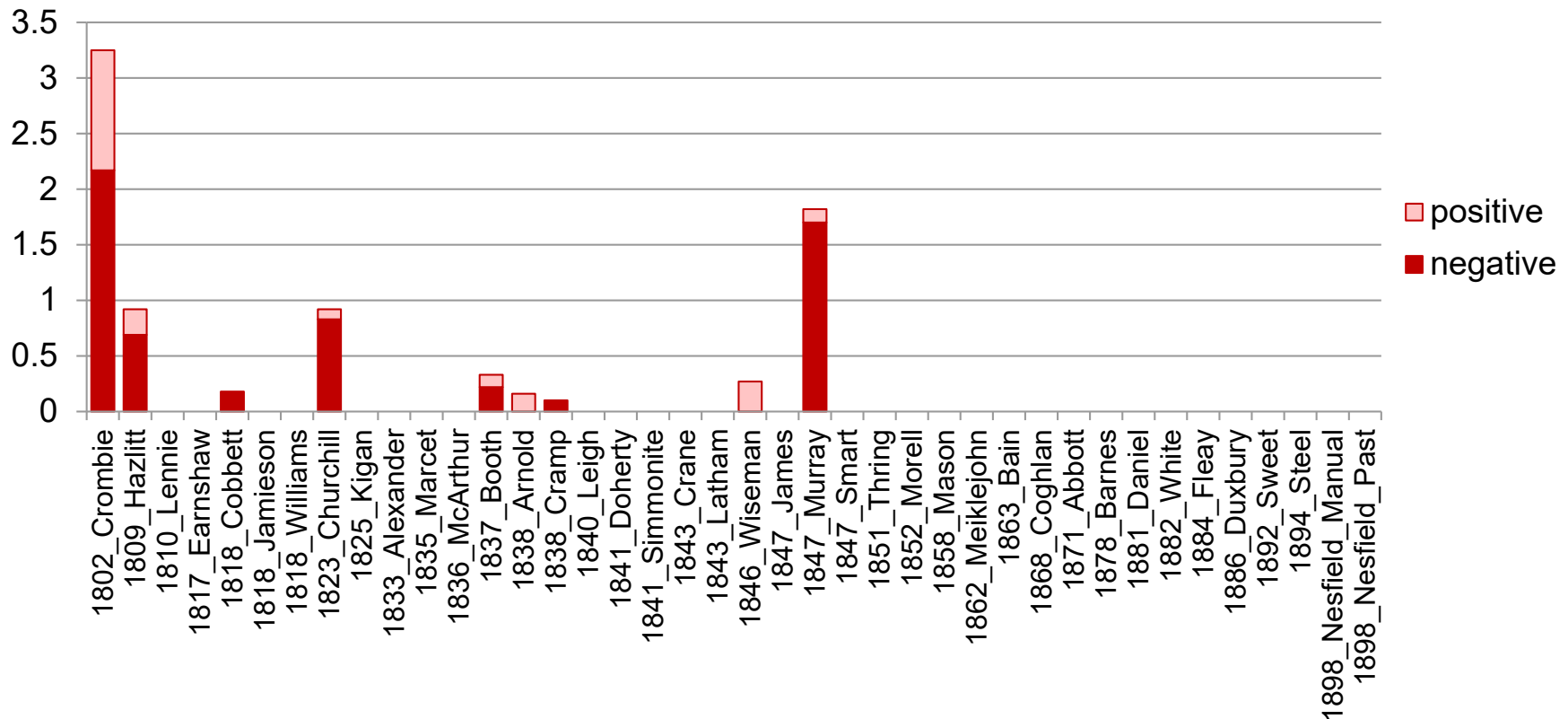
references to **Lowth (1762ff)** (frequencies per 10,000w)



4. Main Results



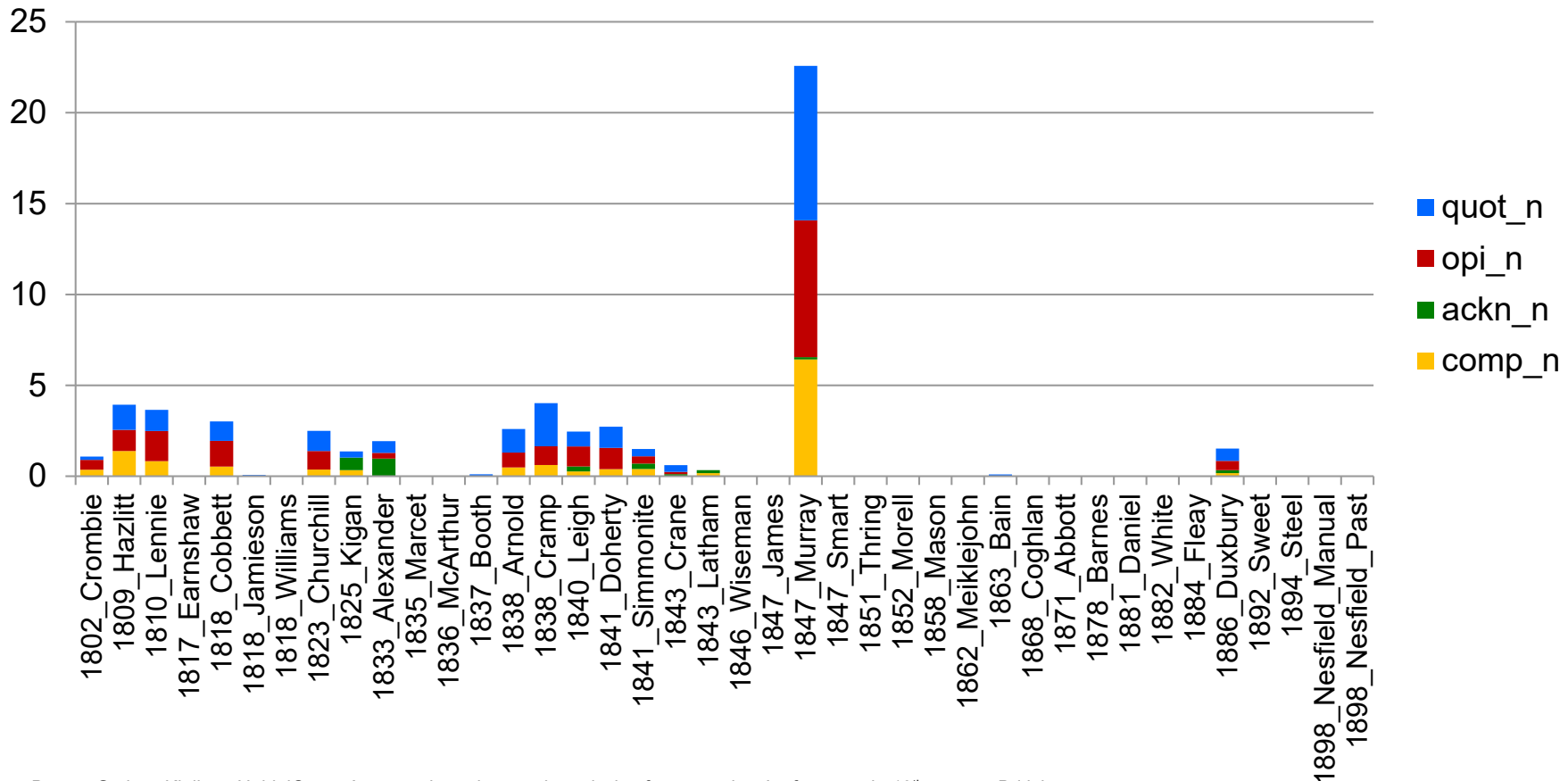
instances of positive/negative opinions with regard to **Lowth (1762ff)**
(frequencies per 10,000w)



4. Main Results



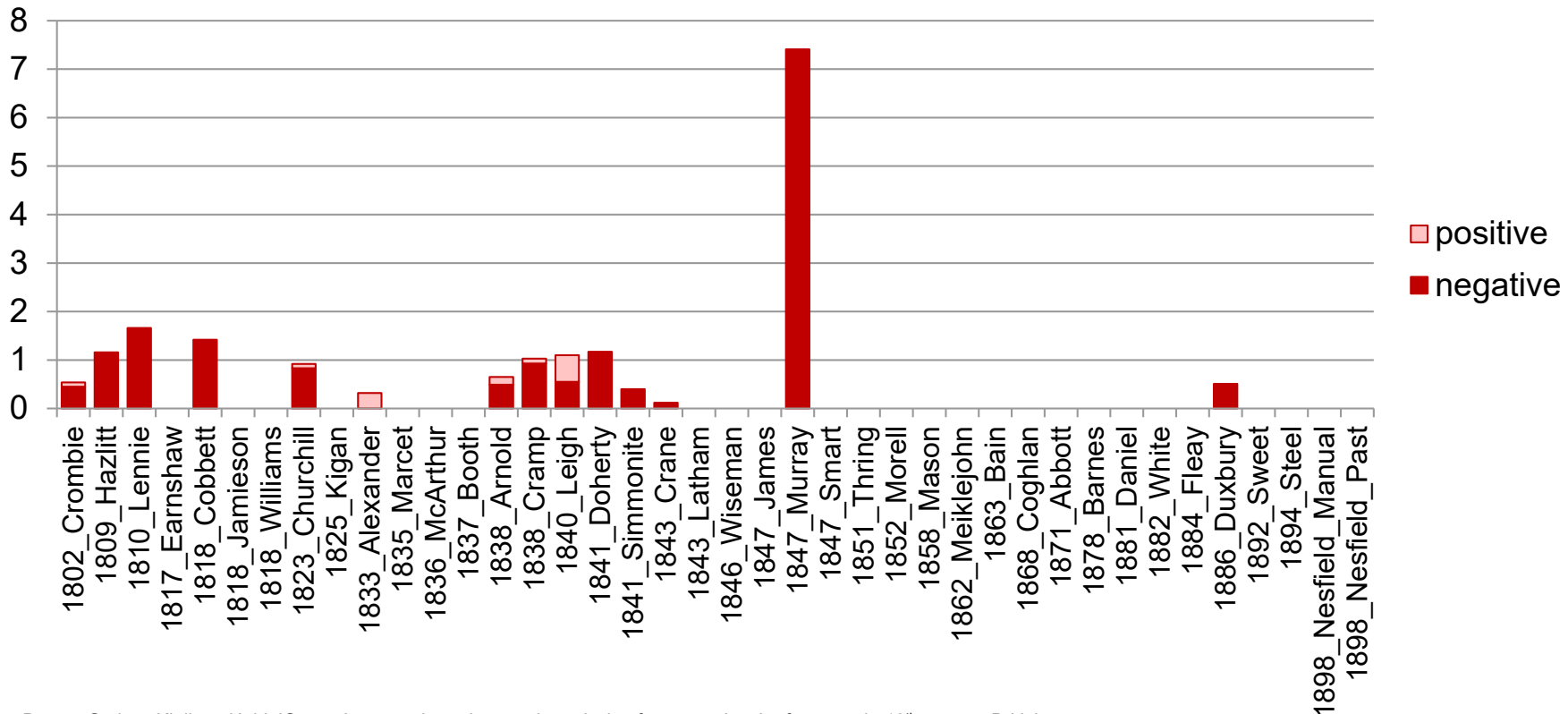
references to **Lindley Murray (1795ff)** (frequencies per 10,000w)



4. Main Results



instances of positive/negative opinions with regard to **Lindley Murray (1795ff)** (frequencies per 10,000w)



4. Main Results



Gerald Murray (1847)

"The definitions of the other tenses which we have examined strongly declare that Lindley Murray wrote without critically investigating the truth or accuracy of his own writings or those which he copied; but the definition of the pluperfect tense shows an unpardonable indifference to perspicuity, consistency, and common-sense."

"Doctor Crombie seemed to vie with L. Murray in inconsistency, and was wonderfully successful in his endeavours."



5. Summary and Conclusion

5. Summary and Conclusion



approx. 2,800 references in 35 out of 40 grammar books

very few references to grammars written between 1640 and 1750

many references to grammarians that wrote their major works in the second half of the 18th century, especially to the prescriptivists Lowth and Lindley Murray, and to the philosophical grammar of Horne Tooke

almost all of these references occur in grammars of the first half of the 19th century

break around 1850:

- ♦ turn away from 'traditional' (esp. prescriptivist) grammar writing
- ♦ references generally less frequent
- ♦ occupation with contemporary grammars

5. Summary and Conclusion



most of the references are quotations, followed by opinions, and instances of comparison and contrast

break around 1850 particularly in these three categories, less so in acknowledgments

three out of 40 grammarians (G. Murray, Cramp, Crombie) refer to other grammarians excessively (but for different reasons), but only Crombie's grammar is referenced by later grammarians

→ balance of positive/negative opinions relevant? (more data needed)

references to the two major 'prescriptivists' Lowth and L. Murray:

- ♦ instances of negative opinions by far outweigh positive remarks
- ♦ almost no acknowledgments whatsoever

→ turn away from prescriptive grammar writing not surprising

5. Summary and Conclusion



in general

- ♦ mixed-methods approach (corpus linguistics and network analysis) leads to noteworthy results
- ♦ findings of previous study could be confirmed and refined
- ♦ consideration of central problems (e.g. with regard to data quality, inter-rater reliability, network visualisation) necessary in order to obtain valuable results



Thank you for your attention!

Project Link and Bibliography



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Link to the *HeidelGram* Project: <https://heidelgram.de>



Bibliography

- Anderwald, Lieselotte. 2016. *Language Between Description and Prescription: Verbs and Verb Categories in Nineteenth-Century Grammars of English*. Oxford: Oxford University Press.
- Görlach, Manfred. 1998. *An Annotated Bibliography of Nineteenth-Century Grammars of English*. Amsterdam/Philadelphia: John Benjamins.
- Michael, Ian. 1987. *The Teaching of English*. Cambridge: Cambridge University Press.
- White, Howard D. 2011. "Scientific and Scholarly Networks". In: Scott, John and Peter J. Carrington (eds.). *The SAGE Handbook of Social Network Analysis*. London: SAGE Publications Ltd., 271-285.