

Universität
zu Köln

Funded by

DFG

Deutsche
Forschungsgemeinschaft
German Research Foundation

Universität
zu Köln



HeidelGram Workshop

Friday, April 8th 2022

Agenda

Block A: (10:00-11:15)

HeidelGram: Network Analyses in 16th- and 19th-Century Grammars

Block B: (11:30-13:00)

Tracing and Automatizing Verbal Hygiene and Grammatical Terminology

Block C: (13:15-14:00)

Associated PhD Research – *A Corpus-Based Analysis of American Grammaticography until 1900*

Purpose of the Workshop

- Presenting our work thus far
- Feedback and suggestions on the project, and some specific issues
- Comments on future endeavours

HeidelGram: Network Analyses in 16th- and 19th-Century

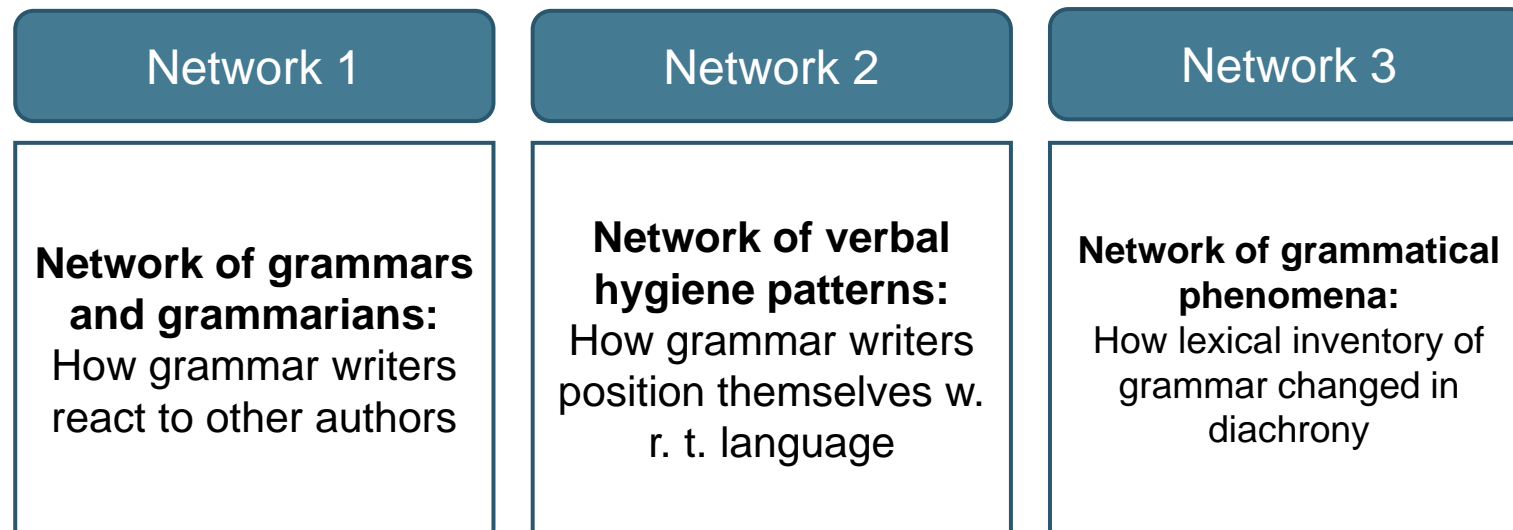
Grammars

Block A

HeidelGram – Project Overview

Overarching Aims

- Compile and analyse a corpus of 16th- to 19th-century historical English grammars
- Innovatively combine corpus-linguistic and network-analytical approaches
- Investigating changing as well as stable discourses of historical English grammar writing



HeidelGram Corpus – Data Selection

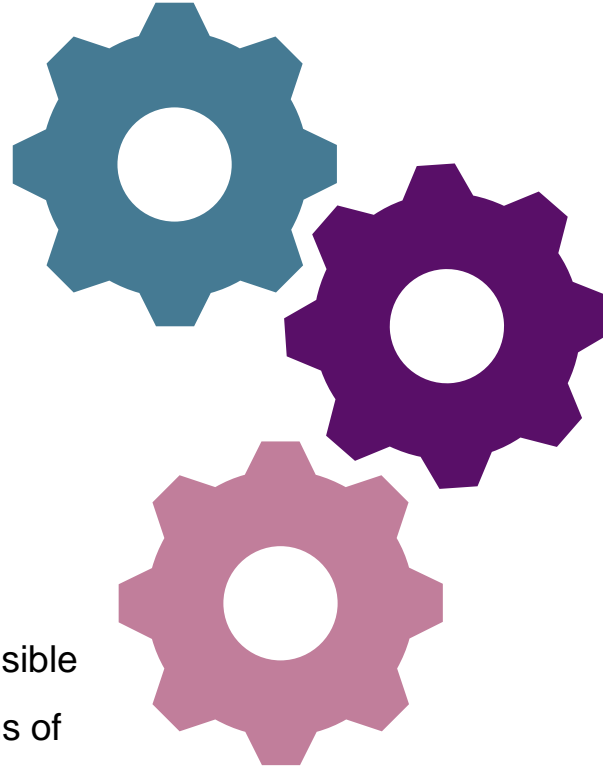
Definition of Grammar:

Combining a conventional approach (“the rules and conventions of everyday language” (McCarthy 2020: 4)) with the idea of verbal hygiene (“practices to ‘clean up’ language” (Cameron [1995] 2012: vii))

Variety of Texts:

Covering the widest possible range of variability in terms of

- Function
- Audience
- Text form



Popularity and Distribution:

- 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.

HeidelGram Corpus – 19th Century

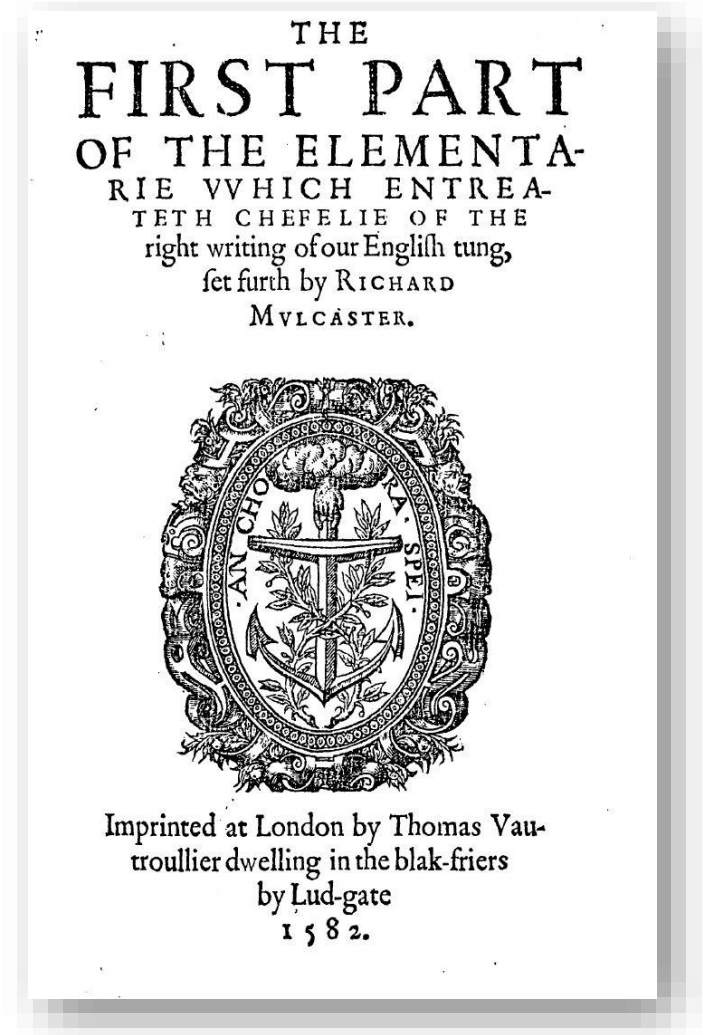
1. Alexander Crombie (1802) <i>The Etymology and Syntax of the English Language, Explained and Illustrated</i>	18. Percival Leigh (1840) <i>The Comic English Grammar</i>	34. Thomas Clarke (1862) <i>Student's Handbook of Comparative Grammar</i>
2. Joseph Taylor (1804) <i>A System of English Grammar</i>	19. Hugh Doherty (1841) <i>An Introduction to English Grammar, on Universal Principles</i>	35. John Meiklejohn (1862-66) <i>An Easy English Grammar for Beginners</i>
3. William Hazlitt (1809) <i>A New and Improved Grammar of the English Tongue</i>	20. Wiliam Joseph Simmonite (1841) <i>The Practical Self-teaching Grammar of the English Language</i>	36. Alexander Bain (1863) <i>An English Grammar</i>
4. William Lennie (1810) <i>The principles of English grammar briefly defined, and neatly arranged</i>	21. George Crane (1843) <i>The Principles of Language; Exemplified in a Practical English Grammar</i>	37. John Coghlan (1868) <i>Reformed English Grammar</i>
5. Dominique St. Quentin (1812) <i>The First Rudiments of General Grammar</i>	22. Robert Gordon Latham (1843) <i>An Elementary English Grammar</i>	38. Edwin A. Abbott (1871) <i>English Lessons for English People</i>
6. Christopher Earnshaw (1817) <i>The Grammatical Remembrancer</i>	23. Thomas John Wiseman (1846) <i>A School Grammar of the English Language</i>	39. Alexander F. Murison (1875) <i>First work in English: grammar and composition taught by a comparative study of equivalent forms</i>
7. David Williams (1818) <i>The Catechism of English Grammar</i>	24. James Hutchinson (1847) <i>Juvenile Grammar</i>	40. William Barnes (1878) <i>An outline of English speech-craft</i>
8. William Cobbett (1818) <i>A Grammar of the English Language, In a Series of Letters</i>	25. Benjamin H. Smart (1847) <i>Grammar on its True Basis</i>	41. Rev. Evan Daniel (1881) <i>The Grammar, History and Derivation of the English Language</i>
9. Alexander Jamieson (1818) <i>A grammar of rhetoric and polite literature</i>	26. J.H. James (1847) <i>The Elements of Grammar, according to Dr. Becker's System</i>	42. James Edward of Kendal Singleton (1882) <i>Notes on lessons on English grammar for the use of teachers in elementary schools</i>
10. T.O. Churchill (1823) <i>A New Grammar of the English Language</i>	27. Gerald Murray (1847) <i>The Reformed Grammar, or Philosophical Test of English Composition</i>	43. Frederick Averde White (1882) <i>English Grammar</i>
11. John Kigan (1825) <i>A Practical English Grammar, agreeably to a new System</i>	28. Edward Shelley (1848) <i>The People's Grammar; or English Grammar without Difficulties for 'the Million'</i>	44. Frederick Gard Fleay (1884) <i>The logical English grammar</i>
12. Levy Alexander (1833) <i>The young lady and gentleman's guide to the grammar of the English language</i>	29. James Douglas (1851) <i>Principles of English Grammar</i>	45. C. Duxbury (1886) <i>A New English Grammar of School Grammars</i>
13. Jane Marcet (1835) <i>Mary's Grammar</i>	30. Rev. Edward Thring (1851) <i>The Elements of Grammar Taught in English</i>	46. Henry Sweet (1892/98) <i>A New English Grammar, Logical and Historical</i>
14. Alexander McArthur (1836) <i>An outline of English grammar for the use of schools</i>	31. John Daniel Morell (1852) <i>The analysis of sentences explained and systematised: with an exposition of the fundamental laws of syntax. After the plan of Becker's German grammar</i>	47. G. Steel (1894) <i>An English grammar and analysis for students and young teachers</i>
15. David Booth (1837) <i>The Principles of English Grammar</i>	32. James Millar (1855) <i>Outlines of English Grammar; Arranged for being taught on the Intellectual System</i>	48. John Earle (1898) <i>A Simple Grammar of English Now in Use</i>
16. William Cramp (1838) <i>The Philosophy of Language</i>	33. Charles Peter Mason (1858) <i>English Grammar; including the Principles of Grammatical Analysis</i>	49. John Nesfield (1898) <i>English Grammar: Past and Present</i>
17. Thomas K. Arnold (1838) <i>An English Grammar for Classical Schools</i>		50. John Nesfield (1898) <i>Manual on English Grammar and Composition</i>

➤ approx. 2.7 million words in total (for 40-book pilot corpus)

HeidelGram Corpus – 16th Century

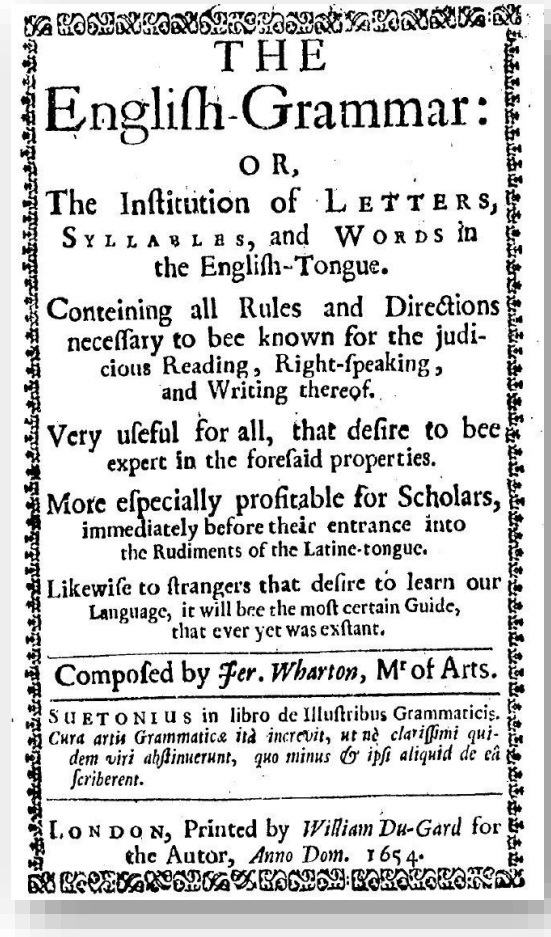
1. Richard Sherry (1577) *A Treatise of the Figures of Grammar and Rhetorike*
2. Richard Mulcaster (1582) *The First Part of the Elementarie*
3. William Bullokar (1586) *Brief Grammar of English*
4. Gabriel Meurier (1586) *The Conjugations in Englishe and Netherdutche*
5. Edmund Coote (1596) *The English Schoole-Maister*

➤ 182,628 words in total



HeidelGram Corpus – 17th Century

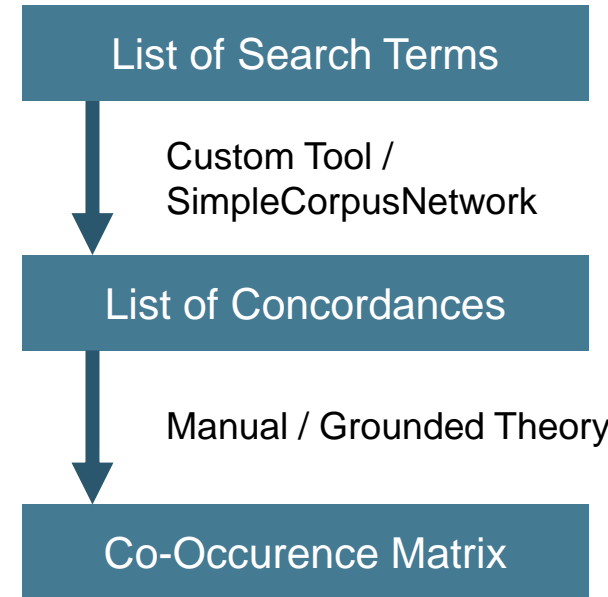
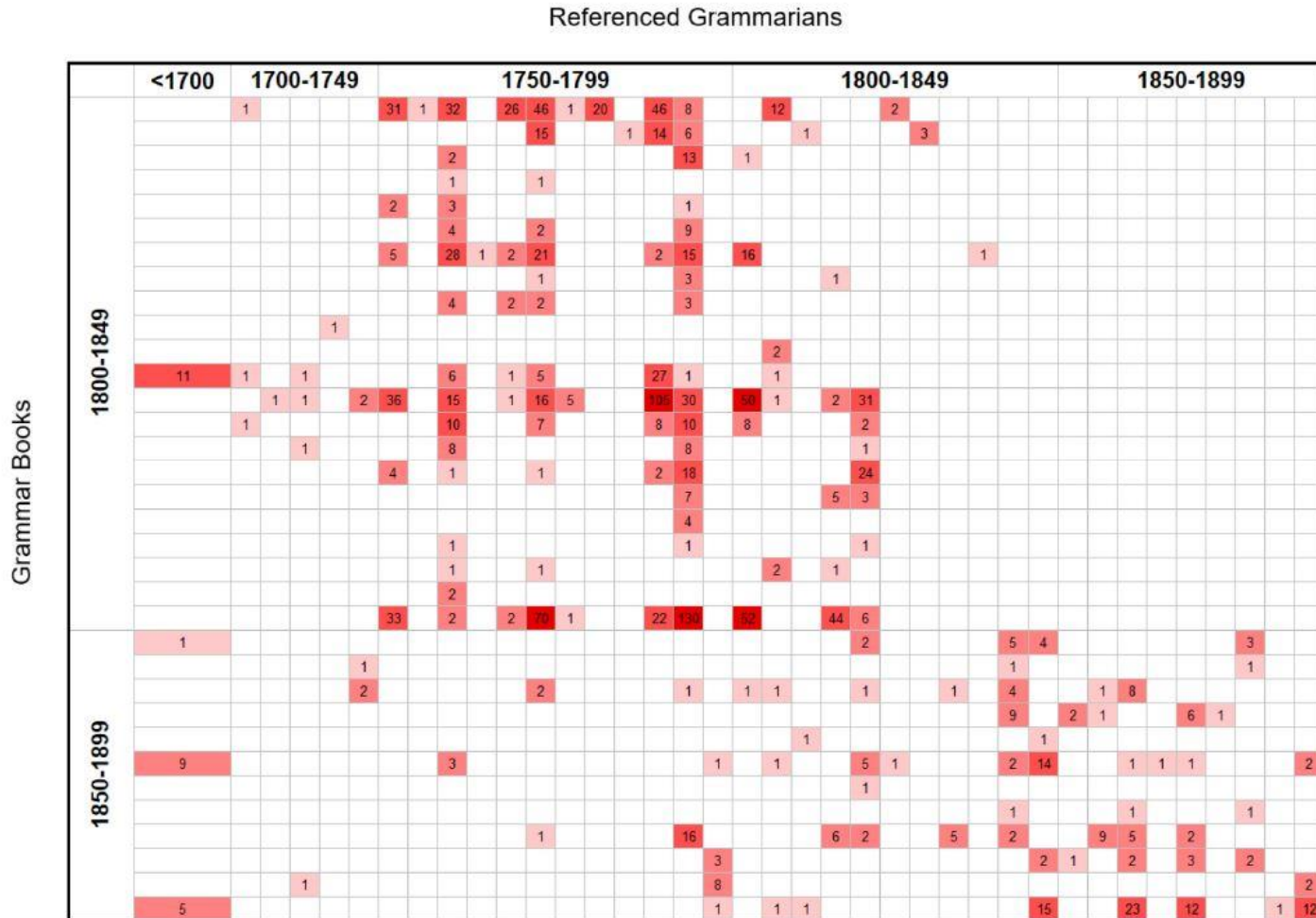
- | | |
|--|--|
| <ol style="list-style-type: none"> 1. Alexander Hume (1617) <i>Orthographie and Congruitie of the Britan Tongue</i> 2. John Hewes (1624) <i>A Perfect Survey</i> 3. John Brinsley (1630) <i>The Posing of the Parts</i> 4. Charles Butler (1633) <i>English Grammar</i> 5. Ben Jonson (1640) <i>The English Grammar</i> 6. Joshua Poole (1646) <i>The English Accidence</i> 7. Francis Lodwick (1647) <i>A Common Writing</i> 8. Jeremiah Wharton (1654) <i>The English Grammar</i> 9. James Howell (1662) <i>A New English Grammar</i> 10. John Wilkins (1668) <i>An Essay towards a Real Character, and a Philosophical Language</i> 11. John Newton (1669) <i>School Pastime for Young Children: or the Rudiments of Grammar</i> 12. M. Lewis (1675) <i>Plain & Short Rules</i> | <ol style="list-style-type: none"> 13. John Newton (1677) <i>The English Academy, or, a Brief Introduction to the Seven Liberal Arts</i> 14. Richard Blome (1686) <i>The Gentleman's Recreation</i> 15. Christopher Cooper (1687) <i>The English Teacher (English translation of Grammatica Ling. Angl.)</i> 16. Guy Miège (1688) <i>The English Grammar</i> 17. Maurice Wheeler (1690) <i>The Royal Grammar</i> 18. Joseph Aickin (1693) <i>The English grammar</i> |
|--|--|



HeidelGram Corpus – 18th Century

<ol style="list-style-type: none"> 1. A. Lane (1700) <i>A Key to the Art of Letters</i> 2. Richard Browne (1700) <i>The English School Reformed</i> 3. John Brightland (1711) <i>A Grammar of the English tongue</i> 4. James Greenwood (1711) <i>Essay towards a practical English grammar</i> 5. Michael Mattaire (1712) <i>The English Grammar</i> 6. Hugh Jones (1724) <i>An Accidence to the English Tongue</i> 7. Daniel Duncan (1731) <i>A New English Grammar</i> 8. Isaac Barker (1733) <i>An English Grammar</i> 9. John Collyer (1735) <i>The General Principles of Grammar</i> 10. Solomon Lowe (1737) <i>English Grammar Reformed</i> 11. Ann Fisher (1745) <i>A New Grammar</i> 12. John Kirkby (1746) <i>A New English Grammar</i> 13. John Wesley (1748) <i>A Short English Grammar</i> 	<ol style="list-style-type: none"> 14. Benjamin Martin (1748) <i>Institutions of Language</i> 15. James Harris (1751) <i>Hermes, or a Philosophical Inquiry Concerning Language and Universal Grammar</i> 16. Thomas Dilworth (1751) <i>A New Guide to the English Tongue</i> 17. Joseph Priestley (1761) <i>The Rudiments of English Grammar</i> 18. Robert Lowth (1762) <i>A short introduction to English grammar with critical notes</i> 19. James Buchanan (1762) <i>The British Grammar; or, an Essay, in Four Parts, towards Speaking and Writing the English Language Grammatically, and Inditing Elegantly</i> 20. John Ash (1763) <i>Grammatical institutes or, An easy introduction to Dr. Lowth's English grammar</i> 21. William Ward (1765) <i>An Essay on English Grammar</i> 22. Samuel Johnson (1766) <i>A dictionary of the English Language... to which is prefixed, a Grammar</i> 	<ol style="list-style-type: none"> 23. Robert Baker (1770) <i>Reflections on the English Language</i> 24. Daniel Fenning (1771) <i>A New Grammar of the English Language; or, an easy introduction to the art of speaking and writing English with propriety and correctness</i> 25. Dorothea Du Bois (1772) <i>Short English Grammar</i> 26. Ellin Devis (1775) <i>The Accidence; or First Rudiments of English Grammar</i> 27. John Horne Tooke (1786) <i>Epea Pteroenta, or The Diversions of Purley</i> 28. Lindley Murray (1795) <i>English grammar adapted to the different classes of learners</i> 29. Benjamin Rhodes (1795) <i>A Concise English Grammar</i> 30. Ellenor Fenn (1799) <i>The Child's Grammar</i>
--	--	--

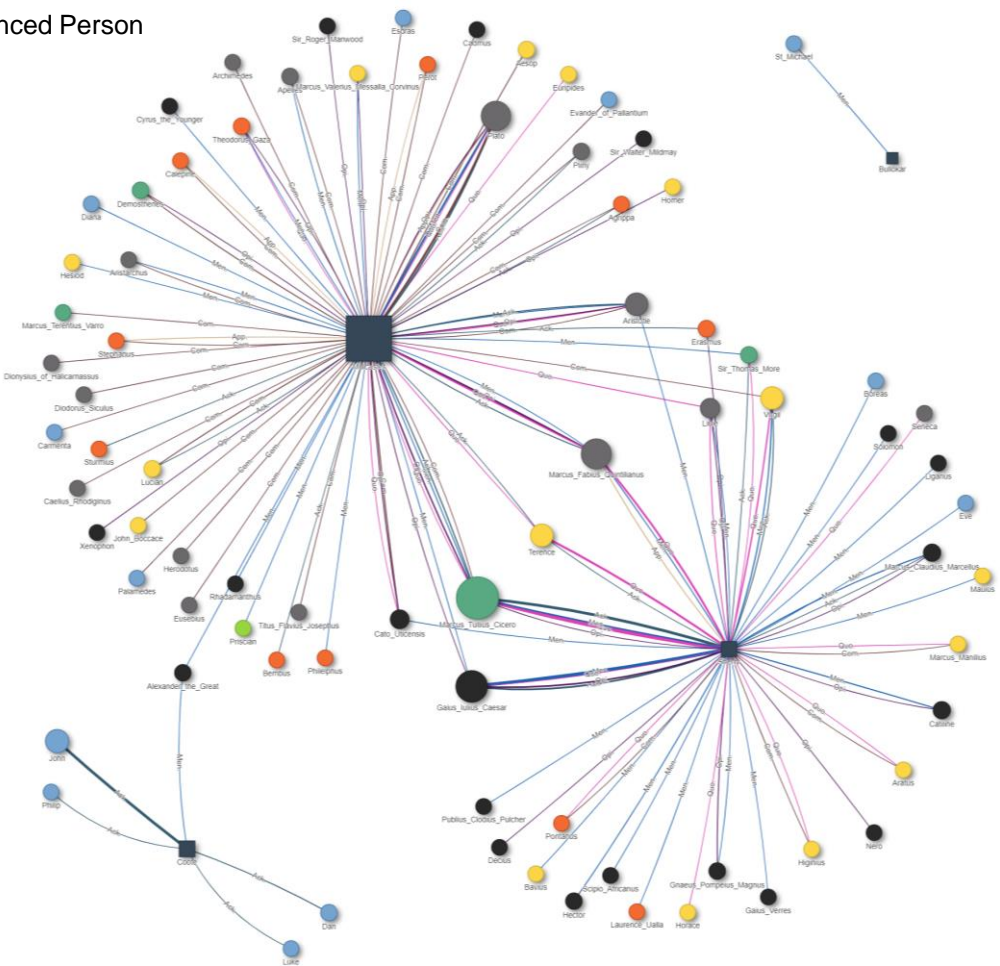
19th-Century Citation Networks: Method



	Grammar 1	Grammar 2
Term 1	2	4
Term 2	8	0
...		

16th- and 19th-Century Citation Networks

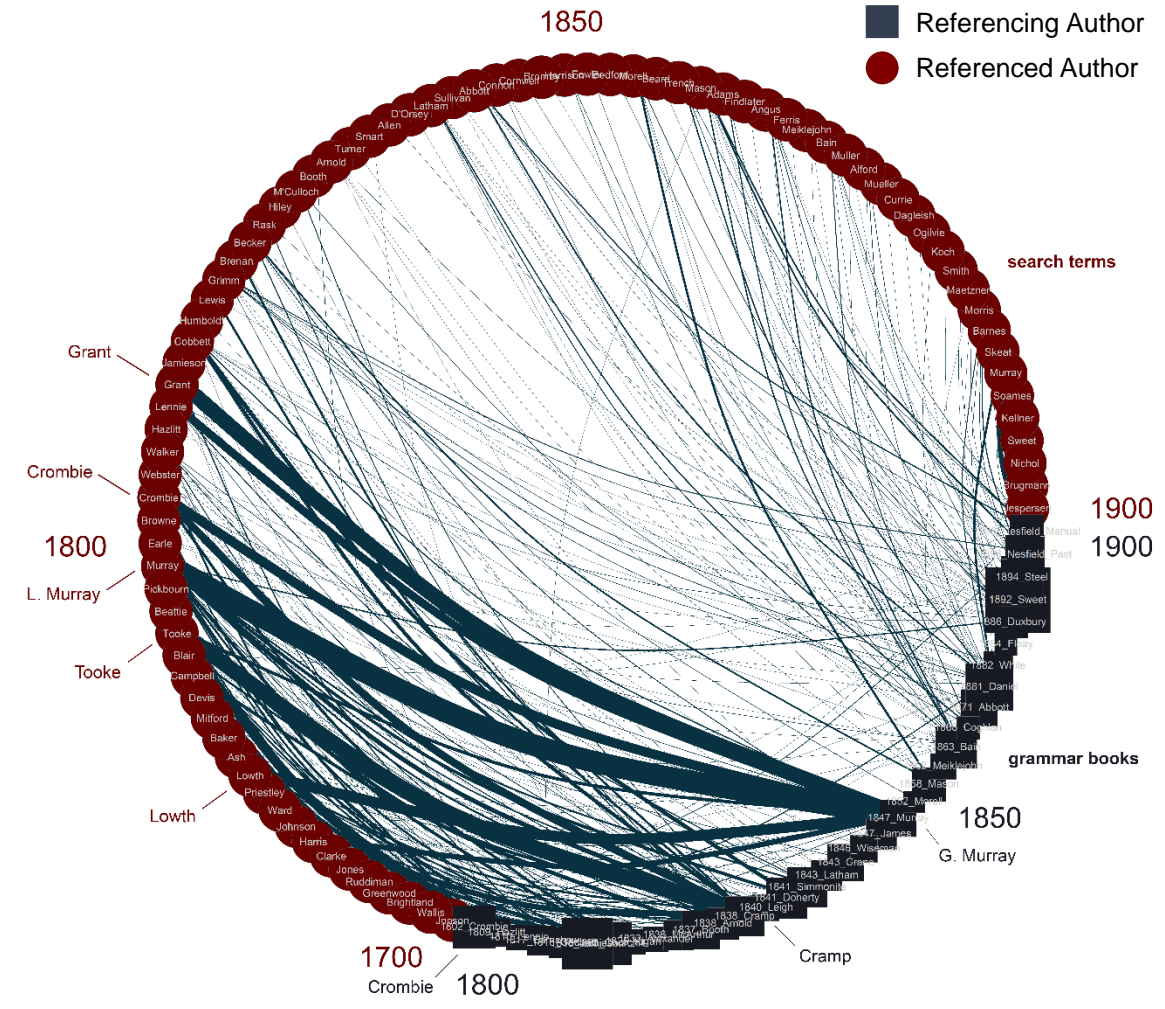
Referencing Author
 Referenced Person



Network of 16th-century references, indicating reference types (edge color) and author types (circular node colors).



Referencing Author
 Referenced Author



Network of 19th-century references, indicating changes in reference frequency over time.

Example References: 19th Century

Opinion

„**Murray's** rule, as it stood, is clearly repugnant to perspicuity, and syntactical correctness.“
(Crombie 1802: 205)

Application

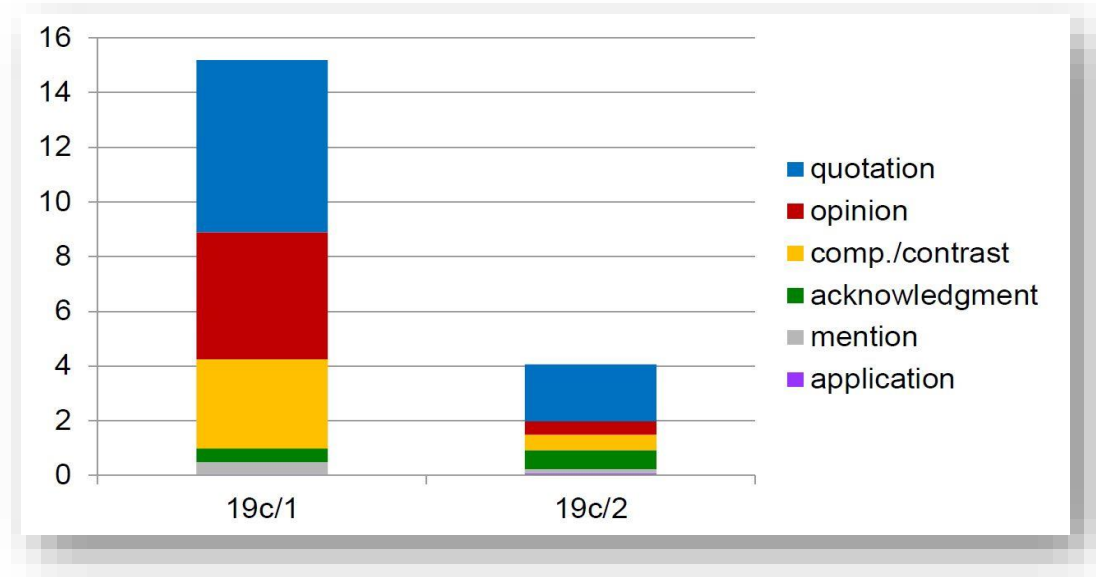
„Speaking of the defender of a fortress , it may be said , 'He *will* die rather than surrender, ' which, by **Dr. Wallis 's** Rule , would be ungrammatical , because *will* is here in the third person [...]“
(Booth 1837: 116)

Opinion + Quotation

“**Lowth** justly observes, that *or* is here improperly used for *nor*, the correlative words being *neither, nor*.”
(Crombie 1802: 332)

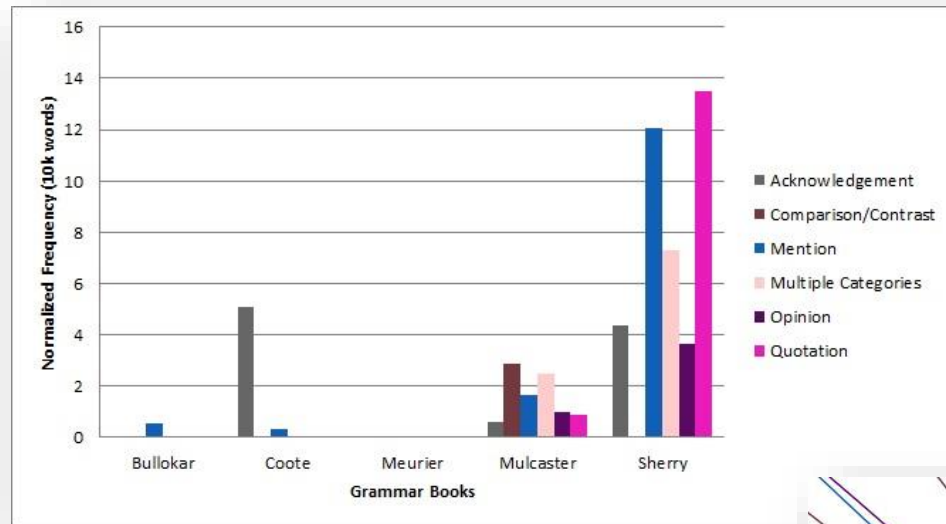


19th-Century Reference Categories



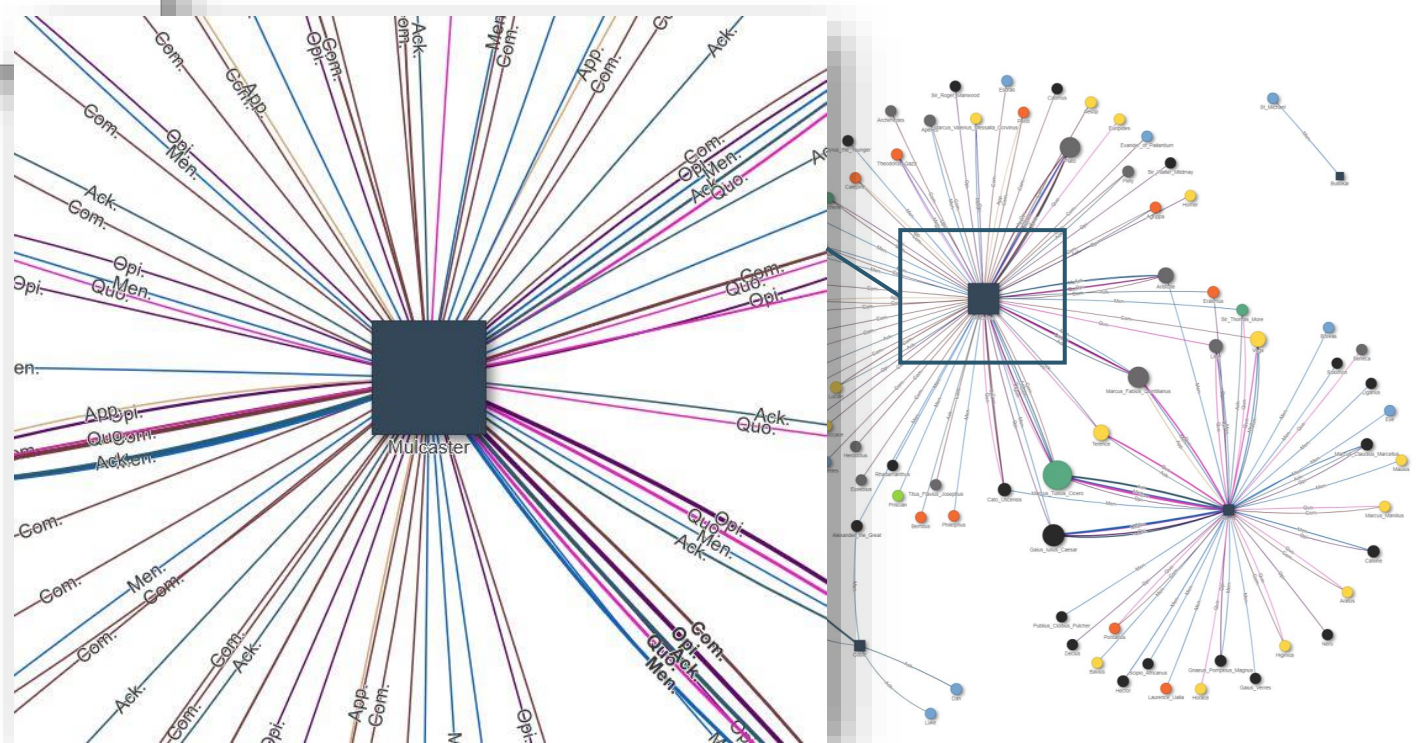
- Based on onomastic references
- Based on grounded theory, achieving high IRR scores
- Sudden decrease in references after 1850, especially to prescriptive grammars
- Categories of quotation, opinion, and comparison/contrast decrease markedly
- Opinions were mostly negative

16th-Century Reference Categories



- Categories for all five books (due to sparsity of data a diachronic split was not useful)
- Sherry and Mulcaster refer most to other authors / persons
- There is no clear pattern
- Rather few Opinions or Comparisons/Contrasts

Reference categories originally established for the 19th-century data were applicable to the 16th-century data with high IRR scores.

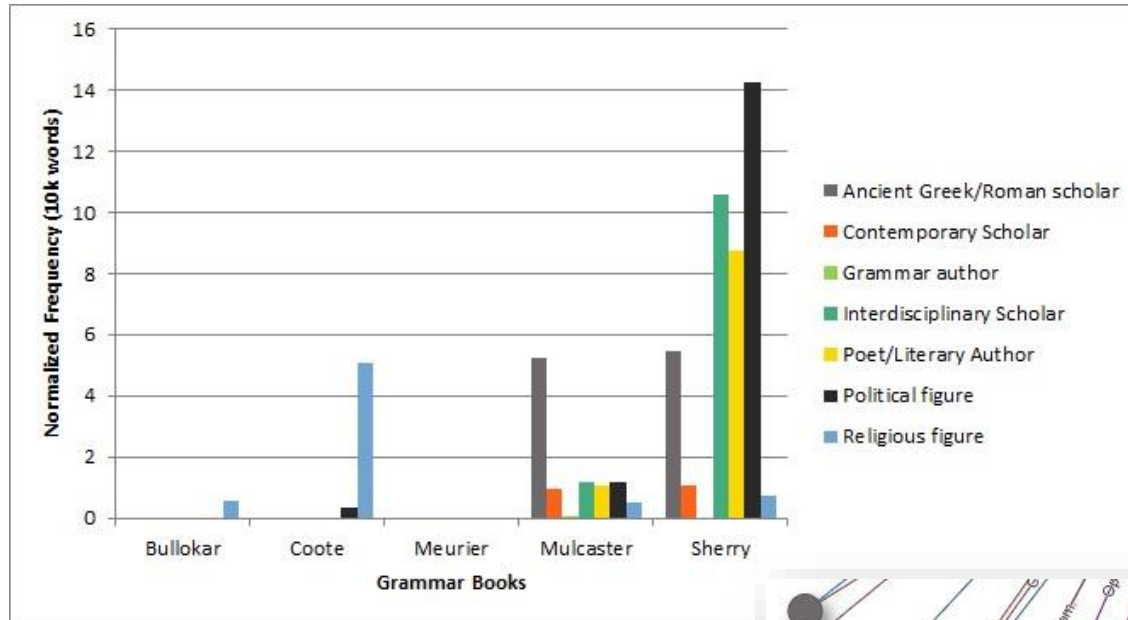


Example References

16th Century	Opinion	19th Century
„In expressing of these among the Latines, Liuius is very cunning.“ (Sherry 1577: xlvi)		„ Murray’s rule, as it stood, is clearly repugnant to perspicuity, and syntactical correctness.“ (Crombie 1802: 205)
Application		
„In the mean while that rule of Aristotle must be preciselie kept, whereby we ar taught, that the best waie to learn anie thing well, which must afterward be don, when it is learned, is full to be a doing, while we be a learning.“ (Mulcaster 1582: 228)		„Speaking of the defender of a fortress , it may be said , 'He <i>will</i> die rather than surrender, ' which, by Dr. Wallis 's Rule , would be ungrammatical , because <i>will</i> is here in the third person [...]“ (Booth 1837: 116)
Opinion + Quotation		
„This is well handled of Cicero in the preface of the third boke of his Offices: that Scipio was wont to saye, he was neuer lesse ydle then whe he was voyde of the common wealthe matters, and neuer lesse alone, the whe he was alone.“ (Sherry 1577: lv.)		“ Lowth justly observes, that <i>or</i> is here improperly used for <i>nor</i> , the correlative words being <i>neither, nor</i> .” (Crombie 1802: 332)

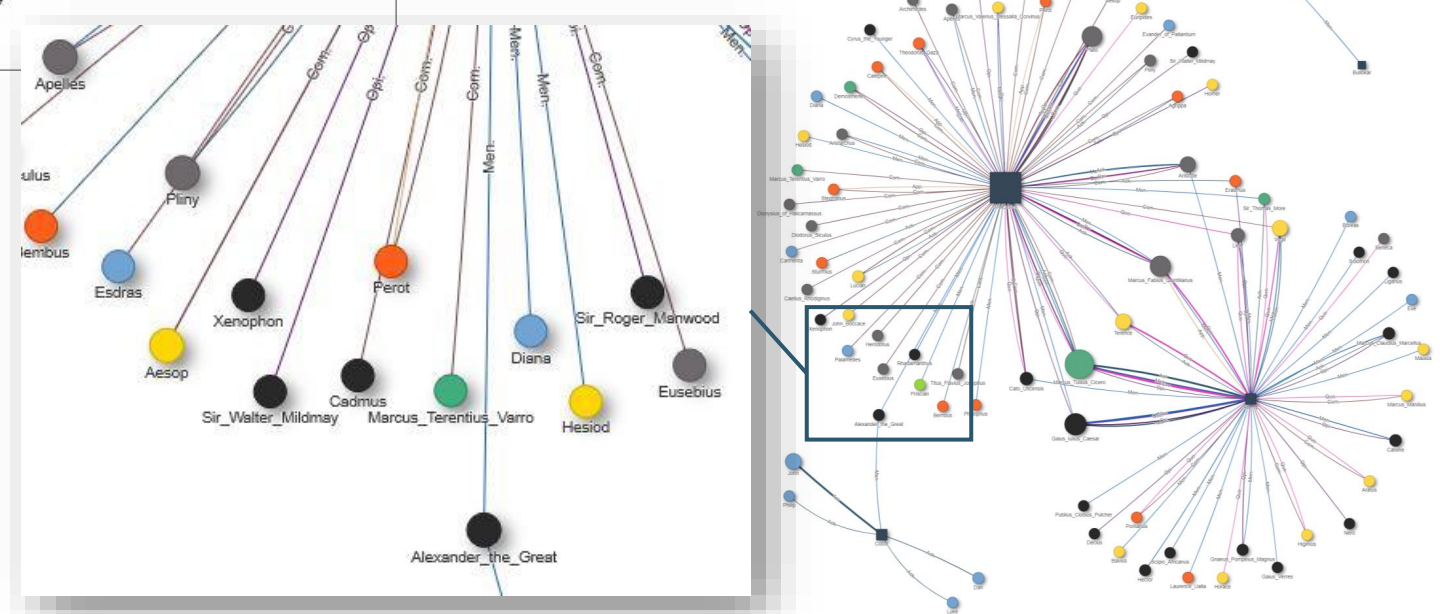


16th Century – Categories of Authors

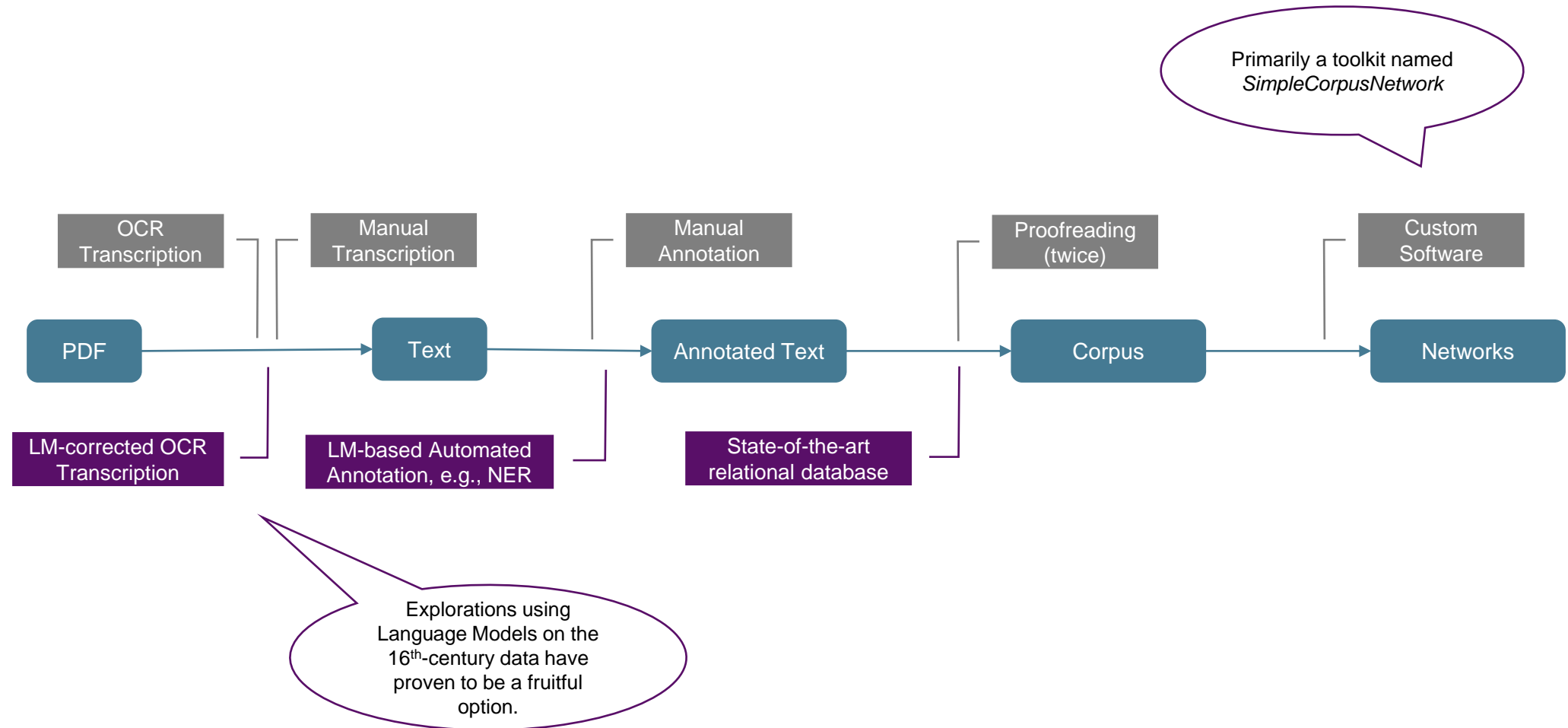


- There is no clear pattern
- No references are made to English grammar authors
- Most references are made to learned scholars from ancient Greek or Rome

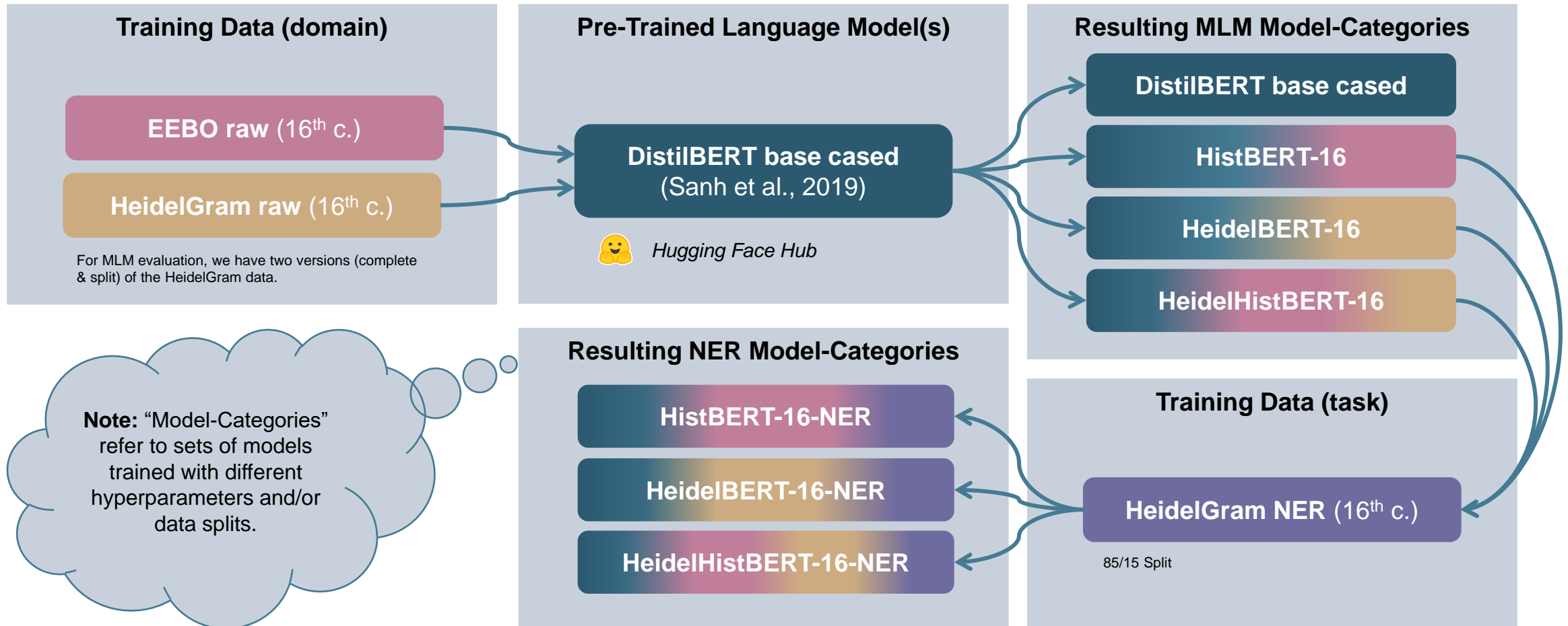
19th-century data will be revisited to identify different author types as well.



Methodology – Corpus Compilation



Methodology – Language Models and NER



Methodology – SimpleCorpusNetwork

SimpleCorpusNetwork is a CLI **corpus (analysis) toolkit** focused on **network/graph approaches**. It's designed to work with **text files** and within (semi-)automated data **processing/analysis pipelines**.

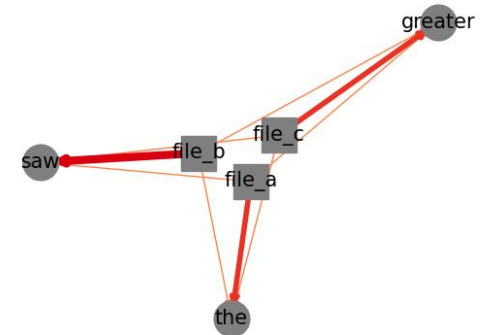
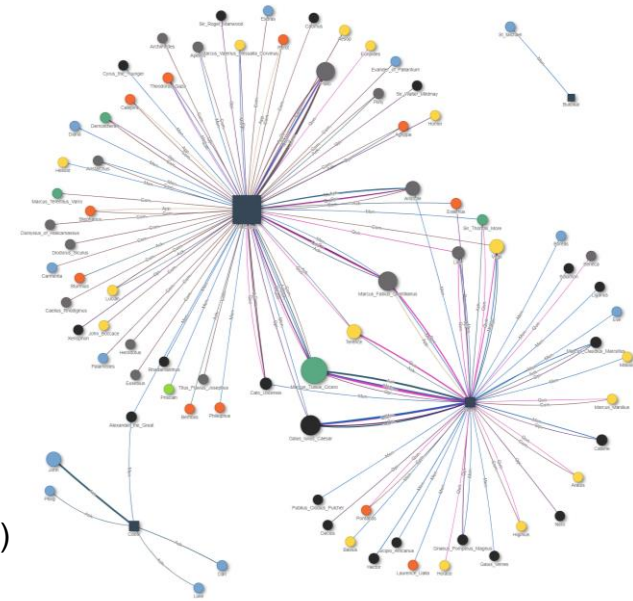
While its primary focus lies on **extracting different types of network/graph data** from corpora, it can also be used to, for example:

- Generate term-document matrices (frequency; concordances) from error-ridden (OCR) corpus data (fuzzy search)
- Automatically fix (OCR) errors in corpora based on complex rulesets
- Work with various taggers (e.g., semantic tagging) and language models
- Manipulate network/graph data

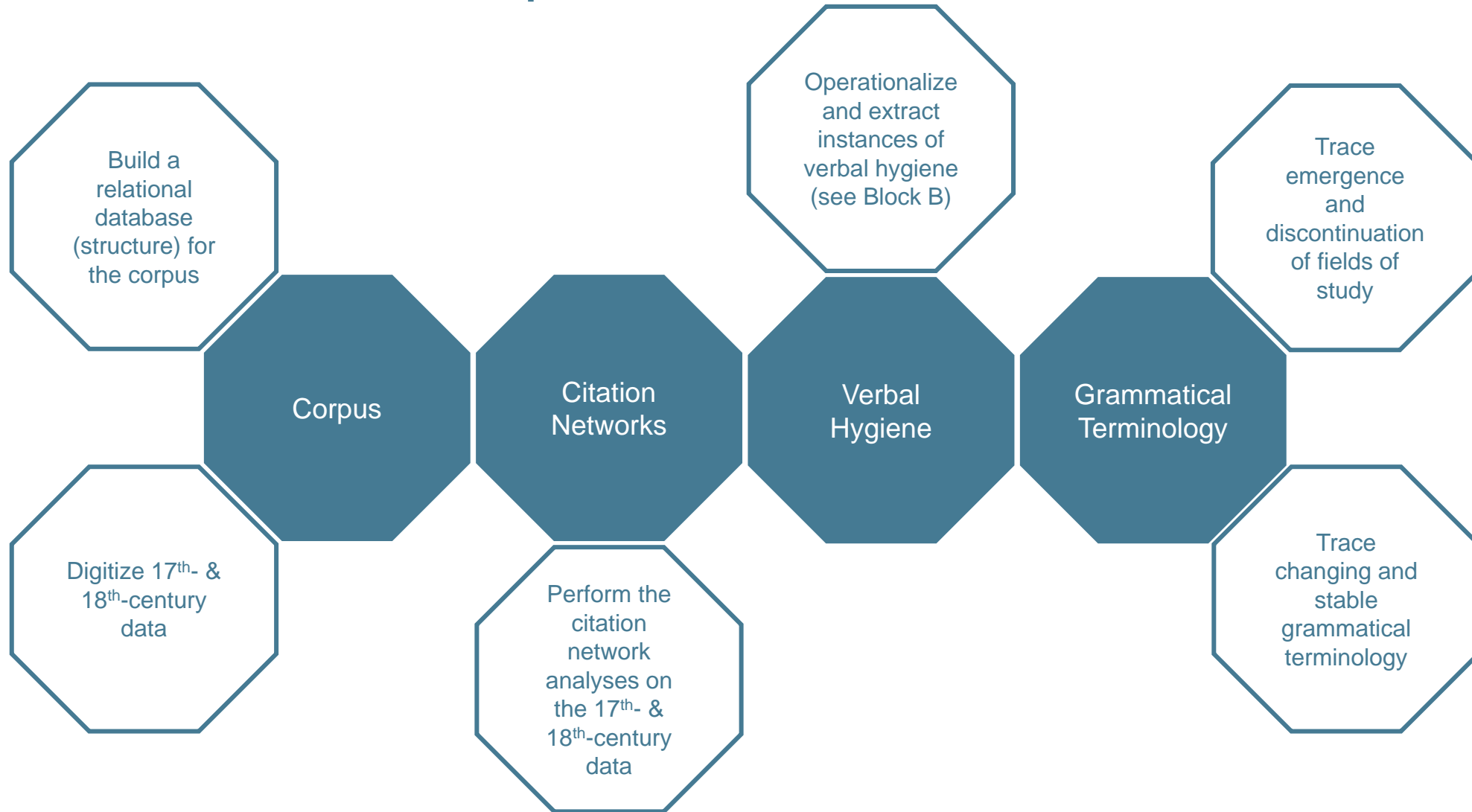
So far, SimpleCorpusNetwork has been used to generate **all networks/graphs used in the HeidelGram project**. Hence, the toolkit and its functionality grew alongside the project.

Challenges:

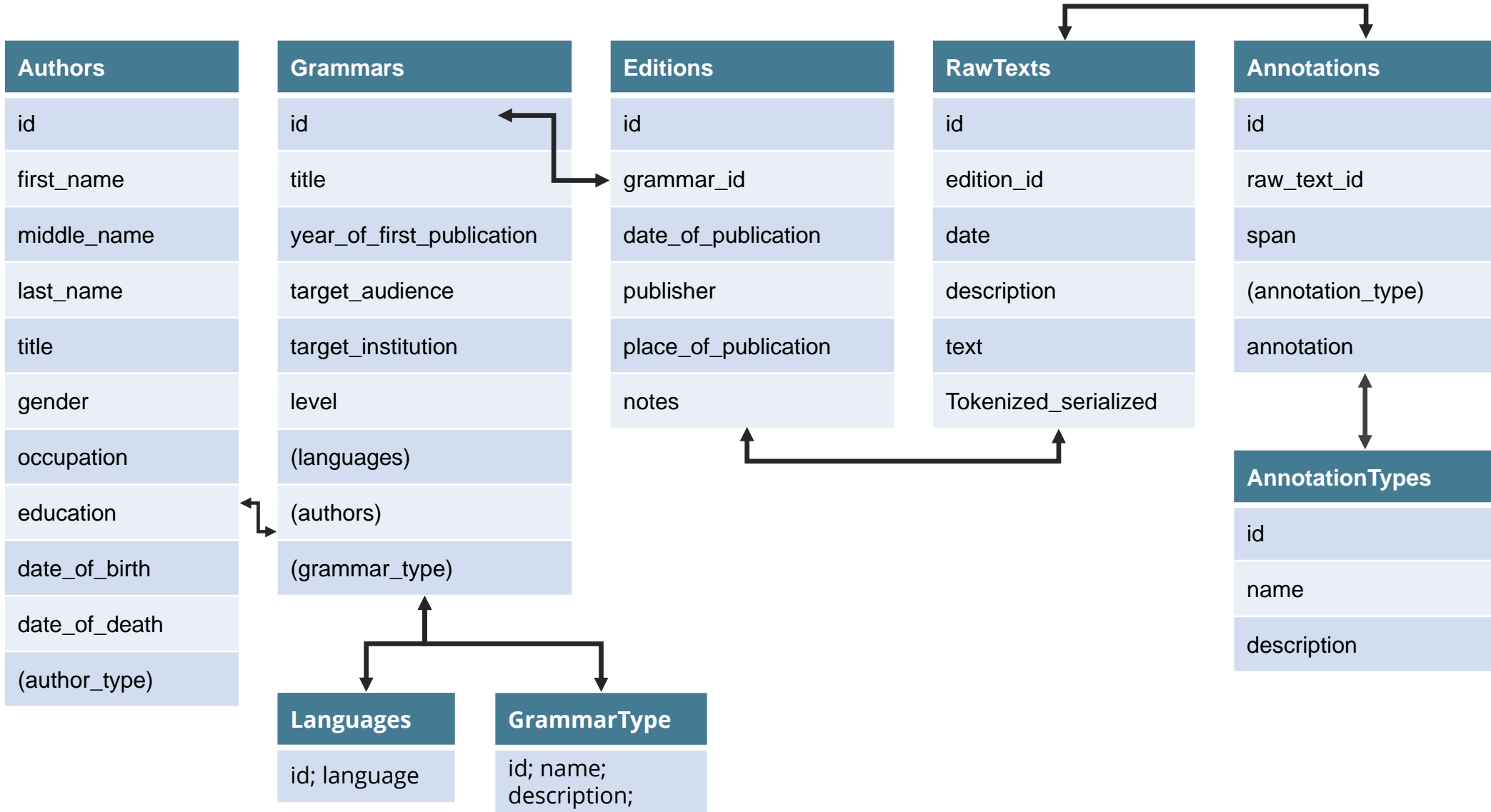
- (1) The toolkit needs to be **rewritten** to be used with a **relational database**.
- (2) As the tool has grown alongside the project, its quite **idiosyncratic**. While this works for the project, it makes documenting and publishing the toolkit hard. → *The plan is to open-source the tool alongside the corpus itself*



Goals and Next Steps



HeidelGram Database



HeidelGram Database – Example

‘Dummy Grammar’ by ‘Dummy Author’

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Pellentesque elementum nibh pulvinar tristique convallis. Quisque turpis metus, mattis id nisi vel, cursus pulvinar nisi. Proin a ultrices est. Nunc sodales leo tempus eros porttitor, a volutpat urna consequat. Maecenas fringilla urna urna, sit amet luctus leo accumsan rhoncus. Aliquam erat volutpat. Etiam ullamcorper facilisis lectus, ut interdum massa laoreet ut. Sed arcu lorem, vulputate vel augue a, sodales porttitor lacus. Ut in tellus sollicitudin, ultrices arcu non, ultricies elit. Proin eu vestibulum leo. In lobortis consequat maximus. Maecenas a enim sed est lobortis lobortis vel nec mi. Nam vel felis.

Edition 0, 2021

Authors

0; Dummy Author

Grammars

12; Dummy Grammar

Editions

0; 12; 2021

RawTexts

0; 0; Lorem ipsum ...

Annotations

0; 0; [0-4]; 0; NN

AnnotationTypes

0; pos

In **text 0** at **position [0-4]** we find an annotation of **type 0** that is ‘NN’

Switching to a database approach will require major changes in the **tooling** and **data-processing pipelines**.

Feedback and Discussion

Database:

- How should one structure sub-textual elements (e.g., tokens, texts, etc.)?
- How would you, technically speaking, 'store' both the documents and the annotations?
- With regard to tooling and data formats, how would you like to access the corpus data?

Data Selection:

- Where and how would you draw the line between a 'grammar' and similar types of books, e.g., usage guides?
- Are there any grammars missing from our list that seem significant to you?

Citation Networks:

- How can we account for non-name-based references, such as anaphora?
- Is there a way to automatically identify different references to the same person i.e. how would you deal with spelling variation?
e.g., Caesar, Iulius Caesar, Julius Caesar, Cesar, etc.



Tracing and Automatizing Verbal Hygiene and Grammatical Terminology

Block B

Early Definition of Verbal Hygiene

“Verbal hygiene –my term for the motley collection of discourses and practices through which people attempt to ‘clean up’ language and make its structure or its use conform more closely to their ideals of beauty, truth, efficiency, logic, correctness, and civility –is not just an unnatural and futile enterprise rooted in a failure to appreciate how language works. Rather verbal hygiene is a *product* of the way language works: it is an outgrowth of the capacity of metalinguistic reflexivity which makes human linguistic communication so uniquely flexible and nuanced” (Cameron [1995] 2012: vii).

Verbal hygiene:

- “comes into being whenever people reflect on language in a critical (in the sense of ‘evaluative’) way” (Cameron 1995: 9)
- “[a]ll attitudes to language and linguistic change are fundamentally ideological” (Cameron 1995: 4)
- “impulse to regulate language, control it, make it ‘better’” (Cameron 1995: 9)
- “complaints about language changes are usually symbolic expressions of anxieties about larger social changes” (Cameron 2012a: 238).

Contemporary Work using Verbal Hygiene as a Framework

- Politically motivated language activism among youth in a high school semiotic landscape (Arman 2020)
- Bishop Oyedepo's crisis-motivated church sermon: the data showed controversial utterances, verbal attacks, utterances associated with physical assault and violence, bad premonitions, curses and unconfirmed reports (Idowu et al. 2021)
- Most of Cameron's examples of VH come from intra-linguistic discourses, i.e., talk about what is considered correct and appropriate (mainly) within the English language, however the concept is arguably applicable to inter-linguistic discourses too, i.e., how English as a lingua franca is prompting a social justice agenda (Hultgren 2020)
- ...

Defining Verbal Hygiene for the Context of Historical Data

Why do we focus on VH rather than distinguishing between descriptivism and prescriptivism?

- Verbal Hygiene ≠ prescriptivism and descriptivism
 - „even normative grammars are rarely entirely pre- or proscriptive in nature“ (Tieken-Boon van Ostade 2006: 545)
 - Verbal Hygiene is dynamic and allows for textual interpretation and for various speaker approaches and variation
 - It can be used throughout the corpus and for a variety of phenomena
 - It is not merely concerned with sociological paradigms such as gendered languages, political correctness in discourse, euphemisms, or age-related linguistic phenomena.
 - In our research we explore genre in more depth and conduct a meta-analysis of how authors of books that discuss language use language themselves
 - We combine Verbal Hygiene with Grounded Theory and expand the term to have a broader meaning
 - By using the Verbal Hygiene concept, we can look at how (sociocultural/political) ideology and rhetoric style is reflected in historical data
- There is an opportunity with using this term is to also investigate the WHY rather than only the WHAT/HOW?

Examples from **Sherry** (1577)

- “Eloquution which the Greekes call Phrase, wherof also the name of Eloquence doth ryse: as of all partes **it is the goodliest, so is it also ye profitablest + hardest**, in whiche is seen that diuine might and vertue of an orator...”
- “And out of this great streame of Eloquution, **not onely must we choose apt + mete woordes, but also take hede of placyng and setting them in order...**”
- “And **well knowen was the counsel that I gaue.**”
- “And in this same kind (as Cicero saieth in his oratour) **some be crafty, but vnpolished, and of purpose like the rude and vnskilful. Other in yt leanes are trim, that is, mery coccited, florishing also, + a litle garnished**”.

Examples from **Mulcaster** (1582)

- “Quintilian also **a Rhetorik master** among the Romanes, and of no lesse account in his cuntrie then he was of cunning, and **so esteemed of among vs**, in the framing of his best orator”
- “in most plane terms, **if I thought it nedefull**”
- “This Elementarie am I now to perform. VVhose particular braches being manie in number, & the book thereby growing to som bulk, **I thought it good** to deuide it into parts, vpo sundrie causes, but chefelie for the printer, whose sale will be quik if the book be not big”

Examples from Basson (1586)

“And although that j for my parte, am willing to pleasure euerie one, according to my simple skil, yet thinking my selfe vnable to accoplisse such worke, j haue bin verie lothe to trouble my selfe there with all for feare of being reprehended, forasmuch as j knowe that manie are apter for to finde faultes, then to amend the same. And yet **seing that some haue taken vppon them the print Englishe by other speaches, hauing not any vnderstanding of the same**, as appeareth by their labour which they (not with standing with great dilligence) haue donne. So have j (out of loue which j beare to my natiue contrie men, and **also being more desirous then able to fulfill the desire of some of my freendes**) **imbouldened my selfe accordinge to my simple skill, to correct certaine of their worke.**”

Examples from **Bullokar** (1586)

- “**Both for the perfect picturing, of spech, and Grammar too:** not laeuing-out old letter, nor bringing new shaps for mo: Nor altering the senc of words, nor of sentenc the phras, but that aech volum, tym too com, may be raedd az it waz: And by my trauel English tryd, a perfect ruled tung, conferabl in Grammar-art, with any ruled long”.
- “A Nown-Adiectiu iz **a word not perfectly vnderstanded** except a nown-substantiu be iooynd with it: which substantiu iz known by the answer whoo?”

Examples from Coote (1596)

- “**O**ther men in their writings (gentle Reader) may iustly vse such stile, as may declare learning or eloquence fit for a scholer: **but I am enforced of necessitie to affect that plaine rudenes, which may best fit the capacitie of those persons, with whom I haue to deale. The learned sort are able to vnderstand my purpose,** and to teach this treatise without further direction, I am now therefore to direct my speech vnto the vnskilfull, which desire to make vse of it for their owne priuate benefit...”
- “**thou shalt learne many very profitable things that thou neuer knewest, yea thou shalt know more for the English tongue,** then any man of thy calling (not being a Grammarian) in England knoweth: thou shalt teach thy scholers with better commendation and profit then any other, not following this order, teacheth...”

Verbal Hygiene in the 19th Century Grammar Books

“Many writers, who ought to know better, blunder terribly in the attempt to turn an adjective clause into the reported form. 'That is the man who was so ill' is often modified into 'That is the man whom I heard was so ill.' **This is altogether wrong.** (See §382.) The only way of meeting the difficulty is to turn 'I heard' into a parenthesis, 'who (I heard) was ill'.”

(Mason 1858: 158, footnote Chapter: Syntax > Adjective Clauses)

Further Examples from 19th Century Grammars

- “greatly erred” (Crombie 1802)
- “shows an unpardonable indifference to perspicuity, consistency, and common sense” (Murray 1847)
- “correctly speaking” (Marcet 1835)

- bad arrangement / English / grammar / sense / spelling / style
- good English / sense / taste / writer / grammar / scholar / reason / quality

Advisor Feedback / Exchange on Proposed Theoretical Frameworks

Theoretical Questions

1. How do you understand verbal hygiene?
2. Do you consider Verbal Hygiene to be a productive theoretical framework within this context or would you instead use a different term, such as language management (Spolsky 2009) or evaluative terminology?



Advisor Feedback / Exchange on Proposed Methodology

1. Grammarians' approaches to verbal hygiene have probably changed throughout the centuries. How would you operationalize the search for VH patterns?
2. Methodologically speaking, how would you go about finding instances of Verbal Hygiene in a given text or corpus?
3. On which level would you focus when approaching Verbal Hygiene – authors, whole grammars, chapters, sentences, words, etc.?
4. Initially, we planned to assess Verbal Hygiene based on a list of evaluative and normative expressions. Do you think that this is a reasonable approach, or would you advise a more comprehensive approach?
5. Grammatical Terminology: Methodologically speaking, how would you go about finding lexemes referring to grammatical phenomena in a given text or corpus?



Automatizing the Search for Verbal Hygiene Patterns & Grammatical Terminology in the Corpus

Challenge 1: How to operationalize and automatize the process of extraction of verbal hygiene patterns historically?

Challenge 2: How to operationalize and automatize the extraction of grammatical terminology?

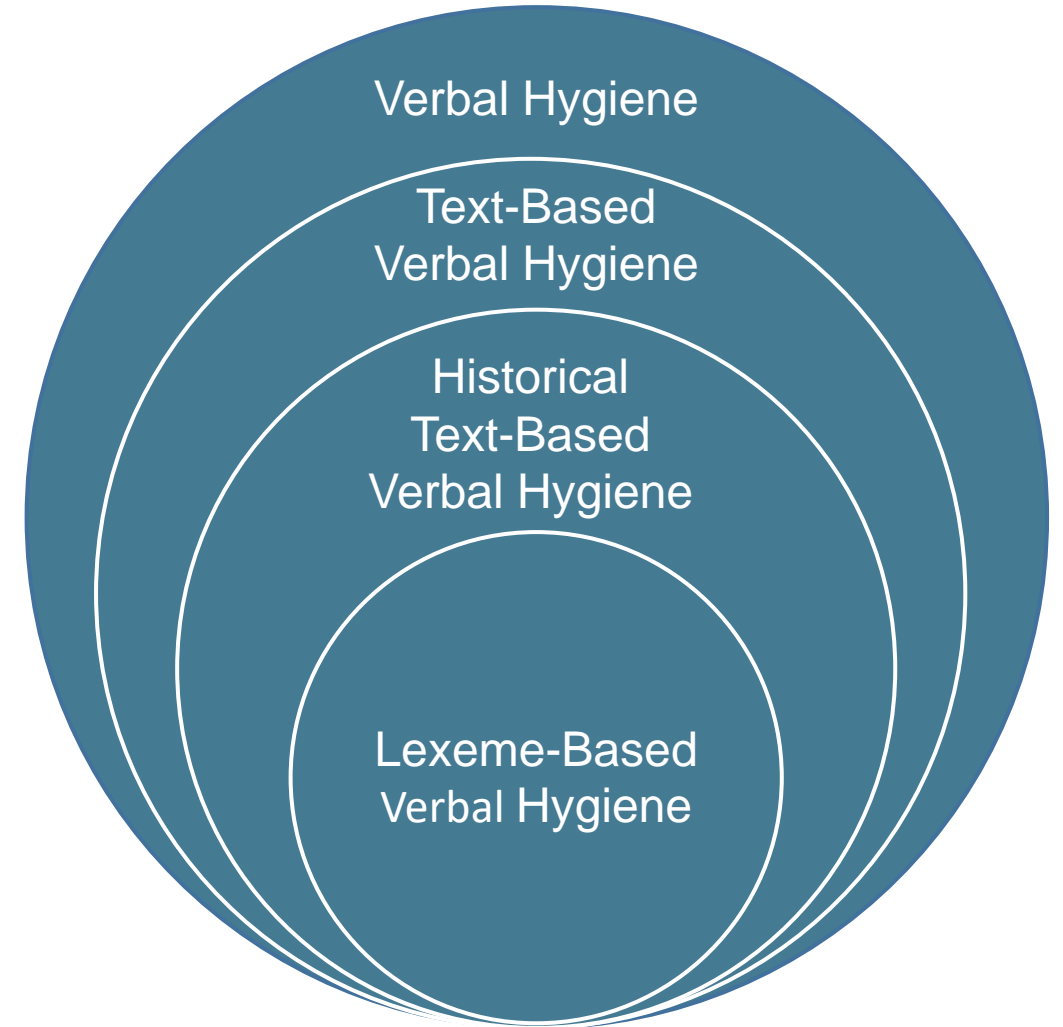
Challenge 1: Operationalizing Verbal Hygiene

Historical English Language Data

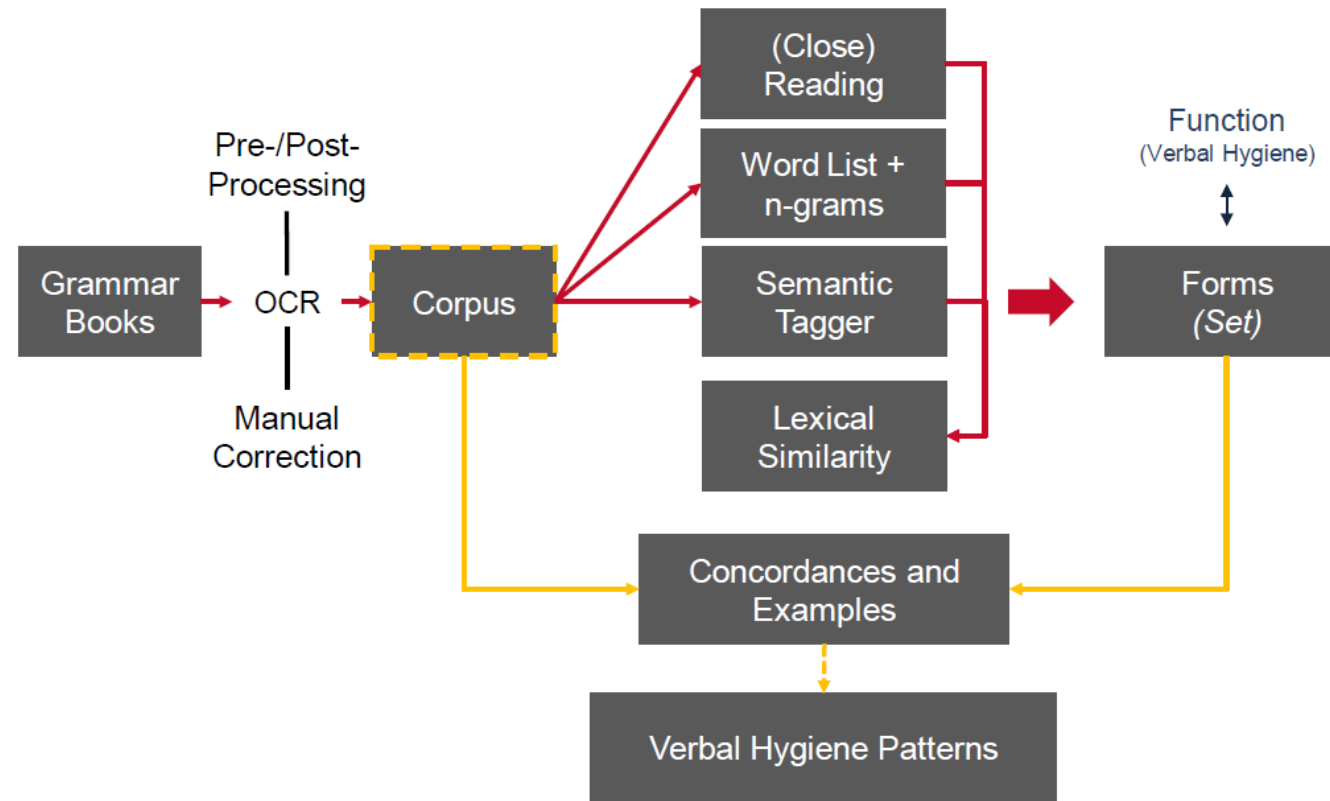
- Historical Pragmatics function-to-form
- Form-to-function mapping

Mixed-Methods Approach Empirical

- Going beyond case-studies
- Set of forms of verbal hygiene



Challenge 1: Approach Towards Automatization as Presented at ICAME39




Basic operationalization:
Verbal hygiene = set of semantic categories (purely lexeme-based approach)

Challenge 1: Approach as Presented at ICAME39 *(Close) Reading and Previous Studies*

Good old **reading** (*or:*
exploratory data analysis) ...

Previous **studies** on grammar
authors and their references to
each other ...

- 
- **Where?** (e.g., Footnote) **What?** (e.g., Example) **How?** (e.g., Critiquing)
 - Salient **forms and patterns**(e.g., critical references to other authors; adjective phrases)
 - “Gold Standard” / **Reference**

Challenge 1: Approach as Presented at ICAME39

Word-List and N-Grams

Word-Lists / n-grams / Clusters

- Manually looking for relevant forms and patterns
- Validity and accuracy
- Finding patterns is relatively straightforward

but:

- Confirmation bias / Availability heuristic / Clustering illusion etc.
- Labour extensive (growing corpus)

- ***wrong** + X**
- wrong to
- wrong and
- wrong in
- wrong construction
- wrong collocation
- wrong choice
- wrongly written
- wrongly placed
- ...

Challenge 1: Approach as Presented at ICAME39

Semantic Tagger

- Extraction of (potential) forms via the *Historical Thesaurus Semantic Tagger / SAMUELS** (Alexander et al. 2014)
 - *SEMTAG1 (USAS)*: A5 (Evaluation)
 - *SEMTAG2 (HTOED)*: e.g., 02.02.07, 02.02.09, 02.02.12, 02.02.13
 - 1203 candidates (337 adjectives)
- e.g., *self-contradictory, inaccurate, commendable, uncivilised, impure, tastelessly, purity, auspicious, provincial(ism), ...*
- Basis for in-depth analysis and Word2Vec approach

02.02.07
Judgement, decision

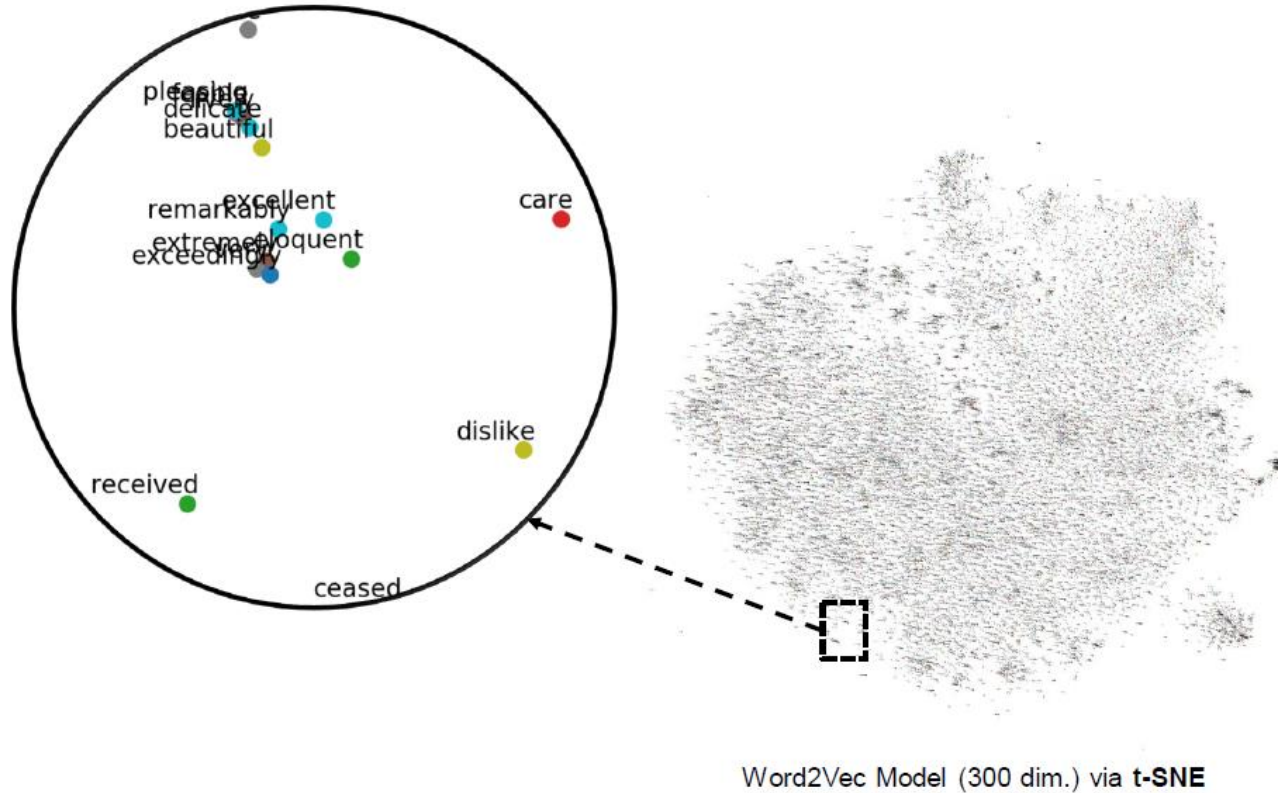
02.02.09
Esteem

02.02.12
Good Taste

02.02.13
Bad Taste

Challenge 1: Approach as Presented at ICAME39

Word2Vec



Word2Vec, as well as similar language models, can „map“ all lexical items in a semantic (vector) space.

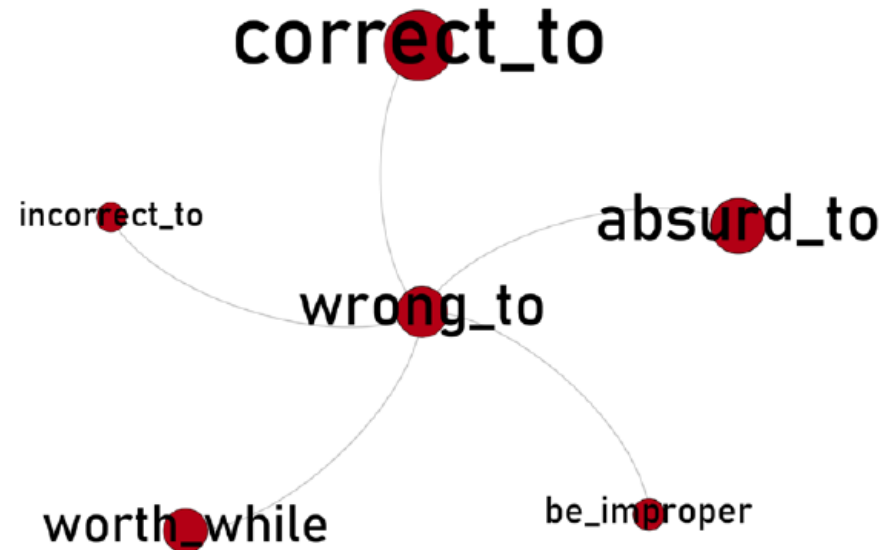
Here, after dimensionality reduction (300 to 2), **semantically similar** items (based on *distributional semantics*) are clustering together.

Challenge 1: Approach as Presented at ICAME39

Word2Vec

Word2Vec model trained on **bigrams**.

→ Predicting similar constructions to *wrong to* without any human intervention



Challenge 2: Extraction of Grammatical Terminology

What is the use of a diachronic form-to-function mapping of grammatical terminology?

- Work on 19th-century treatment of verbs and verb categories: OCR was not applied, but rather manual reading since “terminology in many cases was not unified”, so searchable corpus would not have had many benefits

(Anderwald 2016: 14-15)

- A diachronic form-to-function mapping of grammatical terminology would be useful for this

Challenge 2: Extraction of Grammatical Terminology

Suggested approach:

1. Extract structural cohesive markers (e.g., headlines, chapter headings) from the grammar books
 2. Compare keywords against a set of modern grammatical terms (e.g., Quirk et al. 1985)
 3. Create diachronic form-to-function mapping
 4. Extend the approach to the entire corpus
- Perhaps extend the mapping using the Historical Thesaurus of English to find historical synonyms for grammatical terminology
- “in the *Thesaurus* an average of 3.5 word forms has been used to describe each concept across the history of English”
- (Alexander, Kay 2022: <https://ht.ac.uk/facts-and-figures/>)

Advisor Feedback / Exchange on Proposed Methodology

1. Grammarians' approaches to verbal hygiene have probably changed throughout the centuries. How would you operationalize the search for VH patterns?
2. Methodologically speaking, how would you go about finding instances of Verbal Hygiene in a given text or corpus?
3. On which level would you focus when approaching Verbal Hygiene – authors, whole grammars, chapters, sentences, words, etc.?
4. Initially, we planned to assess Verbal Hygiene based on a list of evaluative and normative expressions. Do you think that this is a reasonable approach, or would you advise a more comprehensive approach?
5. Grammatical Terminology: Methodologically speaking, how would you go about finding lexemes referring to grammatical phenomena in a given text or corpus?



Associated PhD Research
***A Corpus-Based Analysis of American
Grammaticography until 1900***

Sophie Du Bois

Block C

Bibliography I

Alston, R. C. (1965). *English grammar written in English and English grammar written in Latin by native speakers: Vol. 1. A Bibliography of the English Language From the Invention of Printing to the Year 1800. A Systematic Record of Writings on English, based on the Collections of the Principal Libraries of the World.* Leeds: Arnold & Son.

Alexander, Marc and Christian Kay. 2022. Facts and Figures. In *About the Historical Thesaurus of English*. Glasgow: University of Glasgow.
<https://ht.ac.uk/facts-and-figures/>

Anderwald, L. (2016). *Language Between Description and Prescription: Verbs and Verb Categories in Nineteenth-Century Grammars of English.* Oxford: Oxford University Press.

Arman, H. (2021). Affects of verbal hygiene: the impact of language activism at a Swedish high school. *Language Policy* 20:151–171.

Biber, D. (1993). Representativeness in Corpus Design. *Literary and Linguistic Computing*, 8(4).

Busse, B., Gather, K., & Kleiber, I. (2018). Assessing the Connections between English Grammarians of the Nineteenth Century: A Corpus-Based Network Analysis. In Eric Fuß, Marek Konopka, Beata Trawiński, & Ulrich H. Waßner (Eds.), *Grammar and Corpora 2016* (pp. 435–442). Heidelberg: Heidelberg University Publishing.

Busse, B., Gather, K., & Kleiber, I. (2019). Paradigm Shifts in 19th-Century British Grammar Writing: A Network of Texts and Authors. In B. Bös & C. Claridge (Eds.), *Norms and Conventions in the History of English*. Amsterdam: John Benjamins.

Bibliography II

- Busse, B., Gather, K., & Kleiber, I. (2020). A Corpus-Based Analysis of Grammarians' References in 19th-Century British Grammars. In A. Cermakova & M. Malá (Eds.), *Diskursmuster - Discourse Patterns: Vol. 20. Variation in Time and Space: Observing the World Through Corpora*. Berlin: De Gruyter.
- Busse, B., Kleiber, I., Dumrukcić, N., & Du Bois, S. (2021, July 13). A corpus-based network analysis of 16th-century British grammar writing. [Conference session]. Corpus Linguistics International Conference 2021, University of Limerick and Mary Immaculate College, Ireland.
- Cameron, D. (2012). *Verbal Hygiene*. Abingdon and New York: Routledge (Original work published 1995).
- Evans, C. (1903). *American Bibliography: A Chronological Dictionary of All Books, Pamphlets and Periodical Publications, Printed in the United States of America; 1639-1729. American Bibliography: Vol. 1*. Chicago, Illinois: Blakely Press.
- Görlach, M. (1998). *An Annotated Bibliography of Nineteenth-Century Grammars of English*. Amsterdam: John Benjamins.
- Hultgren, A. K. (2020). Global English: From "Tyrannosaurus Rex" to "Red Herring". *Nordic Journal of English Studies* 19(3): 10–34.
- Idowu, O., Aliu, Y. O., & Ikuelogbon, K. (2021). An Appraisal Analysis of Verbal Hygiene in Bishop Oyedepo's Crisis Motivated Sermon. *Voices: A Journal of English Studies* 6: 1-17.
- Kytö, M. (1991). *Variation and diachrony, with early American English in focus: Studies on can/may and shall/will. Bamberger Beiträge zur englischen Sprachwissenschaft: Bd. 28*. Frankfurt am Main, Bern, New York, Paris: Lang.

Bibliography III

McCarthy, M. (2020). *Innovations and Challenges in Grammar. Innovations and Challenges in Applied Linguistics*. Abingdon, Oxon, and New York: Routledge.

Michael, I. (1987). *The Teaching of English: From the Sixteenth Century to 1870*. Cambridge: Cambridge University Press.

Smith, R. N. (1977). A Bibliography of Books on Language and Languages Printed in the United States Through the Year 1800. *Historiographia Linguistica*, 4(2), 207–243.

Smith, R. N. (1979). The Interest in Language and Languages in Colonial and Federal America. *Proceedings of the American Philosophical Society*, 123(1), 29–46.

The Historical Thesaurus of English (2nd edn., version 5.0). 2022. University of Glasgow. <https://ht.ac.uk/>.

Tieken-Boon van Ostade, I. (2006). Eighteenth-century Prescriptivism and the Norm of Correctness. In A. van Kemenade & B. Los (Eds.), *The Handbook of the History of English* (pp. 539–557). Blackwell Publishing Ltd.

Watts, R. J. (2008). Grammar writers in eighteenth-century Britain: A community of practice of a discourse community? In I. Tieken-Boon van Ostade (Ed.), *Topics in English Linguistics: Vol. 59. Grammars, Grammarians and Grammar-Writing in Eighteenth-Century England* (pp. 37–56). Berlin: Mouton de Gruyter.

Yáñez-Bouza, N., & Rodríguez-Gil, M. E. (2013). The ECEG database. *Transactions of the Philological Society*, 111(2), 143–164.

Yáñez-Bouza, N. (2021). Methodological approaches to the study of codification, prescription, and prescriptivism. *Studia Neophilologica*, 1–31.

Bibliography IV – Grammars

Booth, D. (1837). *The Principles of English Grammar*. London: Charles Knight and Co.

Bullokar, W. (1586). *Brief grammar for English*. Edmund Bollifant.

Coote, E. (1596). *The English schoole-maister teaching all his schollers, of what age soever, the most easie, short, and perfect order of distinct reading, and true writing our English-tongue, that hath euer yet beene knowne or published by any. Printed by the Widow Orwin, for Ralph Jackson, and Robert Dextar*.

Crombie, A. (1802). *The Etymology and Syntax of the English Language, Explained and Illustrated*. London: Taylor and Walton.

Humphreys, Daniel. (1792). *The Compendious American Grammar, or Grammatical Institutes in Verse*. Portsmouth, N.H. Spy Printing Office.

Mason, C. P. (1858). *English Grammar, Including the Principles of Grammatical Analysis*. London: Walton and Maberly.

Meurier, G. (1586). *The coniugations in Englishe and Netherdutch, according as Gabriel Meurier hath ordayned the same, in Netherdutch, and Frenche*. Thomas Basson.

Mulcaster, R. (1582). *The First Part of the Elementarie Which Entreateth Chefelie of the Right Writing of our English Tung*. London: Thomas Vautroullier.

Bibliography V – Grammars

Sherry, R. (1577). *A Treatise of the Figures of Grammer and Rhetorike*. London: Ricardi Totteli.

Webster, Noah. (1789). *Dissertations on the English Language: With Notes, Historical and Critical, to Which Is Added, by Way of Appendix, an Essay on a Reformed Mode of Spelling, with Dr. Franklin's Arguments on That Subject*. Boston: Isaiah Thomas and Company.

Thank you for
your attention!

