Exploring Complementation Choices in Eighteenth-Century Late Modern English: A Corpus Study

Valentina Ravest Córdova

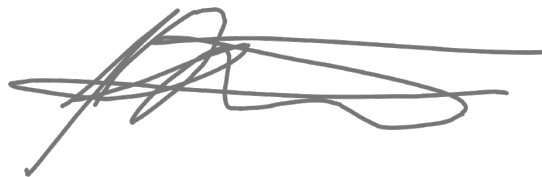Supervisor: Peter Petré

Assessor: Eleanor Smith

University of Antwerp

The undersigned, Valentina Ravest Córdova, student of the Master program in Digital Text Analysis at the University of Antwerp, declares that this thesis is completely original and exclusively written by herself. For all information and ideas derived from other sources, the undersigned has referred to the original sources, both explicitly and in detail.

Exploring Complementation Choices in Eighteenth-Century Late Modern English: A Corpus Study

Valentina Ravest Córdova
University of Antwerp

This paper undertakes a comprehensive exploration of complementation patterns in Late Modern English, specifically as employed by authors of the eighteenth century. Employing a blend of corpus analysis and advanced statistical methodologies like Inferential Trees and Random Forests, this research meticulously examines the range of variables that influence decisions in complementation, contrasting the selection of these clauses and their relation to the cognitive complexity involved. The study not only highlights general behavioral tendencies within the author group but also pinpoints individual preferences.

A significant outcome is the salient role of the denotation variable, which significantly shapes cognitive intricacies within sentences and decisively influences complementation patterns. Authors often opt for simpler structures for more intricate clauses, and vice versa, while denotation disparities result in a preference for finite clauses for clarity and non-finite clauses for cases of sameness.

1. Introduction

This thesis is presented to fulfill the requirements for the degree of Master in Digital Text Analysis. It represents the culmination of extensive research, critical analysis, and scholarly inquiry within the realm of Late Modern English Grammar. The study carried out in this thesis is based on and contributes to the FWO research project 'Complexity in Complementation'[1]. This proposal aims to explore the influence of individual cognitive differences on language change at a population level. By investigating the interplay between individual and community factors, the project contributes to the understanding of language as a complex adaptive system. It emphasizes the significance of individual-level analysis to enhance our understanding of abstract grammatical variation. As Steels (2000) mentions, 'human languages are constantly changing and differ significantly from one speaker to the next and from one context to the next.' (18) This research is an aim to contribute to these efforts and to expand the existing body of knowledge in Corpus Linguistics. Through a comprehensive examination of variation in verb complementation, this research endeavors to shed new light on complementation variation based on syntactic and semantic features and provide valuable insights into the understanding and characterization of Late Modern English. By exploring into this subject matter, we aim to offer practical implications and potential avenues for future research.

1.1 Brief summary

This thesis focuses on exploring the different patterns of sentence complementation employed by English authors of the 18th century, employing meticulous corpus annotation alongside the application of multifactorial classification models like Inferential Conditional Trees and Random Forests. The main objective is to find recurring patterns among the group of authors as well as interindividual differences that show some of the characteristic writing style of the period through the use of sentence complementation. Via the application of analytical techniques, the study delves into the complex interaction of finite and nonfinite clauses, utilized to complement the verbs 'forget' and 'remember.' Despite the apparent semantic similarity of these verbs, they give rise to distinct complementation patterns, allowing us to understand how factors such as cognitive processing and semantic subtleties impact an author's linguistic choices. The objective was to search for complementation patterns through the use of finite and non-finite clauses related to the mentioned verbs, for this purpose a total of four different types of complementation were selected, two finite and two non-finite. To obtain the dataset to work with, we consulted a corpus of authors from the selected period, the ECCO-TCP, a collection containing more than 200,000 texts by eighteenth-century authors who spent most of their lives in London. After obtaining the data, the corpus was manually annotated with a focus on different lexical and syntactic variables, subsequently, the next step was to perform the aforementioned analyses to determine how these variables influenced the selection of clauses, through the inferential conditional trees and random forests. The exhaustive examination of the research findings has revealed significant factors that play a crucial role in molding the observed literary practices, thus revealing distinct trends and unique stylistic attributes specific to each writer and to the general group. We found that the main variable in interfering with complementation patterns was denoising, which indicates the subject's correspondence between TCC and CC. The results showed that in most cases, when finite clauses were preferred subjects tended not to correspond between the two clauses, whereas when finite clauses were not preferred subjects did correspond. It was also possible to extract from the results that while the forget verb exhibited patterns common to the group, the remember verb showed different behaviors for the different authors.

1.2 Objectives

This research endeavors to address two fundamental queries. Firstly, our investigation revolves around identifying the key variables that significantly influence the decision-making process in sentence completion. Through meticulous analysis of an extensive 18th-century English corpus, our objective is to uncover the underlying determinants that guided authors in selecting specific verb complements, shaping the very essence of their written expression. Additionally, we delve into how the choice of particular verbs interacts with the selection of these variables, leading to the emergence of distinct patterns within the corpus. This investigation seeks to discern whether the observed patterns are unique to each author, reflecting their distinct stylistic inclination, or whether recurrent patterns emerge across multiple authors, potentially signifying either shared linguistic tendencies characteristic of the 18th century, and/or constraints on the types of cognitive routes our minds can take.

1.3 Hypotheses

Building upon the groundwork laid out earlier, we formulate some hypotheses that serve as our guiding lights to analyze the complex interactions between language factors related to sentence complementation in the Late Modern English period. Our aim is to uncover connections between these factors and how they influence the way sentences are constructed.

H1: Discernible patterns of complementation will emerge, providing compelling evidence of the preference for finite and non-finite clauses in specific linguistic contexts. These patterns are anticipated to unveil a clear tendency towards the predominance of one type of complement over the other, illuminating how authors strategically employed different structures to convey their intended messages.

H2: No notable differences will be evident in the authors' usage of complementation for the verbs forget and remember . Given the fact that they are antonyms and both are private verbs used in similar ways, we expect them to show similar patterns of finite and non-finite clause utilization

H3: Certain writers will demonstrate particular tendencies in making choices for sentence complementation. By closely examining the individual linguistic choices, we aim to identify characteristic patterns that will distinguish these authors from their contemporaries.

H4: Discernible patterns will be intricately influenced by a spectrum of distinct variables that assume a decisive role during the segmentation process. We anticipate that the chosen variables will exert a discernible impact on the decision-making paradigm, thereby intricately shaping the observed patterns.

1.4 Empirical and theoretical contribution

This work has an empirical and a theoretical contribution to make. Empirically, it contributes to the domain of Late Modern English studies. By investigating complementation patterns in the writings of 18th-century authors, this research offers a nuanced glimpse into language usage during that historical period. The insights garnered from the analysis of finite and non-finite clauses in diverse linguistic contexts provide an addition to the ongoing exploration of language evolution, offering a small piece to the puzzle of understanding linguistic developments over time. This study joins the scholarly conversation on Late Modern English, hoping to provide a contribution to the understanding of the linguistic landscape of that era. As Hundt (2014) mentions in the introduction to Late Modern English Syntax, the shift in focus within historical linguistics towards socio-historical and corpus-based methodologies has resulted in a notable increase in fascination with Late Modern English. This has been particularly evident through the hosting of a triannual conference on this period, which has also witnessed a proliferation of published research studies, among which Cuyckens, D'hoedt, and Szmrecsanyi (2014), Rohdenburg (2014), Fonteyn and Nini (2020), Cuyckens, Fonteyn, and Peter (2021), to name but a few. Theoretically, the thesis aims at contributing to the theory of language as a complex adaptive system. By investigating the intricate interplay between linguistic variables and their impact on complementation patterns, this study offers valuable insights into the dynamic nature of language usage. The

identification of discernible patterns and tendencies reflects the inherent adaptability of language to diverse communicative contexts. This investigation adds to the growing body of knowledge that recognizes language not as a static entity, but as a dynamic and adaptive system influenced by various factors and contexts. The theory of Language as a Complex Adaptive System suggest that language is a complex and dynamic system that evolves and adapts over time in response to various factors and pressures, the 'patterns of use strongly affect how language is acquired, is structured, is organized in cognition, and changes over time' (Group", Beckner, Blythe, Bybee, Christiansen, Croft, Ellis, Holland, Ke, Larsen-Freeman et al. 2009, p. 2). This theory is based on the idea that language is not a static structure with fixed rules, but rather is a constantly changing and evolving system that responds to internal and external factors. From this perspective we understand language as a distributed system, meaning that is not limited to one individual or groups. Instead, language is based and shaped by a wide range of factors, including language users, social and cultural context, historical developments and environmental influences. This research also contributes to the field of corpus linguistics through the utilization of rigorous corpus analysis methods. This approach allows for a comprehensive exploration of linguistic patterns and trends within Late Modern English texts, thereby enhancing our understanding of language evolution during that period.

2. Background

2.1 Socio-cognitive Linguistics

This work is based on socio-cognitivism, since the problem studied focuses on the cognitive complexity of the writing of a particular social group. In linguistics, sociocognitivism is an interdisciplinary science that combines sociolinguistic and cognitivist methods around the phenomenon of language, it focuses on how 'linguistic structures can be explained and even assumed to be motivated by fundamental cognitive principles' (Schmid 2016, p. 544) studying the relationship of language, society and cognition. Language is seen as a dynamic system that emerges from the interaction between individuals, their cognitive processes and the social environment in which they are inserted. More than a means of communication, language is considered a reflection of social and cultural norms, identity and values, articulated through mental representations. Socio-cognitive linguistics takes into account how language use varies across different social groups and contexts. Social factors such as age, gender, ethnicity and social class are considered. In the study of socio-linguistic variables such as pronunciation, vocabulary and grammatical structures in relation to the cognitive process underlying the choice of these variables. As Dąbrowska (2020) mentions,

> It follows that language is not merely a social phenomenon, or merely a cognitive phenomenon. It is clearly both: a language can live only in individual minds, but is learned from examples of utterances produced by speakers engaged in communicative interaction. (Dąbrowska 2020, p. 224)

In this sense we take into account the dynamic interplay between language and society, and our research is grounded in the understanding that language does not exist in isolation but is intricately woven into the fabric of human cognition and social interaction. This approach offers a comprehensive framework that acknowledges the reciprocal

relationship between language and the human mind, providing a deeper insight into the intricate mechanisms that underpin our linguistic expressions.

2.2 Corpus linguistics

As mentioned before, this thesis bases its methodology in Corpus linguistics techniques, which is a branch of linguistics that focuses on the study of language using large collections of written, spoken, or recorded texts known as corpora. A corpus is a carefully compiled and organized database of texts that serves as a representative sample of a language or a specific subset of language, such as a particular genre, time period, or region. As Gries and Berez (2017) mention, it encompasses instances that are prototypical due to their display of commonly acknowledged attributes. Yet, it also encompasses numerous instances that are connected to the prototype or, to a lesser extent, to other instances within the category through family resemblance connections. Corpus studies involve the systematic analysis of linguistic patterns and structures within these corpora to gain insights into how language is used and how it functions in different contexts.

In corpus linguistics, researchers use computer-based tools and techniques to analyze and process the vast amount of data contained within corpora. In order to fulfill the purpose of being used in an automated way, 'that machine readability and interoperability requires some degree of standardization of annotation' (Gries and Berez 2017, p. 393). The texts in a corpus can be in various formats, such as books, articles, transcripts of spoken language, social media posts, or any other written or recorded material. The corpora may cover a wide range of languages and dialects, enabling researchers to investigate language variation and change across different communities and time periods.

2.3 Verb Complementation

In particular, this study contributes to efforts to understand the relationship between sentence complementation and the cognitive complexity involved in literacy processes. Clause verb complementation has been a subject of significant research within both generative and cognitive-functional linguistic frameworks. As Cuyckens, D'hoedt, and Szmrecsanyi (2014) mentions, generative linguists have been intrigued by complementation phenomena since Chomsky seminal work, 'Aspects of a Theory of Syntax' 2014 , and their exploration has continued over the years. Concurrently, functional-typological linguists expanded the scope of complementation research in the 1980s. While generative studies mainly focus on syntactic issues, cognitive-functional research emphasizes the distribution of complement clauses across complement-taking predicates.

However, it is crucial to recognize that synchronic matches between complement-taking predicates (CTP) and complement clauses (CC) can change over time. In recent decades, diachronic studies have emerged, presenting comprehensive accounts of change and variation in complementation patterns over the history of English.

Verbal complementation, the topic of this study, refers to the different ways in which verbs can be combined with complements, such as objects, clauses or prepositional phrases to express meaning and grammatical relations. It encompasses the study of various patterns and structures that verbs exhibit in their semantic behavior.

In accordance with Somers (1987), the concept of complementation encompasses the phenomenon in which the meaning of a sentence remains incomplete until it is fused with specific elements that are anticipated to accompany a particular verb or to enhance its meaning. This notion is echoed by Huang (1997), who asserts that complements serve

as integral elements that contribute to fulfilling the semantic requirements of a sentence as dictated by a verb. In practice, distinct verbs exhibit the propensity to accept varying types of complements, leading to their subcategorization based on the complementation they entail. Huddleston and Pullum (2005) further elaborate that verbs frequently allow for multiple patterns of complementation, albeit subject to semantic constraints that govern the range of acceptable complements.

Numerous scholars, including Quirk and Wrenn (2002), Rohdenburg (2014), and Cuyckens, D'hoedt, and Szmrecsanyi, have used diverse complementation patterns to explore the multifaceted aspects of finite and non-finite clauses, with an emphasis on linguistic complexity. For the purpose of our study, we adopt the framework proposed by Cuyckens, D'hoedt, and Szmrecsanyi as our guide, utilizing two distinct finite and two non-finite complementation patterns for our analysis.

A finite clause, constituting a grammatical unit within a sentence, incorporates both a subject and a finite verb. Notably, a finite clause forms a complete thought and possesses the ability to function autonomously as a standalone sentence. The finite verb within this construct is inflected to convey temporal information, person, and number.

The first finite-clause type involves the '*that-clause*,' characterized by the subordination of a sentence using the conjunction 'that.' This is followed by a sentence that encompasses both subject and predicate, encapsulating the essence of finite clause structure.

Example 1
Remember , **that** your future Welfare depends greatly on your present Behaviour. (Richardson, 1741)

The second finite-clause type is the *0-complementizer*, whose structure is similar to the previous one, but without the use of any connector after the verb. The main clause is immediately followed by the complement clause, with its subject and predicate.

Example 2
I remember I was one morning disturbed at my breakfast (Colman, 1755)

Regarding these two finite clauses, (Rohdenburg 1995) discusses the significance of the word 'that' as a crucial indicator of boundaries closely related with the complexity principle. It is normally present in those contexts that constitute a more challenging construction, while it seems to be omitted when the CC contains a more explicit statement, which translates as a less cognitively complex sentences. What he mentions in his research is that 'the conjunction may be more easily dispensed with in cases of straightforward clause union than in more convoluted combinations' (383).

The non- finite clauses are clauses that lack a subject and tense marking (Aljović 2017). The most common type in our data is the *to-infinitive*, where the complement clause starts with the word to and is followed for a verb in infinitive.

Example 3
Not forgetting to return him her thanks (Trusler, 1788)

The second type of non-finite complement is the use of the gerund in the complement clause. In this case, the CC that follows the CTP contains a verb in gerund, which shares with the *to-infinitive* a lack of tense and person marking. Quirk et al. devote a special section on finite clauses entitled 'Structural 'deficiencies' of non-finite clauses'. He mentions that lacking the structural elements of tense, they tend to be more concise and allow for compactness of information, but more difficult to interpret. He mentions that poets often make use of this resource to fit their ideas within the meter of their

poem. However the benefit of conciseness needs to be weighed against the challenge of an increased processing cost; the omission of a subject creates uncertainty about which nearby nominal element should be considered the subject in terms of meaning. This ambiguity means an increase in complexity, since it is not obvious who the subject of the sentence is without having to resort to co-reference strategies.

Example 4
I don't remember ever tasting any thing so nauseous (Trusler, 1788)

2.4 Complexity

An equally influential principle, referred to as Rohdenburg's complexity principle (1996) or transparency principle, puts forth the proposition that when confronted with more or less explicit grammatical alternatives, those which are more explicit are favored, especially in cognitively intricate contexts. Thus, in sentences composed of finite clauses, there would be a tendency to refer to the *that-clause* over the *0-complementizer*, since the former is simpler and more direct, while the latter raises cognitive complexity. About this topic, Quirk et al. (1985) mentions that:

> the zero *that-clause* is particularly common when the clause is brief and uncomplicated. In contrast, the need for clarity discourages or even forbids the omission of that in complex sentences loaded with adverbials and modifications. (276)

Preferences are observed within non-finite clauses. In the case of non-finite clauses, a tendency to prefer the *to-infinitive* over the *-ing* has been observed. Duffley (2000) points out that in sentences that use the *to-infinitive* there is a coincidence between the subject of the CTP and the CC, while in sentences with gerunds there is no clarity about the identity of the subject. In both cases there is a preference for more explicit clauses, following the complexity principle. This principle rests upon the assumption made by Rohdenburg (1996) that differentiation between more and less explicit options is feasible, with the former often encompassing bulkier elements or constructions, thereby achieving heightened explicitness. As a consequence, the complexity principle may intersect with other factors, particularly stylistic and semantic inclinations. Empirical investigation becomes essential in disentangling the individual contributions of each potentially conflicting factor. Regrettably, due to constraints, this paper can only briefly allude to pertinent semantic or stylistic tendencies. The intricate interplay between language variation and its interaction with multiple factors remains a captivating realm within linguistic exploration (Rohdenburg 1996).

2.5 State of the art

In this section, we will explore the latest research and knowledge related to our topic. In the study of language variation and change there are notorious attempts to vindicate the figure of the individual from the generalization of groups at the population level. Among these is the work done by Fonteyn and Nini (2020), who seek to show how each individual makes his or her syntactic choices in a motivated yet idiosyncratic fashion within the possibilities of variation, focusing on the alternation of gerunds in 17th century authors. (Fonteyn and Nini 2020). The study explores the 'gerund alternation,' where speakers choose between two different types of objects complementing the gerunds. While previous research assumed uniformity across individuals

in this variation, this study uncovers individual diversity within a socially homogeneous population. It finds that individuality's predictive power surpasses group-based external factors like age. The study aligns with usage-based theories, highlighting how individuals construct unique cognitive models based on encountered exemplars. The importance of determiner use in gerund selection is shared among authors, revealing both idiosyncrasy and commonality. This underscores the complex interaction between individuality and shared behavior in language variation, emphasizing the need for comprehensive research in this understudied area. Another relevant study within this field is the one conducted by Cuyckens, Fonteyn, and Peter (2021), in which they seek to account for the impact on language variation caused by differences in cognitive representations between individuals in the long term at the population level. The primary aim of this research is to assess how individual differences in cognitive representations influence broader language changes across populations over time. This study focuses on investigating variations in the English system of clausal verb complementation. While some amount of variation might be attributed to a desire for diversity itself, it has been demonstrated that the selection of complement variants is affected by various functional factors such as semantics, structure, and discourse, including aspects like animacy and complement clause length. However, prior studies have encountered challenges in adequately explaining this variation solely through social variables at the population level. When social cues fall short in explaining usage, cognitive mechanisms may play a role that diverges among individuals and thus cannot be easily generalized. The present project seeks to enhance our understanding of the functional dynamics of such abstract grammatical variations by placing a strong emphasis on individual-level analysis. Another relevant study on complementation variation is the research conducted by Cuyckens, D'hoedt, and Szmrecsanyi (2014). Their primary objective is to present a corpus-based analysis of complement clause variation with the complement-taking predicates remember, regret, and deny. Furthermore, the authors aim to explore the extent to which the varying distributions of finite and non-finite complement clauses can provide insights into more general hypotheses about complement choices. The examination centers on the probabilistic or non-categorical variations in Late Modern English involving the use of finite versus non-finite complement clauses alongside matrix verbs mentioned before. Employing a binary logistic regression model with mixed effects, the study's aim was to scrutinize the intricate nature of choices in complementation strategies. This methodology has provided insights into the factors favoring non-finite complementation and the shifting impact of different conditioning factors during the Late Modern English era. One of the main outcomes of the study is that alongside the emphasis on non-categorical complementation variation, an appreciation for categorical variation remains integral to providing a comprehensive understanding of CC-choice.

3. Methodology

In this section, a detailed exposition of the methodology adopted for the execution of this experiment will be provided. The presentation will commence by addressing the key aspect of data, with a particular focus on the ECCO-CTP corpus, which forms the bedrock of this study. Subsequently, a comprehensive overview of the annotation process will be expounded upon. Lastly, the attention will be directed towards elucidating the utilization of the Conditional Tree and Random Forest methodologies, instrumental in the systematic analysis of the amassed data. Through these successive

stages, a coherent understanding of the methodological framework underpinning this investigation will be established.

3.1 Verbs

In the choice of matrix verbs, following the guidelines of the Complexity in Complementation project, those that permit variation between finite and non-finite complement clauses were selected. The verbs, remember and forget, were carefully chosen from Quirk et al. (1985) (1985) compilation of factual verbs that exhibit this characteristic of allowing both finite and non-finite complement clauses (Cuyckens, D'hoedt, and Szmrecsanyi 2014).

3.2 Data

The data consists of a principled sample retrieved from the ECCO-TCP. ECCO-TCP is a combined effort between the Eighteenth Century Collections Online and the Text Creation Partnership. ECCO is a collection that contains more than 200,000 texts from authors from the 18th century. While ECCO also contains, for instance, authors from North America, to maintain social homogeneity, no such authors were included in our study. The selection was initially made by taking authors named in the New Cambridge Bibliography of English Literature. It provides researchers, scholars, and students with access to a vast array of printed material, including books, pamphlets, essays, and other types of literature from this time period. ECCO is an essential resource for understanding the literary, historical, and cultural developments of the 18th century. The TCP is a collaborative effort between libraries and scholars aimed at producing accurate, fully searchable, and encoded editions of early printed works. The encoded texts are made available to the public for research and educational purposes. (Tolonen et al. 2021).

Author	Birth year	Word count	Label
George Berkeley	1685	483503	1
James Boswell	1740	675187	2
Edmund Burke	1729	1237897	3
George Colman	1732	488978	4
Richard Cumberland	1732	1119378	5
Charles Dibdin	1745	1263055	6
Edward Gibbon	1737	1239403	7
William Hayley	1745	1183655	8
Samuel Johnson	1709	371448	9
John Pinkerton	1758	663747	10
Alexander Pope	1688	1217873	11
Samuel Richardson	1689	2854509	12
John Trusler	1735	2801867	13
Thomas Warton	1728	1080551	14

Table 1
Table of authors.

3.3 Annotations

The authors selected from ECCO for this study are presented in

To access the corpus the Corpus Linguistics tool CosyCat[1] (Collaborative Synchronized Corpus Analysis Toolkit) was used. This is a web based application that allows users to query the corpus through the use of regex-based query syntax, apply filters and display the results in a query pane. The program allows to do annotations in an easy way, providing a different pane for the annotated items. To query the corpus we looked at sentences containing the two verbs under study (remember and forget), filtering by the selected authors, and then we proceeded with a manual annotation. To ensure that all the possible forms of the verbs were captured, we included in the query different spellings and forms of the verbs. The annotation was made based on the following variables:

Meaning of Complement Clause This variable focuses on the meaning of the Complement clause, whether the verb describes an event or a state. Events describe actions, activities, or processes that have a specific beginning and end. An example is (5). These contexts represent actions that can be observed or measured in terms of duration or completion. States describe conditions, states, or situations that are not actions with a clear beginning and end. An example is (6). These contexts represent static, unchanging, or ongoing states of being, feelings, or thoughts.

Example 5
Event: Remember we **sell** oil and candles (Trusler, 1745)

Example 6
State: Sometimes forget that clemency **is** the warrior's best virtue (Dibdin, 1800)

Animacy This variable describes if the subject of the complement clause is animate or not. An animate subject (cf. (7)) refers to living beings or entities capable of performing actions, having emotions, or possessing volition. This category typically includes humans, animals, and sometimes supernatural beings or anthropomorphic objects. An inanimate subject (cf. (8)), on the other hand, refers to non-living things or entities that lack volition or agency. Inanimate subjects may include objects, abstract concepts, natural phenomena, or ideas.

Example 7
Animate: **We** at this moment **have forgot** to relish the admirable humour of the Rehearsal (Dibdin, 1800)

Example 8
Inanimate: Remember, my dear cousin , that **vengeance** is God's province (Richardson, 1748)

Negation of the Complement Clause As the name indicates, this variable expresses if a negation in the complement clause is present or not. A separate value is assigned for this variable if it is the main clause (MC) itself that is negated.

---

1 Arévalo, E. M., Petré, P. (2017, June). Enabling annotation of historical corpora in an asynchronous collaborative environment. In Proceedings of the 2nd International Conference on Digital Access to Textual Cultural Heritage (pp. 9-14).

Example 9
Negation CC: remember, that it is not only dangerous , but the very height of folly (Trusler, 1790)

Example 10
Negation MC: I should not forget to acknowledge your letter sent from Aix (Pope, 1707)

Example 11
Without negation: who the reader may remember had a curious adventure with the Cordeliers (Dibdin, 1800)

Voice: This variable indicates if the complement clause is in active or passive voice. In the active voice (12), the subject of the sentence is the doer of the action. The subject performs the action described by the verb, and the object receives the action. In the passive voice (13), the subject of the sentence is the receiver of the action.

Example 12
Active voice: Remember , you have no Time you can call your own (Richardson, 1741)

Example 13
Passive voice: you forget that Joseph was thus promoted for the same reason (Gibbon, 1796)

Denotation: This variable is about the correspondence between the subject that performs the action in the complement clause and the subject of remember or forget in the CTP. The two subjects can be either the same (14) or different (15).

Example 14
Same: We forgot to mention in its proper place , that we had a friend amongst the mourners (Cumberland, 1795)

Example 15
Different: She never forgets that the raillier is her uncle (Richardson, 1753)

Type of Complementation: This variable indicates which of the different types of complementation is used in the sentence. We can group them by finite clauses and non-finite clauses. A finite clause contains a finite verb, which is a verb that is marked for tense, person, and number. A non-finite clause contains a non-finite verb, which is a verb form that is not marked for tense, person, or number. In the finite clauses we have the *that-clause* (16) and the *0-complementizer* (17).
Finite clauses
that-clause: subordinate construction that uses the word that to introduce the complement clause:

Example 16
Forgot that the people were men (Pinkerton, 1790)

0-complementizer: construction similar to the *that-clause*, but there is no particle that introduce the complement clause.

Example 17
Remember only you have that in charge (Cumberland, 1795)

Non-finite clauses
item*to-infinitive*: Non-finite clauses either contain a *to-infinitive* (18) or a gerundial ing-form (19).

Example 18

They forget to pay their more private and particular (Pope, 1707)

  itemIng-complement clause: the verb in the complement clause is a gerund.

Example 19

whom you remember dancing after me all last winter in London (Cumberland, 1789)

Length of the CC This is a variable that take into account the amount of words present in the CC.

After the annotations were completed, the annotated data was exported to a csv file containing the existing metadata, the annotations and the word count of the complement clauses.

The following graph fig. 1 shows the distribution of complements used by each author.
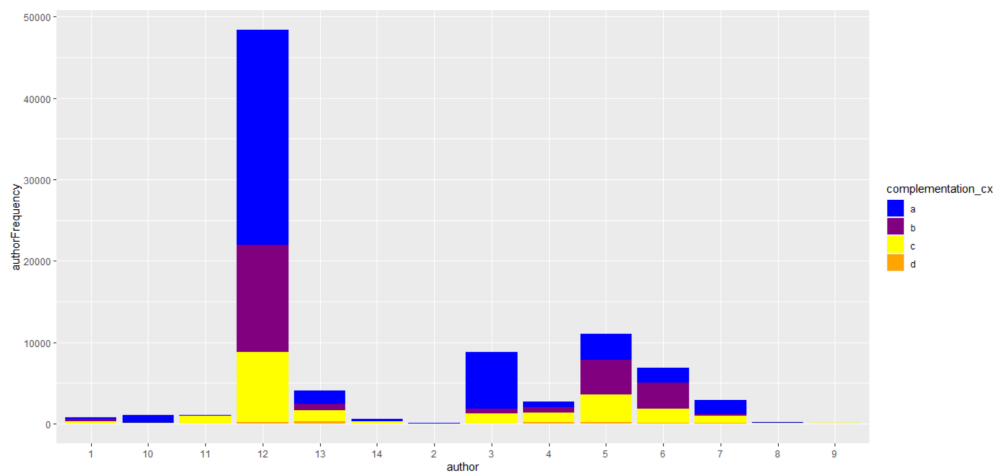
Figure 1
Complementation distribution

To start with the analysis, the only preprocessing necessary was to convert the datatypes into factors, except for the count of words, which was the only non-categorical variable. In order to simplify the visualizations, the name of the authors and the different complement clauses were changed to numbers, as shown in Table 1 and letters (Table 2), respectively.

Complementation pattern	Label
that-clause	a
0-complementizer	b
to-infinitive	c
cc-ing	d

Table 2
Table of complementation patterns' labels.

3.4 Conditional trees and Random Forests

To carry out all the analyses described in this section, the R language was used (R Core Team 2017). In order to analyze the new data and discern patterns in complementation among various authors, we employed several Conditional Trees and Random Forests. Additionally, to gain insights into the most influential variables shaping these patterns Random Forests were generated. The reason for using tree methods arises from the unbalanced nature of the data, as Th Gries (2019) says linear models or their extension to generalized linear mixed-effects models can run into problems especially when applied to observational data such as corpus data' (1) and in these cases the use of random forests and decision trees could help to overcome these obstacles. These Multifactorial classification models prove to be useful in the concerning task, as they enable us to classify data based on multiple variables simultaneously and can deal with the scarcity of the data.

3.5 Conditional Inference Trees

These are a type of decision tree algorithm used in statistical analysis. They are used to identify the most important variables that influence a specific outcome and how they interact with one another. As Stefan Th Gries (2019) mentions:

> Tree-based methods function by repeatedly splitting data sets up into two parts such that the split leads to the best increase in terms of classification accuracy or in terms of some other statistical criterion (such as deviance or the Gini coefficient or others) when it comes to predicting the dependent variable. (2)

Conditional Inference Trees are similar to traditional Decision Trees, which are hierarchical structures of nodes to represent a sequence of decisions that lead to the final classification. However, unlike decision trees, conditional inference trees are designed to handle continuous categorical data, as well as interactions between variables. The algorithm works by iteratively splitting the data based on the variables that have strongest association with the outcome. The algorithm continues to split the data until a stopping criterion is met, such as reaching a predetermined maximum depth or a minimum number of observations in each terminal node. The resulting tree provides a graphical representation of the decision-making process, providing groups that share similar characteristics. One advantage of conditional inference trees is that they can handle missing data and outliers, which can be problematic for other statistical methods. Additionally, conditional inference trees can provide insights into complex interactions between variables, making them a useful tool for exploratory data analysis. To start the experiment and to get an overview of the generation of authors, the first Conditional Tree was performed. To create the tree the ctree() function from was used from the party package (Hothorn, Hornik, and Zeileis 2006) taking all the authors in the study, with all the variables and trying to classify according to the type of complementation used. Then the tree was plotted, showing the decision-making process and the classification, as can be seen in fig. 2.

The next step was to group the data by lemma, obtaining one dataset for all the instances of the verb forget and one for all the instances for the verb remember. With these two new datasets the process was repeated, generating a new decision tree for each lemma. After this process we obtained three decision trees for the general group, one trying to classify by type of complementation and the others doing the same but

separating by lemma. The intention to generate three instances of decision tree for the same group was to observe if there was a significant difference between the behavior of the complementation patterns within the lemma and the authors. The next step is to elaborate individual trees for each author, so individual datasets were created. First one tree was plotted based on the whole data, and then two more for each verb.
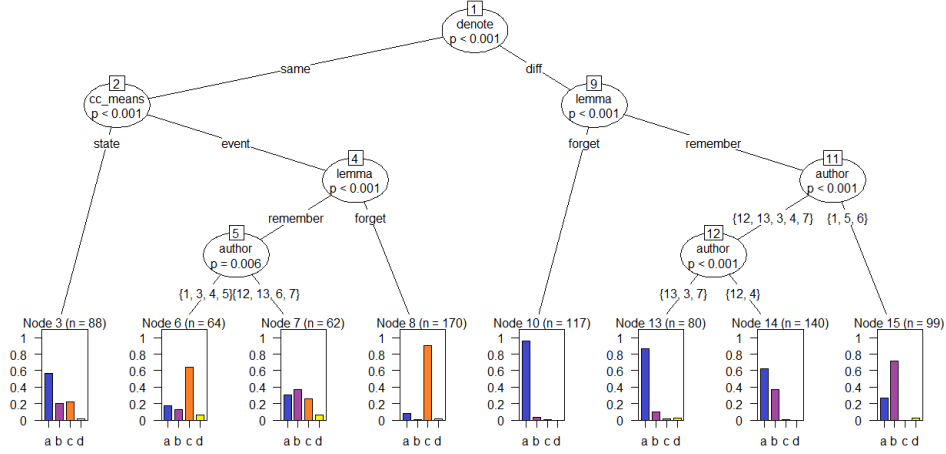

Figure 2
Aggregated tree

3.6 Random Forest

The Random Forests are a type of Machine Learning algorithm that is used for both classification and regression tasks. The Forests provided by the party package construct the forest based on multiple conditional inference trees at training time and "use of a voting scheme. All the trees in the forest contribute a vote based on what each tree thinks is the most likely response outcome" (Tagliamonte and Baayen 2012)(2012) . Each decision tree in a Random Forest is generated using a random subset of features. The goal of this randomization is to reduce overfitting by making each tree less correlated with the others. This technique was used to identify the importance of each variable in every split and to discern which had the most impact in the choice of complementation patterns. In the figure is possible to observe the Forest constructed for the aggregated group, considering both lemmas. At the top is shown the variable that contributes the most to reduce the impurity in the model. This variables are presented in decreasing order.

4. Results

In the following section, we will present the array of results obtained through the method outlined earlier. Our investigation commences with an examination of the general group findings, providing a broader perspective on linguistic trends and common patterns that pervaded the 18th-century English texts. This collective analysis allows us to discern the prevalent syntactic and semantic preferences that characterized the language during this epoch. Subsequently, we venture into a more intricate exploration, scrutinizing each author's distinct linguistic footprint. By delving into the individual
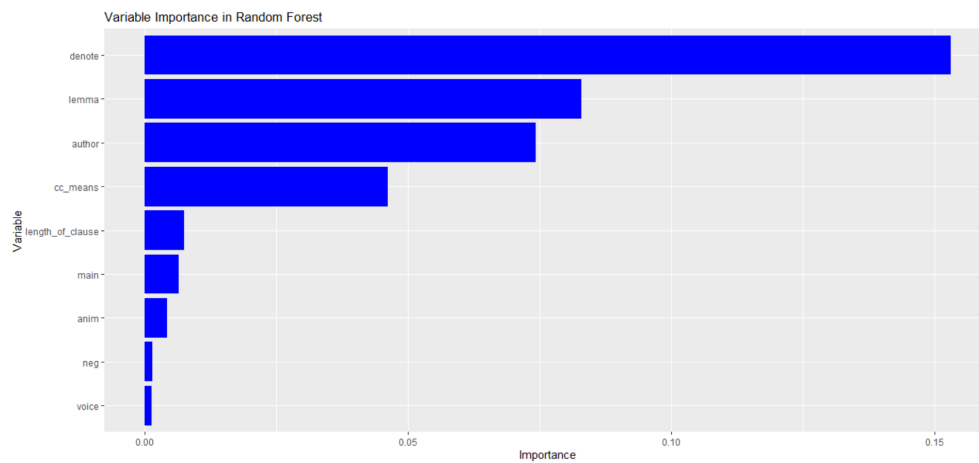
14

Figure 3
Aggregated Forest

writer's choices in sentence complementation and verb usage, we pretend to uncover the stylistic nuances that set them apart. The fusion of these two analyses paints a comprehensive portrait of Late Modern English, showcasing both the unified linguistic features and the diverse authorial voices that coexisted during this era.

4.1 Aggregated group

In this tree fig. 2 we can observe that there is a first split on *Denote*, obtaining two well differentiated groups. When *Denote = Diff*, which indicates that the subject from the CC is different from the one of the CTP, we can observe that the vast majority decides to use a finite clause. The next split in this area of the tree is by Lemma. When Lemma = forget, there is only one terminal node where the majority chooses the *that-clause* construction. When Lemma = remember we face a different scenario, even if all the CCs are finite, we can see that we have two splits by author, getting three groups, one that chooses the *that-clause*, with Trusler, Burke and Gibbon; one mixed, with Richardson and Colman; and one that chooses the *0-complementizer* construction, with Berkeley, Cumberland and Dibdin. Going back to the first split, when *Denote* = same, indicating that the subject from the CTP and the CC are the same, the next split is by the Meaning of the CC. When the verb of the CC is a state (CC-means = state), there is a preference for the *that-clause*, even if some cases of *0-complementizer* and *to-infinitive* are admitted. When the CC-means corresponds to an event we observe a different situation, where the *that-clause* is not the preferred choice and the non-finite clauses start to get relevance. At this level we may observe again a split by Lemma. For forget we observe that the majority of cases opts for the *to-infinitive*, privileging the non-finite clauses, which is the opposite to when the denotation is different between the CTP and the CC. When Lemma = remember we can observe a similar pattern than in the previous case, for the verb remember we have different groups by authors. The first group, conformed by Berkeley, Burke, Colman and Cumberland show a marked tendency to use the non finite *to-infinitive* construction, while the other group conformed by Richardson, Trusler, Dibdin and Gibbon show a slightly tendency to the finite clause *0-complementizer*, but they also accept cases of non finite. Is necessary to remark that the due to the nature of

the data the classes are imbalanced. When we observed at the results of the Random Forest applied to this group we can see that the variable that influence the classification the most is the denotation, followed by the lemma.

4.2 Aggregated forget data

By looking at the tree fig. 4 generated for the classification of the total group, but only to the lemma forget, we can see a much more simplified plot. This one only has three terminal nodes. The main split corresponds, again to *Denote*. When the denotation is different between the CTP and the CC we can see that the preference is for the finite *that-clause*. When the denotation is the same in both clauses we have a second split on the Meaning of the CC. When the verb corresponds with a state we can observe that there is a slight preference for the finite *that-clause*, but the non finite *to-infinitive* also seems to be an option. When the verb exhibits a state we can see a marked preference for the non finite clauses, with the majority being the *to-infinitive*.
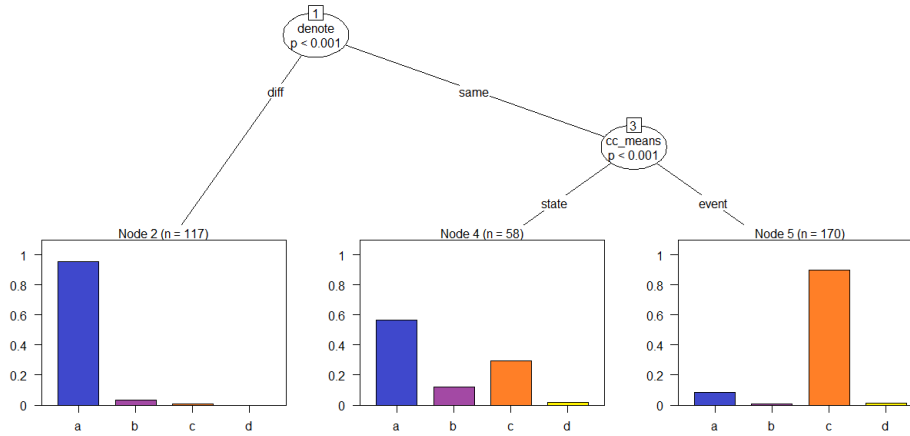

Figure 4
Aggregated forget tree

By looking at the importance of the variables of the Random Forest we can observe that the one with the highest value corresponds to *Denote* as well.

4.3 Aggregated remember data

As we could observe in the first tree fig. 2, the classification depends mainly on the authors. In the tree fig. 6 plotted only for the lemma remember in the aggregated group we can note that the terminal nodes are grouped by authors, showing that with the verb remember exists a major freedom when it comes to use the CC. The first split shows that there are two big groups, one that doesn't privilege the use of the *that-clause* and the other that shows a certain tendency to the use of finite clauses. In both groups we can observe that the next split corresponds to *Denote*. We can see that in all the cases where the denotation is different in the CTP and the CC the tendency is to use finite clauses. One group conformed by Burke, Gibbon, Richardson and Trusler (3, 7, 12 and 13, respectively) have a strong tendency to use the *that-clause*, while in the other group by Berkeley, Colman, Cumberland and Dibdin (1, 4, and 6, respectively) there
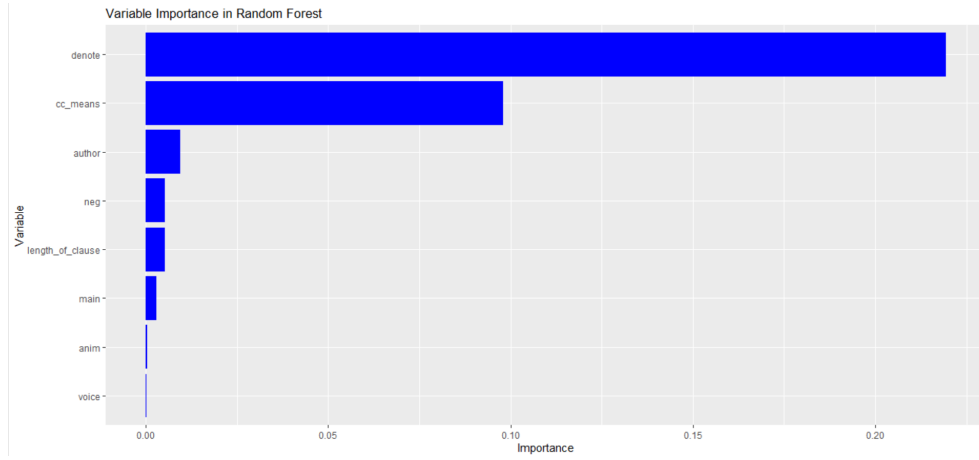
16

Figure 5
Aggregated forget Forest

is a preference for the *to-infinitive*. When the denotation is the same, the first group of authors do different things, but we can observe that the *that-clause* is not the majority. In the second group, when the denotation is the same, the preference is for the *to-infinitive*.
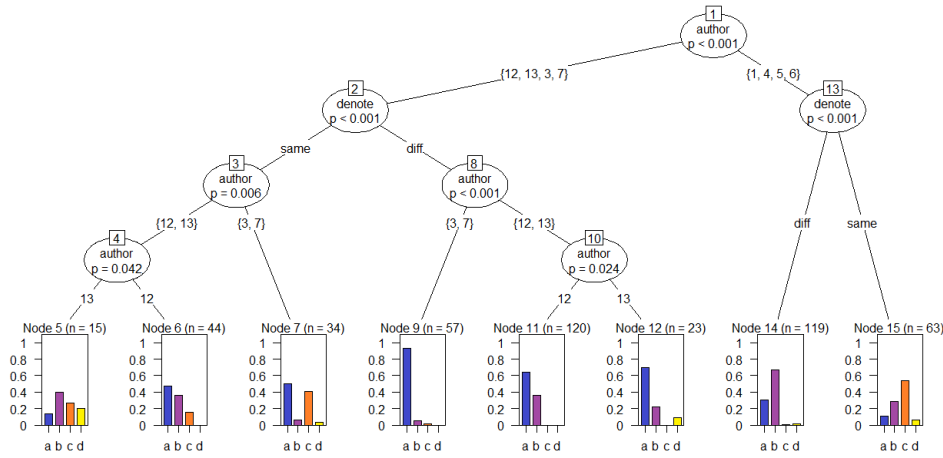

Figure 6
Aggregated remember tree

In this particular case, the most important variable is author, followed immediately by *Denote*, while the other variables are drastically different from the first two.

4.4 Individual trees

Now, we proceed with a focused analysis of specific cases, particularly those that reveal notable distinctions. While some authors align with the general linguistic trends, forming a cohesive group, others deviate, showcasing unique characteristics that set them apart. As it was mentioned in the methodology section, it is important to emphasize
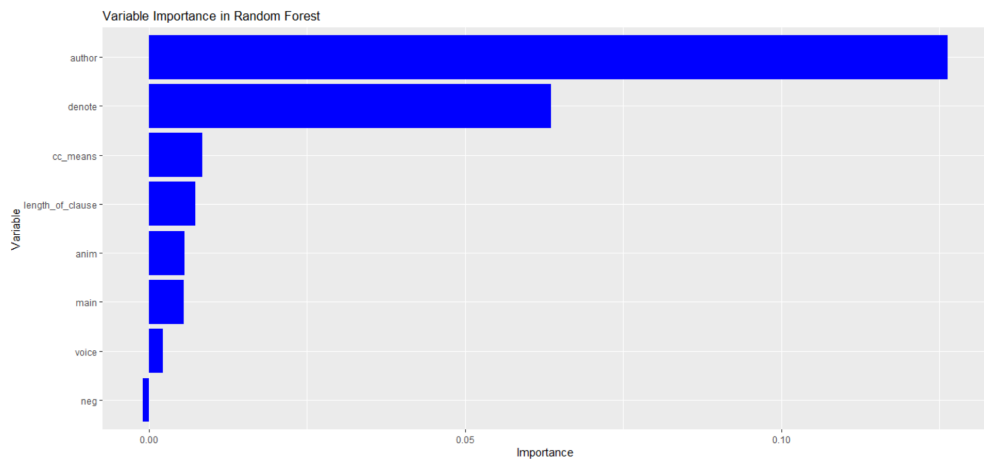
Figure 7
Aggregated remember forest

that only those authors with more than 50 observations were subject to the singular experiments. This decision was made considering the fact that authors with fewer observations lack sufficient data to perform a representative analysis.

4.4.1 Edmund Burke. In Burke's initial tree analysis fig. 8, when both lemmas are encompassed within the classification, a noteworthy observation emerges: there exists no distinct separation based on *lemma*, suggesting that *lemma* alone does not significantly influence the choice of complementation type. However, a notable separation does become evident in the case of *Denote*, where a specific pattern unfolds related to the correspondence between the complement-taking predicate and the complement clause subjects.
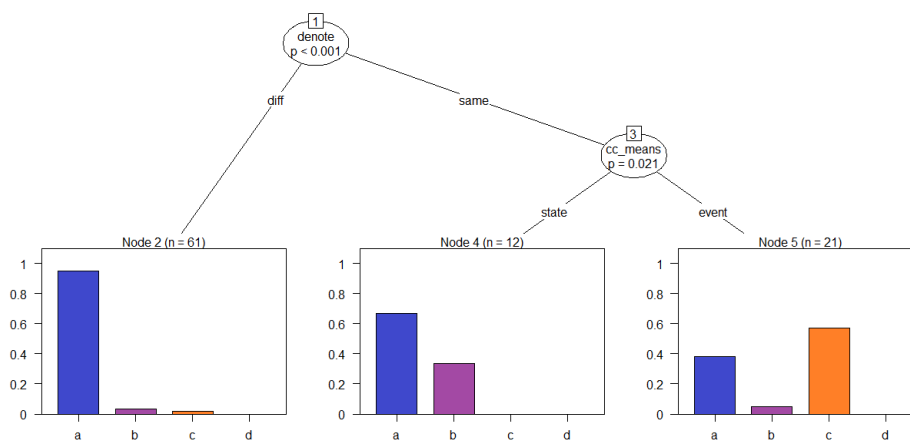

Figure 8
Burke tree

In instances where *Denote = Diff*, a clear preference surfaces for *that-clause*s, indicating a preference for finite clauses. This preference appears to hinge on the subject alignment between the CTP and the CC. Conversely, when *Denote = Same*, the significance of the verb's semantic meaning comes into play. When the verb signifies a state, the inclination remains toward finite clauses, incorporating the utilization of the *0-complementizer.* However, when the verb signifies an event, non-finite clauses take center stage, manifested through the use of the *to-infinitive* structure.

When we transition to the graphical representation of trees organized by lemma, a consistent adherence to the trends observed in the aggregated tree becomes evident. This uniformity of pattern suggests a consistent approach by the author, indicating a lack of differentiation based on lemma.

In essence, the analysis of Burke's trees implies that, for Burke, the most decisive factor governing the choice of complementation lies in the correspondence between the subjects of the CTP and the CC. This emphasis on subject alignment appears to wield a significant influence on the selection of complementation type within Burke's linguistic framework.

4.4.2 George Colman. Turning our attention to Colman's analysis, a familiar trend emerges: upon considering both lemmas for tree construction fig. 9, there remains a relevant absence of differentiation by lemma. This suggests that lemma itself does not have a substantial impact on the choice of the complementation type.
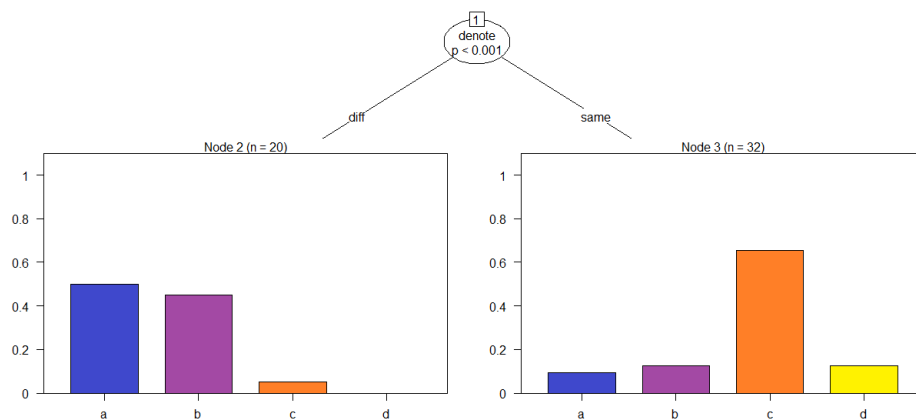

Figure 9
Colman tree

Similarly, in alignment with previous observations, a crucial division materializes first with the *Denote* variable. Upon closer examination, we discern that when *Denote* = Diff, a predominant inclination exists for finite clauses. Notably, this preference manifests itself through the utilization of both the *that-clause* and the *0-complementizer*. By contrast, under the condition *Denote* = Same, an alternative preference surfaces: non-finite clauses take precedence. Upon further scrutinizing the trees created by lemma differentiation, a consistent pattern emerges regardless of the specific lemma. This conformity underscores the author's alignment with the overarching trends shown on the general tree. In sum, the scrutiny of Colman's analysis underscores a recurring motif: lemma-based distinctions offer limited insight into complementation type se-

lection. Instead, it becomes evident that the dynamics of *Denote* and the nature of its connection to finite or non-finite clauses wield more substantial sway over the author's choice of complementation structure. Furthermore, the author's fidelity to the broader trends observed in the general tree underscores a cohesive and consistent approach to complementation within Colman's linguistic framework.

4.4.3 Richard Cumberland. In the context of Cumberland's analysis fig. 10, a similar trend emerges once again: the initial division hinges on the alignment of subjects between the CTP and the CC. Specifically, when scrutinizing instances where *Denote* = Diff, two notable observations come to light. Firstly, a recurring pattern arises: a distinct preference leans towards finite clauses. Secondly, a lemma-based differentiation becomes evident. When the verb is *forget*, the preference veers towards the *that-clause*, whereas when the verb is *remember*, the preference shifts towards the *0-complementizer*.
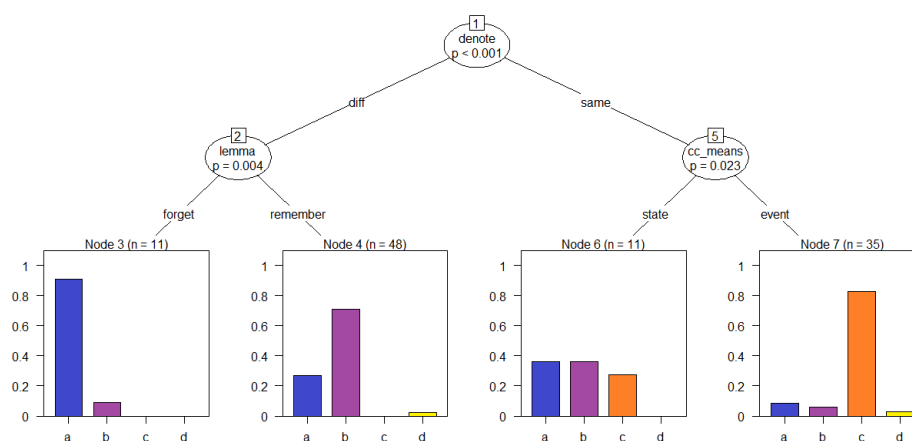

Figure 10
Cumberland tree

Meanwhile, when the condition is *Denote = Same*, the subsequent divergence rests on the nature of the verb's meaning (*CC-Means*). Notably, for eventive verbs, an evident predilection surfaces for non-finite clauses, often characterized by the use of the *to-infinitive* structure. Conversely, when dealing with state verbs, a nuanced approach emerges: both finite and non-finite clauses are employed, with a preference for the use of finite clauses.

Upon closer examination through lemma-based differentiation, a striking consistency emerges. Cumberland's linguistic choices align with the prevailing tendencies observed within the general group. This alignment underscores the author's adherence to overarching patterns, which reflect a coherent approach to complementation strategies.

In essence, Cumberland's analysis reinforces the recurring theme that subject correspondence wields significant influence over complementation choices. Moreover, the interplay between *Denote* and the verb's semantic attributes emerges as a key determinant.

4.4.4 Charles Dibdin. Dibdin's case distinguishes itself from the previously discussed authors. Notably, in the tree fig. 11 constructed by considering both lemmas, a unique

variable takes center stage for the initial split. Specifically, when the lemma is *remember*, the division leads to finite clauses, with a notable preference for the utilization of the *0-complementizer*. Conversely, when the lemma is *forget*, the subsequent division centers around *Denote*. When *Denote = Same*, a clear inclination emerges towards non-finite clauses, often employing the *to-infinitive* structure. On the other hand, when *Denote = Diff*, the predilection once again shifts towards finite clauses, often characterized by the use of the *that-clause*.
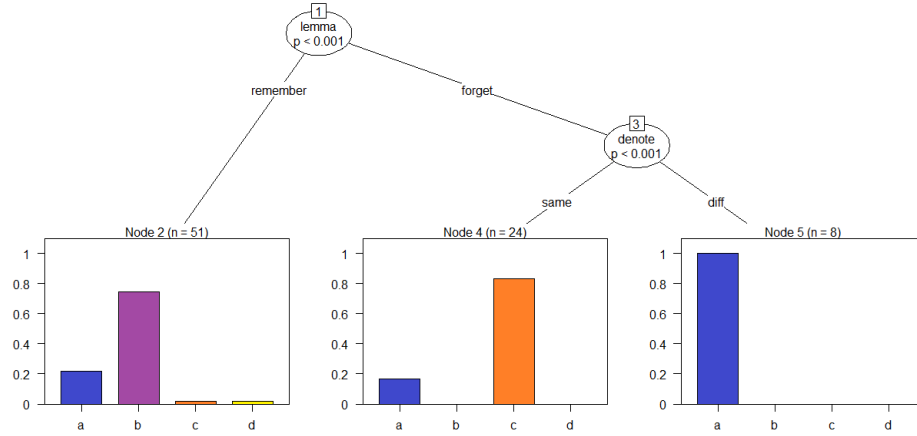

Figure 11
Dibdin tree

In a broader context, Dibdin's approach adheres to the overarching complementation patterns observed within the general group. However, a distinctive feature emerges: the pronounced emphasis on lemma during the differentiation process. This emphasis suggests that Dibdin tends to approach complementation choices more from a lexical perspective rather than strictly adhering to semantic class distinctions.

4.4.5 Edward Gibbon. Gibbon's analysis brings us back to the prevailing pattern, looking at the tree fig. 12 we can see that *Denote* assumes a central role as the primary dividing factor. When exploring instances of *Denote = Diff*, the inclination reverts to a favoring of finite clauses, prominently utilizing the *that-clause* construction. In the context of *Denote = Same*, a distinct preference emerges for non-finite clauses, primarily through the utilization of the *to-infinitive* complementizer. However, noteworthy flexibility exists, allowing for the occasional utilization of finite *that-clauses* as well.

Upon delving into the trees constructed based on different lemmas, a harmonious alignment with the principal trends becomes evident. This concurrence reaffirms the author's adherence to the broader complementation patterns established within the corpus.

In essence, Gibbon's analysis reiterates the familiar theme of *Denote* playing a key role in complementation distinctions. The prevailing preference for finite and non-finite clauses depending on the *Denote* condition underscores the intricate interplay between semantic attributes and grammatical choices. Furthermore, the seamless transition across different lemmas reaffirms Gibbon's consistency in conforming to the overarching complementation tendencies, thus contributing to the cohesion of the linguistic analysis.
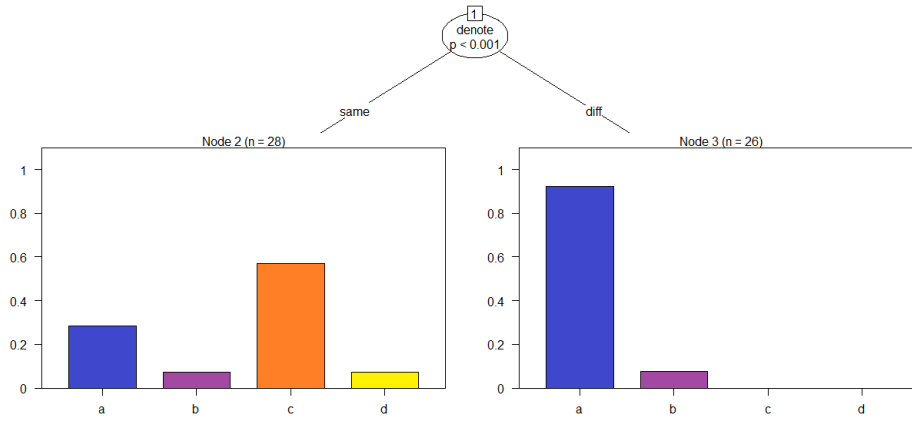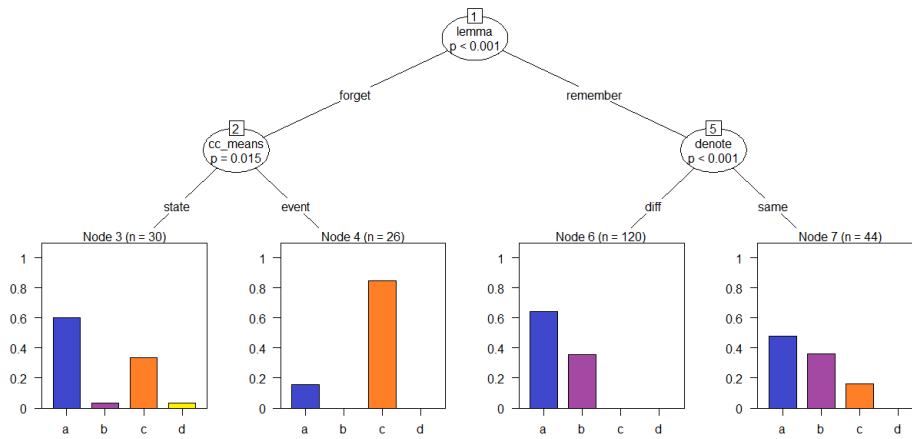
Figure 12
Gibbon tree

Figure 13
Richardson tree

4.4.6 Samuel Richardson. Richardson's analysis introduces the most marked deviations from the approaches of other authors. In this instance, similar to Dibdin, the initial bifurcation arises from the lemma itself. When Lemma = Forget, the subsequent division aligns with the verb's semantic import. Within the context of CC-Means = Event, a notable preference emerges for the non-finite *to-infinitive* clause. Conversely, when the verb conveys a state, the choice pivots between finite and non-finite clauses, with a mild inclination towards finite *that-clause*s. Turning attention to the verb 'remember,' a distinct preference for finite clauses surfaces. This predilection then cascades into a further differentiation based on subject correspondence. Under the conditions of *Denote* = Same, finite clauses, encompassing both the *that-clause* and the *0-complementizer*, assume prominence. Nonetheless, support for non-finite *to-infinitive* structures also remains. In the context of *Denote* = Diff, an exclusive utilization of finite *that-clause*s becomes evident. While Richardson displays a propensity for lemma-driven complementation

selection, a concurrent adherence to the broader group trends is discernible. Notably, when plotted in the aggregated tree under the lemma remember, Richardson occupies a distinct node from fellow authors. This differentiation is attributed to his distinctive allowance of the *0-complementizer* combined with the verb remember. This atypical choice is exemplified in the following illustration:

Example 20
And remember we have been acquainted these hundred years. (Richardson, 1753)

In summary, Richardson's approach introduces a dual inclination: prioritizing complementation based on lemma while concurrently aligning with the general group tendencies. His distinctive embrace of the *0-complementizer* with the verb *remember* sets him apart within the group, contributing to a nuanced understanding of his complementation strategies.

4.4.7 John Trusler. In the context of Trusler's analysis fig. 14, a familiar trend emerges once again: subject alignment between both clauses assumes a key role as the primary division factor. Notably, under the condition *Denote = Diff*, a notable preference emerges for finite clauses, prominently favoring the *that-clause* construction.
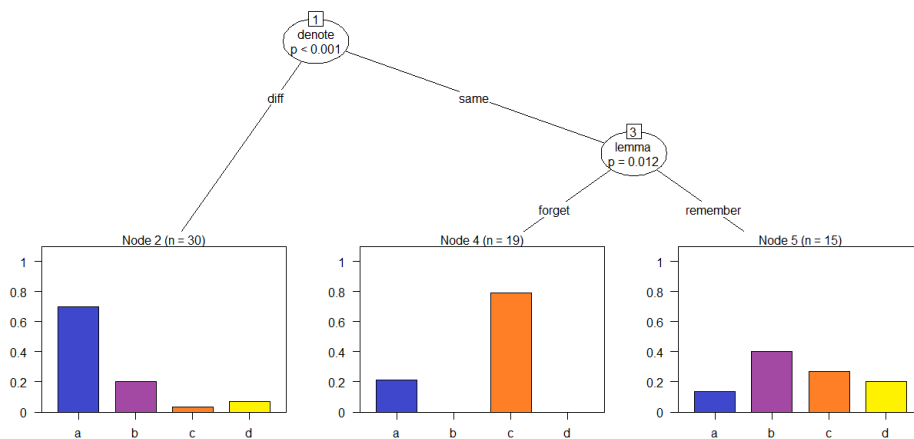

Figure 14
Trusler tree

However, a distinctive shift is observed when *Denote = Same*. Under this circumstance, a novel bifurcation surfaces, this time rooted in *Lemma*. Upon closer examination, for the lemma *remember*, a diverse array of clause types finds application, with finite and non-finite clauses being employed with equal frequency. Conversely, when dealing with the lemma *forget*, a clear inclination emerges for the non-finite *to-infinitive* structure.

Trusler's analysis introduces an intriguing departure from his contemporaries. His preference for non-finite *to-infinitive* structures with the lemma *forget* distinguishes his approach. This notable variation underscores the nuanced individual tendencies within the broader group.

In summary, Trusler's study contributes to the corpus of analyses by reaffirming the centrality of subject correspondence in complementation choices. Furthermore, his distinct behavior, particularly regarding the lemma *forget*, adds a layer of complexity and individuality to the overall understanding of complementation strategies.

5. Analysis

Upon a comprehensive examination of the results, it becomes evident that discernible trends emerge both at the collective group level and at the author level analyses. Primarily, the *Denote* variable emerges as an essential factor exerting its influence consistently across both the broader group and most of the specific cases examined. The role played by *Denote* is directly related with the concordance of the subject in the MC and in the CC. For those sentences in which there was subject agreement between the two clauses, the choice was for finite clauses. Conversely, when the subject did not correspond between the two parts, the complementation patterns pointed towards the use of finite clauses. Both results are in agreement with Rohdenburg (1996) principle of complexity, since for the first case described, when the denotation is the same, there is a lower level of complexity in the sentence, giving rise to more intricate structures such as non-finite clauses. For the second case, in which the subject of both is not the same and the cognitive complexity increases, the use of finite clauses is favored. The finite clauses encapsulate their own subject and tense, rendering them self-contained semantic units. Consequently, the cognitive load on the reader is alleviated, particularly in scenarios necessitating the processing of distinct referents and agents. Drawing from the same idea as above, it is also possible to realize that Rohdenburg (1996) statement on the preference for the most explicit clause for finite and non-finite is also evidenced in the results. When the denotation is the same and space is given to non-finite clauses, we can observe that the to-inifitive quantity is much higher than that of cc-ing. While similarly, when the denotation differs and finite clauses are chosen, the *that-clause* exhibits a much higher frequency than the *0-complementizer*. By focusing on the analysis carried out individually by author, it was possible to identify differentiated trends. Most of the authors (Burke, Colman, Cumberland, Gibbon and Trusler) showed an inclination towards similar compelementation patterns without distinguishing the lemma used. These authors maintained *Denote* as the most important variable when determining the type of complementation to use, regardless of whether the verb used was forget or remember. The second group, composed of Dibdin and Richardson, exhibited a tendency to prioritize the lemma in deciding which form of complementation to use. Taking into account the analysis conducted by lemma, we can see that, in general, the same pattern is repeated. The prevalence of the *Denote* variable as the one that determines the type of complementation is maintained, opting for the use of finite clauses for those clauses with different denotation and using non-finite clauses for those in which the subject of the denotation coincides. However, it is interesting to note that, although the behavior is broadly the same, for the case of remember the author becomes relevant. It is possible to see that, while for the tree of the verb forget there is no split at the author level, for remember there is. This shows that despite sharing the same patterns, authors seem to have more freedom when combining other variables. However, the nature of these results is still vague and further study of this phenomenon is needed to determine whether it is really a determining factor.

6. Conclusion

In this paper, we have comprehensively analyzed complementation patterns in Late Modern English as employed by eighteenth-century authors. Through the integration of corpus analysis and statistical methodologies such as Inferential Trees and Random Forests, we have effectively examined the diverse variables that influence complemen-

tation choices. Our research sheds light on both collective behavioral trends within this group and distinctive individual preferences.

One notable observation is the prominence of the denotation variable, signifying its key role in determining cognitive intricacies within sentences. Denotation emerges as a significant factor impacting the choice of complementation pattern. Authors appear to opt for simpler complementation structures for sentences with higher cognitive complexity and vice versa. In instances of dissimilar denotation, authors lean toward finite clauses, thereby enhancing the clarity of subject representation. Conversely, when denotation is the same, non-finite clauses are favored to avoid redundancy.

Additionally, we noted the lemma's influence on complementation choices. Despite semantic opposition, the verbs *forget* and *remember* manifest distinct preferences during complementation. The former displays a more straightforward approach, whereas the latter exhibits a greater degree of linguistic flexibility, often leading to unique patterns that differentiate authors.

At an individual level, a subset of authors adheres to the denotation variable as a primary determinant in their complementation choices. Conversely, another subgroup showcases variations influenced by lemma distinctions. This intriguing interplay further enhances our understanding of how different variables contribute to authors' linguistic decisions.

In summary, our investigation delves beyond linguistic surface structures, revealing intricate dynamics between cognitive nuances and stylistic preferences. This study not only provides historical insights but also invites a deeper exploration into the cognitive and linguistic complexities that shaped the eighteenth-century literary landscape.

Returning to the hypotheses that were put forward at the beginning of the research, we can see that, in general, the results point to the verification of these hypotheses.

H1: Discernible patterns emerge within the data, effectively grouping either finite or non-finite clauses. In the majority of cases, a clear preference is observed for one type over the other, with minimal instances of mixed patterns.

H2: This assertion holds true to a certain extent. The trend is mostly consistent for the verb *forget*, with authors following a similar path. However, when it comes to *remember*, a more diverse range of choices becomes evident. Despite the semantic closeness of these verbs, their complementation behaviors manifest notable distinctions.

H3: Contrary to a scenario of unique individual behaviors, our analysis reveals a trend within particular author groups. The majority of authors align with consistent tendencies, with minimal variance from the established patterns. Intriguingly, two authors deviate from the norm by introducing a distinct variable into their complementation choices.

H4: Additionally, it's noteworthy to include an unfulfilled hypothesis that posits the limited role of intervening variables in the decision-making process. While denotation often took the lead, and occasionally *cc-means* contributed, the other variables exhibited negligible impact.

Moreover, an intriguing observation emerges regarding the lack of influence from the length variable. Although an expectation existed for possible semantic and syntactic variations, the outcome reveals its insignificance in shaping complementation preferences.

Concluding this study prompts the identification of valuable prospects for potential future endeavors. Primarily, the scope of the analysis could be widened by encompassing a more expansive cohort of authors. While the limitation of available data did necessitate the exclusion of certain authors from our examination, a broader and more diverse selection could yield deeper insights and foster more coherent outcomes.

The eighteenth century's prolific literary output suggests a rich reservoir of works that could significantly expand the corpus. With the application of the utilized techniques, accommodating a larger dataset becomes viable, and the ongoing efforts by scholars to augment existing databases are highly promising.

Secondly, an intriguing avenue for exploration lies in extending this experiment across successive generations of authors. This temporal extension could unravel the dynamic evolution of complementation patterns over time, enabling insightful comparisons and elucidating potential shifts within linguistic choices.

Thirdly, the horizon beckons us to incorporate a more diverse array of verbs into the study. The distinct behaviors observed between *forget* and *remember* raise the tantalizing possibility that a spectrum of underpinning patterns remains yet to be unveiled. The inclusion of additional verbs could unveil an intricate tapestry of complementation choices, providing a more comprehensive understanding of the nuanced intricacies governing these linguistic decisions.

References

Aljović, Nadira. 2017. Non-finite clauses in english. *Properties and Function. Sarajevo.*

Chomsky, Noam. 2014. *Aspects of the Theory of Syntax*. 11. MIT press.

Cuyckens, Hubert, Frauke D'hoedt, and Benedikt Szmrecsanyi. 2014. Variability in verb complementation in late modern english: Finite vs. non-finite patterns. *Late Modern English Syntax*, pages 182–203.

Cuyckens, Hubert, Lauren Fonteyn, and Petré Peter. 2021. Complexity in complementation. *Cognitive sociolinguistics revisited*, pages 264–276.

Dąbrowska, Ewa. 2020. Language as a phenomenon of the third kind. *Cognitive Linguistics*, 31(2):213–229.

Duffley, Patrick J. 2000. Gerund versus infinitive as complement of transitive verbs in english: The problems of "tense" and "control". *Journal of English Linguistics*, 28(3):221–248.

Fonteyn, Lauren and Andrea Nini. 2020. Individuality in syntactic variation: An investigation of the seventeenth-century gerund alternation. *Cognitive linguistics*, 31(2):279–308.

Gries, Stefan Th and Andrea L Berez. 2017. Linguistic annotation in/for corpus linguistics. *Handbook of linguistic annotation*, pages 379–409.

Group", "Five Graces, Clay Beckner, Richard Blythe, Joan Bybee, Morten H Christiansen, William Croft, Nick C Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman, et al. 2009. Language is a complex adaptive system: Position paper. *Language learning*, 59:1–26.

Hothorn, Torsten, Kurt Hornik, and Achim Zeileis. 2006. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, 15(3):651–674.

Huang, James. 1997. Introduction to syntax. *LSA Linguistic Institute*.

Huddleston, Rodnry and Geqffrry Pullum. 2005. The cambridge grammar of the english language. *Zeitschrift für Anglistik und Amerikanistik*, 53(2):193–194.

Hundt, Marianne. 2014. Introduction: Late modern english syntax in its linguistic and socio-historical context.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. A comprehensive english grammar. *London and New York: Longman*.

Quirk, Randolph and Charles Wrenn. 2002. *An old English grammar*. Routledge.

R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rohdenburg, Günter. 1995. On the replacement of finite complement clauses by infinitives in english.

Rohdenburg, Günter. 1996. Cognitive complexity and increased grammatical explicitness in english.

Rohdenburg, Günter. 2014. On the changing status of that-clauses. *Late Modern English Syntax*, pages 155–181.

Schmid, Hans-Jörg. 2016. Why cognitive linguistics must embrace the social and pragmatic dimensions of language and how it could do so more seriously. *Cognitive linguistics*, 27(4):543–557.

Somers, Harold L. 1987. Valency and case in computational linguistics. *(No Title)*.

Steels, Luc. 2000. Language as a complex adaptive system. In *International*

Conference on Parallel Problem Solving from Nature, pages 17–26, Springer.

Tagliamonte, Sali A and R Harald Baayen. 2012. Models, forests, and trees of york english: Was/were variation as a case study for statistical practice. *Language variation and change*, 24(2):135–178.

Th Gries, Stefan. 2019. On classification trees and random forests in corpus linguistics: Some words of caution and suggestions for improvement. *Corpus Linguistics and Linguistic Theory*, 16(3):617–647.

Tolonen, Mikko, Eetu Mäkelä, Ali Ijaz, and Leo Lahti. 2021. Corpus linguistics and eighteenth century collections online (ecco). *Research in Corpus Linguistics*, 9(1):19–34.