Conceptual Engineering and the Royal Society Corpus

Haket, Nina

NULLIUS IN VERBA

CONCEPTUAL ENGINEERING AND THE ROYAL SOCIETY CORPUS

Nina Haket

UNIVERSITEIT LEIDEN

2022

Word Count: 32,085

NULLIUS IN VERBA

CONCEPTUAL ENGINEERING AND THE ROYAL SOCIETY CORPUS

Thesis submitted to the

Leiden University Centre for Linguistics

in partial fulfillment of the requirements for the degree of

Research Master

in

Linguistics

by

Nina Haket

Universiteit Leiden

2022

ACKNOWLEDGEMENTS

First of all, I would like to thank Lauren Fonteyn and Janet Connor for their guidance and support before and during the production of this thesis. Their advice has been invaluable.

Secondly, I would like to thank Jake Stuchbury-Wass for explaining numerous scientific principles and terminology with which I was not familiar, and Bianca Haket-Koolen for her dedication and input in the proofreading process.

Finally, I would like to express my gratitude to my friends and family, both in the Netherlands and in the UK, for their words of encouragement, coffee provision, and endless positivity.

DECLARATION OF ORIGINALITY

I, Nina Haket, certify that

- I am the sole author of this thesis and that I have fully acknowledged and documented in my thesis all sources of ideas and words, including digital resources, which have been produced or published by another person or institution;
- This thesis contains no material that has been submitted or accepted for a degree or diploma in any other educational institution;
- This is a true copy of the thesis approved by my supervisors at Universiteit Leiden, including final revisions required by them.

Signature: Nina Haket

Date: 09/06/2022

ABSTRACT

Conceptual Engineering is the practice of improving the concepts we use for a specific purpose. However, despite involving words and their meanings, this practice has not been looked at from the perspective of linguistics. This paper takes a small, niche scientific community, namely the Royal Society, and investigates to what extent Newton's proposed distinction between *mass* and *weight*, which can be thought of as an instance of Conceptual Engineering, was consistently used in scientific journal articles between 1700 and 1920.

Before Newton's *Principia*, the lemma *weight* referred to both the amount of matter that makes up an object and the force of gravity acting on the matter. In 1687, Newton proposed that the former concept should be referred to as *mass*, and the latter as *weight*. Success, for this project, is differentiation based on sense and not on any other extra-linguistic factors.

To discover whether the project was successful, 1500 tokens of *mass* and *weight* from the Royal Society Corpus were annotated for their lemma, 'sense', 'object', 'subfield', 'author', 'year', 'type', 'plurality', and part-of-speech. This data was analysed by looking at the frequencies of the respective senses over time, along with Ctrees and Random Forests to identify annotations that were the most important in predicting the lemma, and Principal Component Analysis to visually inspect clustering and patterns over time.

The results showed that sense was not an effective predictor of the lemma, but that the extra-linguistic factors of 'object', 'author' and 'subfield' had great predictive power. Furthermore, the Principal Component Analysis confirmed clustering based on 'object', 'author' and 'subfield' rather than based on 'sense'. It was concluded that Newton's Conceptual Engineering of *mass* and *weight* was unsuccessful.

From the perspective of Conceptual Engineering more generally, this work showed future research on more diverse communities with less coherent language ideologies is necessary when investigating cases of Conceptual Engineering, and also that linguists can be important contributors to Conceptual Engineering research in the future.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Conceptual Engineering (CE) is becoming a veritable buzzword within studies on the philosophy of language. CE is the practice of improving the concepts we use for a specific purpose. Rather than taking a concept and analysing what it is and how it is used, as is the case in conceptual analysis, conceptual engineers take a concept, find its defects, and offer an improved concept that addresses these deficits, thereby making it more suitable for purpose. CE is a methodology within philosophy, and concerns about concept suitability can already be found in the work of Nietzsche, who suggests “What is needed above all is an absolute scepticism toward all inherited concepts” (Nietzsche, 1901, pg 220-1). Despite its long history and prominence in philosophy, few linguists are aware of CE.

Upon being told that CE is the practice of improving our concepts, linguists may be perturbed by the subject of discussion. As linguists, we are descriptive rather than prescriptive. We investigate how language is used rather than how it should be used. This is intrinsically in opposition to CE, which aims to provide prescriptive and normative judgement on how we should be speaking. On the surface, this suggests the two are incompatible.

Traditional instances of CE tend to focus on improving concepts required for the progress of academic knowledge. For example, CE projects in philosophy resulted in new ameliorated concepts of *truth* (Scharp, 2021), *knowledge* (Fassio and McKenna, 2015), and *belief* (Clark and Chalmers, 1998). However, the recent wave of CE projects goes beyond these academic and logical roots. These new projects aim to improve concepts pertinent to society. These include projects to improve concepts such as *woman* (Haslanger, 2000), *fake news* (Habgood-Coote, 2020), and *misogyny* (Manne, 2017).

The criteria for the success of CE projects, especially those aimed at more societal concepts, have changed. It no longer suffices to simply suggest an improved concept. Instead, the aim is society-wide adoption and reform. Haslanger’s political reengineering of *woman*, for instance, might be conceived and even picked up in sociological academic work, but the project will not be truly successful until society as a whole uses the new concept.

The practice of CE has, therefore, introduced an implementation problem (Cappelen, 2018; Deutsch, 2020; Jorem, 2021; Koch, 2021). Any re-engineered or ameliorated concepts must spread through society, but little work has been done on how to spread them, or what patterns we may expect to see. However, this question lies outside the domain of philosophical thought and crosses into the domain of linguistics. Previous investigations of this implementation problem conclude that the success of CE, and thus the worthiness of continued study and funding, depends on whether these projects can ever be successfully spread or not. Philosophy cannot answer this question, but linguists potentially can. While linguistic knowledge surrounding language change is not predictive (it cannot predict whether a change will be successful), it has an established toolbox of strategies to analyse ongoing and past semantic change.

For centuries, (socio-)linguists have studied how linguistic change is introduced and diffuses over time and communities. As will be argued in this thesis, linguists have tools to analyse the potential or actual spread of these terms (Labov, 2001, 2010; Paul, 1890; Osthoff and Brugmann, 1878).

1.1 This work

This work aims to take the first step in analysing the spread of conceptually engineered terms with two intentions. The first is to apply linguistic tools to the issue of concept diffusion, and the second is to draw the attention of linguists and hopefully inspire more work from linguists in CE.

Rather than tacking a large, societal instance of CE, many of which are still ongoing and hard to evaluate, this study looks at CE in the diachronic scientific domain where ameliorating the concept aimed to improve scientific precision. It will take a community with clear criteria for whether a concept is defective, and definitive authority on who decides what constitutes improvement. More specifically, it will look at the scientific community and the concepts *weight* and *mass* as first differentiated by Newton (1687).

1.1.1 Research questions

This study investigates the usage patterns of *mass* and *weight* between 1700 and 1920 using data from the Royal Society Corpus (RSC) (Fischer et al., 2020). Influences on lemma choice may include the author, scientific sub-discipline, and various linguistic factors such as sense, plurality, and part-of-speech. This study aims to use these factors to create a first picture of how engineered terms diffuse, including how long it takes for adoption to occur and what linguistic and social factors can encourage or hinder uptake. These patterns may help evaluate the likelihood of success of CE projects. The specific research questions addressed by this work are the following:

RQ1. To what extent are *mass* and *weight* successfully differentiated by sense after Newton proposed separating the concepts in 1687?

RQ2. What time frame and trends are there in the diffusion of *mass* and *weight*?

RQ3. What effect do extra-linguistic variables such as author and subfield have on the choice of lemma?

Research question 1 aims to identify whether the sense is the most important conditioning factor in lemma selection. If the engineering of *mass* and *weight* was successful, the sense should ultimately be the best explanatory factor. Research question 2 aims to provide insight into how long it may take for a CE project to be implemented, and research question 3 assesses the influence of external factors. If these extra-linguistic factors have a large impact on the choice of lemma, then the CE project has not been successful.

1.1.2 Outline

This thesis will be structured as follows. Chapter 2 will look at Conceptual Engineering and its past. Chapter 3 will look at language ideologies and other linguistic theories, approaches, and research that will aid in the linguistic contribution to CE. Chapter 4 provides insight into the Royal Society and contextualises the ideologies and philosophical approaches towards science and language that may have contributed to the change in usage of *mass* and *weight*. Chapter 5 introduces the history of the concepts of *mass* and *weight*. Chapter 6 details the methodology, and Chapter 7 presents the results. After the discussion in Chapter 8, Chapter 9 will conclude and comment on potential future directions.

CHAPTER 2

CONCEPTUAL ENGINEERING

Before we can investigate *mass* and *weight* from the viewpoint of linguistics and Conceptual Engineering, it is necessary to explore what Conceptual Engineering is.

2.1 Amelioration in Conceptual Engineering

The definition of CE is hard to pin down. As a form of introduction to the topic, the characterisation of Eklund (2014) is a good place to start:

> *[W]hile philosophers often have been concerned with our actual concepts or the properties or relations they stand for, philosophers should also be asking themselves whether these are the best tools for understanding the relevant aspects of reality, and in many cases consider what preferable replacements might be.*
> (Eklund, 2014, pg 293)

Eklund's statement reveals two things about CE, the first being that it involves finding 'preferable' replacements, also called improvements or ameliorations, for our concepts. Secondly, it shows that CE contrasts with a purely descriptive approach, which aims to find what our concepts are and how we use them. An 'ameliorative' approach, on the other hand, looks to improve these concepts based on any defects we find. Amelioration in CE contrasts with the linguistic definition of amelioration, which refers to a semantic change in which a word gains a more positive meaning. For example, the adjective *nice* previously meant foolish, whereas now it means pleasant (Altakhaineh, 2018; "† nice, 2021). In CE, amelioration refers only to improving a concept for a specific purpose.

The assumptions of CE are, therefore, that the concepts we use can be defective, and if found to be defective, we should create alternatives to fix them (a sentiment echoed in the title of Cappelen (2018), *Fixing Language*). However, what is not clear from these definitions is who decides what constitutes a defect and what constitutes an amelioration within CE. For whom are certain concepts preferable, and who has the power to determine that we should change the way we speak? Answering these questions requires us to appeal to language ideologies and the notion of authority over language, something which requires attention in and of itself, and so they will be discussed in depth in the next chapter. What I aim to show by mentioning this here is that CE is not uncontroversial. It is not a clearly defined and accepted process. This work does not intend to defend nor promote CE as a process but simply to show possible contributions from linguistics.

However, if we accept that someone has the power to determine what counts as a defect in language, the next question is to ask what these defects can be, and there is no singular answer to this. One way to characterise a defect is to identify the ideal concept. In terms of the relation between form and concept, this could be a single concept that maps onto a single lexical item, as shown in Figure 1. It is important to note that this figure is simplistic since synonymy, polysemy, and homonymy cause form and meaning to deviate from this perfect template (see section 6.1).

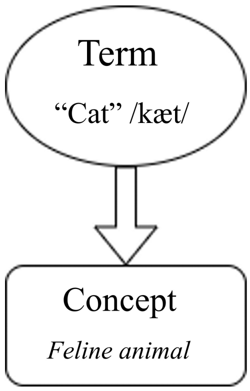

Figure 1: The ideal form-meaning mapping

Regardless, such a straightforward form-meaning mapping does not guarantee that the concept itself is not 'defective'. While multiple concepts mapping onto a single lexical item is the focus of this work, Cappelen and Plunkett (2020) offer a selection of other defects mentioned in the literature, including cognitive defects, moral/political defects, epistemological defects, and metaphysical defects.

To contextualise these broad statements and aims, it may be useful to mention some CE projects. Staying close to the domain of philosophy, Scharp (2007, 2013) improved the concept of *truth* to fix the Liar's Paradox (amongst other problems). Scharp created 'ascending truth' and 'descending truth'. Clark and Chalmers (1998) engineered *belief* to create a more unified concept, and Fassio and McKenna (2015) and Nado (2021) both looked at *knowledge*, with the latter concluding that we need multiple concepts of *knowledge* to be able to account for all its functions. Outside of philosophy (or still within philosophy but with a more social/political motivation) are Haslanger (2000, 2012), who proposed changing the concepts associated with *woman*, *race* and *parent*, Jenkins (2016) and her response and consequent improvement of Haslanger's concepts, and Habgood-Coote (2020), who advocates for the elimination of the concept of *fake news*. These examples are in no way exhaustive but were chosen to highlight the applicability of CE to several different domains, and the numerous possible outputs, as shown in Figure 2.

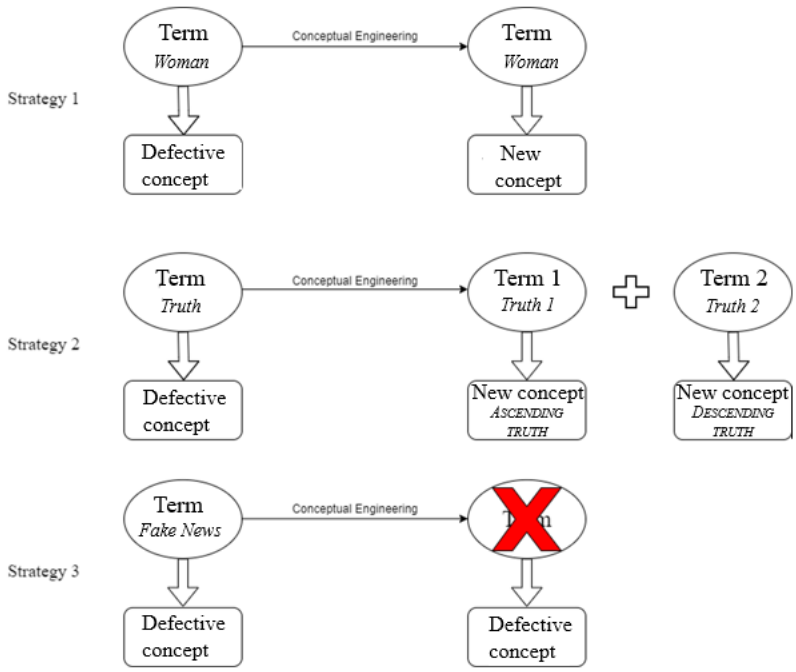

Figure 2: The three CE strategies

Haslanger (2000), Clark and Chalmers (1998), and Jenkins (2016) all maintain the original lexical item but choose to improve the concept behind it (strategy 1), whereas Scharp (2013) and Nado (2021) propose multiple concepts to replace the single original "jack of all trades concept" (Nado, 2021, pg 2) (strategy 2). Habgood-Coote (2020) represents the last option, which is to admit the concept is beyond saving and propose getting rid of it altogether (strategy 3). These represent the three strategies mentioned in Cappelen (2018, pg 35).

2.2 Origins and Carnapian Conceptual Engineering

While there has been an increase in interest in CE since the beginning of the 21st century, most philosophers agree that the idea has existed for a long time. Nietzsche (1901) writes of being dubious about our inherited concepts.

One approach discussed in the literature as being one of the first instances of CE is Carnap's explication (Carnap, 1947, 1950, 1962). Carnap was concerned with the usage of language within science, both in the sense of creating artificial languages and exploring notions of precise language. The latter idea originated from the fact that he believed ordinary concepts to be too vague for precise scientific enquiry (Cull, 2021). Carnap defines explication as the following:

> By the procedure of explication, we mean the transformation of an inexact, pre-scientific concept, the explicandum, into a new exact concept, the explicatum.
> (Carnap, 1947, pg 3)

Put more simply, Carnap proposes replacing vague and imprecise ordinary concepts with more precise ones that fit into a "broader scientific picture" (Cull, 2020, pg 13). Carnap, therefore, is engaged in a form of CE. Carnap also proposes four requirements for explication. The first concerns similarity to the *explicandum*, meaning that in cases where the *explicadum* has been used, the *explicatum* can replace it. The second requirement is the exactness of the *explicatum*, meaning it must be specifiable and fit into a unified system of scientific concepts. The third is fruitfulness, meaning that the *explicatum* must be useful in formulating universal and scientific statements (see Carus (2007) and Dutilh Novaes and Reck (2017) for differing accounts of what fruitfulness could entail). The last is simplicity, where the concept must have a simple definition, and link simply to existing concepts. Carnap did not include social, moral, or political considerations in his explication.

An example given by Carnap himself is that of *fish* and *pisces* (Carnap, 1947). He suggests that *fish* is the pre-scientific *explicandum* improved by zoologists to create the *explicatum pisces* which removed whales and other cetacea. This was because *pisces* was able to generate more universal laws and observations; it is a much more fruitful concept for zoologists. *Pisces* is a more exact term, and is similar to *fish* in the sense that zoologists will use *pisces* in mainly the same contexts they would have used *fish*. Simplicity is a complicated desideratum, but in removing whales and cetacea from the category, one could argue for simplicity in a more homogeneous extension/intension.

2.3 A social turn

Since Carnapian scientific explication, CE projects have not only been concerned with concepts relating to the exact sciences. Even Carnap's contemporaries, such as Neurath (1944) already suggested that non-epistemic values warrant consideration when engineering concepts, such as the needs and wants of the community, thereby introducing moral, political, and social influences. The inputs to CE thus become plural in the addition of this new dimension. This newer CE usually has an aim broader than scientific precision. Instead, it appeals to ontology and linguistic relativity because what many conceptual engineers are interested in changing is not only the concept but phenomena in the real world (Nado, 2021).

The belief that we can change the world through conceptual and lexical change is reminiscent of linguistic relativity, which is the idea that the language you speak affects the way you think about the world (Lucy, 2001). While this idea is commonly applied to grammar, the same principles can be applied to lexical items. For example, Russian has two separate words for *light blue* and *dark blue*. Some scholars suggested that Russian speakers are more astute in finding subtle differences in the colour blue than speakers of languages who only have a single word for blue or those who do not have a word for blue at all (Winawer et al., 2007).

The stance one takes on linguistic relativity affects how one views CE. If linguistic relativity is not real, then we cannot change things by talking about them differently. The concept or referent does not change if we change how we speak, and therefore CE of these more social terms has no future. This is not to say that explication in the Carnapian sense has no future but simply that engineering the terms *rape, woman, misogyny* or *crime* to mean something different will not stop people from committing crimes or rapes or referring to women in a derogatory way.

Conversely, those who subscribe to linguistic relativity may be more optimistic about the potential of CE to change things within society. This view allows language to change how speakers view the world. Often, the purpose of CE is not to completely change the referent of a concept, but to change ideas and ideologies surrounding a concept. Take mosquitoes, for example. One can imagine a CE project that aims to cause speakers to discuss mosquitoes in a more positive light, emphasising their contribution to the ecosystem, rather than negatively, by discussing them concerning malaria. The referent of the concept of *mosquito* does not change, only our ideas about them.

This newer hope to change the world through changing concepts is not as new as one might think, since Carnap (1963) already believed that more precise scientific concepts "must be regarded as one of the most valuable instruments for the improvement of life" (pg 83).

2.4 A note on concepts

Across the literature, it seems as if every article takes a different stance on what concepts are in CE. Deutsch (2020) takes concepts to simply be "the meaning of terms" (Deutsch, 2020, pg 3938), which is similar to Richard (2020), who thinks the topic of enquiry is word meanings. Cappelen (2018) proposes a theory of CE without any concepts, namely *The Austerity Framework*, in which CE aims to engineer the intentions and extensions of words. CE, therefore, applies "to the world" (pg 46).

Cull (2021) takes a view similar to that in this work. He suggests that it is possible to sidestep any discussion of concepts if we take CE to be the "improvement of our representational devices (whatever those are)" (Cull, 2021, pg 232). In his view, it is simply not relevant to the sort of project he pursues. This work will remain neutral in this debate and looks at word meanings without making any theoretical claims as to the structure of the underlying concepts. In a historical empirical study, it is difficult to investigate the possible intensions and extensions of words across time.

2.5 The implementation problem

While the issues above are important and fascinating, they are currently (and rigorously) being debated. Now that the reader is broadly aware of these issues, it is possible to turn to the issue that will form the frame of this thesis.

Authors who are pessimistic about CE often express doubts due to the implementation problem. Depending on your meta-sematic viewpoint, this problem can lie in different places. For internalists such as Chomsky (2000), the problem lies in targeting enough of the population to change concepts, since concepts reside solely in the mind of users. CE is much easier to combine with internalism since we as individuals have full control of our concepts. Externalists, on the other hand, believe that the meaning of words is at least partly determined by speaker-external factors (Koch, 2021), which could include their usage history and reference 'baptism' (Kripke, 1980), expert's beliefs (Burge, 1979), or simply our environment (Putnam, 1973). Coming back to CE, this means changing concepts is much harder since we have little to no control over these external factors. Not only do we have to change internal representations, but also how the world is (though conversely, CE projects aim to change the world, creating a chicken-and-egg scenario).

CE, therefore, has an implementation problem. Numerous philosophers discuss this problem, with both positive and pessimistic conclusions. Much of the rest of the literature focuses either on potential strategies in the implementation process or on whether implementation is even possible.

Jorem (2021) examines several approaches for implementation, looking at how feasible each of them is and if they are worth pursuing, and concludes that changing the 'standing meaning' (the meaning that stays stable across speakers, times, and places) is difficult or impossible. Even if we were to take an internalist viewpoint, the number of internal representations that would have to change is a barrier to CE implementation. Interestingly, what logically follows is that either smaller languages or sociolects, such as "language spoken by experimental scientists at the University of Pittsburgh" (Jorem, 2021, pg 198) are easier to change. Despite 'standing meaning' being too difficult to change, there are other types of meaning that do fall within our control. For example, we can modulate the meaning of a word in a specific conversation or use Gricean pragmatic principles (Grice, 1989), such as the difference between what-is-said and what-is-meant, to speaker-refer to meanings different to the standing meaning. Jorem suggests that if the design stage of CE is done well, then all necessary materials for changing what-is-meant should already be in place.

Other philosophers have also provided specific options for implementation. Sterken (2020) discusses the possibility of transformative communicative disruptions as a method for implementing CE, which involves the hearer simply deciding to understand the concept in a new way. While this temporarily breaks the communicative chain, it forces self-reflection on one's conceptual usage. Andow (2021a) is one of the only works that looks at implementation more concretely and performs a sort of cost-benefit analysis of implementation, concluding that he also believes that, given enough time and resources, it is possible to implement a concept.

In the group looking at the possibility of CE implementation is Koch (2021), who concludes that changing concepts is not beyond our collective long-range control, likening it to climate change. Cappelen (2018) expresses the same viewpoint, suggesting that the mechanisms that underlie CE are too complex to grasp, and that we have little to no control over meaning change, but that we should not give up on CE. On the other hand, Richard (2020) and Deutsch (2020) are among the philosophers who believe CE projects are either incredibly difficult to spread, or altogether impossible.

The problem with all these analyses (except for perhaps Andow (2021b)) is that they stay in the theoretical domain; they look at potential ways of implementing conceptual changes and debate what is meant by implementation but do not look at concrete examples. As stated by Jorem, "Neither Cappelen (2018), Koch (2018) nor Deutsch (2020) assesses particular languages for the feasibility of effecting semantic change in them" (Jorem, 2021, pg 192). Much of the debate concerns whether implementation is even possible, and for many authors, the future of CE depends on whether it is possible.

Assessing a concrete instance of CE is a necessary contribution to these debates. While many of the discussions concern whether CE can be implemented, or some vague strategies to do so, many more aspects are important. For example, how long implementation takes, what factors govern it, who adopts the term early, forces of resistance and preferable contexts. Such a study is beyond the scope of philosophy, but within the grasp of linguists.

2.6 Linguistics and Conceptual Engineering

What has been discussed up to now is largely prescriptive. CE is about what concepts should be and how we can improve them, something that is strongly against the descriptive principles of linguistic norms. Linguists describe how language is used, how it is structured, and what words mean. It, therefore, seems hard to find anywhere in CE that linguists could contribute. After all, “normative projects are not addressed by descriptive answers” (Andow, 2020, pg 3).

However, recent work by Nado (2021) and Andow (2020) has shown that descriptive processes have a place in CE. They do not mention linguistics but do discuss how experimental philosophy can contribute to the discussion in CE. In these discussions, experimental philosophy can be used to find out what our existing concepts are, establish the relevant normative constraints, and establish what kind of concept would best meet these constraints for the relevant group (Andow, 2020, pg 14). If the descriptive processes of experimental philosophy have a place in CE, it seems logical that the descriptive processes of linguists could do too. More specifically, linguists can use the descriptive and investigative frameworks they have built to address the implementation problem.

In this solution, linguists need not offer any normative judgements. Language change and how it spreads are major themes in linguistics (Labov, 2001, 2010), and there are multiple facets of the implementation problem that linguistics could help address. This work takes a first step in encouraging linguists to participate in CE. The approach taken here consists of a speculative historical case of CE and an analysis of its spread. The investigation here will target which groups and individuals use the new concepts, how fast and complete the spread is, and in which contexts they are favoured. Of course, this is not everything that linguists can do, but it is a start.

If, as Cappelen (2017) argues, questions of natural language outside of a normative project can best be left to linguists with training in linguistics, then analysing the implementation of concepts is a job for linguists.

CHAPTER 3

LANGUAGE IDEOLOGIES

The previous chapter touched on ideas of language ideologies and language authority in discussing who decides which concepts are preferable, what constitutes amelioration, and who these changes will benefit. This chapter will look at these ideas in more depth.

3.1 Language ideologies

Language ideologies were brought forward in the seminal work of Silverstein (1979) and have been used in linguistic anthropology, sociolinguistics, and applied linguistics (Costa, 2018). In terms of linguistics, Silverstein argued that understanding these ideologies is crucial in understanding how linguistic structures change (Silverstein, 1985, pg 220), making these ideologies important to the study of CE. Language ideologies are defined in several different ways in the literature, and some interesting definitions are below.

1. "Sets of beliefs about language articulated by users as a rationalization or justification of perceived language structure and use" (Silverstein, 1979, pg 193)

2. "Self-evident ideas and objectives group holds concerning roles of language in the social experiences of members as they contribute to the expression of the group" (Heath, 1989, pg 53)

3. "The cultural system of ideas about social and linguistic relationships, together with their loading of moral and political interests" (Irvine, 1989, pg 225)

All these definitions differ in terms of where the emphasis is. Definitions can be oriented more towards linguistic structure, activism, or the social sphere, and embody different theoretical stances and epistemological concerns (Valdez, 2011). In Silverstein's characterisation, language ideologies create a three-way link between linguistic forms, social use, and human reflections on these forms in use (Woolard, 2008). Most importantly, "definition of language is always, implicitly or explicitly, a definition of human beings in the world" (Williams, 1977, pg 21) and is therefore also about their values, morals and identities, and what they want to achieve by using it (Woolard and Schieffelin, 1994).

The social structures and viewpoints reflected in ideology can include ethnic, gender, indigenous, and national identities (e.g. Bunte (2009)). Since people can choose to identify with many different groups, their ideologies can often become contradictory and complex (Woolard, 1998), and indeed can even be contradictory within the same person or group. For example, Ruíz (1984) categorised language ideologies as viewing language as a resource, a problem, or a right, but it is easy to imagine that ideologies can take aspects of each category to form their ideology or change their view across time. They are not a “fixed, abstracted, and circumscribed set of beliefs” (Briggs, 1992), but are constantly in flux, remoulded and refashioned based on social, cultural and linguistic changes (Gal and Irvine, 2019).

Bucholtz and Hall (2005), Eckert (2000, 2004) and Schilling-Estes (2002) are all advocates for the inclusion of language ideologies in linguistic study, especially sociolinguistic study. This viewpoint can also be called the ethnography of speaking. Instead of taking predefined social variables and investigating linguistic patterns within these selected labels, researchers cannot predict which groups speakers choose to show alliances to, nor which linguistic features they find particularly salient (Lesley, 2004). Hymes (1974, 1964) suggests focusing on the attitudes and values of speakers as well as which contexts or forces of authority influence these attitudes. Researchers are thus encouraged to link any linguistic variation to ideological principles. The same is true of CE. One cannot investigate an instance of CE without considering the underlying language ideologies.

The ideologies you have will constrain what you consider to be a successful CE project and influence what will be perceived as an amelioration or ‘preferable’ concept in a particular scenario. There is no singular concept of ‘better’, or ‘defect’, but these notions instead depend on the groups in which these ideologies originate. This makes the study of CE even harder.

3.2 Authority over language

Authority over language is a complicated idea to explore. Following the stance of language ideologies, to change language we must change ideology, which is linked to people. To do so, we must target different publics. Publics are constructs created by ideologies, or sets of ideologies (Spitzmüller, 2019), and can be used to refer to “a kind of social totality” (Warner, 2002, pg 65), or more specifically to a social domain. Publics are created and maintained by ideologies, and crucially stand in contrast to another sphere, for example, the general population contrasting with the scientific domain.

Habermas (1996) suggested that publics are largely reason-driven, universalist and deliberate, and thus normative. Decisions occur through rational debate among communicative equals and “the power of the better argument” (Habermas, 1996, pg 54) in a Habermasian society (Deumert, 2019). This kind of public would be ideal for CE since they would be able to accept the reason-driven amelioration of certain concepts. However, publics have also been characterised as wild (Young, 1987; Gardiner, 2004) or Carnivalesque (Bakhtin, 1982), and while this may be less ideal for CE, they are more representative of real society. Rather than using logic and reasoning, publics use many rhetorical devices, including metaphor, irony, and figure of speech. These publics consist of a “world of multiple and unregimented voices” (Gal, 2019, pg 452). The public becomes a sort of “Bakhtinian marketplace” (Deumert, 2014, pg 165); the loudest and most powerful forces tend to win.

If publics are indeed wild and Bakhtinian rather than Habermasian, the question of authority becomes much more complex since ideologies might not change simply through the power of a better argument. Authority is constantly in flux and affected by numerous other social factors.

Linguistic forms can have many potential sources of authority (Gal, 2019). For example, a standardised language has authority because it is ideologically linked to the idea that the standard they prescribe is inherently better than any alternatives. Ideological studies look at the reasoning behind these notions of linguistic correctness, how a hegemonic standard will inevitably favour certain groups over others (Kroskrity, 2010), and how they relate to other social or cultural ideas of power and identity (Woolard, 1998; Silverstein, 1985).

Standard languages are not the only linguistic forms that have authority. Counterpublics (Fraser, 1992; Warner, 2002) are a kind of talk not encoded in grammars or dictionaries but are created informally. These counterpublics could consist of a specific social group, an occupation, a political orientation, or a scientific viewpoint.

This notion of authority over language is crucial for CE. Ensuring that an engineered concept becomes used in the intended community in the way you intend requires authority. In a Habermasian society, this authority may come from simply having the better argument, but in (the arguably more realistic) Bakhtinian society, power and authority come from a complex interweaving of several social factors. In CE, the question remains as to what kind of authority is needed to meet the project goals, and how to position oneself to get it.

3.3 Register

In a study centred around one specific type of language used for one specific period, it is important to introduce the idea of registers. Biber and Conrad (2009, pg 6) define a register as "a variety associated with a particular situation of use (including particular communicative purposes)". Since a single person can use language for infinite functions and in infinite situations, each person, therefore, has command of an array of registers which they can adopt based on the situational context, depending on whether they are participating in academia, writing a letter, speaking to their children, and so forth. Different registers can have different structural and lexical features that characterise them (Gray and Biber, 2018) and are often functionally motivated (Agha and Frog, 2015). Registers are therefore examples of contextual variation. Enregisterment is a term employed by Agha (2007) to refer to the creation of registers and involves reflexive processes to differentiate the register from others. Each register is therefore created by different people for a different purpose, and all have different diachronies.

The register of interest here is academic English, a register that has previously been defined as a register that is resistant to linguistic change (Hundt and Mair, 1999). This register differs from others in that it is usually written, subject to revision, non-interactive, and aims to convey information (Agha and Frog, 2015).

CHAPTER 4

SCIENCE, IDEOLOGY, AND CONCEPTUAL ENGINEERING

Of all the places a linguist could start with investigating CE, *mass* and *weight*, or scientific literature and concepts more generally, between 1700-1920 do not form an intuitive starting point. However, it provides an ideological background suited to a CE study. As discussed, language ideologies and authority provide a complicated barrier in understanding what is considered an amelioration and who has the power to propose ameliorated concepts. The scientific community has one overarching goal: scientific progress. The scientific register has a definitive authority, namely those who can prove their theories, and the viewpoints of many scientists within this sphere are well-documented. An ameliorated concept is thus one that allows for scientific progress, potentially by being more precise or accurate, and one has authority over language if one can provide the things that the scientific community needs, mainly by using logic and reasoning (in a Habermasian sense).

While the next chapter will look at specifically *mass* and *weight* and their usage history, this chapter examines the language ideologies and authorities of the scientific sphere in the area in question. This allows for the creation of an ideological background onto which the results of this study will be projected.

4.1 17th century

Between 1600 and 1900, language came to the forefront of public concern due to the emergence of new kinds of discourse that triggered new ideas of inclusion and exclusion (Woolard and Schieffelin, 1994). Starting from the very beginning, the 17th century was a time of great interest in language.

4.1.1 Bacon and Locke

In the *Novum Organum* (Bacon, 2011), Bacon (1561-1626) writes about the idol of the marketplace, which consists of words that "lead men away in numberless empty controversies and idle fancies" (Bacon, 2011, pg 55). One thing Bacon focussed his attention on were the superficial distinctions or similarities produced by words, such as words that grouped things that should not be grouped (whales as fish), or separated things that were essentially the same (water, ice, steam) (Urbach et al., 2022). This had a significant impact on the sciences, and Bacon stated that "words retort and turn their force back upon the understanding and this has rendered philosophy and the sciences unproductive" (Bacon, 1883, 1620, pg 48). In fact, according to Bacon, language in its current state was the greatest hindrance to scientific progress (Bauman and Briggs, 2003). The Baconian methodology of the 17th century consisted of induction, which meant inducing generalities from observations (Gauch, 2012), and this meant that observation and facts had priority over theory and knowledge. Theorising and hypothesising before analysing the data were said to be dangerous (Yeo, 1986).

Locke (1632-1704) wrote his work *An Essay Concerning Human Understanding* (Locke, 1975, 1690) three years after becoming a member of the Royal Society. He refers to the "abuse of words" (Locke, 1975, 1690, pg 508), and the notion that people "make them stand sometimes for one thing, and sometimes for another" is "great folly, or greater dishonesty" (Locke, 1975, 1690, pg 495-2). In his view, the words one used had the power to shape thought, and therefore faulty words could lead to faulty thoughts (Losonsky, 2021). These faulty thoughts could then affect scientific, moral and political progress (Losonsky, 2021). His *Essay* was, therefore, an attempt "to make language and human understanding safe for science – and society" (Bauman and Briggs, 2003, pg 31).

Ruíz (1984) suggested ideologies of language view language as either a resource, problem, or right. Both Locke and Bacon seem to fall within the category of seeing language as a problem ("the cheat and abuse of words") or something that is hindering scientific and human progress. Bacon especially falls into this category, but Locke takes a slightly more positive outlook, suggesting that language can become a resource once we have fixed it and made it fit for purpose. CE generally appears to fit into the problem ideology and therefore aligns with the viewpoints on language throughout time. As explored in Chapter 3, linguistic changes can be caused by language ideologies, and thus the study of language change must also be the study of the individuals and societies that contribute to that change. It seems logical that an ideology of language as a problem, as well as an ideology that we have the power to fix it, is a driving force behind CE.

4.1.2 Science in the 17th century

An important notion for this study is that of 'modernity'. The work of Bauman and Briggs (2003) investigated important philosophical, anthropological and scientific figures through time, introducing a diachronic dimension into the study of scientific ideologies. In doing so, they give an insightful look into the scientific revolution, and the development of the ideology of 'modernity'. From the 17th century, national and global power no longer necessarily came from brute strength and warfare, but also through innovation and ideological reorientation (Bauman and Briggs, 2003), part of which came from small groups of elites, or scientists, whose ideologies involved aiming to both understand and dominate the rest of the world (Bauman and Briggs, 2003). The emergence of 'modernity' was doubtlessly influenced by the development of modern science in the 17th century, which motivated the turn towards secularism and naturalism (Bauman and Briggs, 2003).

4.1.3 The Royal Society

The Royal Society started as an 'invisible college' of scientists in 1660 (McDougall-Water et al., 2021). Soon after, it gained royal approval and became *The Royal Society of London for Improving Knowledge*. This society allowed fellowships for scientists and had prominent members, including Einstein, Newton, and Darwin. However, especially in the beginning, the membership of the Royal Society was not only based on academic merit but also social status and connections. It represented the social networks of prestigious individuals and had a large amount of "gentlemanly" culture (Menzel, 2020). The Royal Society was, especially at its inception, a sort of gentleman's club. An important part of a gentleman's training was the art of rhetoric (Dawson, 2007), and part of rhetoric was the idea of words as loosely tied to meaning, allowing them to be manipulated by both the speaker and the audience. It could be expected that this culture transferred to the Royal Society, but precisely the opposite is true. Sprat (1667) reported that the consensus of the Royal Society was "to reject all the amplifications, digressions, and swellings of style" (Sprat, 1667, pg 111-13).

The motto of the Royal Society was *Nullius in Verba* (take nobody's word for it). This motto applies in many ways to the study of CE. Firstly, it encouraged members not to accept scientific discoveries at face value, encouraging verification and replication. Secondly, it promoted the distrust in language and our words that CE relies upon. Trusting nobody's word is reminiscent of Nietzsche's statement to be sceptical of all inherited concepts. Thirdly, it can be read as the encouragement to look beyond the surface value of words. Just because somebody uses the word *X* does not mean they mean *X*. Instead, hearers need to be sceptical about the meanings of the words they hear and try to recover the intended meaning from the lexical meaning. It is for these reasons that the title of this work includes the motto of the Royal Society.

The first issue of the journal *The Philosophical Transactions of the Royal Society* was published in 1665 by Henry Oldenburg, the first secretary, and is now the oldest continuously published scientific journal in the world (McDougall-Water et al., 2021). This journal was published monthly and contained scientific correspondence, reviews, experiments, observations, and discussions. Alongside this journal, the society also published other scientific works, including Hooke's *Micrographia*, Newton's *Principia*, and Franklin's kite experiment. From its inception to the 1800s, the journal nearly had a monopoly over scientific discourse in Britain and America (Atkinson, 1996).

Those in control of the society had large amounts of influence, and the whole network consisted of a tight circle of both professional and personal relations (Røstvik and Fyfe, 2018). For example, from 1707 to 1723 Newton was president of the society, and the whole society was largely under the control of Newtonians (Sun et al., 2021). As a result, during this period, the journal focused on issues that Newtonians found important and had the descriptive, plain, experimental style championed by the Newtonians (Atkinson, 1998).

Eventually, the journal was officially taken over by the society in 1752 and publishing became less frequent. Rather than the published articles being selected personally by the president (and the bias that this created), a committee was created to select and edit articles, and the practice of peer-review started in '1835 responding to increasing competition from other journals. The Royal Society was at the forefront of introducing the notion of peer-review (Baldwin, 2019), standardisation, dissemination, and archiving. In the 1830s, a new journal was introduced, named *The Proceedings of the Royal Society*. Both journals became more specialised and eventually split into A and B in 1887. Journal A covers Engineering, Mathematical and Physical sciences, whereas B covers Biological sciences (Menzel, 2020).

The Royal Society was arguably at the forefront of the change to 'modernity'. They aimed to promote empirical scientific study in the pursuit of knowledge and to be free from the confounding factors of tradition and superstition (Bauman and Briggs, 2003). In establishing London and other cosmopolitan areas as the home of this 'modernity', and by allowing only the elite, connected, wealthy and intelligent into the community, they defined themselves as differentiated from the "lower class, ignorant, old-fashioned, indigenous" public (Bauman and Briggs, 2003, pg 2). Their linguistic practices will have both reflected this ideology and formed part of how the Royal Society created this community. Language, after all, does not only index existing groups but helps create, maintain, and transform their identities.

4.2 18th century

Aside from Condillac in France, there were few works produced during this time that were considered important for considering language in the philosophy of science (Laudan, 1968). Condillac (1714-1780) believed when languages expanded, either through coining new words or borrowing foreign words, they kept adding more words that less closely corresponded to ideas, which hindered philosophical progress (Knowlson, 2019). While he saw the merits of algebra and some physics and chemistry terminology, the rest of scientific language was "riddled with defects" (Knowlson, 2019).

4.3 19th century

While before the 19th century, the focus of scientific writing was to be clear, straightforward, and accessible, this began to change around 1830 (Yeo, 1986). Fields of study started to become more specialised, and writings started to turn away from concrete observations that were understandable to non-specialists to theoretical discussion. Rather than data collection that could be done by anyone, the focus started to shift onto a combination of both data and theory. This drove science into the arms of the elite few that possessed this theoretical and mathematical knowledge. Science became genius rather than hard work, and facts and theorising could no longer be kept separate. Science moved away from general society and into its special domain.

Most pertinent to our investigation is the scientific sphere's relationship with the public sphere since the external viewpoint of the scientific community is created by the social value they both have and exude. This creates the necessary environment for the construction of their scientific and academic identities (Spitzmüller, 2019), which could be considered to consist of authority and knowledge.

While at first science was an activity for all, it quickly became hypo-deductive and theoretical and shut itself off from the observation of the public. As observed by Spitzmüller (2019), this followed the more general European pattern of separating science from society from the end of the 16th century, crucial for the development of 'modernity' (Bauman and Briggs, 2003).

The scientific community has turned inwards so strongly that even its linguistic features change in different ways from normal language (Degaetano-Ortlieb and Teich, 2019). Scientific concepts are renowned for changing in specialised ways, such as reconceptualisation, recombination and relabelling in the process of evolution (Bradie, 1986) making them prime targets for change in the sense of borrowing, extension or narrowing. These approaches have not been applied to scientific terminology as frequently as other types of language. This may be partly due to the lack of specialised Natural Language Processing tools for scientific language, something which is now being combatted through the creation of SciBERT (Beltagy et al., 2019), a resource based on word embeddings.

The development of the scientific register has been looked at by numerous linguists. These types of work are common, and much of previous literature is based on the development and change of the scientific discourse style over time. Shapin (1984) looked at the first seven years of *The Transactions of the Royal Society* and found that it focused on the rhetoric of immediate experience rather than the more prominent framework of late scholasticism at the time (Atkinson, 1996). Biber's 1988 study found that scientific writing became more informational, non-narrative, elaborated, and impersonal over time, but later found that these trends did not always proceed linearly or consistently, finding large amounts of fluctuation (Biber and Finegan, 1997). Biber and Gray (2016) found that rather than increased explicitness, the trends seen in passivisation and nominalisation contributed to greater opacity during the development of the register.

An increasing number of studies are turning to information-theoretic notions to investigate how scientific language has changed over time rather than frequency-based approaches, such as Degaetano-Ortlieb et al. (2018) and Degaetano-Ortlieb and Teich (2019), who used entropy and surprisal to study the conventionalisation and specialisation of scientific literature. They found that scientific literature gradually became more differentiated from everyday language. Lexical items became more complex and specialised and appeared in patterns of initial innovative usage and expansion, followed by eventual conventionalisation. Conversely, the grammar had a simple trajectory of conventionalisation. Degaetano-Ortlieb and Teich (2018) pinpoint the start of this change to the 1750s and note that this pattern occurs in waves, where increased lexical complexity usually leads to a subsequent reduction in grammatical complexity. Similar results were found by Bizzoni et al. (2020), who also found topic fluctuation over time using a diachronic topic model. Sun et al. (2021) found a similar pattern using word embeddings and entropy, but also investigated the impact of external factors, comparing times of change in the external environment to changes in concreteness and imageability of scientific vocabulary. They found that when *Transactions of the Royal Society* was under by Newtonian control, there was an increase in imageability and concreteness, thereby showing scientific evolution was not linear.

CHAPTER 5

MASS AND WEIGHT

Now that it is clear what the aim of this investigation is, what the ideologies of the people within the scientific register are, and what CE more generally is, it is possible to look at the specific words that will form the base of this study. *Mass* and *weight* are often used interchangeably in everyday vernacular, but, in the scientific literature, they have crucial differences.

5.1 *Mass*

The word *mass* was borrowed into Middle English as *masse* through language contact with Anglo-Norman. This respectively came from Old French, which had inherited it from Latin *mass*a "lump/dough". Even Latin *massa* was borrowed from Ancient Greek *μᾶζα* (mâza, "barley-cake, lump (of dough)") ("mass n.2", 2021, OED).

Modern dictionaries detail numerous different senses of *mass*. The definition concerning the Eucharist and its liturgical celebration ("mass n.1", 2021, OED) is not relevant to this thesis, but the definition of *mass* as "the quantity of matter a body contains" ("mass n.2", 2021, OED) is highly relevant. Table 1 shows some of the entries in the OED, with non-relevant or obsolete entries removed, along with their date of the first attestation.

The scientific sense that is in the principal interest of this paper is first attested in the works of Isaac Newton in 1687 in the Latin form *massa*, and later in the work of Harris in 1704 for the first time in an English context.

(1) *Hanc autem quantitatem sub nomine corporis vel massæ in sequentibus passim intelligo.*
(Newton, 1687, pg 15)

(2) *Masse, this Word is used by the Natural Philosophers to express the Quantity of Matter in any Body.*
(Harris, 1704)

The use of *mass* to refer to the quantity of matter contained within a body is a relatively new development. All other usages detailed in the OED, except for its application to people, are older than the scientific usage. Scientific *mass*, referring to a scientific property of

Category	Definitions from OED
Scientific	The quantity of matter which a body contains, as measured by its acceleration under a given force or by the force exerted on it by a gravitational field; an entity possessing *mass*. (1687/1704)
People	The *mass*: the generality or majority of mankind; the main body of a people, nation, etc.; the ordinary people. (1621) A large number of human beings collected closely together or viewed as forming an aggregate in which their individuality is lost. (1814)
Collective	Dense aggregation of objects having the appearance of a single, continuous body. Also figurative. (1382) A large amount, number, or quantity of a thing or things, material or immaterial (often with the sense of oppressive or bewildering abundance). Now frequently (colloquial) in plural, sometimes with singular agreement. (1566)
Matter	A coherent body of matter of unspecified or indeterminate shape, and usually of relatively large bulk; a solid and distinct object occupying space. (1425)
Domain-specific	Metal, especially gold or silver, in a lump or lumps; bullion. See also in (the) *mass* at Phrases 1. Obsolete. (1477) A quantity of amorphous matter used in or remaining after a chemical reaction; (also) a quantity of a substance from which medicinal pills are made. (1559) Geology and Mining. An irregularly shaped deposit or layer of ore, mineral, or rock. (1815)
Phrases	In (the) *mass*: In a lump or block. In early use: (of gold or silver) in the form of bullion. (1423) Without distinction of individuals or component parts; bodily, collectively, as a whole. (1631)

Table 1: Definitions of *mass* from the OED

objects, developed after the other senses of *mass* referring to a collection of things, or simply a coherent body of matter. Scientific *mass* is the amount of matter within an object, regardless of its size. Mass is an inherent property of an object, and one cannot change the mass unless adding or subtracting some of the matter. According to the international system of units, the base unit of *mass* is the kilogram.

5.2 *Weight*

The word *weight* comes from Middle English *weight* (*weiȝte, weght, wight*) which in term comes from Old English *wiht, ġewiht*. There are a large number of cognates in other

Germanic languages, such as German *wicht, gewicht* and Dutch *gewicht* ("weight n.1", 2021, OED).

Like *mass*, there are multiple senses of the word *weight*. Table 2 shows meanings listed in the OED ("weight n.1", 2021, OED).

Category	Definitions from OED
Figurative	A burden (of responsibility, obligation, suffering, years, etc.). (1380) Burden (of proof), onus. (1824) To pull (one's) weight: see pull v. Phrases 10; To throw (chuck, etc.) one's *weight* about or around: to assert oneself or one's authority, esp. in an objectionable way; to act officiously. Colloquial. (1617)
Scientific	The amount that something weighs; the quantity of a portion of matter as measured by the amount of its downward force due to gravitation; the amount of resistance offered by a body to forces tending to raise it. (1385)
Quantity	Portion or quantity weighing a definite amount. Often preceded by an expression indicating the amount: in Old English in the genitive, as *anes pundes, þreora punda wiht*; now in attributive or appositive form, as one pound, three pounds weight. Often abbreviated wt.(1000) Its, his, etc. *weight* in or of gold, silver, etc.: a quantity of gold, silver, etc. of the same weight. Chiefly in hyperbolical statements of value. (1275) In phrases stating how much a thing weighs, as of two pounds weight. (1389)
Heaviness	A heavy *mass*; usually, something heavy that is lifted or carried; a burden, load. Also figurative. (1374)
Domain-specific	Physics. The relative *weight* of the atom of an element or the molecule of a compound (see ATOMIC *weight* n., molecular *weight* n. at MOLECULAR adj. Compounds). (1836)
Body	To lose weight: to become thinner or less corpulent; to put on weight: see to put on 6a at put v. Phrasal verbs 1. (1961)

Table 2: Definitions of *weight* from the OED

Again, the sense of greatest import to this investigation is the scientific sense. Under this definition, *weight* is a force, and not a property of an object. *Weight* is proportional to *mass* and acceleration, where acceleration is the acceleration of a free-falling object in that gravitational field. *Weight* can therefore be found by multiplying *mass* and this acceleration, which on Earth is 9.80665 m/s^2 However, this can vary slightly in different places on Earth (Hirt et al., 2013) and is different per planet, or more specifically, per gravitational field.

Objects always have *mass* but do not always have weight. For example, an object in free fall is weightless but still has *mass*.

According to the international system of units, the base unit of *weight* is the Newton, which was introduced in 1948. A force of 1N accelerates a *mass* of 1 kg at a rate of 1 m per second squared.

Example (3) shows the usage of *weight* to refer to "A portion or quantity weighing a definite amount" ("weight n.1", 2021, OED) from as early as 1000. By 1385, *weight* had gained its scientific meaning, in the sense of "the quantity of a portion of matter as measured by the amount of its downward force due to gravitation" ("weight n.1", 2021, OED), as shown by example (4). Finally, example (5) shows a piece of scientific literature from 1830, using the scientific meaning of *weight* to refer to the differing weights a single *mass* can have on Earth.

(3) *Genim of ælcere þisne wyrte xx penega wiht.*
(Old English Leechdoms, c1150)

(4) *Sakkis ful of gold of large weyghte.*
(Chaucer and McMillan, 1987, pg 1118)

(5) *The weight of all bodies is diminished by the centrifugal force, so that the weight of any body is greater at the poles than it is at the equator.*
(Brewster, 1831, pg 163)

To illustrate, imagine a cube with a *mass* of 100 kg. This cube is made of 100 kg worth of matter. On Earth, where the acceleration due to gravity is 9.81 m/s^2, we can calculate the *weight* of the cube by multiplying the two. This gives us 981 N, which is the *weight* of the cube on Earth. If the cube were on the moon, the acceleration would be 1.62 m/s^2, meaning the cube weighs 162 N. If the cube were on the sun, the acceleration would be 275 m/s^2, meaning the cube weighs 27500 N.

To summarise, *mass* is the amount of matter that makes up an object. *Weight* is the force of gravity acting on the *mass*.

5.3 Conceptual Engineering?

As can be seen from the dates above, *mass* in its scientific sense is a relatively new concept. *Weight* is a word with much more history and much more varied usage. However, the splitting of *weight* into *mass* and *weight* does not necessarily mean it is an instance of CE. This work, however, takes the opinion that it is, for the reasons specified below.

5.3.1 Carnapian explication

The everyday vernacular does not need differentiation between *mass* and *weight* because normal human beings never experience other gravitational fields, and since *weight* is proportional to *mass*, the distinction is unnecessary.

However, in science differentiation between *mass* and *weight* is crucial. It is needed to measure acceleration using the famous $F = m\,a$ equation (force equals *mass* times acceleration). Newton's Law of Universal Gravitation (Newton, 1687) also depends on *mass* to measure the attraction between particles of matter. The study of astronomy (and some branches of physics) necessarily involves different gravitational fields, making *mass* versus *weight* an important distinction. Engineering concepts such as *breaking weight* and *weight-specific resistance* depend on the concept of *weight*, and research into flying, floating, or falling bodies will require reference to buoyancy and air resistance, and by extension *weight*.

However, despite this usefulness, the usage of these terms often varies. The time between when *mass* and *weight* started being used with their respective scientific senses and the 20th century was rife with scientific discovery. Some things that were research targets in that time have developed names that we could consider as fossilised since they reflect this inconsistent usage. Take the definition of *weight* in the domain-specific category in Table 2: "The relative weight of the atom of an element or the molecule of a compound". If we take our working definitions at face value, this refers to the force acting on the *mass* of the atom/molecule in the working gravitational field. However, this is not the case. Atomic *weight* refers to the *mass* of an atom (Bievre and Peiser, 1992) and molecular *weight* refers to the *mass* of a given molecule. However, they have been used so frequently now that their usage has become thoroughly entrenched (Siyanova-Chanturia et al., 2011), making them resistant to change (Ishizaki, 2012). Another usage that blurs these boundaries is *bodyweight*. Whilst the usage of this phrase may refer to the *weight* of the body, it is more likely to be referencing the *mass*. When you want to lose *weight*, you want to reduce the amount of matter that makes up your body, consequently decreasing your *weight*.

To recall explication, Carnap writes “By the procedure of explication, we mean the transformation of an inexact, pre-scientific concept, the *explicandum*, into a new exact concept, the *explicatum*” (Carnap, 1947, pg 3), and offered criteria of similarity, exactness, fruitfulness and simplicity. Taking this framework, we can evaluate *mass* and *weight*. In this case, *weight* is the pre-scientific lemma, used for both concepts until 1687. The similarity criterion is fulfilled since *mass* and *weight* can be used in the same contexts as *weight* used to be used. The separation into these two concepts rather than a single one allows for more exact scientific enquiry, as shown by Newton’s laws and discoveries after he explicitly separated them. This also lends credence to the idea that separating these concepts leads to increased fruitfulness. Simplicity is a difficult criterion, but it could be argued that *mass* and *weight* allow for increased simplicity since they separate matter and the forces acting on it, rather than considering both at once. Newton’s differentiation of *mass* and *weight* can therefore be considered as an instance of Carnapian explication, one of the most frequently discussed forms of CE.

5.3.2 Fixing a problem

In a first unfinished draft of the *Philosophiæ Naturalis Principia Mathematica*, henceforth *Principia*, Newton defined *weight* as “the quantity or amount of matter being moved, apart from considerations of gravity, so long as there is no question of gravitating bodies” (Newton, 1999, pg 87). Newton, therefore, is making explicit that while he is using the term *weight*, he is looking for a way to refer to *mass*. He just simply does not have a lexicalisation for the concept yet. He is searching for a way to discuss the invariable property of the quantity of matter that makes up an object, and not the variable property of *weight*. Newton knows of experiments involving pendulums and clocks at different points on Earth and knows that *weight* is something that varies whereas *mass* remains constant. He is unhappy with the use of *weight* for this concept and knows that it is unsuitable for discussions that do not involve gravity or “heaviness” but has to make do due to “the want of a suitable word” (Newton, 1999, pg 87).

By the first finished draft, around 1685, Newton has settled on a lexicalisation for this concept, namely *mass*. Definition 1 in the *Principia* states that *mass*, and also what he coins *body*, is “the quantity of matter” that stems from “its density and volume jointly”. He emphasises that *mass* can be found through someone’s *weight* or gravitational force.

Newton found the ambiguity of *weight* to be problematic, and therefore offered a separation into *weight* and *mass* as an amelioration. As for attempts to implement, in the *Principia* he used this distinction to discuss inertia, the definition of force, displacement, momentum, and acceleration. Anybody who wanted to study or verify these claims, therefore, had to use the terminological difference that Newton proposed. *Mass* and *weight* therefore seem to fit nicely into the framework of CE.

5.4 Interim summary

To summarise the previous few chapters, the Royal Society was a small, close-knit scientific community united in its pursuit of scientific advancement, making it an ideal community to test linguistic methods of CE.

Authority was held by those in charge of the society, and also those who conducted respectable scientific research, thereby creating a somewhat Habermasian society in that it responded to “the power of the better argument”. The Royal Society participated in the creation of the scientific register, which stood in contrast with the vernacular of the time. Philosophical approaches towards language at the time stressed the unreliability of language, especially in regard to its precision concerning scientific study.

The engineering of *mass* and *weight* within this framework fits with the ideological conceptions of amelioration and defects within the Royal Society. The scientific register provides the data used to identify whether Newton’s suggestion to differentiate *mass* and *weight* by sense was successful, or whether other intra or extra-linguistic factors, such as the author or the subfield, had a greater effect.

CHAPTER 6

METHODOLOGY

Before detailing the methodology of this study, it is worth repeating the research questions.

RQ1. To what extent are *mass* and *weight* successfully differentiated by sense after Newton proposed separating the concepts in 1687?

RQ2. What time frame and trends are there in the diffusion of *mass* and *weight*?

RQ3. What effect do extra-linguistic variables such as author and subfield have on the choice of lemma?

To operationalise these research questions, the method chosen needs to identify the factors that condition the choice of lemma and rank these factors based on their significance. Through doing this, it is possible to identify whether sense conditions the usage of *mass* and *weight*, or whether extra-linguistic factors have a larger influence. In the former case, one could argue for successful CE, and in the latter case, idiosyncratic usage continues to overrule any CE proposals made.

6.1 A note on concepts, words, and written texts

Concepts and words are not the same. It is often the case that multiple concepts are represented by a single word or vice versa. This can occur in polysemy, homonymy, and synonymy, as seen in Figure 3.

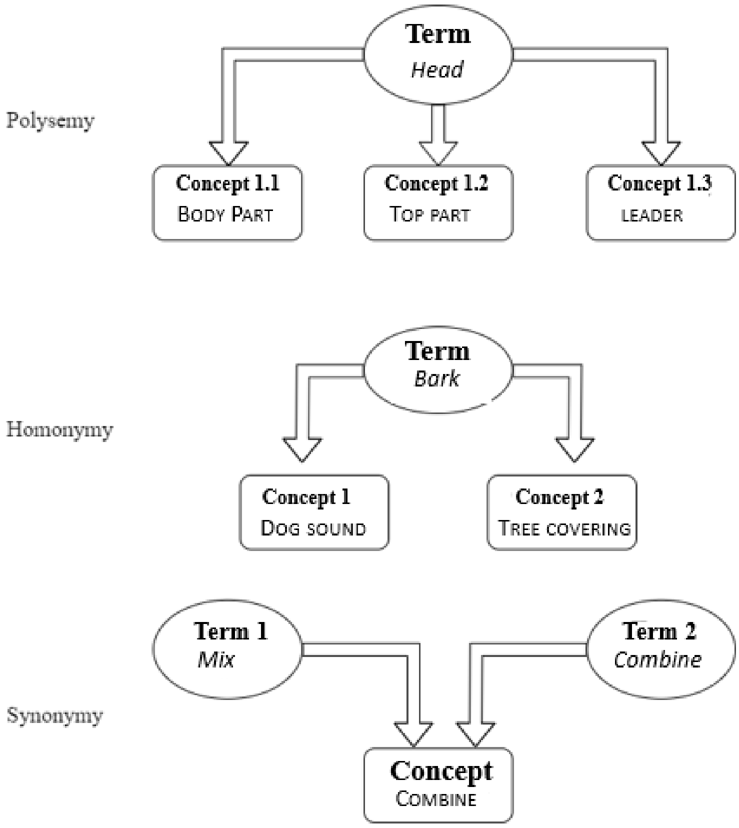

Figure 3: Schematic representation of polysemy, homonymy, and synonymy

It may also be the case that a term is incorrectly assigned as a hyponym/hypernym of another, as is the case with *fish* and *whale*. Figure 4 shows this schematically, where *feline* is a hypernym of *house cat*, which in turn is a hyponym of *feline* but if we were to discover that house cats are not felines, this categorisation would be incorrect. The concept of HOUSE CAT could not be defective, but how we refer to it would be.

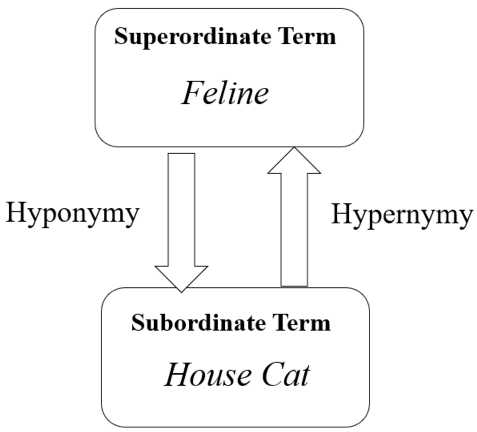

Figure 4: Schematic representation of hyponymy and hypernymy

The case we examine here is arguably a case of synonymy, where before Newton's CE, the term *weight* could be used to mean either MASS or WEIGHT (see Figure 5).

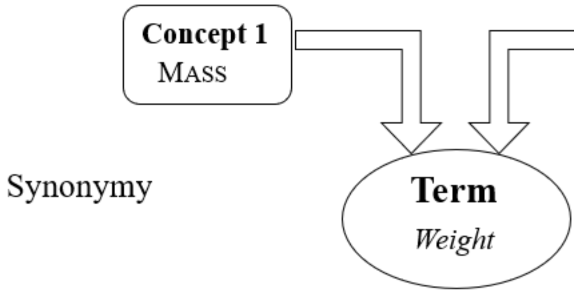

Figure 5: Schematic representation MASS and WEIGHT

Chapter 2 introduced the various attitudes towards concepts in CE and stated that this work remains neutral when it comes to conceptual debates. Concepts exist in cognition, not in language, making the direct, quantitative study of concepts difficult without access to human cognition (Meyns, 2020). With historical data, only the linguistic manifestation of these concepts remains, which may not be reliable. This study looks at the linguistic uses of *mass* and *weight* to try and investigate how and when the concepts became separate. Rather than doing this through a study of cognition, this will occur through a study of usage. There is the assumption that if usage patterns start to diverge between these two terms, there is evidence of the acceptance of the differential lexicalisations of these concepts.

Turning now to the type of data used, it is commonly thought that the vernacular is the primary and best object of linguistic research, and that written sources are secondary to this spoken data (Gray and Biber, 2018). However, registers differ in important ways that are largely contextually determined.

Each community has different communicative goals, and these goals produce a certain register that ultimately becomes differentiated from the vernacular. The scientific community aimed to share information between its members, and during this time the main medium for doing so was through written language, especially journal articles and letters. The public associated with the Royal Society was a reading public as opposed to a speaking public.

Written communication was thus the main medium and register used by the scientific community during this time. Another common misconception is that scientific language is slow when it comes to innovation, and that written language cannot be the locus for linguistic change, both of which have been argued against by Gray and Biber (2018). Therefore, in the case of the scientific register from 1700 to 1920, written language is not the next best thing, it is the best option.

6.2 Sub-corpus creation

This study uses the Royal Society Corpus 6.0 (Fischer et al., 2020) to investigate these questions about *mass* and *weight*. The Royal Society Corpus (RSC) is a linguistic corpus based on *The Philosophical Transactions of the Royal Society of London* and the other journals of the society. The most recent release (6.0) was in 2020 (Fischer et al., 2020), and spans from 1665 to 1996, of which the data from 1665 to 1920 is available via the CQPweb interface (Hardie, 2012) from Saarland University. The RSC 6.0 consists of 47,837 texts, most of which come from *The Philosophical Transactions of the Royal Society of London* and *The Proceedings of the Royal Society of London*, of which 17,520 texts are available in the open release, giving a total of 78,605,737 tokens. It is homogenous in nature (Menzel et al., 2021) and covers a diachronic period of great transformation in the scientific domain, spanning the transitional period between Early Modern and Late Modern English to contemporary scientific writings.

Furthermore, detailed metadata is available for the corpus (Menzel et al., 2021), and each token contains a link to the source texts on journal websites, including JSTOR and the Royal Society website. The number of texts and tokens in each period of the RSC 6.0 Open are in Table 3, taken from Menzel (2020, pg 285).

Time Period	No. of Texts	Tokens
1665-1699	1325	2,582,856
1700-1749	1686	3,414,795
1750-1799	1819	6,342,489
1800-1849	2774	9,112,274
1850-1900	6754	36,993,412
1900-1920	6849	45,271,473

Table 3: Number of texts and tokens in the Unrestricted RSC Version 6.0 Open

The corpus was searched online through CQPweb (Hardie, 2012). All instances of *mass*, *masses*, *weight*, and *weights* were identified, and downloaded into a .txt file along with a context window of 20 words on either side, and metadata such as their part-of-speech (POS), the year, author, journal, title, type, and URL. The extremely low frequencies of some of the non-standard spelling variants can be found in Table 4, and these were excluded from the data. This gave a total of 56,813 results in the new *weight-mass* sub-corpus.

Spelling	Frequency	Comment
Masse	11	
Massa	29	
Weght	2	
Wight	262	252 of these refer to "Isle of Wight"
Weite	1	1 from a German text
Weit	11	10 of which are in German texts
Waite	11	10 of which are the surname "Waite"
wiht/wayte/waight/waighte	0	

Table 4: Spelling variations and their frequencies in the entirety of the RSC 6.0

Since the annotations for sense and object had to be done by hand, it was necessary to reduce the 56,813 results, which were filtered by the frequency with which each author occurred in the sub-corpus. Only authors who used *mass* and *weight* over 100 times were included in the present study. This meant that fewer authors were present, and so the chances that a single author contributes more than one token to the dataset when it is randomly sampled is maximised. Authors that occur with a high frequency in the sub-corpus are more likely to use either *mass* or *weight* often in their work, suggesting that these concepts form the theoretical core of their research or that they discuss these terms in some depth. Furthermore, this methodological choice may skew the results since authors studying and focussing on these concepts may be more likely to separate the two concepts than other authors. If even this group is found to have incomplete or inconsistent separation of the concepts, then it is unlikely the lower frequency authors would have used them consistently.

The sub-corpus consisted of 15,249 total instances after the application of the filter, consisting of *mass* = 4,843 (31.7%), *masses* = 1,475 (9.7%), *weight* = 7,119, (46.7%), *weights* = 1,820 (11.9%). Since 15,249 tokens were too much to annotate by hand, roughly 10% of this amount was chosen for annotation, thus 1,500 tokens. To keep the 1,500 tokens representative of the original sub-corpus, the filtered sub-corpus contained 476 random tokens of *mass*, 145 of *masses*, 700 of *weight* and 179 of *weights* to occur with equal frequency compared to the total corpus and to ensure a representative distribution of each author and year.

6.3 Annotation

The annotation for these 1500 tokens occurred by hand, and included annotations for ‘subfield’, ‘object’, and ‘sense’, in addition to existing metadata from the RSC 6.0 Open.

6.3.1 Subfield

Firstly, tokens were annotated according to 7 ‘subfield’ categories using their titles and the original text: Astronomy, Biology, Chemistry, Engineering, Geology, Physics, and Meta. The first six are self-explanatory, but the last requires explanation. Meta texts discuss the creation of standard weights and measurements and the weight of ancient coinage and do not fall under any of the other categories posited.

It is worth noting that these are not all mutually exclusive categories, and there is much overlap between them. For example, a text such as *IV. On the stresses caused in the interior of the earth by the weight of continents and mountains* (Darwin, 1882) includes aspects of Geology, Physics, Engineering and even Astronomy within a single paper. Texts were annotated by what was perceived to be the main aspect of the contents, and, most importantly, these annotations were consistent, meaning that all tokens from the same article were annotated with the same ‘subfield’. The labels given for the ‘subfield’ is, in a sense, arbitrary and were not designed to be definitive splits between the disciplines. The boundaries we have in place today simply do not apply to historical scientific enquiry, and scientists often worked on subjects that crossed modern boundaries. The modern divisions did not become established until the 19th century. Rather, these ‘subfields’ aim to show the communities of scientists working on roughly similar ‘subfields’, meaning there is a high chance they would have known each other and interacted. This, therefore, created a sort of social co-influence that is interesting when examining the spread of a scientific concept. Furthermore, they provide a macro-topic of the token in question.

6.3.2 Linguistic features

Linguistic features were included in the analysis to rule out any other linguistic influences apart from the sense. The POS of the word before and after was derived from the POS tagged contexts provided in the metadata.

Plural/singular was annotated by using the target words themselves; *weights* and *masses* are plural whereas *weight* and *mass* are singular. These were annotated as belonging to the lemma *mass* or *weight*.

6.3.3 Sense

The annotations become more complex when it comes to the sense, and each shall be explained below in turn. The tokens were annotated for 'sense' using the context window, and if it was still unclear, by finding the token within the original text itself.

6.3.3.1. Notation

Before starting the section on the different senses, a note must be made about how these will be referred to throughout the rest of the paper. To avoid confusion between the lemmas and the senses, the lemmas will be referred to by italics, such as *mass* and *weight*, whereas the senses will be referred to by small capitals, such as MASS and WEIGHT, or the letter M or W. The word *mass* can therefore be used with both the senses MASS and WEIGHT.

6.3.3.2. N

The annotation N represents the usage of the words *mass* and *weight* in a nominal sense. Rather than referring to the MASS or WEIGHT of an object, it instead refers to the object itself. One way this was judged was whether the target word could be replaced by "stuff" or "thing" without losing the intended meaning.

(1) *the mass on the filter was treated with boiling alcohol*
(Schunck, 1853)

(2) *a flat circular weight nicely turned, and pierced in the direction of its diameter to receive the bar, was slid upon it*
(Kater, 1819)

Example (1) shows the word *mass* being used in a nominal sense. In this case, it is used to refer to a chemical compound that had appeared on a filter following a chemistry experiment. *Mass* in this instance is used in the sense of "compound" and acts as a noun. In Example (2), the word *weight* is used to refer to an object used in scientific experiments, in a similar way that we speak about "weights" at the gym. While these nominal weights do possess WEIGHT in the scientific sense, the usage as shown in 2 is not referring to the property of WEIGHT within another item, but instead uses *weight* as the noun itself. A nominal usage can often be identified using adjectives or modifiers; in the case of Example (2), the weight is described as "flat circular". The concept of WEIGHT being the force upon the MASS cannot be either "flat" or "circular", but a thing can be.

6.3.3.3. M

The annotation M is used when a token shows a sense of MASS. To revisit the previous discussion, this refers to how much matter is within an object, rather than the force generated by MASS in a certain gravitational field. The MASS of an object would not change if we changed the gravitational field, but the weight would. Both the words *mass* and *weight* can be used with the 'sense' M, as is highlighted below.

(3) *We are thus led to inquire how the stresses are distributed in the earth's mass and what are their magnitudes*
(Darwin, 1882)

(4) *In the third, the weight of the principle bones of a selected number of species (27) is stated*
(Davy, 1865)

In Example (3), Darwin is talking about the amount of matter within the earth. While the mention of stresses within this mass may imply he is speaking of force, he is in fact writing of the stresses within the matter of which the earth is comprised. Example (4) uses the word *weight* but with the ‘sense’ M. In a study titled *“Some Observations on Birds, Chiefly in Relation to Their Temperature, with Supplementary Additions”*, it is unlikely that he is speaking about the downwards force that these bones are generating in earth’s gravitational field. Instead, he writes of the amount of matter within them and compares them across species.

6.3.3.4. W

W is the category for the ‘sense’ WEIGHT. As mentioned in the previous section, this refers to the interaction of MASS with a specific gravitational field; it is a measure of force rather than a measure of matter. As such, tokens with the ‘sense’ W explicitly refer to force, balancing, counterpoises, or the amount of effort required to lift something. While the current official unit of WEIGHT is Newtons, this convention was only introduced in 1948, after the end of the corpus. Before this time, the same units were used for both *mass* and *weight*, meaning this had no contribution to the annotations.

(5) *fig. 3 is only 40 feet from the bow, and that the excess of weight over buoyancy on this length is only 45 tons*
(Reed & Stokes, 1871)

Within the instances annotated in the corpus, there are no instances of *mass* having ‘sense’ W. This interesting one-sided polysemy of *weight* but not *mass* will be addressed in the discussion section. Example (5) shows the word *weight* with ‘sense’ W. Reed & Stokes are referring to the balance between the upwards force of buoyancy within the water and the downwards force generated by the interaction of a ship’s *mass* with the gravitational field. There were a few cases that were difficult to judge as to whether they had the ‘sense’ W or N, as is the case in Example (6).

(6) *Placing quantities of two simple bodies… in the orometer, and equilibrating by weights placed in the upper pans*
(Joly & Fitzgerald, 1889)

It would be possible to say that the *weights* spoken of are nominal, just as it would also be possible that they refer to the property of WEIGHT. In cases like this, it was decided that the 'sense' would be W. The reasoning behind this is that even if the *weights* spoken of referred to nominal weights, as would be depicted by 'sense' N, what is important in this context is their downwards force. *Weights* could not be replaced by 'things' in this context, instead, it is the WEIGHT of the weights that is important.

6.3.3.5. W/M

This annotation is used when it is unclear whether the token refers to the 'sense' M or the 'sense' W. Again, there are no examples of this 'sense' for *mass*; it appears as if it is always clear which 'sense' is being used, which is almost always N or M. It does, however, occur with *weight*.

(7) *The Commissioners for the Restoration of the Standards of weight and measure, in their Report dated December 21, 1841, recommended that*
(Miller, 1856)

Example (7) is an instance of what has previously been described as a Meta text. Miller is writing about the construction of the new imperial standard pound and refers to the commissioners of "standard weight and measure". In this case, *weight* is not used as a noun interchangeable with "object" but does refer to a property that objects have. However, they may be talking about standardising the unit of MASS or standardising the unit of WEIGHT. It is not made clear which they speak of. Therefore, the category reflects this ambiguity.

6.3.3.6. COL

The annotation COL is used to refer to a collection of objects rather than the objects themselves, or the property of either MASS or WEIGHT. As opposed to some of the previous categories, only tokens of the word *mass* in both singular and plural forms can occur with this 'sense'.

(8) *A glacier is not a mass of fragments*
(Forbes, 1846)

In Example (8), Forbes is stating that a glacier is not composed of a collection of fragments, aiming to stress its unity. The word *mass* in this usage is interchangeable with “collection” and does not refer to the MASS or WEIGHT of a glacier.

6.3.3.7. MET

Both the words *weight* and *mass* can be used in a metaphorical sense as well as a descriptive or nominal sense.

(9) *The next thought is that I may have assigned too great a mass to the doubt*
(Pratt, 1855)

(10) *The contact theory has long had possession of men’s minds, is sustained by a great weight of authority*
(Faraday, 1840)

Both Examples (9) and (10) from Pratt and Faraday respectively refer to the weight in a metaphorical sense. No literal WEIGHT is being described, but rather the importance or impact of a certain authority, statement, or opinion.

6.3.3.8. Q

Finally, Q is a ‘sense’ that does not occur frequently within the corpus and might be a calque or translation of the French phrase *en masse* which is also used in English, referring to a single body or group operating collectively. Interestingly, the tokens that have this ‘sense’ do not use the modern conventionalised spelling (which is identical to the French spelling) but instead make slight adaptations.

(11) *The pores of artificial graphite appear to be really so minute, that a gas in mass cannot penetrate the plate at all*
(Graham, 1863)

(12) *In the iron ship, there are no actual joints, for the whole being bound together en mass*
(Fairbairn, 1850)

Example (11) completely anglicises the French phrase, rendering it “in mass”, whereas (12) only anglicises the second part, leaving the French preposition intact.

6.3.4 Object

The ‘object’ of the token is what the property of MASS or WEIGHT refers to, more specifically what kind of object has the specified mass or weight. In cases where the ‘sense’ is neither MASS nor WEIGHT, the ‘object’ still refers to what the token represents, or what kind of ‘object’ the nominal, metaphorical, collective, or calqued sense is referring to. While this remains vague, some of the examples previously used to illustrate the various senses will again be used to explain what is meant by the ‘object’.

(13) *The mass on the filter was treated with boiling alcohol*
(Schunk, 1853)

The *mass* in this case is a chemical of sorts, whether it be the outcome of a chemical reaction, an atom, or a powder used in chemical experiments. Therefore, this token gets the annotation C, for “chemical”.

(14) *a flat circular weight nicely turned, and pierced in the direction of its diameter to receive the bar, was slid upon it*
(Kater, 1819)

As mentioned in the section speaking about the ‘sense’ category W, it is easy to confuse *weight* the property with *weight* the noun. While the object has the property of WEIGHT, in these instances it is the object itself they are discussing. The label for such an ‘object’ is WO (Weighted Object). It is possible to refer to these with the ‘sense’ N, but also possible to refer to them with the ‘sense’ W and speak about the WEIGHT of these weighted objects.

(15) *We are thus led to inquire how the stresses are distributed in the earth’s mass and what are their magnitudes*
(Darwin, 1882)

The *mass* here is the earth’s MASS, namely a geological object made up of the crust, mantle, and core. Therefore, this token gets the annotation GEO for “geological object”, a category shared with the likes of glaciers and lava.

(16) *The Commissioners for the Restoration of the Standards of weight and measure, in their Report dated December 21, 1841, recommended that the avoirdupois pound of 7000 grains.*
(Miller, 1856)

In the case of Example (16), the 'object' refers to a standard of weight or measure. For example, many texts write about what should be the standardised WEIGHT/MASS of the pound or kilogram. Miller is not speaking about a specific WEIGHT or MASS, but rather what a standard unit of either WEIGHT or MASS should be, and is therefore in the 'object' category STAN.

(17) *The next thought is, that I may have assigned too great a mass to the doubt*
(Pratt, 1855)

This metaphorical usage has no physical object, rather the object is abstract, in this case, Pratt's doubt. Therefore, instances like this are assigned to the 'object' category ABS.

The total number of 'object' categories used in the annotation is shown in Table 5. A note must be made about nesting in the annotation. Many of the 'object' annotations detailed above will be more likely to occur within certain 'subfields'. For example, the 'object' GEO in Geology, and S in Astronomy. At first glance, this may make the 'object' annotations seem redundant. However, for many of the 'subfields', the 'objects' discussed are diverse. In Physics, one might use *mass* to refer to matter, a weighted object, a vessel or machine, force, fluid, equipment, a shape, a concept, and so forth. While some 'objects' may be more likely in certain 'subfields' than others, specific 'subfields' do not guarantee the usage of specific 'objects' token in question, standing in contrast to the 'subfield', which functions as the macro-topic.

Thus, once all the annotation had occurred, the metadata that will be used in the analysis is as follows: lemma, 'sense', 'object', 'subfield', 'author', 'year', 'type', 'plurality', 'POS before', and 'POS after'. These were chosen to be a mix of linguistic and extra-linguistic factors, to analyse the change in the usage diachronically.

Abbreviation	Object	Description
ABS	Abstract	Not physical, such as importance and impact
BIO	Biological	Biological items such as cells, cancers, and blood
BO	Body	The human body
C	Chemical	Of a chemical nature, whether it be a chemical compound, or the product of reaction
COINS	Coins	Either contemporary or ancient forms of currency
CONCEPT	Concept	The concept of what *weight* or *mass* are
EQ	Equipment	Equipment in scientific experiments, such as test tubes or pendulums
F	Fluid	Fluid, for example water or alcohol
FOR	Force	Downwards force, for example breaking weight or weight specific resistance
GEN	General	Generalising statements about a group of entities
GEO	Geological	Geological in nature, including glaciers, lava, and volcanoes
MAT	Material	The material something is made up of, for example electricity passing through a sheet of iron
MATTER	Matter	Referring specifically to matter
S	Space	Astrological in nature, namely planets and stars
STAN	Standard	A standard of weight or measure
VM	Vessel or Machine	Forms of transportation, or mechanical devices. Examples include ships and engines
WO	Weighted Object	Objects that have weight and are used in experiments/observations. When the object is WO, the word can be replaced with "thing"

Table 5: Object categories in annotation

6.4 Data analysis

Different data analyses will be conducted to answer different research questions.

6.4.1 Frequency

Firstly, a simple comparison of frequency over time will be used, both to track the percentage and absolute frequency of each of the terms and all their senses, and to check the engineered terms for 'correctness of usage' . Generally speaking, a usage-based approach to language does not subscribe to the general separation of *langue* and *parole* (de Saussure, 1986) or *competence* and *performance* (Chomsky, 1969). Instead, it takes grammar to be the culmination of one's language experience (Bybee, 2006). Frequency has been found to have a large effect on many processes within linguistics and therefore is an important inclusion in the study.

The frequency with which the lemmas *mass* and *weight* occur therefore do not only give us a synchronic snapshot of how often these words are used at a specific time. Rather, they form the linguistic input for the next generation of scientists and readers. A higher frequency of usage of *weight* with a certain ‘sense’, or a certain ‘object’, will therefore contribute to the entrenchment and salience of certain types and tokens, affecting their diachronic development.

Therefore, including imbalances in the frequency of usage of *mass* and *weight* is crucial to model the language knowledge of speakers at the time. If *mass* occurs more frequently than *weight*, more instances will be encountered and stored (depending on your theoretical viewpoint) and therefore contribute more to any mental representations created. The factors that condition the choice between the two lemmas depend on the experience of the writer.

6.4.2 Conditional Inference Trees and Random Forests

Conditional Inference Trees (Ctrees) and Random Forests (Breiman, 2001) are the second forms of data analysis used in this research. Both methods work in a decision tree fashion, by recursively splitting data into two parts so that the split optimises the classification accuracy of the dependent variable (Gries, 2020). Put more simply, it uses the independent variables specified to predict which outcome will occur in a specific dataset. Ctrees use regression and p-values to identify the optimal split, which has the advantage of reducing the need for pruning and cross-validation that other decision trees have (Gries, 2020). Furthermore, they produce a hierarchical visual structure to represent this. Random Forests are more recent, stemming from Breiman (2001), and work by fitting multiple decision trees on random subsets of the data, and therefore overfit less than individual decision trees (Gries, 2020) and also allow for predictions of data that were not included in its training set. However, since a Random Forest includes so many different trees and datasets, visualisation and interpretation are difficult. To remedy this, the Random Forests compute variable importance scores, which quantify and rank the importance of the independent variables in the model in predicting the dependent variable.

Random Forests will be used alongside Ctrees to provide the statistical significance of the Random Forests with the visualisation and clarity of Ctrees. Ctrees are not representative of the Random Forest they are used to build but are instead an alternative and complementary visual and statistical cue for the conditioning factors in the variation. This is built on the caution of Gries (2020), who says that while Random Forests are useful for reducing some of the risks concerning the accuracy, using a single Ctrees to represent or interpret an entire Random Forest is dangerous. This analysis aims to capture the factors that condition the variation between *mass* and *weight* diachronically, in other words, what linguistic and non-linguistic factors influence the choice of word.

The Ctrees and Random Forests will be made in R using the party package (Hothorn et al., 2006a,b). The code is based on that provided by Fonteyn and Nini (2020).

6.4.2.1. Models

The Ctrees and Random Forests will be used to create several models. The first kind of model will only involve cases where the 'sense' annotation is either W or M, to investigate the instances where these senses are used, and whether there is a match or mismatch between the 'sense' and target word used. These are chosen since these are the choice contexts; authors can choose to use either the word *weight* or the word *mass* in these contexts interchangeably. For example, "the *weight/mass* of the table is 12kg" is a context in which authors have the choice between the two terms.

The second kind will include all the different senses that have been analysed. This will result in contexts in which there is not necessarily a choice, for example, "the *mass* of cancerous cells" is not interchangeable with "the *weight* of cancerous cells". This is because the 'senses' N, MET, COL and Q are being compared in the same model as W and M. While this is not as effective at identifying the factors that influence the choice of the target word, it is effective at identifying the patterns in the larger lexical network. If it is found that the word *weight* becomes more closely associated with the 'sense' W over time, and the word *mass* becomes more closely associated with M, it is still not known how this affects the other senses of the terms *mass* and *weight*. We may see entrenchment effects as described by De Smet (2017), such as the increase in metaphorical or nominal contexts. While *mass* and *weight* may have been engineered to reduce ambiguity in one specific area of scientific discourse, the result may be increased ambiguity or polysemy in another place in the lexical network. These models aim to extract these effects.

6.4.3 Principal Component Analysis

Ctrees and Random Forests are useful for determining how the terms are distinguished but will not provide much insight into how similar the terms are at different periods. In addition to the decision trees, a Principal Component Analysis (PCA) will be used. This is a dimensionality reduction method that can bring out patterns in a dataset. All the data can be plotted in a single graph using principal components as the axes. To find these principal components, a PCA uses the eigenvectors of a covariance matrix of all the variables in a dataset. The eigenvector with the largest eigenvalue is the first principal component (the X-axis), and the eigenvector with the second largest eigenvalue is the second principal component (the Y-axis).

PCA has several benefits, including its ease of computing and ability to reduce high dimensionality data to something interpretable. However, the issues include the fact that the principal components are often hard to interpret. While the principal components are linear combinations of features in the high dimensionality dataset, they are not as easy to understand as these features, making their identification difficult. Furthermore, as with any dimensionality reduction technique, it is important to be aware of the compromise between dimensionality reduction and information loss.

The PCA was done using Python. The foundations of the code come from Fonteyn (2021). The dimensionality reduction occurred through first creating count vectors. This is done by merging all the columns containing the annotations into a single "sentence". Using 'CountVectorizer', the co-occurrence matrix is transformed into binary features. This is because PCA is sensitive to feature scale, and this bias can be controlled by assigning binary features for whether a feature is present or absent. ‘CountVectorizer’ works through a bag-of-words model, simply assigning values for the presence or absence of certain features, giving no preference to annotations that may happen to be in certain positions. These count vectors then undergo dimensionality reduction through PCA. This results in a two-dimensional graph where each token is plotted according to two principal components. As mentioned, the main aim of conducting a PCA is to create a visual representation of the data that can show patterns in the usage of *weight* and *mass* over time. The annotations of each of the data points will be analysed based on their coordinates and clustering.

6.5 Hypotheses

In using this methodology to operationalise the research questions, we can now concretely state what we expect to find concerning the research questions.

If the Conceptual Engineering of *mass* and *weight* was successful, then the Random Forests and Ctrees will show ‘sense’ to be the factor with the highest importance when splitting the data. Furthermore, the PCA should show visual differentiation based on ‘sense’. There would also be an increase in the usage of *mass* and *weight* with their ‘correct’ respective senses MASS and WEIGHT.

If the Conceptual Engineering of *mass* and *weight* was unsuccessful, the Random Forests and Ctrees will instead show that extra-linguistic factors such as ‘author’ or ‘subfield’ have higher importance than ‘sense’ when splitting the data. Furthermore, the PCA will show visual differentiation based on these external factors rather than ‘sense’. There would also be no consistent usage of *mass* and *weight* with their ‘correct’ respective senses MASS and WEIGHT over time.

CHAPTER 7

RESULTS

This section will use the data and its annotations described in the previous chapter to investigate the factors that determine the choice of lemma.

7.1 Frequency

The first piece of data we can investigate is the frequencies of each lemma over time and its dispersion within texts in the RSC. Table 6 shows the dispersion and frequency of *weight(s)* in the entire RSC corpus, and Figure 6 shows a visualisation of frequency over time.

Category	Words in category	Hits in category	Dispersion (texts with 1+ hits)	Frequency per million words
1650	2,582,856	1,040	306/1,325 (23%)	402.66
1700	3,414,795	1,553	349/1,686 (21%)	454.79
1750	6,342,489	3,483	469/1,819 (26%)	549.15
1800	9,112,274	3,411	612/2,774 (22%)	374.33
1850	36,993,412	13,612	1,870 /6,754 (28%)	367.96
1900	20,159,911	8,754	1,166 /3,162 (37%)	434.23
Total:	78,605,737	31,853	4,772 /17,520 (27%)	405.22

Table 6: *Weight(s)* distribution in RSC 6.0

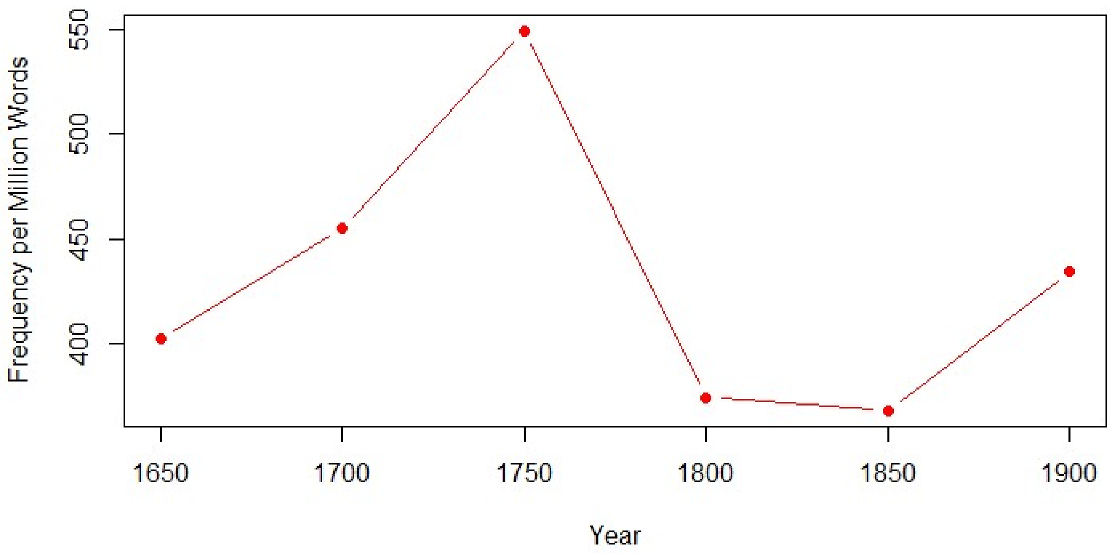

Figure 6: Frequency per million words of *weight* over time

Looking first at the dispersion, there is little variation over time until the 20th century. The number of texts with at least a single instance of *weight* remains relatively stable until 1900, ranging between 21% and 28%. Only after 1900 does the dispersion increase to 37%. The frequency per million words in Figure 6 shows more variation, increasing between 1650 and 1800, before seeing a drop in the 19th century to the lowest frequency in the corpus. It then gradually climbs again but does not reach the frequencies it did between 1650 and 1800.

Moving on to *mass*, the dispersion and frequency can be seen in Table 7, and the frequency is visualised in Figure 7.

Category	Words in category	Hits in category	Dispersion (texts with 1+ hits)	Frequency per million words
1650	2,582,856	179	104 /1,325 (8%)	69.3
1700	3,414,795	274	120 /1,686 (7%)	80.24
1750	6,342,489	915	275 /1,819 (15%)	144.27
1800	9,112,274	3,217	668 /2,774 (24%)	353.04
1850	36,993,412	14,174	2,345 /6,754 (35%)	383.15
1900	20,159,911	6,200	1,164 /3,162 (37%)	307.54
Total:	78,605,737	24,959	4,676 /17,520 (27%)	317.52

Table 7: *Mass(es)* distribution in RSC 6.0

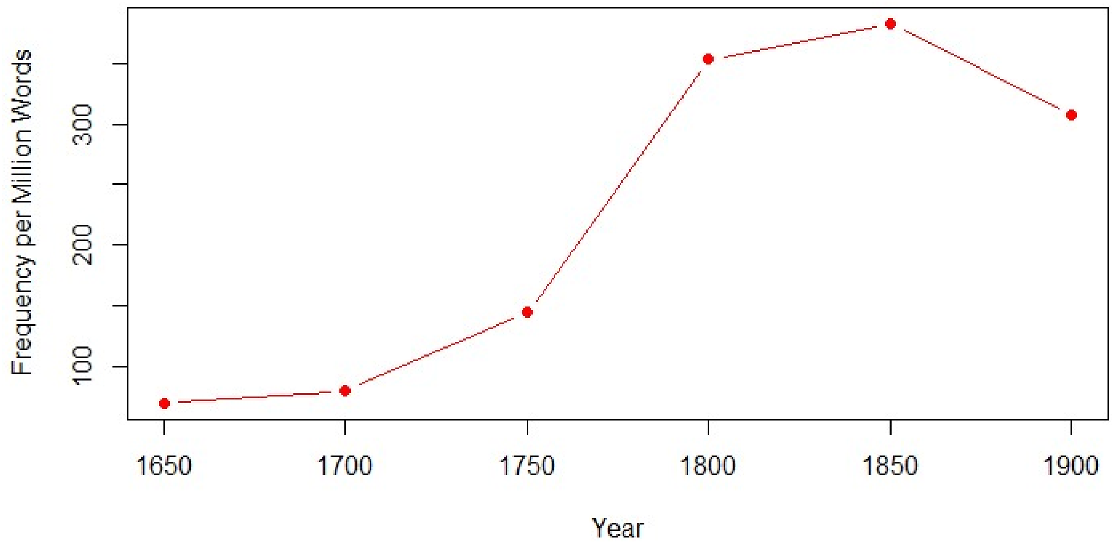

Figure 7: Frequency per million words of *mass* over time

Looking first at the dispersion, there is an increase in the proportion of texts that include the word *mass* over time from only 8% in 1650-1700 to 37% in 1900-1920. A similar pattern can be seen in the frequencies per million words. Over the different periods, *mass* becomes more frequent, except for in 1900-1920, when a decrease can be seen. The most significant increase occurs around 1800 when the frequency jumps from 144.27 to 353.04 instances per million words.

The above figures pertain to the number of tokens in the total RSC 6.0 corpus, but this is not the corpus used for the statistical analysis in this paper. Rather, a random sample of a filtered sub-corpus was used. Turning now to the structure of this sub-corpus, Table 8 shows the number of tokens within the sub-corpus for both *mass* and *weight*.

Time Period	No. of words within the texts with tokens in the sub-corpus	Tokens of *mass*	Tokens of *weight*
1650-1700	0	0 (0%)	0 (0%)
1700-1750	25,719	1 (6%)	16 (94%)
1750-1800	325,353	3 (2%)	123 (98%)
1800-1850	1,283,560	131 (42%)	178 (58%)
1850-1900	3,422,541	406 (48%)	447 (52%)
1900-1920	536,638	80 (41%)	115 (59%)

Table 8: Tokens of *mass* and *weight* in the filtered sub-corpus used for analysis

Noticeable in Table 8 is the lack of tokens in the 1650-1700 category. Once the filter was applied, there were no instances of *weight* or *mass* from authors with a frequency above 100 before 1700. For this study, this is not problematic, since the first instance of *mass* with the sense MASS in English dates to 1704, meaning the diffusion and spread of this term will have only begun after this time. For all other years (except 1900-1920 due to the shorter period), the number of tokens increases as time goes on. *Weight* remains more frequent in each time period, but *mass* tokens spike around 1850-1900.

7.1.1 Frequency of senses

Turning now to the frequency of the different senses over time, Figures 8 and 9 show the normalised frequency of each sense per 100,000 words for both *mass* and *weight*.

This normalisation occurred using equation 7.1 to find f_{norm} (the normalised frequency per 100,000 words).

$$f_{norm} = \frac{N_{sense} T_{period}}{A_{lemma}} \frac{100,000}{S_{subcorpus}} \tag{7.1}$$

Where N_{sense} is the number of tokens of lemma with sense in question in a particular time period, T_{period} is the total number of tokens of lemma in time period, A_{lemma} is all the tokens of the lemma in the entire corpus in time period, and $S_{subcorpus}$ is the total number of words in time period in the annotated sub-corpus taken from Table 8.

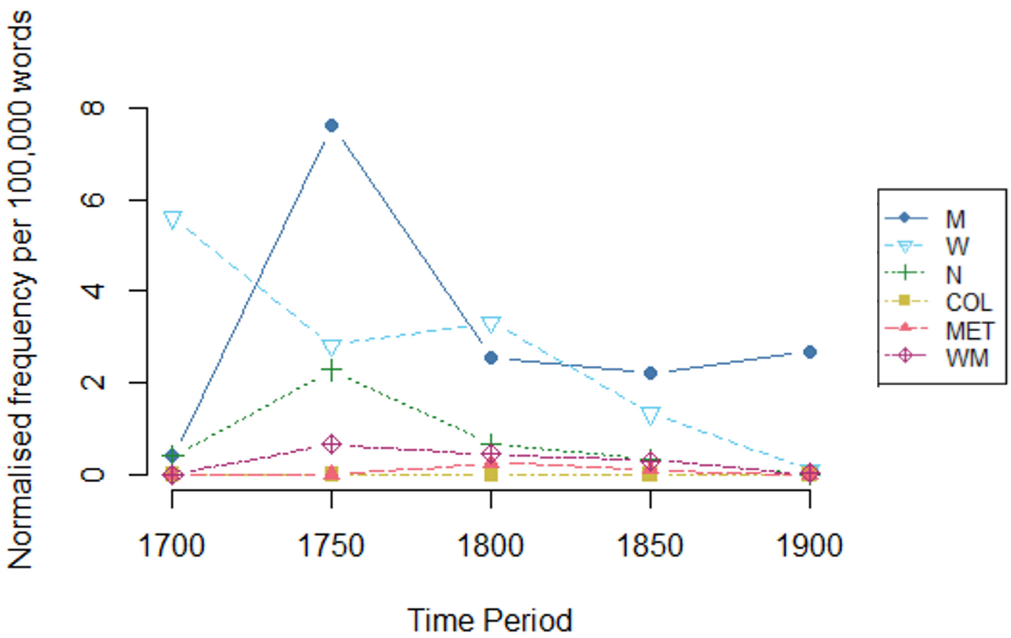

Figure 8: Frequency of all senses of *weight* over time

First noticeable from Figure 8 is that WEIGHT is not the sense used most frequently with *weight*. If this was a successful instance of CE, it would be expected to see an initial phase of fluctuation of senses before *weight* and *mass* were engineered by Newton. Even for a short while afterwards, fluctuations would be expected. However, in the long run, it would be expected to see a steady increase in the proportion of *weight* used with the sense WEIGHT. This is not the case.

The period 1700-1750 looks like an anomaly on the graph in comparison to the trend over time. While MASS is the dominant sense throughout nearly the whole period, in this initial phase it appears with a much lower proportion, and WEIGHT is dominant. However, this is because there are very few tokens in this period in the sub-corpus, and all of them were written by Desaguliers. It is, therefore, not possible to draw statistically sound conclusions that this was indeed the dominant trend at the time.

However, it does allow for an interesting observation. Desaguliers was an avid Newtonite, even writing a book named *The Newtonian System of the World, the Best Model of Government: An Allegorical Poem* in 1728. The uptake of Newton's engineered terms by Desaguliers could reflect Desaguliers' language ideologies, framing Newton as an authoritative inspiration to scientific progress. The period at the start of the 18th century coincides with the time that Newton was the president of the Royal Society (1703-1727), meaning he would have had large amounts of influence, authority, and control over what got published, especially when it came to the representation of his ideas and theories. Banks (2008) notes that Newton controlled the Royal Society in a highly authoritarian manner.

Desaguliers uses *weight* with sense WEIGHT, showing that at least within this one individual, Newton's CE seems to have found root.

(2) *I added 17 Pounds and a half before the Power could go down and raise the weight. Experiment 2 . Two Hundred and an half being balanced by half an Hundred , the Addition of 28*
(Desaguliers, 1931)

This is not the only sense of *weight* that Desaguliers uses, since Example (3) also shows he used it with sense N. Desaguliers is referring to a weighted object, one that reflects one conception of a standard of weight and measure. He is not speaking of a downwards force generated by MASS, but rather a physical object compared with another.

(3) *I sent to a curious Gentleman for some Paris Weights exact to the Standard weight at the Chatelet; and found upon trial, the Paris Ounce , which contains 576 of their Grains*
(Desaguliers, 1920)

It is correct to be sceptical that all tokens from this period are from Desaguliers. However, this is not an error due to the random sampling, but upon closer inspection, Desaguliers is the only author to appear in the corpus for this period once the filter was applied to find authors who used the words *mass* or *weight* over 100 times.

The data around 1800-1825 is the only time that it can be considered that Newton's engineered terms are being successfully and accurately used. For this period only, *weight* is used most frequently with the sense WEIGHT rather than with MASS. However, after this period, the proportion of WEIGHT drops again, and the differences in the proportions become pronounced. Between 1850 and 1900, the frequency of MASS is once again rising whereas WEIGHT is falling. Thus, if we measure the success of this CE project by the proportion of *weight* used with the sense WEIGHT, this instance of CE did not succeed.

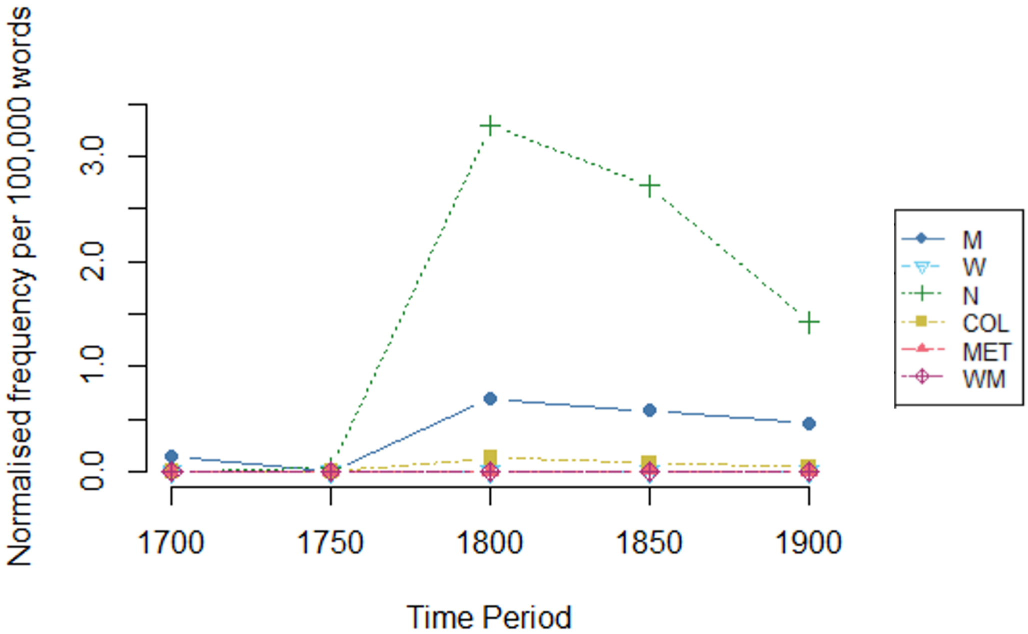

Figure 9: Frequency of all senses of *mass* over time

Figure 9 shows the same data for *mass*. At the start of the periods in question, the data is sporadic. In 1704, *mass* was first used with the sense MASS in English, after Newton first used it in Latin in 1687. This original token is not part of the annotated corpus used for analysis, and instead, the earliest token of *mass* is from Desaguliers in 1724 and will be looked at in some more detail.

(4) *For since the Gravity at a : Is to the Gravity at d : : As dn d C : to ab a c , it will follow (from the Nature of the Ellipse) that the Gravity at A : will be to the Gravity at d : : As A C4 : to d C4 : and therefore the Forces , with which the Columns of Fluid AC and dC tend towards C , will be as their* <u>*Masses*</u> *drawn into the Forces driving towards C , that is , as AC A C4 to dC d C4.*
(Desaguliers, 1724)

This token was annotated as having the sense MASS because it appears that *masses* refer to the MASS of the columns of fluid, meaning that the matter that makes up the fluid will be "drawn into the forces driving towards C". However, it is possible to analyse the same token as having a nominal sense rather than MASS.

Only after 1750 does the lemma *mass* become frequently used with any of the senses analysed. From this time onward, the nominal sense is by far the most frequent, displaying a slight downwards trend but remaining significantly above the other senses. Since this nominal sense of *mass* was established before Newton's CE, this appears to be a continuation of the way it was used before the *Principia*.

Furthermore, while *weight* is used with both MASS and WEIGHT, *mass* is never used with the sense WEIGHT. As mentioned in the methodology section, there is some kind of one-sided polysemy occurring.

7.1.2 Frequency across journals

Another factor worthy of consideration is the fact that *The Philosophical Transactions* split into two journals in 1887, with journal A dealing with Mathematical, Physical and Engineering sciences, and B dealing with Biological sciences. As discussed in Chapter 5, certain scientific disciplines may have more need for the *mass/weight* distinction than others, namely Physics, Astronomy and Engineering. It is potentially the case that these journals show very different patterns once they were split. Figures 10 and 11 show the percentage of each 'sense' used with both *mass* and *weight* between when these journals were established to 1920.

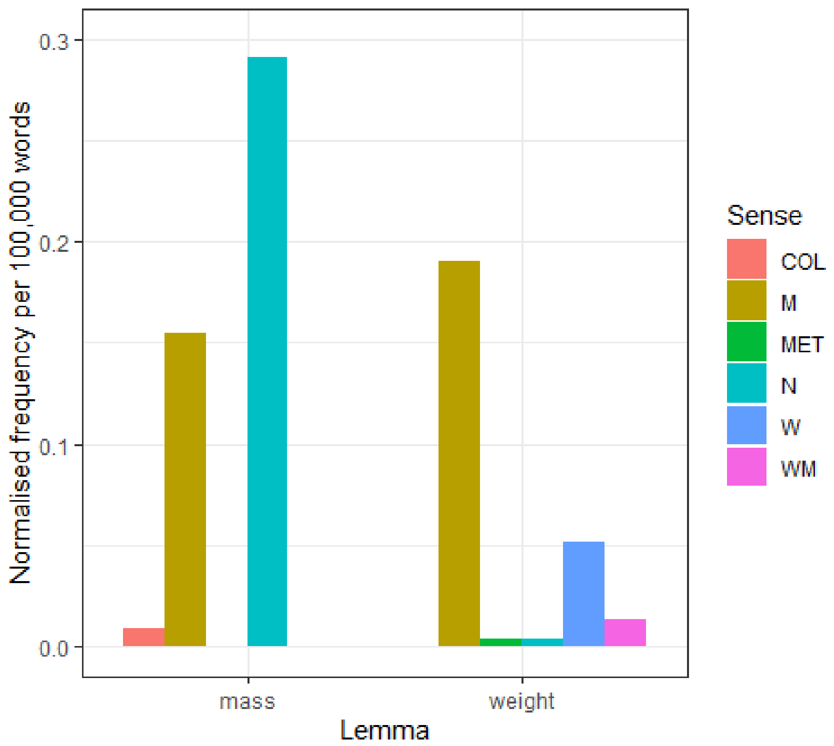

Figure 10: Sense frequencies for journal A: Mathematical, Physical and Engineering Sciences

In journal A, *mass* is most frequently used with the sense N, but the sense MASS is the second most frequent, being around 35% of cases. *Weight* is used most frequently with the sense MASS, with only around 20% of all instances being used with WEIGHT.

In journal B, *mass* is used exclusively with the sense N. *Weight* is used almost exclusively with the sense MASS, and very infrequently with a metaphorical sense.

Thus, there seems to be a clear difference between the journals. While neither consistently use *mass* with MASS or *weight* with WEIGHT, there are differences in distribution. The lack of *mass* with any of the relevant senses in journal B might suggest that utility is an important factor in determining CE projects. Biology (usually) does not need to consider the different forces acting on biological substances, and so the *mass/weight* distinction is of little use. However, in journal A, where the distinction is very useful, the terms are still not used consistently.

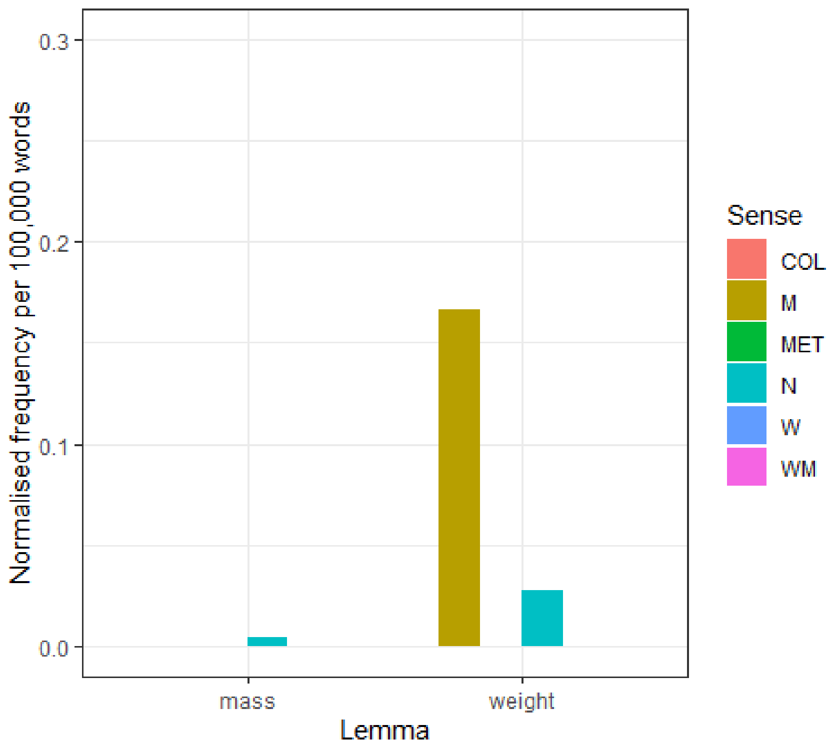

Figure 11: Sense frequencies for journal B: Biological Sciences

7.2 Ctrees and Random Forests

While the frequency data is useful in indicating whether the CE project was successful, another metric we can look at to measure success is the factors that determine the choice between *mass* and *weight*. To restate the hypotheses, if the CE project was successful, 'sense' should be the most important factor when determining lemma choice. If the CE project was unsuccessful, then other factors such as 'author', 'subfield', 'object' or language-internal factors like 'POS' or 'plurality' should be more influential.

To get more insight into the diffusion, models were made to check, for each period, which factors help determine the choice between *mass* and *weight*.

First, a general picture will be shown of the whole dataset. A Ctree and Random Forest were produced on the entirety of the 1500 tokens in the sub-corpus, the model fits of which can be seen in Table 9.

	All senses		Relevant senses	
train(75%) + test(25%)	Random Forest	Ctree	Random Forest	Ctree
Area under the curve	0.9719	0.9597	0.8889	0.8226
Accuracy	0.9751111	0.9626667	0.9670846	0.9498433
F1-score	0.9695652	0.9545455	0.8609272	0.7714286

Table 9: Model fits

The evaluation metrics show that the models created are a good fit. Accuracy refers to the percentage of correct predictions made by the model. In both the whole dataset and in the data of the relevant senses, this is above 94%. F1-score is the harmonic mean of recall and precision. The trade-off between recall and precision means that as precision increases, recall tends to decrease. The models here have an F1-score of above 0.77 and get up to 0.96 for the Random Forest for the entire dataset. The area under the curve (AUC) can be interpreted as the ability of the model to distinguish between the different classes it aims to classify. Again, this value ranges from 0.82 to 0.97. Random Forests usually have higher explanatory power compared to Ctrees, meaning that the evaluation metrics show a better model fit for the Random Forest than for the Ctree (Fonteyn and Nini, 2020; Tagliamonte and Baayen, 2012). This difference in model fits is visible in the current study.

7.2.1 Models for all data

The Variable Importance ranking (VIR) generated can be seen in Figures 12 and 13.

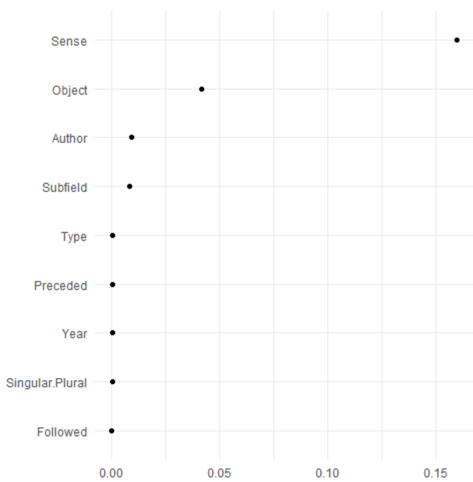

Figure 12: VIR all senses

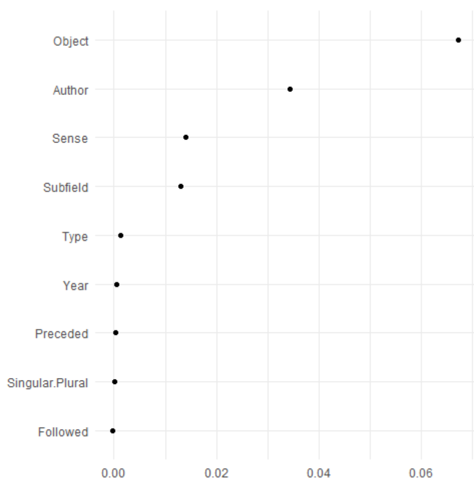

Figure 13: VIR relevant senses

The VIR shows that the ‘sense’ is the most important predictor when all senses are considered, followed by ‘object’ and ‘author’. When only the relevant senses are considered, ‘object’ ranks most highly, followed by ‘author’ and then ‘sense’. This is an interesting discovery since, as is mentioned in the methodology section, there is some nesting when it comes to the ‘subfield’ and the ‘object’. Regardless of this nesting, ‘object’ seems to be more important than ‘subfield’ when it comes to predicting whether *mass* or *weight*, both for all senses and only the relevant senses.

Since ‘sense’ ranks more highly in the VIR for all senses than the VIR for the relevant senses, this suggests that the senses COL, MET, Q and N contribute to this importance more than M and W, and the addition of the other senses to the model makes ‘sense’ a much more powerful predictor. While the VIR is opaque when it comes to which labels cause these rankings, the Ctree can give more insight, as shown in Figure 14.

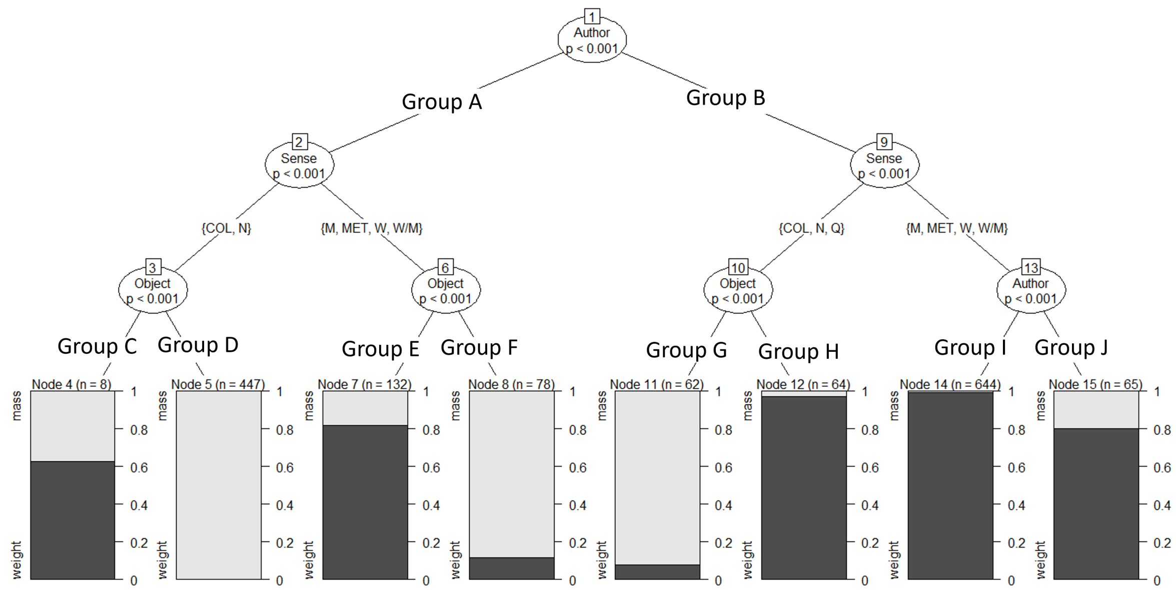

Figure 14: Simplified Conditional Inference Tree of all annotated data

The groups shown in the Ctree in Figure 14 are available in Appendix A. The Ctree confirms the idea that MASS and WEIGHT are not split by the model, but rather that there is a contrast between N, COL, and Q in particular and the relevant senses for this study, W and M. If one were to look only at VIR all senses, it is tempting to suggest that that CE was successful, since it is the sense of the word that governs which word is chosen. However, since MASS and WEIGHT are not the driving forces behind this predictive importance, this suggestion is not borne out.

‘Sense’ ranks third in the VIR for the relevant senses only, and since only the senses MASS and WEIGHT were included in this mode, it must be the case that there is some predictive power in these senses. Indeed, this is the case, since the p-value for this node is <0.001 (Ctree available in Appendix E).

The high ranking of ‘author’ in both models suggests that individual authors influence which sense is used with which word. Rather than becoming a standard usage, *mass* and *weight* seem to be subject to large amounts of individual variation. If these individuals can choose which sense and word they want to use, then the idea of an engineered improved concept that is used throughout the community to improve scientific progress has not been achieved.

‘Object’ and ‘subfield’ provide micro and macro perspectives on the topic of the article and lemma respectively. Both are found to be important in both models, suggesting variation between disciplines and between topics under discussion. Again, variation based on these factors suggests that consistent differentiation based on sense has not been achieved.

‘Year’, surprisingly, has no predictive power according to this model. The factor ‘year’ shows whether time helps predict whether *mass* or *weight* is chosen. If *mass* is replacing *weight* in the sense MASS, then the likelihood of *mass* should go up in a function of time. However, the data suggest that such a temporal effect is very weak if it occurs at all. It is possible that ‘year’ still interacts with *mass*, but that it was not included in this set-up.

The POS that precedes or follows the tokens, the type of text, and plurality are ranked low on the VIR. Only the POS following the lemma appears in any of the Ctrees. They will thus receive no further consideration.

In sum, the models created on the entirety of the dataset show that ‘sense’ is important in the model containing all senses, but that it had lesser importance when it came to the relevant senses. Other factors that rank highly are ‘author’, ‘object’ and ‘subfield’.

7.2.2 Time period specific analysis

At first glance, it appeared that time ('year') had little predictive importance. The general modal does not, however, indicate whether there may have been any smaller shifts in factor importance across time. To more effectively explore how *mass* and *weight* changed over time, a closer examination of each period is required. This investigation takes intervals of 50 years and looks both quantitatively at the C-trees and Random Forests created for those years, and qualitatively at any sentences that are of interest.

This analysis will mainly be based on the VIRs for each period, for both relevant senses and all senses, which are shown in Figure 15. These time-specific analyses only cover the periods 1800-1920. In 1700-1750, there is one token of *mass* in contrast to 16 instances of *weight*. There are, therefore, too few tokens to be subject to multifactorial statistical analysis. In 1750-1800, there are only 3 instances of *mass* opposed to 123 of *weight*. While the total number of instances is reasonable, the data is unfortunately affected by extreme class imbalance, namely that there are significantly more of one token to predict than another. This means that the model is likely to predict the lemma *weight* in all cases due to the high percentage of accuracy it would deliver. A Ctree or Random Forest is thus too reliant on the class imbalance, even with a weighting. This might be due to the sampling technique, or simply because there were very few tokens in this time period regardless of sampling. For these reasons, the discussion below will apply only to the periods 1800-1850, 1850-1900, and 1900-1920.

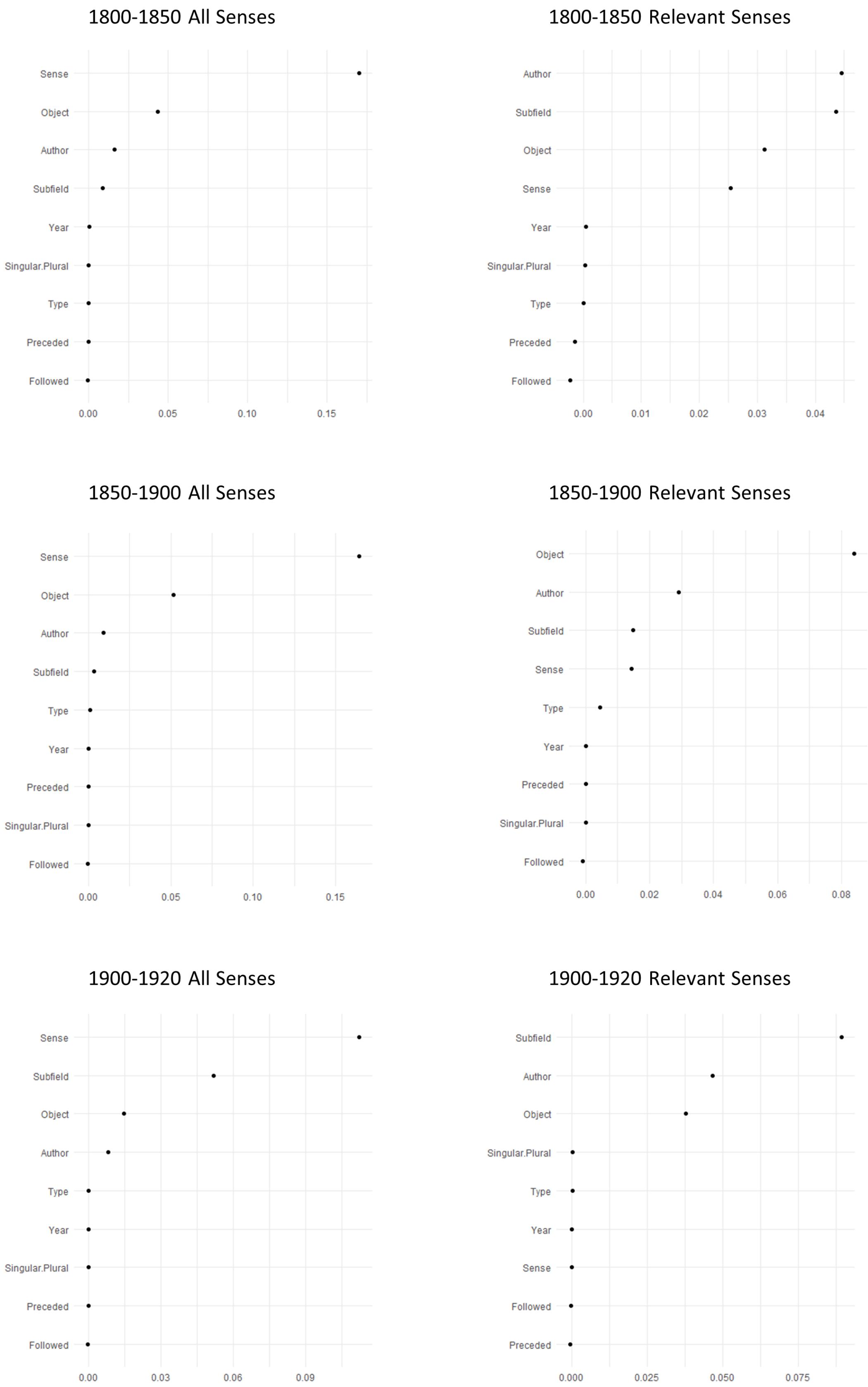

Figure 15: VIRs for each period

7.2.2.1. All senses

As can be seen from Figure 15, in the models containing all senses, ‘sense’ remains by far the most important variable in all three time periods. However, as was the case in the models generated for the entire dataset, this split is based on the ‘senses’ N, Q, COL versus W and M. This can be seen in the Ctrees generated for each period, of which one (1850-1900) is shown in Figure 16. Nodes 2 and 9 split the data based on ‘sense’, but base this split on N, Q, COL versus W and M. The Ctrees for the other periods show the same groupings, so will not be shown here for reference. They can be found in Appendix F. Full details of the groups in figure 16 are shown in Appendix B.

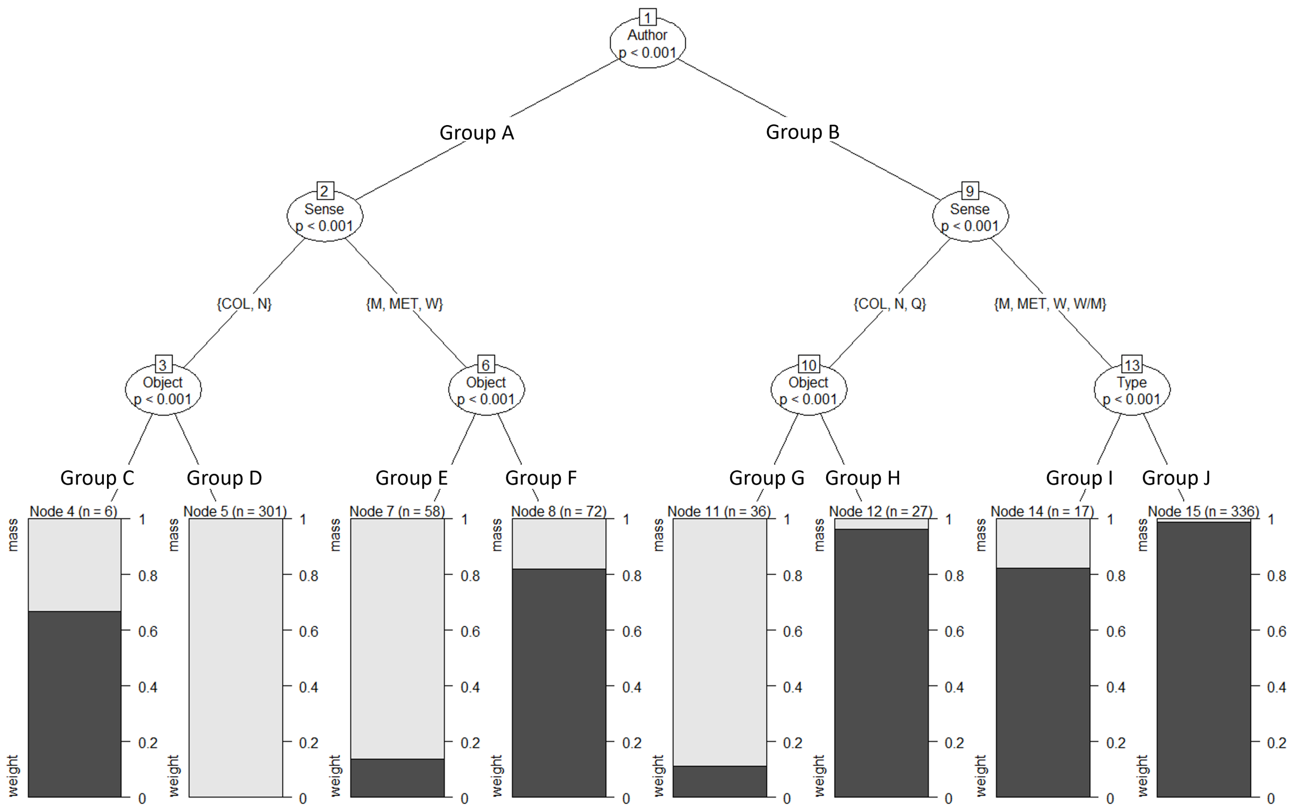

Figure 16: Ctree of all senses between 1850-1900

This supports the observation made from Figure 14 that N, COL and Q (and MET) contribute to the high ranking of ‘sense’ in the VIR and that there is less differentiation between W and M. This stands in opposition to what we would want in a CE project; we want the ‘senses’ W and M to become more differentiated over time, and yet this is not seen in the data.

The second and third most important factors in the VIR do fluctuate over time. In 1800-1850 and 1850-1900, they are ‘object’ and ‘author’, and in 1900-1920 they are ‘subfield’ and ‘object’. For explicit discussion of these, see section 7.2.2.2.

7.2.2.2. Relevant senses

Turning now to the models containing only the relevant ‘senses’ W and M, much more variation is found. Between 1800 and 1850, ‘author’ is the most important, followed by ‘subfield’ and ‘object’. Between 1850 and 1900, ‘object’ is most important, followed by ‘author’ and ‘subfield’, and between 1900 and 1920, ‘subfield’ is the most important, followed by ‘author’ and ‘object’. Each period, therefore, has a different factor as the most important, but the same three factors are consistently found important in each period.

‘Author’ always ranks in the top three in each period, which shows that individual preference or choice governs lemma choice more than the ‘senses’ of the words themselves. The fact that ‘author’ is ranked more highly than ‘sense’ suggests that the two-way distinction between *mass* and *weight* is only applicable to certain authors. ‘Author’ is ranked as the most important factor in 1800-1850, and as second for the other periods. This suggests that individual choice has become a less crucial feature in predicting lemma choice, and thus that some more regularity emerges between authors as time goes on. However, ‘author’ remains ranked second, and thus this idiosyncratic author choice still governs lemma choice, which stands in contrast to what we would expect if the CE project was successful.

To exemplify this, node 1 in Figure 16 above makes a split based on ‘author’. Group A consists of Hofmann, Griffiths & Glazebrook, Abel, Bower, Darwin, and Henessy among others. These authors have little in common, embodying a number of disciplines including Chemistry, Botany, Astronomy and Physics, and numerous different nationalities, including German, Irish, and British. Group B consists of Brodie, Heycock, Mendel, Hodgekinson, Reed, Reed & Stokes, Sabine, Airy, and Roscoe among others. Again, there is a wide variation of disciplines and nationalities in this group. This suggests that lemma choice goes beyond simply belonging to a certain discipline or nationality. For full author lists for each group see Appendix B.

A good indicator that an author utilises the difference is the usage of both lemmas in the same sentence. Example (5) shows the use of both *mass* and *weight* in the same sentence. Only the token *weight* with 'sense' WEIGHT was included in the analysed sub-corpus, but on closer inspection, *mass* here is used with the 'sense' MASS. This shows that the same author can use the same word with different 'senses' within the same text.

(5) *To this we believe there is only one answer. The motive power is supplied by the weight of the mass.*
(Anderson, 1903)

The fact that 'subfield' is so high suggests some differentiation within the different scientific sub-communities. 'Subfield' seems to have gained importance over time, being second in 1800-1850, third in 1850-1900 and first in 1900-1920. This could suggest that the specific scientific subdomains are slowly gaining more power over the choice of lemma. The increasing importance of 'subfield' implies that the different disciplines within science are becoming more differentiated in terms of the language they use. Put more simply, each discipline could be creating its own specialised register.

Intuitively speaking, the differentiation of *mass* and *weight* is most crucial in the domains of Physics, Astronomy and Engineering. This is because there are certain equations, ideas, and approaches in these fields that require a differentiation between the two 'senses'. Certain scientific sub-communities may find the *mass/weight* distinction communicatively helpful and looking from a functionalist viewpoint, this would suggest these specific 'subfields' would be the most consistent in their usage of the 'senses' and lemmas.

(6) *the unit of force being that which, acting on the unit of mass, through the unit of time, generates in it the unit of velocity.....For the unit of mass, then, we may take a grain*
(Sabine, 1843)

In Example (6), Sabine distinguishes between MASS and WEIGHT, explicitly stating that force acts upon the mass of an object. Since he uses Newton's laws of motion, this differentiation is required to calculate velocity. One could therefore say, that due to the topic he is writing about, the correct use of *mass* and *weight* is not a choice but a necessity to reach the mathematical conclusions he wishes to reach. This introduces an interesting dimension to the study since to discuss these things and further academic research on the topic, one must accept and use Newton's lexicalisations. This is similar to the need to use newly created terminology to discuss a new concept. If one wants to ensure that a CE project spreads, creating a context in which the distinction is crucial seems to be a successful way to ensure success.

One could, therefore, expect that in periods where 'subfield' has a high contribution to lemma prediction, these disciplines form a distinct group in opposition to the other disciplines. However, since the workings of a Random Forest are opaque, it is not possible to see which topics are using which words more frequently. For this, it is necessary to turn to the Ctrees. The Ctree for 1900-1920 is shown in Figure 17.

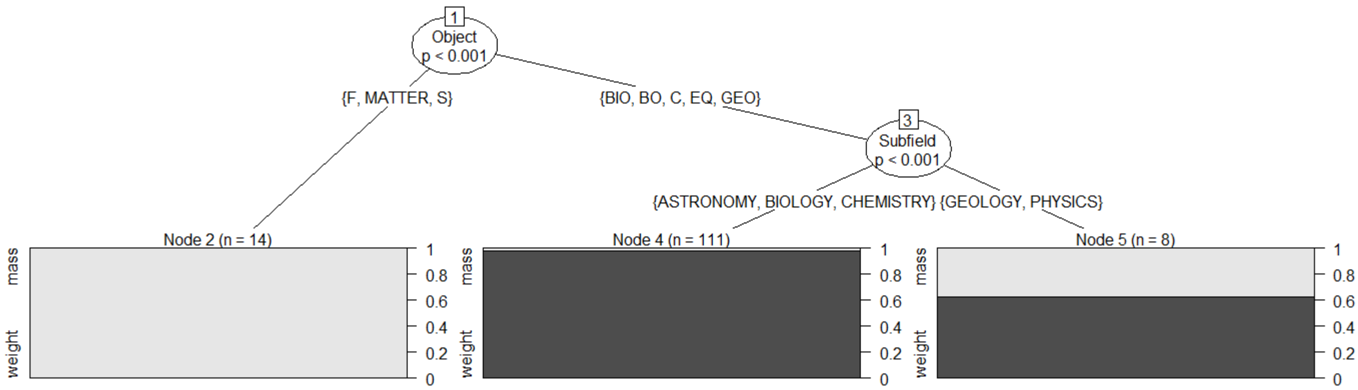

Figure 17: Ctree of relevant senses between 1900-1920

The split seen in node 3 for 'subfield' is of particular interest. Astronomy, Biology, and Chemistry are grouped in opposition to Geology and Physics, which is an unexpected division. Both Astronomy and Physics are more reliant on the *mass* and *weight* split for their analysis, whereas Biology, Geology, and Chemistry (usually) do not need this distinction. From observing the patterns during the annotation, Biology and Chemistry usually consistently use *weight* with the 'sense' M and often use *mass* with the 'sense' N, whereas Astronomy and Physics more often use these words with their respective 'correct' senses. However, the Ctree shows that Astronomy, Biology, and Chemistry are more likely to always use *weight*, whereas Geology and Physics use *weight* and *mass* more equally. This binary split is therefore unexpected.

'Object' is ranked highly across all models, but this may be because of the nesting with 'subfield'. Certain subfields are more likely to discuss certain objects, and so the high ranking of 'subfield' in this model may have affected the ranking of 'object', meaning its ranking must be taken with scepticism.

However, in 1850-1900 'object' ranks higher than 'subfield', meaning that this micro-measurement of topic was more important than the overall sub-discipline that the lemma is used in. While in 1900-1920 'object' does not rank more highly than 'subfield', it does appear higher up the Ctree, which implies greater importance in splitting the data in that single model (the VIR is created through a collection of many Ctrees, so this single tree does not have to correspond with the rankings found in the VIR). We see in Figure 17 that the objects BIO, BO, C, EQ, and GEO lead to the split between Astronomy, Biology, and Chemistry versus Geology and Physics. The data is thus first split by 'object', and then by 'subfield'.

While the object is an important conditioning factor, there does not appear to be any consistency over time concerning which objects particularly condition which lexical item. The Ctrees are shown to split the data based on different objects in each period. For example, Figure 16 shows that the next period, 1850-1900, splits 'object' differently. Node 2 splits EQ, F and S from ABS, BIO, BO, C, FOR, GEO, MAT, and WO. Nodes 6 and 10 have splits based on different objects. Other periods, for example 1800-1850, split 'object' based on CONCEPT, GEO AND MAT versus BIO, C, EQ, FOR, GEN, S, STAN VM and WO. No single 'object' always splits from the others in these Ctrees. It is therefore very time-dependent which 'objects' condition which terms. This goes against the intuition that certain 'objects' may become more associated with certain words and senses through frequency effects; if certain combinations appear more frequently, they will become more entrenched in the minds of speakers. However, no combination seems to become entrenched. Usage is in such constant fluctuation that consistency is not achieved.

The VIR created on the entire dataset did assign some significance to 'sense' in the model containing the relevant senses, but this is not reflected here. However, this does not mean that there was no explicit discussion about how the words should be used. In Example (7) Crookes is making the statement that he will be using the term *weight* to refer to the 'sense' MASS.

(7) *The weight in grammes is another name for the mass.*
(Crookes, 1873)

Rather than simply using the term *mass* in this case, Crookes has specified that he will be using one term to refer to the other. Crookes explicitly states what he will be using the term *weight* to convey, leaving none of the ambiguity that the CE project was designed to combat.

Brodie's 1866 paper provides an explicit discussion of his ideas about the meanings of *mass* and *weight* in the first part of his paper, which is devoted to the definitions he uses. The quote below is long but interesting concerning the present investigation.

> *"A weight" is a portion of ponderable matter of any specified kind considered as regards weight. This application of the term a weight is only a slight extension of its ordinary use. A gramme of platinum which serves for the purposes of weighing is termed a weight, this being the only property of that portion of matter of, which it is necessary to take cognizance*. Now the weight of matter, from a different and far wider point of view, in the laws of its composition and resolution, is the special subject of this investigation; and the aspect in which every chemical substance, every portion of ponderable matter, will be here regarded is exclusively as a weight. In speaking of such weights we habitually employ, by a tacit convention, the terms by which the chemical substances, of which the weight alone is referred to, are usually designated. But this is not a strictly accurate use of language; and it is necessary to observe that in the following pages, where chemical substances, such as chlorine or alcohol or water, are mentioned, or the term "a portion of matter" is employed, the objects referred to are certain weights of the sub-stances under consideration, to the exclusion of all other properties.*
> (Brodie, 1866)

In this passage, Brodie discusses how he uses the term *weight* in a nominal sense, referring to a "portion of ponderable matter", something which would usually require the 'sense' MASS. He uses the word *weight(s)* to refer to both the nominal sense and the 'sense' MASS within the text, and states that the former is "a weight" while the latter speaks of "the weight of matter". He states that this is not a "strictly accurate use of language".

In a footnote below this passage, Brodie goes on to state the following.

> The term weight is, in ordinary language, used with two distinct meanings. (1) In what may be termed its abstract signification, as denoting a certain measurable property of matter, as when we inquire, "What is the weight of that loaf?" (2) In its concrete sense, as denoting certain objects, which we discriminate from others by naming them from their most essential property, as when we say "Bring me that box of weights." It is in an extension of this concrete meaning that the word is here employed.
> (Brodie, 1866)

Brodie here discusses the ambiguity of the term *weight* and describes the 'sense' N alongside the sense that denotes "a certain measurable property of matter". He still uses the lemma *weight* for this definition, while we know from the previous discussion that he is referring to the MASS rather than the WEIGHT.

The 'sense' MET does not occur with any particular frequency, and so the frequency analyses from section 7.1 do not shed much light on the metaphorical usage of the two lemmas. However, the usage of both these terms in metaphors is a particularly interesting case of growing polysemy that is theoretically very interesting when it comes to investigating CE. Example (8) shows the sole instance of *mass* being used with a metaphorical sense.

(8) *I may have assigned too great a mass to the Doubt* Some idea may be formed of the amount of these deflections*
(Pratt, 1855)

This kind of construction would occur more frequently with *weight*, such as "What I more especially lay weight upon is this" (Hermann Kopp, 1865). More instances use *weight* in this construction, all between 1822 and 1894, and this irregular metaphorical usage of *mass* can be found right in the middle of this period of high usage of *weight* with this sense. This may potentially be an instance of a kind of transfer; since *weight* was being used in this way frequently, and there were many contexts where *mass* and *weight* are used interchangeably, the concepts may have started to be perceived as linked (or at least by this author). Changes affecting *weight*, namely the increase of metaphorical usage, are spread to *mass*, potentially in a network-like fashion, in a similar way to De Smet (2017). This is an interesting area of future research to be done with the entirety of the RSC corpus.

7.2.2.3. Summary

Returning to the hypotheses, we can see that 'sense' is not the most important variable when it comes to the models containing the relevant senses only. In the models containing all senses, 'sense' is the most important, but only due to a clear separation between N, COL, and Q as opposed to W and M. Instead, extra-linguistic factors 'author', 'subfield', and 'object' have the most influence. This part of the data analysis points to the CE project being unsuccessful.

7.3 Principal Component Analysis

The results of the multifactorial analysis can be supplemented through the PCA. In the hypotheses, it was said that if the CE project was successful, then we would expect visual separation of tokens by 'sense' rather than by other external factors. The VIRs and Ctrees showed that especially where the relevant senses are concerned, 'sense' is not the most important predictive factor. The PCA aims to determine how different each lemma is in each period rather than investigating the conditioning factors of lemma choice. PCA may show that they also become distributed more regularly over time, since PCA is a dimensionality reduction method, and so it becomes easier to visually interpret changes over time.

Once again, the earlier periods do not have enough tokens to show any interesting groupings or clusters, so the PCA is restricted to the later periods.

Figure 18 shows that between 1800 and 1850 there is little regularity in the distribution. Rather, both *mass* and *weight* have a random distribution over the space. 1850-1900 is a turning point, and from there on we begin to see organised clusters for both *mass* and *weight*. To find out what is causing the clustering, the clusters can be analysed in terms of their annotations to find the principal components, or the annotations that influence the clustering.

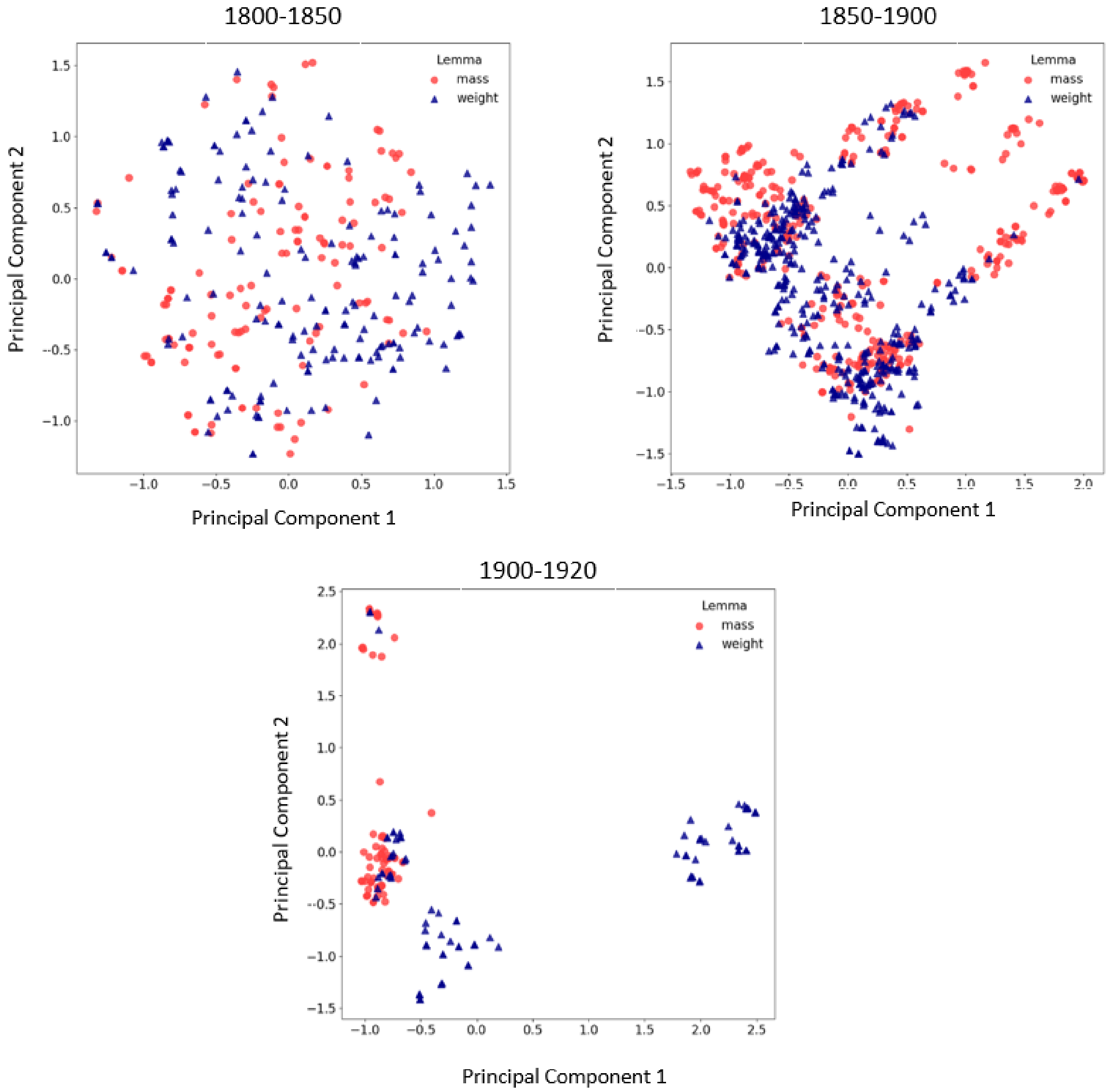

Figure 18: PCA across all time periods

The first time period analysed was 1850-1900, and Figure 19 visually clarifies these clusters. From Figure 19, it is clear that the clustering is not based on lemma, since both *mass* and *weight* tokens occur in each cluster. The clusters identified through the coordinates of each token are shown in Tables 10 and 11.

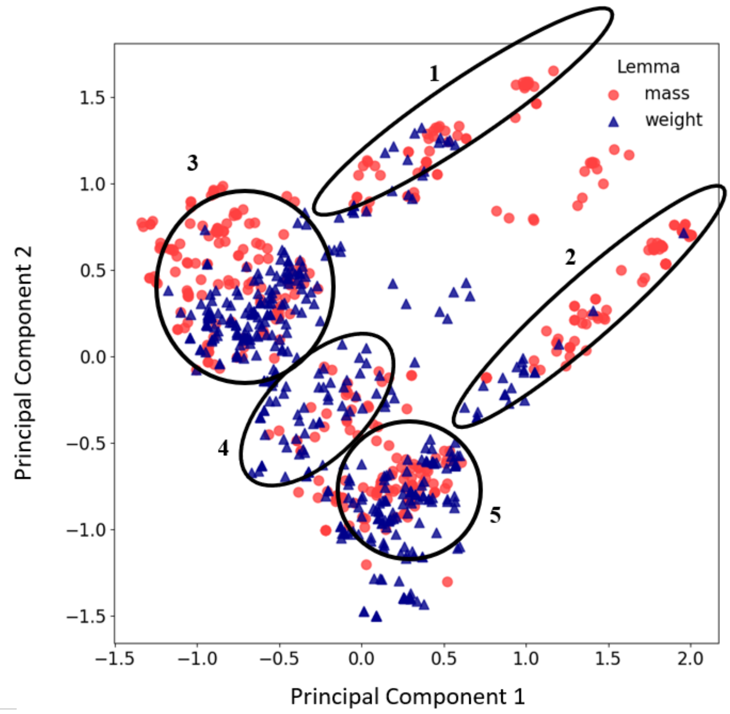

Figure 19: PCA clusters 1850-1900

Clusters 1 and 2 are very similar, and it is possible to infer that the positive values of principal component 1 are associated with the 'subfield' Biology, the 'object' BIO, the 'sense' N, and a preceding adjective. What differentiates clusters 1 and 2 is the type of text and the authors that contributed tokens to the clusters.

	Cluster 1	Cluster 2	Cluster 3
Subfield	Biology	Biology	Geology, Meta
POS before	JJ	JJ	DT
POS after		IN	IN
Plurality		Singular	Singular
Sense	N	N	N
Object	BIO	BIO	GEO, CONCEPT, GEN
Lemma	*Mass*	*Mass*	
Type	fla	article	fla
Author	Davy, Beale, Clarke, Parker	Parker, Bower, Williamson, Watney, Marcet	Tyndall, Hopkins, Pratt, Thomson, Hennessy, Brodie

Table 10: Clusters 1-3 in the 1850-1900 PCA

Cluster 3 is associated most strongly with Geology, but also Meta. The lemmas are frequently preceded by a determiner and followed by a preposition, are singular in form, and have the ‘sense’ N. In terms of the ‘object’, GEO, CONCEPT AND GEN are the most frequent within the cluster. The authors associated with it are John Tyndall, William Hopkins, and John Henry Pratt, amongst others.

	Cluster 4	Cluster 5
Subfield		Astronomy, Engineering
POS before	JJ?	
POS after		VBD
Plurality	Singular	Singular
Sense	M?	M
Object		VM, FOR
Lemma		*Weight*
Type	abstract, paper-read	article
Author		Abel, Crookes, Joly & Fitzgerald, Tomlinson & Adams, Reed & Stokes, Mendel, Herschel, Gray, Henderson & Kelvin, Thorpe & Roger, Mallet

Table 11: Clusters 4-5 in the 1850-1900 PCA

Cluster 4 is remarkable in that it does not contain any annotation with a particular frequency, apart from the types ‘abstract’ and ‘paper-read’, and the singular form.

Cluster 5 is more homogeneous, characterising the ‘subfields’ Astronomy and Engineering, with the item in question frequently being followed by a third-person singular present verb. The tokens typically have the ‘sense’ MASS, but the lemma *weight*, and appear with ‘objects’ VM and FOR. In terms of ‘author’ and ‘type’, they occur in articles written by F.A. Abel, William Crookes, and D. Mendel, amongst others.

The PCA for this period, therefore, identified ‘subfield’ as the most important factor in the creation of the clusters. Furthermore, ‘author’ is also influential, especially in clusters 1 and 2, where most other variables are the same, except for ‘author’. ‘Sense’ has some influence, but only N and MASS seem to affect clustering.

Turning now to 1900-1920, there are fewer tokens but there are still 4 visible clusters, as shown in Figure 20 and Table 12. Compared to the previous period, the clustering starts to separate the lemmas *mass* and *weight*.

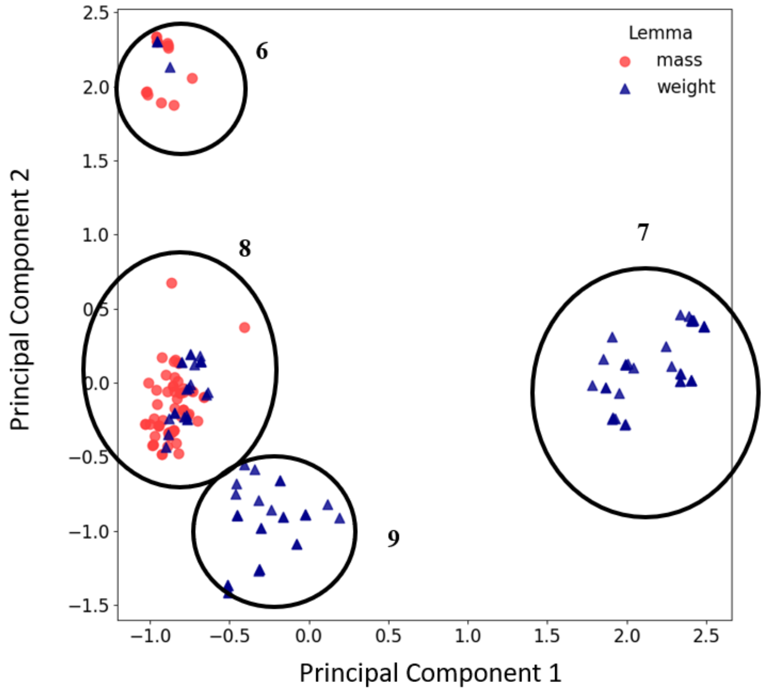

Figure 20: PCA clusters 1900-1920

As seen in Table 12, cluster 6 consists of the ‘subfield’ Geology, where the item in question is proceeded by a determiner, and used with ‘sense’ N. The ‘object’ is usually GEO, and the authors are Anderson & Flett.

Clusters 7 and 9 are remarkably similar. Both are based on the ‘subfield’ Biology, where the object is BIO, and the sense MASS but the lexical item *weight*. The differences lie in the preceding POS, but also ‘object’ since cluster 7 contains ‘object’ BO along with BIO. Furthermore, tokens from cluster 9 are more likely to be experimental texts. Finally, the authors differ, with cluster 7 containing tokens from Dreyer, Ray & Gotch, and Dreyer, Ray, Walker & Gotch. Cluster 9 contains tokens from Matthaei & Darwin, Thoday & Blackman, and Halnan, Marshall, Yule & Langley.

Finally, cluster 8 consists of tokens from the ‘subfields’ of Chemistry, Astronomy and Physics. They are also frequently followed by an adjective, preposition or the verb *be* in the third-person present singular. The ‘sense’ is usually N or W/M even though the lemma is usually *mass*. Lectures frequently appear in this cluster, and some notable authors are J.H. Jeans, G.H. Darwin, and A.E.H. Tutton.

	Cluster 6	Cluster 7	Cluster 8	Cluster 9
Subfield	Geology	Biology	Chemisty, Astonomy, Physics	Biology
POS before	DT	NN	JJ, VVF, IN	SENT
POS after			JJ, IN, VBZ	JJ
Plurality				
Sense	N	M	N. W/M	M
Object	GEO	BIO, BO	F, S, C	BIO
Lemma		*Weight*	*Mass*	*Weight*
Type			Lecture	Experiment
Author	Anderson & Flett	Dreyer, Ray Gotch, Walker	Jeans, Darwin, Hardy, Tutton Haycock & Neville, Manley & Poynting,	Matthaei & Darwin, Thoday & Blackman, Halnan, Marshall, Yule & Langley

Table 12: Clusters in the 1900-1920 PCA

Again, the PCA has identified 'subfield' and 'author' as influential when determining clusters. 'Sense' has some influence, but again only N and MASS seem to affect clustering.

7.3.1 Summary

In sum, the PCA analysis shows similar results to the multifactorial analysis. Despite clusters starting to form from 1850 onward, these are largely based on language-external factors such as 'author' and 'subfield' rather than 'sense', or even the lemmas themselves. Some clusters are showing some differentiation based on 'sense'. In 1850-1900 this is usually between N and MASS, but from 1900 on WEIGHT becomes more important.

CHAPTER 8

DISCUSSION

Now that we have a picture of the quantitative and qualitative features of the corpus, it is possible to look for patterns, conclude how successful this instance of CE was, and evaluate the applicability of the current method to other CE projects in the future.

The research questions addressed by this work are the following:

RQ1. To what extent are *mass* and *weight* successfully differentiated by sense after Newton proposed separating the concepts in 1687?

RQ2. What time frame and trends are there in the diffusion of *mass* and *weight*?

RQ3. What effect do extra-linguistic variables such as author and subfield have on the choice of lemma?

8.1 Patterns in frequency and senses

From the results, it appears *weight* continues to be used more frequently to refer to MASS. There is a brief period at the start of the 19th century where *weight* is used most frequently with WEIGHT, and since *mass* is never used with sense WEIGHT, it could be considered successful in this fleeting period. However, this was not maintained, and the usage with sense WEIGHT sees a sharp decline after this time. This goes against what the CE project was striving to achieve and suggests that it did not reach completion or successful implementation.

Research question 1 asked whether *mass* and *weight* were successfully differentiated by sense, but the frequency data shows that this did not happen. Furthermore, research question 2 aimed to find time frames and patterns in the diffusion, but this study found that even after 220 years the lemmas were still not used with their corresponding senses. In sum, this data shows that the CE project was not unsuccessful, and the question remains as to why.

Since N remains the most prominent sense of *mass*, it is possible that this established usage of *mass* with nominal sense (N), which pre-dates the usage with sense MASS, provides resistance to the allocation of MASS to the same lexical item.

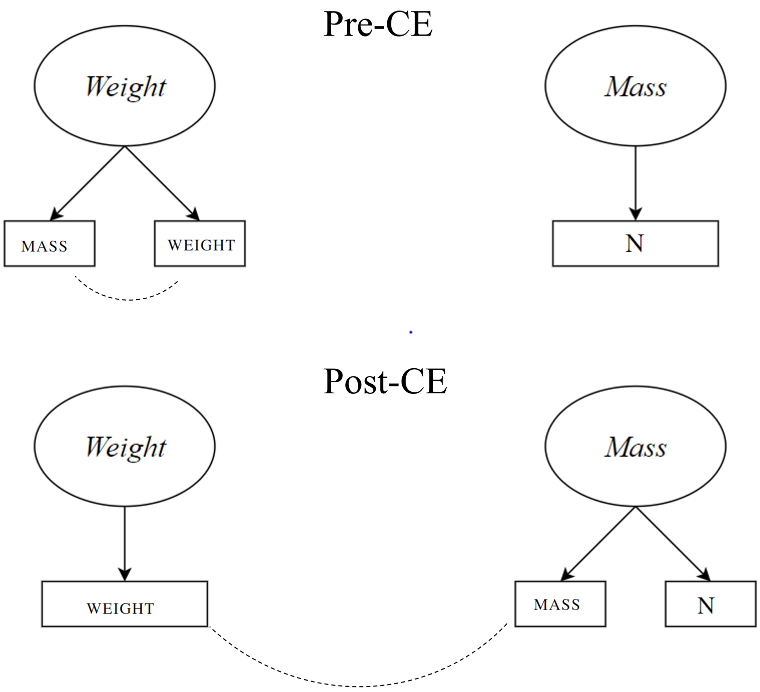

Figure 21: Idealised visual representation of the sense changes

As can be seen from Figure 21, the word *weight* initially had two related but distinct senses, namely MASS and WEIGHT, in a way that is reminiscent of polysemy. After the re-allocation of sense MASS to *mass*, this led to two unrelated senses, namely MASS and N, belonging to the same lexical item. These two senses are not as closely related as MASS and WEIGHT, but do coexist in the same lemma, in a way that is reminiscent of synonymy.

The sense WEIGHT never belonged to the lemma *mass* at any stage in the process, which may explain the one-sided use of *weight* to refer to both MASS and WEIGHT, but *mass* never referring to WEIGHT.

As was mentioned earlier in this thesis, a project that aims to reduce ambiguity in one area may result in increased ambiguity or polysemy in another place in the lexical network, and this is indeed what we see in the case of *mass*. Synonymy and polysemy are not uncommon and occur frequently in words such as *bank* or *head* but in this case, it has been transferred from one lexical item to another.

One proposed reason for the lack of success in this project is the conflict between the two senses now contained within *mass*. Rather than the two senses being closely related, they are further apart semantically. Regardless, they are still often used together within the same text, since the sense N is part of a formal, academic register, in contrast with the informal 'stuff', 'thing', 'object'. This means there is a high rate of co-occurrence in the same texts or registers. Rather than *weight* occurring in scientific texts with two closely related meanings, under Newton's proposal, *mass* would occur in scientific texts with two less closely related meanings. There would therefore be synonymy in *mass* rather than polysemy in *weight*. However, since the project was unsuccessful, the polysemy was maintained rather than creating new synonymy.

Newton did not create a new lexical item for the concept MASS but rather transferred this sense to another existing, established word with existing, established senses. These competing senses within the same lexical item may therefore have led to resistance to the CE project. However, it is important to reflect on whether synonymy and polysemy are legitimate blocking forces within CE. Synonymy and polysemy are rife in language, and in most cases, it is not a problem. To investigate this notion more, it is worth looking at another instance of CE.

Some projects, such as Haslanger (2000), choose to keep the lexical item but change the sense associated with it. In this case, the lexical item *woman* remains the same, but the meaning associated with it changes. Other projects such as Scharp (2013) ameliorate a lexical item by moving certain meanings or senses it may have to another lexical item. This lexical item may already exist, as is the case in *mass/weight*, or a new lexical item may be created altogether.

A case of the latter is explained by Baron (2020), who details the evolution of a gender-neutral, inclusive pronoun. Certain new lexical items have been put forward to fulfil this role, including *shee, hasher/himer/hiser, idn, e/es/em, ir/e/thon, one, ve/vim/vir, ze, han/hans/hanself, sim, ti, le/lis/lim*. These new lexical items were created because some people were uncomfortable using an existing word (such as *they*) with an existing function for this new gender-neutral concept. This mirrors the idea that *mass* with sense MASS was not successfully implemented due to discomfort with using an existing word. None of these innovations have survived into everyday usage. Instead, this function has been fulfilled by *they*, where the third-person pronoun can be used as a second-person pronoun where the gender identity of an individual is unknown, or they prefer to be referred to in neutral terms.

The lexical item *they* is also closely related to the previously existing *he/she*, being part of the same pronoun paradigm. This lends doubt to the idea that it is the close relationship between the pre-CE lexical item and post-CE lexical item and sense distribution that creates conflict in the general acceptance of a CE project.

In the case of pronouns, an existing term was eventually selected over a new term. The inverse occurred in the instance of *mass/weight*, where the transfer of the sense to a different existing lexical item did not work successfully. Generally speaking, appealing to the notions of polysemy and synonymy is not (always) applicable. However, the constraints imposed by synonymy and polysemy may depend not only on the lemmas and senses involved but also on the language ideologies of the communities they affect.

If we were to suggest that polysemy and ambiguity were at play in hindering or promoting conceptual change, we need to reflect on what makes the scientific community, and scientific language, different from regular language. One way to analyse this difference may be through the usage of language ideologies.

The case of pronouns had a different target domain than the case of *mass/weight*. Instead of targeting the scientific domain to increase precision, it targeted the social domain with the intention of social activism. These two different aims may lead to different constraints and strategies.

The general population, or the social domain, is much more diverse than the scientific community. With the varying backgrounds, ideologies, and goals within the social domain, it is easy to imagine it as Bakhtin's Carnivalesque, or a "world of multiple and unregimented voices" (Gal, 2019, pg 452). Precision is not necessarily the most important goal and everyday language is full of ambiguities, along with creative uses of language, such as rhetoric and poetry. Language in the real world is not precise, and we often do not say what we mean, as is explained by theories such as Gricean implicature (Grice, 1989). Language is therefore not (always) seen as a problem, but rather a resource or right (Ruíz, 1984). Furthermore, the specific project of a gender-neutral pronoun is not driven by scientific progress, but rather social inclusion, and to reach this goal ambiguities or polysemy may be accepted since the overall positives outweigh the overall negatives. The social goal we set out to achieve has been attained, and society simply must deal with slightly more ambiguity, which is already rife.

Scientific communication, however, does not work on the same principles. The Habermasian view of society as reason-driven, universalist and deliberate is more likely to apply to the scientific community than to the general population. One can imagine that a debate can be won through "the power of the better argument" (Habermas, 1996, pg 54), even if this may be idealistic. These communities are also groups of "unregimented voices", but perhaps groups where reason has a slightly stronger hold. Furthermore, in these groups, it is easy to see how language can come to be thought of as a problem (Ruíz, 1984) since any imprecision or ambiguity in terminology can lead to problems in communicating complex ideas, replicating studies, or incorrect teaching. An increase in ambiguity elsewhere in the lexical network once a concept has been conceptually engineered therefore means that the goal has not been fully attained; it is simply another word now that has the issue of ambiguity or imprecision. Instead, one can imagine that the ultimate goal, as expressed by Bacon, Locke, and Leibniz, is linguistic isomorphism, which is a perfect form-meaning mapping.

Scientific communities may prefer to create new words rather than re-purpose existing terms, since this allows for a singular form-meaning mapping, as can be seen by the extensive creation of new terminology within the scientific domain. Nevertheless, this author could not find any research comparing the rate of terminology creation in the scientific domain with any other domain.

If the goals of the scientific community mean that the transfer of sense MASS from *weight* to *mass* is non-optimal, the question remains as to why it appeared to be a 'better' solution to continue using *weight* with both senses MASS and WEIGHT. The costs of ambiguity in this domain do appear to outweigh ambiguity between MASS and N, and yet this is not what the data shows. The terms of success and the constraints placed on a specific project depend on the goals and ideologies of the communities in which the change is intended to be spread.

For these reasons, it may be the case that synonymy and polysemy are stronger driving forces in scientific language than in everyday language, but this will require further investigation. The power of synonymy and polysemy to block or inhibit conceptual change will be crucial to future projects of CE.

8.1.1 Transfer in the metaphorical domain

In a project that aimed to decrease ambiguity, it is interesting that new ambiguity arose in new places. In 1855, Pratt used *mass* with a metaphorical sense, something which is usually restricted to *weight*. While this is an isolated instance within the annotated corpus used for this study, it suggests that another sense traditionally associated with *weight* has been transferred to *mass*, along with the sense MASS.

(1) *I may have assigned too great a mass to the Doubt* Some idea may be formed of the amount of these deflections*
(Pratt, 1855)

A change that moves a single sense from one lexical item to another may therefore have wider implications than we may initially believe. To speak metaphorically, it is as if a river was dug from one lake to another, but now we cannot stop the water flowing freely between them. This does not appear to be the goal of CE, which seems to want a one-time permanent transfer rather than a continued relationship.

8.1.2 Factors in selection

While the analysis of the frequencies over time helped partially answer research questions 1 and 2, the rest must be addressed through reference to the statistical analyses. It was hypothesised that if the CE of *mass* and *weight* was successful, then the Random Forests and Ctrees will show 'sense' to be the most important factor, but if it was unsuccessful, then extra-linguistic factors such as 'author' or 'subfield' have a higher importance than 'sense' when splitting the data.

The most unexpected result from these statistical analyses is the lack of influence of 'year' in the model based on the entire annotated data set. If Newton's CE project was successful, we would expect to see a separation of the senses MASS and WEIGHT over time, especially in the models created on the entirety of the data, therefore making the period from which a token is sampled an important variable when deciding which sense a lexical item is likely to have.

This raises the question of time frames within CE projects. When someone engineers a term for amelioration and intends for it to be spread beyond a single individual, what time frame do they have in mind for the implementation? Idealistically speaking, they would want their new concepts to be successfully spread as quickly as possible since the aim of improving these concepts in the first place is to solve a problem. If the new concept is not used, the problem is not solved. However, we know from linguistic research that semantic change often takes a very long period to diffuse, and it is rare that meaning changes overnight or in a short period. It is possible but would take extreme circumstances. Take the word *quarantine*, for example, the meaning of which radically changed in a short period during the Coronavirus pandemic. Conceptual engineers must therefore be prepared for their proposals to take time to spread.

In the data examined in this study, it is possible that the period looked at was too small to track the change to completion. Only 220 years of data were examined, and linguistic changes may require more time than this to spread. Once the data from 1920 onwards becomes available, the study can be extended into the 21st century.

Turning to the time-specific analyses, Table 13 repeats the highest-ranked variables for each of the multifactorial statistical analyses performed.

	All senses	Relevant senses
1800-1850	Sense, Object, Author	Author, Subfield, Object
1850-1900	Sense, Object, Author	Object, Author, Subfield
1900-1920	Sense, Subfield, Object	Subfield, Author, Object

Table 13: Highest ranked variables in each time period

Research question 1 concerned the importance of ‘sense’ in predicting lemma choice over time. The data shows that in the models containing all senses, ‘sense’ stays the highest ranked by the VIR in all periods. However, from inspecting the Ctrees, we know that the splits based on 'sense' are usually made between MASS and WEIGHT and the other senses. If the CE project is successful, as time goes on the VIRs should have shown an increase in the importance of 'sense' over time in the model containing the relevant senses. While it need not necessarily have to become the most important, since the change could still be ongoing by the end of the analysis, it would be expected that it should eventually rank more highly than 'author'. However, this does not occur. The individual choice remains consistently more important than 'sense' in lemma choice. In fact, 'sense' becomes less important over time.

So, if not 'sense', what does affect lemma choice? Research question 3 asked whether extra-linguistic variables have any effect, and the results in Table 13 clearly show that they do. In the models of the relevant senses, 'author', 'object', and 'subfield' are consistently found to be the most important predictive factors, consistently ranking above 'sense' in models containing only relevant senses. We can therefore answer research question 3 by saying that extra-linguistic factors have an overwhelming effect on lemma choice.

Research question 2 aimed to look at the time frame of implementation, and diffusion patterns. From Table 13, we see that there is a remarkable consistency in terms of which factors influence lemma choice over time. The same three factors remain the highest-ranked in the models of the relevant senses only, and 'sense' remains the highest-ranked in the model containing all senses. This suggests there was little change in what factors condition choice. Lemma choice, therefore, varied as a function of individual variation, the 'subfield' in which the author was writing, and the 'object' about which they are writing. This instance of CE diffused socially and marked discipline boundaries. The hypotheses stated that if extra-linguistic factors were found to be the most important, then this instance of CE could not be considered successfully implemented.

8.1.3 PCA

The influence of 'sense' and other external factors was also explored through PCA, which found there is some kind of structure emerging in the use of *mass* and *weight*, but not necessarily in the way that Newton would have imagined. The clusters that emerged after 1850 are based largely on certain 'subfields', 'authors' and 'objects', namely splitting Biology and Geology from the other disciplines, and even creating splits within the Biology tokens based on the author. These three major dimensions align with the variables identified as important by the VIRs. In terms of diachronic patterning, we do see structure emerging, but not based on ‘sense’, but on extra-linguistic variables.

One way of viewing this increasing clustering over time and the differentiation of the scientific sub-disciplines is that the corpus spans over an important period of ideological shift, namely the shift towards ‘modernity’. One part of this was establishing the scientific identity as separate from the public, through the creation of specialist vocabulary and making science increasingly inaccessible to the public. Science was establishing itself and turning away from society at large. Part of defining themselves was the creation of a separate and idealist language, where they have influence and authority. It was also turning away from itself, creating different disciplines and individual registers corresponding to each, which should shift according to scientific need.

8.2 Overall success

Taking into consideration the results of the data analysis, it appears as if Newton’s CE project was not successful. If it were successful, there would have been increasing usage of lemmas with their ‘correct’ respective senses, and that sense would have been an important conditioning factor in the statistical analysis. Although we see growing structure in how *mass* and *weight* are used in the RSC, this seems to come from the ‘subfield’, ‘author’ and ‘object’.

However, this leads to more questions. For example, who were the people and subfields that innovated the change? Who helped spread it? What happened after 1920? What we can definitively say is that Newton’s proposal was not accepted at large. More polysemy was created, more individual variation emerged, and more division between disciplines developed.

A major reason for the failure of the CE project is the permeating idea that authors idiosyncratically choose which words and senses they are using, in an ad-hoc fashion. Sometimes, they explicitly discuss what they mean by each term, and sometimes they do not. It was also found to depend on their ideological position, as was the case with Desaguliers, who used the terms according to Newton's suggestions due to his personal and professional alignment with him. Authors can, and do, define their terms as they see fit. Therefore, a single conceptualisation of what a word should mean, or how we should use it, is difficult to achieve.

Authority, as has been discussed from a theoretical viewpoint, was found to be more complex than initially thought. Despite Newton having almost complete control of the Royal Society between 1703 and 1727, the usage of *mass* during this time is marginal. Since this is hard to quantitatively measure, and the historical nature of the study makes it hard to find qualitative information, it may be the case that there were other authorities at play.

The usage of *mass* and *weight* also depend on the 'subfield' and 'object' with which they are used. Since the disciplines, and therefore the topics they discuss evolved through the process of enregisterment, it is to be expected that certain uses of *mass* and *weight* come to be associated with certain fields and discussions. It has also been mentioned that particular disciplines may find the distinction communicatively helpful, and so, from a functionalist viewpoint, these are the domains that the senses should have been the most carefully differentiated. This is not what was found. Therefore, even within the same discipline (or 'subfield'), there appears to be individual variation. At this point, it is unknown how distinct these disciplines are in the minds of the authors themselves, and so no definite conclusion can be drawn from this.

With the instances of *mass* and *weight*, we could expect the outcome of this study to also apply to the wider public. The vernacular does not separate these words by sense, preferring *weight* for both. This study shows that the scientific world does not deviate from the population as a whole. We only use *weight* as WEIGHT and *mass* as MASS if it is absolutely needed in the context.

In summary, this instance of CE failed if we look at the consistency of usage and differentiation based on 'sense'. However, it did create a way to explicitly discuss MASS and WEIGHT should the author need, or wish, to do so, either for the requirements of the topic they were discussing, or their personal choices.

8.3 The future of Conceptual Engineering

Thus, this instance of CE was unsuccessful. It is tempting to jump to the conclusion that this suggests the outlook for CE is bleak. However, this was only a small study conducted on a single, homogeneous, niche community. Therefore, it is difficult to generalise the results to all future and past instances of CE. Conversely, given this group has a clear common goal of scientific progress, and the case of CE appealed to this goal, the chances of success in a larger and more diverse community may be small.

The implementation problem has been frequently discussed by philosophers in a theoretical sense, or in terms of what society would have to be like for it to spread, but to my knowledge, no empirical study has been conducted on an instance of CE until this one. What was found is that individuality plays a significant role in the usage of engineered concepts and that communities should not be viewed as a homogeneous group who will all accept a certain concept because of a well-argued explanation as to why they should.

8.3.1 Future avenues of research

This work has found that concepts “stretch, shrink, or refigure what exactly we are talking about” (Haslanger, 2012, pg 225). This works, and will continue to work, as long as the hearer/reader can identify the intentions of the speaker/writer and infer the correct sense from what is uttered. This notion of speaker and hearer inference brings us to an avenue for possible further research into CE. Semantics and pragmatics have much to offer to the study of CE, both theoretically and experimentally. For example, when a concept is engineered, does it gain the engineered meaning when the speaker starts using it in a new way, or when the hearer starts understanding it in a new way? This is only one of many pragmatic principles that could, and should, be investigated when it comes to CE. Another is the idea of common ground and background knowledge. If the hearer/reader has shared common knowledge with the speaker/writer and knows which topics they are interested in, which theories they support, and what their ideologies are, then this could cause them to infer from the speaker’s utterance something different than what is said. For example, in the context of Biology, two conversing biologists may have an implicit understanding that in saying *bodyweight*, they are in actuality referring to the MASS of the human body. This depends on there being this common ground, much like other pragmatic phenomena that have been investigated. This is likely to be a fruitful future endeavour.

One interesting idea to explore further is whether the changes that have been observed may be naturally occurring, organic examples of linguistic change. While the CE project was not successful, how *mass* and *weight* were used did change over time. An interesting point of comparison might be other scientific concepts, and whether there are any common patterns to be found in the changes discovered in this study and those other concepts. If certain commonalities are found, or no other external influence can be found on the changes identified, then it is possible to posit that the changes described in this study are simply part of natural semantic change or the natural evolution of (scientific) concepts. Perhaps we do not need to do anything, and concepts will naturally drift towards what we need them to be.

Another important question is whether people who do not know they are doing CE are truly doing CE. While the changes of *mass* and *weight* may not have been successful, imagine momentarily that they had been. Newton did not know he was participating in CE; he was simply pondering the meanings of words that he needed for his scientific writings. If consciousness is a prerequisite for CE, then there are very few CE projects. Many groups and people who are campaigning for the amelioration of concepts have never heard of the term CE since it currently is limited to the academic sphere, and even then, it is little known beyond philosophy. While difficult to answer, future research into CE must decide exactly which projects do fall within the boundaries of CE so that linguists and philosophers alike can guide their investigations.

This work has focused on scientific concepts, and this meant it has also focused on a niche, more Habermasian, group of individuals. While this was necessary for the scope of the current study, there is also the potential that looking at the public more generally, and more social concepts, could result in a vastly different outcome. In a project that aims to target a wider community, more factors could contribute to the uptake of, or resistance to, the ameliorated concept. Many, more diverse, factors could contribute to such a change, including social and political factors that may not be pertinent to a small scientific community. A larger, more varied group leads to even more variation in ideologies and goals, serving to amplify the Carnivalesque nature of publics that has been previously discussed.

These kind of projects are also dependent on the status we give linguistic relativity, and the extent to which we subscribe to it. The present study has highlighted a third option with regards to linguistic relativity rather than being ardently for or against it, and thus ardently for or against CE. By engineering concepts, they become more discussable, which means that they can be combated. Take *sexual harassment*, a term which some have considered to have been conceptually engineered by Lin Farely in 1974 (Baker, 2007) to address what her female students had identified as a semantic gap. The progress this concept has allowed for since 1974 is remarkable, not in that it has caused the disappearance of *sexual harassment*, or that *sexual harassment* did not exist before the word was introduced, but in that, it has allowed for a wider discussion of the problem since it could not be discussed until it was pinpointed and labelled. This is like the instance of *mass* and *weight*, where the explicit splitting of the two terms and concepts allowed for the development of scientific principles such as acceleration.

Naturally, the issues surrounding CE and linguistic relativity are much larger in a project that tackles a societal concept such as *sexual harassment* than in a study focused on scientific terminology. The methods and theories put forward in this work can be applied to more societal changes in the future, and linguistic relativity will become more important.

One last thing to mention would be the sampling technique used in this study. The sample of the corpus that was annotated was first filtered to only include authors that use *mass* or *weight* more than one hundred times. This was done to guarantee that the tokens in question come from texts where MASS and WEIGHT were important concepts. It is possible that authors who use *mass* or *weight* fewer than one hundred times, or low-frequency authors, may have different patterns of usage than high-frequency authors, and therefore different conditioning variables for their usage. Furthermore, the sampling technique used meant that statistical analyses could not be performed on the earlier periods. If this study were to be replicated, then this aspect of the sampling may need to be reworked. Therefore, future research on the same data is required.

CHAPTER 9

CONCLUSION

Newton's Conceptual Engineering project had two criteria for success; the consistent usage of *mass* with sense MASS and *weight* with sense WEIGHT, and that the choice of lemma was governed by sense alone. This study found that the case of Conceptual Engineering proposed by Newton concerning *mass* and *weight* was unsuccessful. Instead, there was much idiosyncratic variability, and variation based on the subfield and object under discussion. Structure emerged in the usage of *mass* and *weight* over time, but this did not correspond to the categories proposed by Newton.

To return to the questions asked at the start of this study, research question 1 concerned the extent of the differentiation of *mass* and *weight* by sense and the implications for other CE projects. As mentioned, differentiation based on the senses did not successfully occur. Future CE projects must be aware that implementation is not a simple task. Outside of the more theoretical questions such as internalism versus externalism, conceptual engineers will find themselves battling with ideologies, social groups, iconisation, or a simple uncooperative, Carnivalesque public.

Research question 2 asked what time frames diffusion of CE has, and which trends are present in the diffusion. This study found an initial period of adherence to the CE project, if only by a single author who was strongly allied to Newton himself. After this, the only potential period of success was from 1825 to 1949, when *weight* was used most frequently with the sense WEIGHT. This occurred around 140 years after Newton suggested the differentiation, and after this the senses and lemmas diverged again. While usage clusters appeared over time, this was correlated more with 'subfield', 'author', and 'object' than with 'sense', and may have been due to natural diachronic development rather than conscious human intervention. Conceptual engineers should be prepared for their projects to take time to implement fully, despite wishes for quick diffusion.

The third question concerned the influence of extra-linguistic variables, and the author played a significant role in the choice of the lemma. In predicting *mass* and *weight*, the model found it most useful to split the data into 'author' groups, which one would not expect should the CE project have been successful. ‘Subfield’ and ‘object’ were also influential, suggesting that items and disciplines under discussion call for different sense-lemma mappings.

As for the linguistic variables, these had little to no influence on the choice of lemma. The sense was of higher importance, but only when all senses were considered, and only in a single period for the relevant sense. This suggests more pronounced differentiation of these non-scientific senses than of scientific senses.

This is, to my knowledge, the first study of its kind that analyses the implementation of a CE example and as such shows that this work needs to continue to understand how CE (deliberately or not) works to change reality.

The results and discussion in this work have exemplified that linguists have much to offer to CE when it comes to understanding and discussing the implementation problem, and that they can contribute without making any normative judgements. By considering sociolinguistic theories and language ideologies in addition to an empirical study, it is possible to evaluate the success of CE projects and contextualise the results of the models. Through similar projects, it may become possible to create metrics to measure the success of such projects and understand which factors need to be considered in the implementation of an engineered term.

The approach of this study is by far not the only contribution that linguists can make, and some of the other ways have been mentioned in passing. These include an in-depth analysis of the semantics and pragmatics of engineered terms and their usage in different contexts, along with consideration of different kinds of sociolinguistic relations and social networks. This is not to say that linguists and philosophers will always see eye-to-eye when it comes to CE. Some philosophers will maintain that adoption is not a criterion for success and that in simply identifying and fixing a problem in a concept, the philosopher will have achieved what they set out to do, namely, to solve a theoretical problem. Spreading this concept beyond the philosopher, or a small circle of philosophers, is unnecessary. If this is the case, then there is not much that linguists can offer for these CE projects.

However, if philosophers believe that at least one instance exists where success requires widespread adoption, then there is a place for linguistics within CE.

REFERENCES

Agha, A. (2007). *Language and Social Relations*. Cambridge University Press.

Agha, A. and Frog (2015). An Introduction to Registers of Communication. In Agha, A. and Frog, editors, *Registers of Communications*, volume 18, pages 13–24. Finnish Literature Society.

Altakhaineh, A. R. M. (2018). The semantic change of positive vs. negative adjectives in Modern English. *Lingua Posnaniensis*, 60(2):25–38.

Andow, J. (2020). Fully experimental conceptual engineering. *Inquiry*, 0(0):1–27.

Andow, J. (2021a). Conceptual engineering is extremely unlikely to work. So what? *Inquiry*, 64(1-2):212–226.

Andow, J. (2021b). Fully experimental conceptual engineering. *Inquiry*.

Atkinson, D. (1996). The "Philosophical Transactions of the Royal Society of London," 1675-1975: A Sociohistorical Discourse Analysis. *Language in Society*, 25(3):333–371. Publisher: Cambridge University Press.

Atkinson, D. (1998). *Scientific Discourse in Sociohistorical Context: The Philosophical Transactions of the Royal Society of London, 1675-1975*. Routledge.

Bacon, F. (1883). The New Organon. In *The works of Francis Bacon*, volume 4. Longman, London. Original date: 1620.

Bacon, F. (2011). The New Organon. In Spedding, J., Ellis, R. L., and Heath, D. D., editors, *The Works of Francis Bacon: Volume 4: Translations of the Philosophical Works 1*, volume 4 of *Cambridge Library Collection - Philosophy*, pages 37–248. Cambridge University Press, Cambridge. Original date: 1620.

Baker, C. N. (2007). The emergence of organized feminist resistance to Sexual Harassment in the United States in the 1970s. *Journal of Women's History*, 19(3):161–184.

Bakhtin, M. (1982). The Dialogic Imagination. In Holquist, M. and Emerson, C., editors, *The Dialogic Imagination: Four Essays by M. M. Bakhtin*. The University of Texas, Austin.

Baldwin, M. (2019). Peer Review. *Encyclopedia of the History of Science*, 1(2). Number: 2.

Banks, D. (2008). *The Development of Scientific Writing: Linguistic Features and Historical Context*. Equinox.

Baron, D. (2020). *What's Your Pronoun?: Beyond He and She*. Liveright.

Bauman, R. and Briggs, C. L. (2003). *Voices of Modernity: Language Ideologies and the Politics of Inequality*. Cambridge University Press, 1 edition.

Beltagy, I., Lo, K., and Cohan, A. (2019). SciBERT: A Pretrained Language Model for Scientific Text. *arXiv:1903.10676 [cs]*. arXiv: 1903.10676.

Biber, D. and Conrad, S. (2009). *Register, Genre, and Style*. Cambridge Textbooks in Linguistics. Cambridge University Press, Cambridge.

Biber, D. and Finegan, E. (1997). Diachronic relations among speechbased and written registers in English. In Nevalainen, T. and Kahlas-Tarkka, L., editors, *To explain the present: Studies in the changing English language in honour of Matti Rissanen*, pages 253–276. Société Néophilologique, Helsinki.

Biber, D. and Gray, B. (2016). *Grammatical Complexity in Academic English: Linguistic Change in Writing*. Studies in English Language. Cambridge University Press, Cambridge.

Bievre, P. d. and Peiser, H. S. (1992). 'Atomic weight': The name, its history, definition, and units. *Pure and Applied Chemistry*, 64(10):1535–1543. Publisher: De Gruyter.

Bizzoni, Y., Degaetano-Ortlieb, S., Fankhauser, P., and Teich, E. (2020). Linguistic Variation and Change in 250 Years of English Scientific Writing: A Data-Driven Approach. *Frontiers in Artificial Intelligence*, 3:73.

Bradie, M. (1986). Assessing Evolutionary Epistemology. *Biology and Philosophy*, 1:401–459.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1):5–32.

Brewster, D. (1831). *The life of Sir Isaac Newton.* J. Murray, London. OCLC: 1317634.

Briggs, C. L. (1992). Linguistic ideologies And the naturalization of power in warao discourse. *Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA)*, 2(3):387–404.

Bucholtz, M. and Hall, K. (2005). Identity and interaction: a sociocultural linguistic approach. *Discourse Studies*, 7(4-5):585–614. Publisher: SAGE Publications.

Bunte, P. (2009). "You keep not listening with your ears!" Language ideologies, language socialization, and paiute identity. *Native American Language Ideologies*, pages 172–189.

Burge, T. (1979). Individualism and the Mental. *Midwest Studies in Philosophy*, 4(1):73–122. Publisher: Oxford University Press: Oxford.

Bybee, J. (2006). From Usage to Grammar: The Mind's Response to Repetition. *Language*, 82(4):711–733. Publisher: Linguistic Society of America.

Cappelen, H. (2017). Why Philosophers Shouldn't Do Semantics. *Review of Philosophy and Psychology*, 8(4):743–762.

Cappelen, H. (2018). *Fixing Language: An Essay on Conceptual Engineering*. Oxford University Press, Oxford.

Cappelen, H. and Plunkett, D. (2020). Introduction: A Guided Tour of Conceptual Engineering and Conceptual Ethics. In *Conceptual Engineering and Conceptual Ethics*, pages 1–34. Oxford University Press.

Carnap, R. (1947). *Meaning and Necessity: A Study in Semantics and Modal Logic*. University of Chicago Press, Chicago, IL.

Carnap, R. (1950). *Logical foundations of probability*. Logical foundations of probability. University of Chicago Press, Chicago, IL, US.

Carnap, R. (1962). *Logical Foundations of Probability*. University of Chicago Press.

Carnap, R. (1963). Intellectual Autobiography. In Schilpp, P. A., editor, *The Philosophy of Rudolf Carnap*, volume XI of *Library of Living Philosophers*, pages 3–83. Open Court, Illinois.

Carus, A. W. (2007). *Carnap and Twentieth-Century Thought: Explication as Enlightenment*. Cambridge University Press, Cambridge.

Chaucer, G. and McMillan, A. (1385/1987). *The legend of good women*. Rice University Press, Houston, Tex.

Chomsky, N. (1969). *Aspects of the Theory of Syntax*. MIT Press.

Chomsky, N. (2000). *New Horizons in the Study of Language and Mind*. Cambridge University Press.

Clark, A. and Chalmers, D. (1998). The Extended Mind. *Analysis*, 58(1):7–19.

Costa, J. (2018). Introduction: Regimes of language and the social, hierarchized organization of ideologies. *Language & Communication*, 66.

Cull, M. J. (2020). *Engineering Genders: Pluralism, Trans Identities, and Feminist Philosophy*. PhD Thesis, University of Sheffield, Sheffield.

Cull, M. J. (2021). Engineering is not a luxury: Black feminists and logical positivists on conceptual engineering. *Inquiry*, 64(1-2):227–248.

Darwin, G. H. (1882). IV. On the stresses caused in the interior of the earth by the weight of continents and mountains. *Philosophical Transactions of the Royal Society of London*, 173:187–230. Publisher: Royal Society.

Dawson, H. (2007). *Locke, Language and Early-Modern Philosophy*. Cambridge University Press, Cambridge.

de Saussure, F. (1986). *Course in General Linguistics*. Open Court Publishing.

De Smet, H. (2017). Entrenchment effects in language change. In Schmid, H. J., editor, *Entrenchment and the psychology of language learning: How we reorganize and adapt linguistic knowledge.*, pages 75–99. American Psychological Association, Washington.

Degaetano-Ortlieb, S., Kermes, H., Khamis, A., and Teich, E. (2018). An Information-Theoretic Approach to Modeling Diachronic Change in Scientific English. *From Data to Evidence in English Language Research*, pages 258–281. Publisher: Brill.

Degaetano-Ortlieb, S. and Teich, E. (2018). Using relative entropy for detection and analysis of periods of diachronic linguistic change. In *Proceedings of the Second Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 22–33, Santa Fe, New Mexico. Association for Computational Linguistics.

Degaetano-Ortlieb, S. and Teich, E. (2019). Toward an optimal code for communication: The case of scientific English. *Corpus Linguistics and Linguistic Theory*, 18(0).

Deumert, A. (2014). *Sociolinguistics and Mobile Communication*. Edinburgh University Press.

Deumert, A. (2019). Sensational signs, authority and the public sphere: Settler colonial rhetoric in times of change. *Journal of Sociolinguistics*, 23(5):467–484.

Deutsch, M. (2020). Speaker's reference, stipulation, and a dilemma for conceptual engineers. *Philosophical Studies*, 177(12):3935–3957.

Dutilh Novaes, C. and Reck, E. (2017). Carnapian explication, formalisms as cognitive tools, and the paradox of adequate formalization. *Synthese*, 194(1):195–215.

Eckert, P. (2000). *Language Variation as Social Practice: The Linguistic Construction of Identity in Belten High*. Wiley.

Eckert, P. (2004). Elephants in the room. *Journal of Sociolinguistics*, 7(3):392–397.

Eklund, M. (2014). Replacing Truth? In *Metasemantics*. Oxford University Press, Oxford.

Fassio, D. and McKenna, R. (2015). Revisionary Epistemology. *Inquiry*, 58(7-8).

Fischer, S., Knappen, J., Menzel, K., and Teich, E. (2020). The Royal Society Corpus 6.0: Providing 300+ Years of Scientific Writing for Humanistic Study. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 794–802, Marseille, France. European Language Resources Association.

Fonteyn, L. (2021). Varying Abstractions: a conceptual vs. distributional view on prepositional polysemy. *Glossa: a journal of general linguistics*, 6(1).

Fonteyn, L. and Nini, A. (2020). Individuality in syntactic variation: An investigation of the seventeenth-century gerund alternation. *Cognitive Linguistics*, 31(2):279–308. Publisher: De Gruyter Mouton.

Fraser, N. (1992). Rethinking the Public Sphere: A Contribution to the Critique of Actually Existing Democracy. In Calhoun, C., editor, *Habermas and the public sphere*, pages 109–142. MIT Press, Cambridge, MA.

Gal, S. (2019). Making registers in politics: Circulation and ideologies of linguistic authority. *Journal of Sociolinguistics*, 23(5):450–466.

Gal, S. and Irvine, J. T. (2019). *Signs of Difference: Language and Ideology in Social Life*. Cambridge University Press, Cambridge.

Gardiner, M. E. (2004). Wild publics and grotesque symposiums: Habermas and Bakhtin on dialogue, everyday life and the public sphere. *The Sociological Review*, 52(s1):28–48.

Gauch, H. G. (2012). *Scientific Method in Brief.* Cambridge University Press, Cambridge.

Gray, B. and Biber, D. (2018). Academic writing as a locus of grammatical change: The development of phrasal complexity features. In Whitt, R. J., editor, *Studies in Corpus Linguistics*, volume 85, pages 117–146. John Benjamins Publishing Company, Amsterdam.

Grice, P. (1989). *Studies in the Way of Words*. Harvard University Press.

Gries, S. T. (2020). 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. Publisher: De Gruyter Mouton.

Habermas, J. (1996). *Between facts and norms: Contributions to a discourse theory of law and democracy*. MIT Press, Cambridge, MA. Original date: 1992.

Habgood-Coote, J. (2020). Fake news, conceptual engineering, and linguistic resistance: reply to Pepp, Michaelson and Sterken, and Brown. *Inquiry*, pages 1–29.

Hardie, A. (2012). CQPweb - combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics*, 17(3):380–409. Publisher: John Benjamins.

Harris, J. (1704). *Lexicon Technicum: Or, An Universal English Dictionary of Arts and Sciences:: Explaining Not Only the Terms of Art, But the Arts Themselves*. Brown, Dan and Goodwin, Tim and Walthoe, John and Newborough, Thomas and Nicholson, John and Benskin, Thomas and Tooke, Benjamin and Midwinter, Dan and Leigh, Thomas and Coggan, Francis.

Haslanger, S. (2000). Gender and Race: (What) Are They? (What) Do We Want Them To Be? *Noûs*, 34(1):31–55.

Haslanger, S. (2012). *Resisting Reality: Social Construction and Social Critique*. Oxford University Press.

Heath, S. (1989). Language Ideology. *International Encyclopedia of Communications*, 2:393–395.

Hirt, C., Claessens, S., Fecher, T., Kuhn, M., Pail, R., and Rexer, M. (2013). New ultrahigh-resolution picture of Earth's gravity field. *Geophysical Research Letters*, 40:4279–4283.

Hothorn, T., Buehlmann, P., Dudoit, S., Molinaro, A., and Laan, M. V. D. (2006a). Survival Ensembles. *Biostatistics*, 7(3):355–373.

Hothorn, T., Hornik, K., and Zeileis, A. (2006b). Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics*, 15(3):651–674.

Hundt, M. and Mair, C. (1999). "Agile" and "Uptight" Genres: The Corpus-based Approach to Language Change in Progress. *International Journal of Corpus Linguistics*, 4:221–242.

Hymes, D. (1964). Introduction: Toward Ethnographies of Communication1. *American Anthropologist*, 66(6):1–34.

Hymes, D. (1974). *Foundations in Sociolinguistics: An Ethnographic Approach*. University of Pennsylvania Press.

Irvine, J. T. (1989). when talk isn't cheap: language and political economy. *American Ethnologist*, 16(2):248–267.

Ishizaki, Y. (2012). A usage-based analysis of phrasal verbs in Early and Late Modern English. *English Language & Linguistics*, 16(2):241–260.

Jenkins, K. (2016). Amelioration and Inclusion: Gender Identity and the Concept of *Woman*. *Ethics*, 126(2):394–421.

Jorem, S. (2021). Conceptual engineering and the implementation problem. *Inquiry*, 64(1-2):186–211.

Knowlson, J. (2019). *Universal language schemes in England and France 1600-1800*. University of Toronto Press.

Koch, S. (2021). The externalist challenge to conceptual engineering. *Synthese*, 198(1):327–348.

Kripke, S. (1980). *Naming and Necessity: Lectures Given to the Princeton University Philosophy Colloquium*. Harvard University Press.

Kroskrity, P. V. (2010). Language ideologies: Evolving perspectives. *Society and Language Use*, pages 192–211.

Labov, W. (2001). *Principles of linguistic change, Volume 2: Social Factors*. Blackwell, Oxford. OCLC: 47901638.

Labov, W. (2010). *Principles of Linguistic Change, Volume 3: Cognitive and Cultural Factors*. Wiley.

Laudan, L. (1968). Theories of Scientific Method from Plato to Mach: A Bibliographical Review. *History of Science*, 7(1):1–63. Publisher: SAGE Publications Ltd.

Lesley, M. (2004). Language ideologies and linguistic change. In Fought, C., editor, *Sociolinguistic Variation: Critical Reflections*, page 161–177. Oxford University Press.

Locke, J. (1975). An Essay Concerning Human Understanding. In Nidditch, P., editor, *An Essay Concerning Human Understanding: Clarendon Edition of the Works of John Locke*. Oxford University Press. Original date: 1690.

Losonsky, M. (2021). The History of the Philosophy of Language before Frege. In Stalmaszczyk, P., editor, *The Cambridge Handbook of the Philosophy of Language*, pages 51–70. Cambridge University Press, 1 edition.

Lucy, J. (2001). Sapir-Whorf hypothesis. In *International Encyclopedia of the Social & Behavioral Sciences*, pages 13486–13490. Elsevier.

Manne, K. (2017). *Down Girl: The Logic of Misogyny*. Oxford University Press, New York.

"mass n.1" (2021). Oxford English Dictionary. In *OED Online*. Oxford University Press. *https://www.oed.com/view/Entry/114665* [Last accessed 17/02/2022].

"mass n.2" (2021). Oxford English Dictionary. In *OED Online*. Oxford University Press. *https://www.oed.com/view/Entry/114666* [Last accessed 17/02/2022].

McDougall-Water, J., Moxham, N., and Fyfe, A. (2021). History of the Philosophical Transactions. *https://royalsociety.orgs/journals/publishing-activities/publishing350/history-philosophical-transactions* [Accessed: 16/02/2022].

Menzel, K. (2020). Exploring Academic English through the Royal Society Corpus. In Tinnefeld, T., editor, *The Magic of Language: Productivity in Linguistics and Language Teaching*, pages 283–305. htw saar.

Menzel, K., Knappen, J., and Teich, E. (2021). Generating linguistically relevant metadata for the Royal Society Corpus. *Research in Corpus Linguistics*, 9(1):1–18. Number: 1.

Meyns, C. (2020). 'Data' in the Royal Society's *Philosophical Transactions* , 1665–1886. *Notes and Records: the Royal Society Journal of the History of Science*, 74(3):507–528.

Nado, J. (2021). Conceptual engineering via experimental philosophy. *Inquiry*, 64(1-2):76–96.

Neurath, O. (1944). *Foundations of the social sciences.* University of Chicago Press, Chicago, Ill. OCLC: 525800.

Newton, I. (1687). *Philosophiæ Naturalis Principia Mathematica.* Apud G. & J. Innys, Londini.

Newton, I. (1999). *The Principia: Mathematical Principles of Natural Philosophy*. University of California Press.

Nietzsche, F. (1901). *The Will to Power*. Random House, New York, 1 edition.

Osthoff, H. and Brugmann, K. (1878). *Morphologische Untersuchungen auf dem Gebiete der indogermanischen Sprachen.* S. Hirzel, Leipzig. OCLC: 4995014.

Paul, H. (1890). *Principles of the history of language*. McGrath., College Park, MD.

Putnam, H. (1973). Meaning and Reference. *Journal of Philosophy*, 70(19):699–711. Publisher: Journal of Philosophy Inc.

Richard, M. (2020). The A-project and the B-project. In *Conceptual Engineering and Conceptual Ethics*, pages 358–378. Oxford University Press.

Ruíz, R. (1984). Orientations in Language Planning. *NABE Journal*, 8(2).

Røstvik, C. M. and Fyfe, A. (2018). Ladies, Gentlemen, and Scientific Publication at the Royal Society, 1945–1990. *Open Library of Humanities*, 4(1).

Scharp, K. (2007). Alethic Vengeance. In Beall, J. C., editor, *Revenge of the Liar: New Essays on the Paradox*. Oxford University Press.

Scharp, K. (2013). *Replacing Truth.* Oxford University Press.

Scharp, K. (2021). Conceptual engineering for truth: aletheic properties and new aletheic concepts. *Synthese*, 198:647–688.

Schilling-Estes, N. (2002). American English Social Dialect Variation and Gender. *Journal of English Linguistics*, 30(2):122–137. Publisher: SAGE Publications Inc.

Shapin, S. (1984). Pump and Circumstance: Robert Boyle's Literary Technology. *Social Studies of Science*, 14(4):481–520. Publisher: SAGE Publications Ltd.

Silverstein, M. (1979). Language structure and linguistic ideology. In Cline, P., Hanks, C., and Hofbauer, C., editors, *The elements: A parasession on linguistic units and levels*, pages 193–246. Chicago Linguistics Society, Chicago, IL.

Silverstein, M. (1985). Language and the culture of gender: At the intersection of structure, usage, and ideology. In Mertz, E. and Parmentier, R. J., editors, *Semiotic Mediation*, pages 219–259. Elsevier.

Siyanova-Chanturia, A., Conklin, K., and van Heuven, W. J. B. (2011). Seeing a phrase "time and again" matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(3):776–784.

Spitzmüller, J. (2019). Sociolinguistics going 'wild': The construction of auratic fields. *Journal of Sociolinguistics*, 23(5):505–520.

Sprat, T. (1667). *The History of the Royal Society*. T.R, London.

Sterken, R. K. (2020). Linguistic Intervention and Transformative Communicative Disruptions. In *Conceptual Engineering and Conceptual Ethics*, pages 417–434. Oxford University Press.

Sun, K., Liu, H., and Xiong, W. (2021). The evolutionary pattern of language in scientific writings: A case study of Philosophical Transactions of Royal Society (1665–1869). *Scientometrics*, 126(2):1695–1724.

Tagliamonte, S. A. and Baayen, R. H. (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. Publisher: Cambridge University Press.

Urbach, P. M., Quinton, A. M., Quinton, B., and Lea, K. M. (2022). Francis Bacon. *https://www.britannica.com/biography/Francis-Bacon-Viscount-Saint-Alban* [Last accessed 05/03/2022].

Valdez, J. R. (2011). Linguistic Ideologies and the History of Linguistic Ideas. In Valdez, J. R., editor, *Tracing Dominican Identity: The Writings of Pedro Henríquez Ureña*, pages 33–62. Palgrave Macmillan US, New York.

Warner, M. (2002). *Publics and Counterpublics*. Princeton University Press, Princeton.

"weight n.1" (2021). Oxford English Dictionary. In *OED Online*. Oxford University Press. *https://www.oed.com/view/Entry/26891* [Last accessed 17/02/2022].

Williams, R. (1977). *Marxism and Literature*. Oxford University Press.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., and Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, 104(19):7780–7785.

Woolard, K. A. (1998). Introduction: Language Ideology as a Field of Enquiry. In Bambi B. Schieffelin, Kathryn A. Woolard, and Paul V. Kroskrity, editors, *Language Ideologies : Practice and Theory*, Oxford Studies in Anthropological Linguistics. Oxford University Press, New York.

Woolard, K. A. (2008). Why dat now?: Linguistic-anthropological contributions to the explanation of sociolinguistic icons and change. *Journal of Sociolinguistics*, 12(4):432–452.

Woolard, K. A. and Schieffelin, B. B. (1994). Language Ideology. *Annual Review of Anthropology*, 23(1):55–82. Publisher: Annual Reviews.

Yeo, R. R. (1986). Scientific Method And The Rhetoric of Science In Britain, 1830–1917. In Schuster, J. A. and Yeo, R. R., editors, *The Politics and Rhetoric of Scientific Method: Historical Studies*, Australasian Studies in History and Philosophy of Science, pages 259–297. Springer Netherlands, Dordrecht.

Young, I. M. (1987). Impartiality and the civic public: Some implications of feminist critiques of moral and political theory. In *Feminism as critique*, pages 57–74. Polity Press, Cambridge, UK.

"† nice, n. (2021). Oxford English Dictionary. In *OED Online*. Oxford University Press. *http://www.oed.com/view/Entry/126731* [Last accessed 08/04/2022].

A Figure 14 group data

Group	Split Based on	
A	Author	A.W.Hofmann, C.Hatchett, Griffiths & Glazebrook, F.A.Abel, F.O.Bower, G.B.Airy, G.H.Darwin, G.Pearson, H.Hennessy, H.Watney, J.D.Forbes, J.H.Jeans, J.H.Pratt, J.Ivory, J.Lockheart, J.L.Clarke, J.N.Lockyer, J.Tyndall, L.S.Beale, M.Barry, M.Faraday, R.Mallet, T.Anderson, W.B.Hardy, W.C.Williams, W.Hopkins, W.K.Parker, W.Thomson
B	Author	A.E.H.Tutton, B.Count, B.C.Brodie, B.Thompson, C.T.Heycock, D.Mendel, D.Thoday, E.Hodgkinson, Halnan, Marshall, Yule &Langley, E.J.Reed, Reed & Stokes, E.Sabine, F.Baily, G.Dreyer, Matthaei & Darwin, G.Rennie, G.Shuckburgh, H.Cavendish, H.Davy, H.E.Roscoe, H.Kater, H.Moseley, H.Tomlinson, J.Barnard, J.Davy, Gray, Henderson & Kelvin, J.Herschel, J.Joly, Joly & Fitzgerald, Lawes & Gilbert, Manley & Poynting J.P.Joule, J.T.Desaguliers, J.W.Mallet, M.Raper, R.Kirwan, T.Graham, T.Thomson, Thorpe & Rodger, W.Crookes, W.Fairbairns, W.H.Miller, W.Marcet
C	Object	GEN, STAN, WO
D	Object	BIO, C, EQ, F, GEO, MAT, MATTER, S
E	Object	ABS, BIO, BO, C, CONCEPT, EQ, FOR, GEO, MAT, STAN, WO
F	Object	F, MATTER, S
G	Object	BIO, C, EQ, F, GEO, MAT, MATTER, S, VM
H	Object	CONCEPT, STAN, WO
I	Author	A.E.H.Tutton, B.Count, B.C.Brodie, B.Thompson, C.T.Heycock, D.Mendel, D.Thoday, E.Hodgkinson, Halnan, Marshall, Yule & Langley, E.J.Reed, Reed & Stokes, E.Sabine, F.Baily, G.Dreyer, Matthaei & Darwin, G.Rennie, G.Shuckburgh, H.Cavendish, H.E.Roscoe, H.Kater, J.Barnard, J.Davy, Gray, Henderson & Kelvin, Joly & Fitzgerald, Lawes & Gilbert, J.P.Joule, J.T.Desaguliers, J.W.Mallet, M.Raper, R.Kirwan, T.Graham, T.Thomson, Thorpe & Rodger, W.Crookes, W.Fairbairs, W.H.Miller, W.Marcet
J	Author	H.Darwin, H.Moselet, H.Tomlinson, J.Joly, Manley & Poynting

Figure 22: Group data for the abbreviated Ctree of all data for all senses in Figure 14

B Figure 16 group data

Group	Split Based on	
A	Author	A.W.Hofmann, Griffiths & Glazebrook, F.A.Abel, F.O.Bower, G.H.Darwin, H.Henessy, H.Watney, J.H.Pratt, J.Lockhart, J.L.Clarke, J.N.Lockyer, J.Tyndall, L.S.Beale, M.Faraday, R.Mallet, W.C.Williamson, W.Hopkins, W.K.Parker, W.Thomson
B	Author	B.C.Brodie, C.T.Heycock, D.Mendel, E.Hodgekinson, E.J.Reed, Reed & Stokes, E.Sabine, G.B.Airy, H.E.Roscoe, H.Kopp, H.Mallet, H.Tomlinson, J.Barnard, J.Davy, Grey, Henderson & Kelvin, J.Joly, Joly & Fitzgerald Lawes & Gilbert, J.P.Joule, J.W.Mallet, T.Graham, Thorpe & Rodger, W.Crookes, W.Fairbairn, W.H.Moorby, W.Marcet
C	Object	EQ, F, S
D	Object	ABS, BIO, BO, C, FOR, GEO, MAT, WO
E	Object	BIO, C, EQ, F, GEO, MAT, S, VM
F	Object	CONCEPT, STAN, WO
G	Object	GEN, MATTER, WO
H	Object	BIO, C, EQ, F, GEO, MAT, S
I	Type	experiment, lecture
J	Type	abs, abstract, article, fla

Figure 23: Group data for Ctree all senses 1850-1900 in Figure 16

C Code

https://surfdrive.surf.nl/files/index.php/s/1oj3zdhyUZUxtqF

D Data

https://surfdrive.surf.nl/files/index.php/s/ObLNWUfFY1fqg6o

E Ctrees of the whole dataset

https://surfdrive.surf.nl/files/index.php/s/H4HFSlKfmQ3FwN3

F Time-specific Ctrees

https://surfdrive.surf.nl/files/index.php/s/ebsh7UpPM5NYJZQ

G Model fits across time periods for all senses

https://surfdrive.surf.nl/files/index.php/s/L8E43sECEFxHCOB

H Model fits across time periods for relevant senses

https://surfdrive.surf.nl/files/index.php/s/pGylOT5fjwDPdFS