PHD

The electrification of language: Computer-assisted language analysis of the construction of Michael Faraday's ideas

Smith, David W.

Award date:
1998

Awarding institution:
University of Bath

The Electrification of Language: Computer-Assisted Language Analysis of the Construction of Michael Faraday's Ideas

submitted by **David W. Smith**

for the degree of PhD

of the University of Bath

September 1998

David W. Smith

Contents

Acknowledgements

I am grateful to Professor David Gooding and Mr Graham Cox for their incisive, constructive, and detailed comments. The completion of my thesis would have taken a far greater toll on my sanity if it had not been for the many pleasurable hours cocooned in the Little Theatre Cinema in Bath, being steered towards decent, cask ale by CAMRA, and for the warmth and support of Claire Wardell.

Note on Citing of World Wide Web Sources

Even though there are a number of styles available for citing WWW sources they tend to recommend a style built around similar features, such as author's name and the URL (Universal Resource Locator). This is reflected in how WWW sources are cited in my thesis. Also URL's are cited on one line (unless too long) and not surrounded by angular brackets (< >).

Thus where a WWW source is used the following form will be adhered to:

i) Footnotes:

1. Name of author(s) (if any)
2. URL/Internet address
3. Date of publication/Date of modification (if available)
3. Last date of access

ii) Bibliography:

1. Name of author(s) (if any)
2. Title of the text
3. URL/Internet address
4. Date of publication/Date of modification (if available)
7. Last date of access

iii) Variation: Bibliography:

1. Name of author(s) (if any)
2. Title of publication
3. Title of the text
4. Pages (if available)
5. Date of publication/Date of modification (if available)
6. URL/Internet address
7. Last date of access

The Electrification of Language: Computer-Assisted Language Analysis of the Construction of Michael Faraday's Ideas

> Usually, when a new world view takes over from a previous one, the previous language continues to be used, but it is used metaphorically. The whole of language is a continuous process of metaphors, and the history of semantics is an aspect of the history of culture: language is both a living thing and a museum of fossils of life and past civilizations.
>
> Antonio Gramsci; Prison Notebook 11

Abstract

The main task of this thesis is to study language use and change in the construction of scientific concepts. This research embodies a 'linguistic turn' in science studies and is part of a larger project concerned with source contained modelling of the cognitive and social processes of science.[1] The thesis argues that science should not be seen as a body of facts transmitted via language, but as a linguistic and social construct, mediated by the variant use of textual features. The close analysis of certain linguistic attributes of scientific texts is used to expose the development of ideas over time, instead of, as with humanities-literary studies, author attribution. This involves a quantitative-qualitative study of the social and linguistic processes at work in scientific discovery and the development of ideas through computer-assisted analysis of nineteenth-century scientific documents. These documents constitute private and public writings of Michael Faraday and Charles Darwin.

I think it is important to clarify the relationship between the qualitative and quantitative aspects of this research. Despite the widespread use of numerical data and statistical analyses, this project is not primarily quantitative. The quantitative data do not stand on their own, but are contextualised and made sense of within a qualitative framework. In my view, it is this framework that provides the perspectives and meanings crucial to the success of computational

[1] Undertaken by David C. Gooding and Tom R. Addis funded by the JCI and ESRC.

linguistic analysis, and that enables a more sophisticated, contextual analysis to be carried out. The marriage of qualitative and quantitative approaches, in which the different forms of material are analyzed together to draw conclusions, greatly enriches my findings and is what makes this project unique.

The two principal questions asked by this research are: 1) whether it is possible to investigate the question: 'how do new expressions and ideas arise in language?', using quantitative methods of text analysis, and 2) whether it is possible to: i) to expose the differences in laboratory activity and the experimental recounting in published papers,[2] and ii) describe and reconstruct the human activity of scientific work through the study of the micro-history contained in private writings. This is important because published papers do not tell the full, unpurged story of scientific work.[3] The overriding question of this research is: 'how is language used as a medium for thinking and communication'?

Central to the research is the argument that 'private' documents can be studied to recover scientific laboratory practice, and that language change can be exposed in the difference between the account of a finished experiment and the description of actual practice as we move along the private-public continuum. The potentially limiting factor in terms of scholarly research is how amenable linguistic meanings are to quantitative-computational analysis. One aspect of this research is to explore the potential and the limitations of this type of computational analysis. Furthermore, this research is in part an investigation of the sensitivity of word-frequency based text analysis.

[2] Gooding, David C; 1989 (a): 63-82.
[3] Medawar, Peter B. 1987: 220-221.

1 Introduction and Summary

Abstract. This chapter aims to achieve several tasks. First I describe the issues and questions my research is concerned with. Next I discuss standpoints within the philosophy of language which inform this research. This is followed by two sections on Faraday, the first of which is a short biography and the second discusses language issues surrounding Faraday. The next main section explores relevant work done in the fields of analysis of scientific discourse and quantitative discourse analysis. I then discuss critiques of the quantitative analysis of language, followed by a brief look at how qualitative and quantitative aspects of research are conjoined in my research. Lastly, an overview of the contents of my other chapters is presented.

1.1 Research Issues, Questions and Theoretical Concerns

I am interested in: 1) the relationship between the production of new knowledge and ideas, the scientific techniques developed to produce and demonstrate them, and the language of description and argument, and 2) evaluating the effectiveness of new text analysis tools in studying this relationship.

Though there have been quantitative text-based studies of innovation in science,[4] there is a need for better methods for analyzing scientific innovation through the quantitative study of language change. Quantitative methods of text analysis have been used to classify different types of discourse,[5] but they do not address the ways in which language changes or is transferred as part of the process of scientific innovation. I am especially interested in the process of articulating new ideas and knowledge into a verbal medium of communication. It is widely argued that we know more than we can say, so language never completely captures what there is to a skilled performance.[6] However, there is scope for a close analysis of how the language of science changes as skills, technologies, and concepts develop over time. It is also important at the

[4] Leydesdorff, Loet; 1990, 1991, 1995, Callon, Michael and Courtial, Jean P.; 1989, and Callon, Michael ;1991.
[5] Gilbert and Mulkay; 1984.
[6] Collins, Harry M.; 1985 and Dreyfus, Hubert L.; 1986.

beginning to state that a collection of private and public writings, as I will be analyzing, will not provide an entire record of experimental activity or language use. This is because historical documents will inevitably be incomplete in three senses. First, every account of experimental work, whether laboratory notes or published papers, will be, to different degrees, a selective account. All writing is interpretation; too many factors, such as the intended audience, intervene to allow pure description to be anything but wishful thinking. As Gooding points out:

> "Private records are less rhetorical in character than published accounts, yet they already contain an element of reconstruction".[7]

Secondly, in a physical sense, pages or whole papers might be missing. A third form of incompleteness arises from omission of tacit knowledge[8] and taken-for-granted knowledge: that is:

> "[T]he techniques and technologies of scientific research are only made to work because of researchers' experience in trying to make them work. That is, they develop skills and understanding which are hidden or 'tacit' and which, without direct contact, are difficult to transfer to other scientists working elsewhere."[9]

Two principal questions are asked by this research. The first is whether it is possible to investigate the question: 'how do new expressions and ideas arise in language?', using quantitative methods of text analysis. The second question is whether it is possible to: i) expose the differences in laboratory activity and the experimental recounting in published papers,[10] and ii) describe and reconstruct the human activity of scientific work through the study of the micro-history contained in private writings. This is important because published papers do not tell the full, unpurged story of scientific work.[11]

The overriding question of this research is: 'how is language used as a medium for thinking and communication'? (given that language and thought are inseparably intertwined, and writing is part of thinking). Very pertinent to this research is Wittgenstein's assertion that:

> "The purpose of language is to express thoughts."[12]

---

[7] Gooding, David; 1989 (a): 64.
[8] Polanyi, Michael; 1964.
[9] Webster, Andrew; 1991: 44-45. Also see Collins, Harry; 1974 and 1985.
[10] Gooding, David C; 1989 (a): 63-82.
[11] Medawar, Peter B. 1987: 220-221.
[12] Wittgenstein, Ludwig; 1967 (third edition): 139[e] (remark 501).

Before leaving this part of the introduction I want to acknowledge that the digital files, as raw ASCII text, for Faraday's correspondence for 1819-40, were provided by Frank James of the Royal Institution. They were checked and put into a minimally indexed form by David Gooding of the University of Bath. The text analysis programs were written by David Gooding and Tom Addis, in the FAITH functional language.[13]

1.2 Philosophical Positioning

I want now to discuss the philosophical position on language which informs my research, which, broadly speaking, lies within activist[14]/constructivist and post-structuralist language theory. First I wish to briefly define a key term in this research, that of 'text'.

Texts[15] are not stable, natural structures, but are complex historical, psychological, affective, and cultural constructions, embedded in contexts. It can be said that: "versions of social life are actively constructed in different kinds of discourse".[16] The complexity and diversity of texts can be grasped by considering Saussure's description of language:

> "Taken as a whole, language is many-sided and heterogeneous; straddling several fields - physical, physiological, and psychological - it belongs both to the individual and to society; we cannot put it into any category of human facts, for we cannot discover its unity."[17]

This view of a text as complex and engaging with the reader, as almost 'organic' in nature, is expressed by Barthes, who believes that:

> "[A]ny text is an intertext; other texts are present in it, at varying levels, in more or less recognisable forms: the texts of the previous and surrounding culture. Any text is a new tissue of past citations. Bits of code, formulae, rhythmic models, fragments of social languages, etc. pass into the text and are redistributed within it, for there is always language before and around the text."[18]

---

13 The manual: 'Using FAITH Text Search & Analysis Programs' describes the programs developed with support from the Joint Research Councils' Initiative on Cognitive Science/HCI (1991-1993) and the ESRC (1994-1996).
14 Wavell, Bruce; 1986.
15 'Text' originates from the Latin word *textus*, which means 'something woven together'. Hodge, Robert and Kress, Gunther; 1988; 6.
16 Potter, Jonathan *et al*; 1984: 85.
17 Saussure, Ferdinand de; 1966: 9.
18 Barthes, Roland; 1981: 39.

No language is able to express all the complexities, subtleties, infinite variations and myriad particular situations of 'real life'; no language can express everything. If this was the case the result would be 'unspeakable' - a cacophony; there would be so much 'noise' that effective communication would be prevented. Thus we are unable to represent the full texture of things by language. This applies to art, literature, political representation, scientific theories, and so on. Similarly, this research rejects the view of language by logico-positivism which assigns a referential function to language; a claim is not true because the words correspond to the world. Instead, as Nietzsche believed, language is primarily figurative or rhetorical, not referential. This is just as true for scientific language use. Furthermore, meaning is never transparent and 'cleanly' communicated in language from sender to receiver. Through deconstructing the work of scholars Derrida has challenged the idea that a text has an unchanging, unified meaning, and attempted to show that language is constantly shifting.

Language is not a mirror of reality. There are no language-independent (or theory-independent) causal or factual relationships detectable by natural science.[19] Neither is scientific discourse neutrally representing 'natural' facts. Nietzsche said that when we are confronted with a text we should ask: 'Who is speaking?" In scientific papers it appears as if language is only the:

> "silent, cautious deposition of the word upon the whiteness of a piece of paper...where it has nothing to say but itself, nothing to do but shine in the brightness of its being"[20]

I argue that we cannot step outside of language, just as we can never stop communicating. This view of language is expressed by Humbolt:

> "Man lives with his objects chiefly - in fact, since his feeling and acting depends on his perceptions, one may say exclusively - as language presents them to him. By the same process whereby he spins language out of his own being, he ensnares himself in it; and each language draws a magic circle round the people to which it belongs, a circle from which there is no escape save by stepping out of it into another."[21]

This is important when considering the role of language in the construction of concepts and ideas, for they are never neatly packaged, fully articulated 'building-blocks' of reality represented in language. Rather concepts and ideas

[19] Rorty, Richard; 1991.
[20] Foucault, Michael; 1970: 300.
[21] Humboldt; Karl Wilhelm von, cited by Cassirer, Ernest; 1946: 9.

have histories and they emerge, and are communicated, through thought and language over time. Language and thought are not mechanical vehicles for the delivery of pre-given ideas, instead language, thought, ideas, and interpersonal and social processes are bound up in a creative interplay. This is one reason why is it considered a very central part of this research to follow the development of certain ideas and concepts for Faraday in his writing as a whole (that is, his personal and public writings), within a wider qualitative and analytical framework which brings to bear the (inter)personal, social, and historical conditions of Faraday's researches. Furthermore, I believe that with journal articles language is being used positivistically to create a direct connection between the author and the reader. The author is seen as communicating certainty and facts with no room for interpretation, thus the reader is enrolled and consensus produces 'truth'. Deconstructionists, such as Derrida, have challenged and abandoned the assumptions of such a direct connection, so multiplying the number of legitimate interpretations of a text.

Language, and more broadly, discourse do not only construct thoughts and ideas, but as Foucault believes, the author is a function of discourse itself:

> "In this sense, the function of an author is to characterize the existence, circulation, and operation of certain discourses within a society."[22]

There is a holism inherent in my research, in how science is perceived, and how ideas and concepts are constructed. This constitutes a rejection of the Cartesian, representationalist approach to reality. I argue against Sartrean dualism of subject and object, for which the:

> "world is not projected by a subjectivity, nor even by a language (as Humboldt argued) but by a system common to all languages, and a very simple one at that - binary discrimination."[23]

Thought and language do not exist in a state of discontinuity with the environment, nor with science or ideas. Lakoff argues that:

> "there is no unbridgeable gulf between language and thought on the one hand and the world on the other. Language and thought are meaningful because they are motivated by our functioning as part of reality."[24]

---

[22] Quoted in Lambropoulos, Vassilis and Miller, David Neal; 1987: 124-42, http://mh.cla.umn.edu/ebibjd1.html; accessed September 19, 1997.
[23] Surette, Leon; 1994: 3.
[24] Lakoff, George; 1987: 292.

The shuffling texture of the world is continually forming and unravelling, and language has an active role in ordering the world and mediating between people and 'reality', rather than existing as structure that represents a 'true' reality.[25] So our knowledge of the world is always fragmentary, approximate, and shaped by personal involvement and experiences. Knowing and knowledge are not in our minds from the beginning, but come about through experiences and language. Similarly, meaning and intentions are: "embedded in [their] situation, in human customs and institutions."[26] An experiment, just like a painting, does not exist prior to the performance; to the application of skills and imagination. There is no empirical orchard in which are displayed, ready for picking, firm and mature ideas and facts. As Mair remarks (discussing Polanyi; 1958):

> "[T]he logical premises of factuality are not known to us or believed by us *before* we start establishing facts. They are recognized rather by *reflecting on the way we establish facts.*"[27] (emphasis original).

We have to be aware of a danger that can emerge from this position. Too much reflexivity may cast: "the language dweller adrift in a Kafkaesque world of endless, meaningless displacement."[28] What you do not want is a "textless text" in the words of Geoff Dyer of the Observer, who, reflecting upon his experience of two of Umberto Eco's books remarks:

> "People may have bought Foucault's Pendulum by Umberto Eco with no intention of reading it but it was still a book in the sense that its nominal purpose was to be read. Foucault/ Blanchot was something else; the truth is that the last thing you would do with this book ... would be to read it. To have read it would have violated its essence. It was, if you like, pure signifier: an almost textless text whose meaning was inscribed in its virtual textlessness."[29]

In conclusion, the world of science and ideas come about through cultural, literary, and artistic modes of expression, through experiences and imagination, through the use of language. This is antithetical to the logico-positivistic belief in the existence of objective knowledge that is divorced from context and contingencies; from social and personal involvement. But it is very relevant to

[25] Wavell, Bruce; 1986.
[26] Wittgenstein, Ludwig; 1967 (third edition): 108[e] (remark 337).
[27] Mair, Miller; 1989: 236.
[28] Surette, Leon; 1994: 4.
[29] Quoted in Private Eye; Jan14, 1994: 9.

the study of the personal and public writings of Michael Faraday and Charles Darwin, which constitute the primary case-study material used in my research.

What is ironic is that computational methods and tools emerged out of the belief that knowledge could be isolated from its socio-cultural context and was thus rendered more objective and meaningful. Greenstein makes the point that: "historians in the US seem wedded to an entirely outmoded notion that computers necessarily and only entail number crunching".[30] This notion has resonance with the assumption that language is separable from content and context:

> "that it is little more than an embellishment that can be changed at will: the wallpaper, furniture and carpet adorning the rooms in an otherwise solid edifice, rather than essential to its architecture and fabric."[31]

The numbers and values computer software produce after scanning of a digitized text are only meaningful and have interpretative value within the wider qualitative context of the text production. Similarly, language construction and use are social processes mediating, and mediated by, a person's place in the world. In the next section I give biographical substance to the name 'Michael Faraday', from which I hope the reader gains a sense of Faraday's place in the world.

1.3 Michael Faraday: Scientist and Dissenter[32]

Faraday possessed a "meticulous, energetic personality".[33] There was an interplay between the "obsessional, rigid aspects" of his character and the ability to "freewheel in clearly defined spaces."[34] Faraday needed to create the psychological security afforded by an ordered, predictable, and safe environment. He found this in his basement laboratory of the Royal Institution,

---

[30] Greenstein, Daniel; 1996/97: 360.
[31] Woolf, Daniel; http://home.cc.umanitoba.ca/~sprague/woolf.htm#*; note 38; August 29, 1997; accessed October 4, 1997.
[32] I have depended chiefly upon, and am indebted to, four biographical studies of Faraday, namely:
Cantor, Geoffrey; Michael Faraday: Sandemanian and Scientist; MacMillan: London; 1991.
Jones, Bence Henry; The Life and Letters of Faraday; 2 volumes; London; 1870.
James, Frank A. J. L. (Ed.); The Correspondence of Michael Faraday, volumes 1-3; 1991, 1993, 1995, and Williams, Pearce L; Michael Faraday; Basic Books: New York; 1965.
[33] Lance J. W.; 1973: 78, Raskin, N. H. and Appenzeller, O.; 1980: 172-84.Quoted in, Cantor, Geoffery; 1991: 282.
[34] Cantor, Geoffery; 1991: 284.

with "the door firmly closed",[35] and under the dominion of God. What follows is a short biography of one of the most prolific and important experimental scientists of the nineteenth-century, who saw scientific discovery as a "major source of progress and truth for mankind."[36]

1.3.1 Childhood

Little is known about Faraday's childhood. What we do is that Faraday was one of four children, and was born on September 22, 1791 in Newington Butts, Surrey, during the turbulent times of the French revolution. His father, James Faraday (married to Margaret), a migrant blacksmith and a devoted Sandemanian in failing health, had moved earlier that year with his family to London from Mallestang near Kirby Stephen in Westmorland (now Cumbria), in the north of England. The family lived in small lodgings over a coachhouse in Jacob's Well Mews, Charles Street, Manchester Square. Faraday was too weak to follow his father's trade (Faraday was often hungry; sometimes surviving on one loaf of bread a week[37]). His formal education was virtually non-existent; he was practically self taught and it is believed Faraday was dyslexic. 'My education', he is quoted as saying, 'was of the most ordinary description, consisting of little more than the rudiments of reading, writing, and arithmetic at a common day school'.

1.3.2 Apprenticeship and a System for Learning

In 1804, at 13 years of age, Faraday was apprenticed to a bookbinder and bookseller, Mr. George Ribeau of 2 Blandford Street, a Frenchman and refugee from the French Revolution. It can be argued that this employment proved to be the motivation for Faraday to seek employment in science for two main reasons. First, Faraday did not find the experience of working with George Ribeau a convivial one,[38] relating later that:

> "My desire [was] to escape the trade, which I thought vicious and selfish, and to enter into the Service of Science, which I imagined made it pursuers amiable and liberal."[39]

[35] Cantor, Geoffery; 1991: 287.
[36] Cantor, Geoffery *et al*; 1991: 14.
[37] Lovie, A. D; 1992: 92.
[38] Cantor, Geoffery *et al*; 1991: 9.
[39] James, Frank A.J.L; 1991: letter 419.

The other reason was that while serving his apprenticeship Faraday's real education began. In the shop, between jobs, he had time to look at the books that came in and to begin reading some of them. As Faraday said to a friend many years later, the two most influential things he read at this time were parts of the Encyclopaedia Brittanica and Jane Marcet's Conversations on Chemistry:

> "I entered the shop of a bookseller and bookbinder at the age of 13, in the year 1804, remained there eight years, and during the chief part of the time bound books. Now it was in those books, in the hours after work, that I found the beginning of my philosophy. There were two that especially helped me, the "Encyclopaedia Brittanica," from which I gained my first notions of electricity, and Mrs. Marcet's "Conversations on Chemistry," which gave me my foundation in that science."[40]

As he learned to bind the books he also read them. Here we arrive at the first critical influence of Christian thinking on Faraday's energetic mind. For someone with no formal education the formation of an astute intelligence required a system. Faraday found this system in a book called *On The Improvement of the Mind* by an author whose name most Christians will still know today: Isaac Watts (1674-1748). Watts was one of the most prolific authors of evangelical Christian hymns of the early eighteenth century, writing more than 600 in all. Besides his Christian vocation as a one of the most influential dissenting ministers in London, Watts was also an accomplished philosopher (of the British empiricist tradition). In 1809 a new edition of his *On The Improvement of the Mind* was published, and in it Faraday found a common-sense guide to learning, with elaborate pointers on how best to benefit from lectures, reading, conversation, and observation. On such piece of advice was to keep a 'commonplace' notebook. In 1809 Faraday began to keep such a log book, in which he recorded all the good ideas he read about. Most importantly Watts provided a philosophical framework for learning that emphasized the importance of observed facts and warned against being 'too hasty to erect general theories from a few particular observations'. A philosophy that Faraday was to remain true to throughout his scientific career.[41]

---

[40] Faraday to Auguste De la Rive, September 2 1858.

[41] David, Gooding C. and James, Frank A.J.L.; 1985: 11; Hutchinson, I. M; http://venus.pfc.mit.edu/Faraday.html (*Background*); accessed September 10, 1997.

1.3.3 Faraday's Arrival at the Royal Institution

James Faraday died on October 30th, 1810. In the same year Faraday was introduced to the City Philosophical Society, which was founded in 1808 to help artisans and apprentices gain wider access to scientific knowledge. Two years later, aged 21, Faraday's great opportunity came when he was offered tickets to attend chemical lectures by Humphrey Davy[42] in London. Faraday went and later sent a bound copy of his notes to Davy asking for employment. In February 1813 the Chemical Assistant (or "fag and scrub") at the Royal Institution, William Payne, was dismissed by the Management ("the committee that used to run the Royal Institution"[43]) for brawling with the instrument maker John Newman. Faraday was offered the opening and was appointed Davy's personal assistant (Chemical Assistant) at the Royal Institution, on March 1st of that year, where he spent the next fifty years carrying out experiment after experiment.

1.3.4 1821: Marriage and Electromagnetism

In 1820 Hans Christian Oersted and André Marie Ampere discovered that an electric current produces a magnetic field. Faraday's own ideas about conservation of energy led him to believe, that since an electric current could cause a magnetic field, a magnetic field should be able to produce an electric current. The following year was to prove to be one of great achievement for Faraday, both personally and professionally. On June 12th, 1821 Faraday married Sarah Bernard (their marriage was to be childless) and into a prominent Sandemanian family. The following month, on July 15th, he made his 'confession of faith' before the Sandemanian church, and thereby entered into full membership of the congregation.

In the same year Faraday made one of his most important discoveries - electromagnetic rotations. He was also appointed Superintendent of the House at the Royal Institution. Though it was to turn out that 1821 was to bring one of Faraday's most distressing incidents. A charge of plagiarism was brought by

[42] Professor of Chemistry at the Royal Institution, 1802-1813. Knighted in 1812. Secretary of the Royal Institution, 1807-12, and President 1820-27.

Humphry Davy. Faraday was accused of not accrediting enough William Wollaston's role in the discovery of electro-magnetic rotations. One effect of this was to estrange Faraday from wider society. This spectre loomed again in 1834 when Sir John Davy accused Faraday of plagiarizing from Sir Humphry Davy.

1.3.5 The Royal Institution and Public Communication of Science

Royal Society elected Faraday as a Fellow on January 8th 1824, and then in 1825 he was appointed Director of the Laboratory of the Royal Institution, when his colleague, W.T. Brand, took up a post at the Royal Mint. The Institution's famous series of 'Friday Evening Disclosures' were begun by Faraday a year later, and they continued until 1862, Faraday giving in excess of 100 discourses. He was very committed to bringing the results of science to the wider public:

> "Faraday held a firm conviction that the truths of science were not just for the *cognoscenti* but had to be shared with others, one reason being that public lectures 'facilitate our object of attracting the world, and making ourselves with science attractive to it.'"[44] (emphasis original).

Most notably this commitment took the form of public lectures and scientific demonstrations. These lectures were designed to communicate the newest scientific discoveries to the public, and later, in 1827, Faraday started giving scientific lectures for children, of which he gave nineteen.[45] Faraday gave Friday evening discourses until 1841. Another manifestation of Faraday's belief that science "should not be the sole property of the scientific community"[46] was his publication of science books for children.

1.3.6 1830's: A Decade of Great Achievement

The 1830's were Faraday's most productive decade in terms of research. In 1829 Humphry Davy died removing, many believe, an impediment to the advancement of Faraday's career. Faraday's research position became increasingly independent due to rising income. During this time Faraday

---

[43] James, Frank A. J. L; 1997: 279.
[44] Cantor, Geoffery; 1991: 147.
[45] Hutchinson, I. M; http://venus.pfc.mit.edu/Faraday.html (*The Brotherhood*); accessed September 10, 1997.
[46] Cantor, Geoffery; 1991: 295.

showed that electrochemical reactions follow all normal chemical stoichiometric relations, but in addition follow certain stoichiometric rules related to charge. These are known as Faraday's laws of electrolysis. Then on August 29th 1831, Faraday demonstrated the effect of electromagnetic induction. Though this did not happen 'out of the blue' for Faraday's electromagnetic experiments were part of a larger 'network of enterprises', a term coined by Gruber.[47] Faraday's first attempts to find electromagnetic induction are written in his notebook of 1822, where he outlined an experiment similar to ones he would conduct in 1831.[48] The principle of induction made possible the dynamo, or generator, which produces electricity by mechanical means. In November of that year Faraday proposed the concept of 'lines of force' to help explain how magnetism could produce electricity. Faraday did keep working relentlessly into the 1840's, and in 1845 he demonstrated the magneto-optical effect and diamagnetism. This led Faraday to formulate the field theory of electromagnetism.

Other achievements during this decade included being made a Deacon of the Sandemanian church July 1832. Between 1832 and 1836 Faraday produced seven volumes of the *Diary*. His humanitarian side was evident when he became a subscriber to the London Orphan Asylum in 1831. A pension was offered by the Whig Lord Melbourne, which Faraday declined, in October 1835. The following year Faraday was appointed Scientific adviser to the Corporation of Trinity House to conduct experiments in light. In 1839 the first volume of his *Experimental Researches in Electricity* was published (three in all were published: 1839, 1844, 1855, and a fourth was started).

1.3.7 Sandemanianism

Historians have argued that science is opposed to theology, but the examples of Newton, Pascal, and Faraday effectively refute this viewpoint. Here I want to consider Faraday's commitment to Christianity. Throughout his long and productive life, Michael Faraday was also a committed Christian fundamentalist. He was not a social church-goer and belonged to a distinctly nonconformist

---

[47] Gruber, Howard; 1981.
[48] Gorman, Micheal E.; http://jefferson.village.virginia.edu/~meg3c/id/id_disc/id_disc_13.html; accessed September 11, 1997.

denomination, which: "demanded from its members an extremely high level of commitment and devotion".[49] Faraday was a member of a small fundamentalist sect called the Sandemanians, who were fully committed Christians who grounded their religious faith on a literal interpretation of the Bible.[50] They adopted the Bible: "as their sole authority in all issues of principle and practice",[51] took with the utmost seriousness Paul's plea (1 Cor 1:10):

> "I appeal to you brethren, by the name of our Lord Jesus Christ, that all of you agree and that there be no dissensions among you, but that you be united in the same mind and the same judgement."[52]

Faraday believed:

> "[T]hat in his scientific researches he was reading the book of nature, which pointed to its creator, and he delighted in it: 'for the book of nature, which we have to read is written by the finger of God'."[53]

It is important not to underestimate the influence of Sandemanianism on Faraday's life and work for he viewed his Sandemanian membership and its Christian beliefs, practices and fellowship more important than his career in science.[54]

The Sandemanian, also known as Glasite, church, arose from the experiences of a Scot, John Glas (1695-1773). Glas, a thoughtful minister in his Church of Scotland parish near Dundee, found it increasingly difficult to reconcile his understanding of the scriptures with the political, and national, role of the established (covenanting) church.[55] The covenant was anti-Catholic in intent and its acceptance:

> "amounted to a proclamation of allegiance to both Church and State, and a pledge to extirpate popery, prelacy and dissent."[56]

In 1725 he and nearly 100 members of his congregation joined together to found their religion on the Bible alone and reject the political covenant. Glas

---

[49] Hutchinson, I. M; http://venus.pfc.mit.edu/Faraday.html (*Introduction*); accessed September 10, 1997.
[50] Cantor, Geoffrey; 1991: 10.
[51] Cantor, Geoffrey; 1985: 70.
[52] Quoted in Hutchinson, I. M; http://venus.pfc.mit.edu/Faraday.html (*Nonconformist*); accessed Septemer 10, 1997.
[53] Hutchinson, I. M; http://venus.pfc.mit.edu/Faraday.html (*Philosophy and Nature*); accessed September 10, 1997.
[54] Cantor, Geoffery; 1991: 72.
[55] Cantor, Geoffery; 1991: 19.
[56] Cantor, Geoffery; 1991: 19.

was challenging the legitimacy of the Church Of Scotland in respect of the words of God.[57]

The independent congregation that Glas subsequently founded in Perth attracted Robert Sandeman (1717-71), who eventually married one of Glas's daughters. A partner in a flourishing weaving firm, Sandeman spent a great deal of time involved in work for the church, and, through the publication of his *Letters on Theron and Aspasio*, which became a key source for Sandemanian teaching, became its most influential spokesman.[58] The church order created was one of: "rule by elders, independent congregations, pastoral function of elders and mutual edification."[59]

The gradual decline of the Sandemanian church had begun by the time Faraday made his 'confession of faith' in 1821. Sandemanians considered themselves set apart from the world. An impression that is enhanced by a remark Faraday made to Ada, Countess of Lovelace, that he belonged to: "a very small and despised sect of Christians, known, if known at all, as Sandemanians".[60] What the Sandemanians wanted to do was to live out a Christian life based on a literal interpretation of the New Testament, faithful to orthodox, primitive Christianity, and unfettered by the political accretions of the established churches.

Faraday would leave the Royal Institution every Sunday morning and Wednesday evening to go to the meeting house in the Barbican.[61] He would have been accustomed as a child to the: "long, even tedious services, the prayers and extensive quotations from the Bible."[62] Among the London brethren he performed numerous pastoral duties, 'such as visiting those in need and tending to them, both materially and spiritually.'[63]

Faraday's membership of the Sandemanian church was not untroubled. Faraday became an Elder in the church on October 15 1840, until four years

---

[57] Cantor, Geoffery; 1991: 20.
[58] Cantor, Geoffery; 1991: 22-24.
[59] Garrett, Leroy; http://www.dm.net/~ginb/scotland.htm; accessed September 10, 1997.
[60] Cantor, Geoffery; 1991: 34.
[61] Cantor, Geoffery; 1991: 65.
[62] Cantor Geoffery; 1991: 58.
[63] Cantor, Geoffery; 1991: 66.

later, on March 31 1844, when he suffered exclusion from the London congregation for a brief period. The reason usually given is that his brethren thought that Faraday demonstrated too little remorse when he was absent on a Sabbath after he accepted Queen Victoria's invitation to dine at Windsor. Also Faraday was reinstated a few weeks later (exclusion lasted from March 31 to May 5) but did not regain his position of Elder for another sixteen years. This exclusion affected Faraday grievously, not only bringing him "low in health and spirits,"[64] but fearful of a second, and final, exclusion. Faraday came to the brink of such an exclusion in Autumn of 1850 and he believed that: "this precipitated a mental crisis."[65] This situation arose out of Faraday's personal interpretation of his exclusion which was disharmonious with the: "sect's disciplinary code and the biblical passages on which it is based."[66] Thus:

> "the sect's strict discipline over exclusion and Faraday's worry that he might be excluded from the church threatened to precipitate a second exclusion."[67]

Crisis number three occurred in the summer of 1864 directly related to Faraday's science. He was offered the Presidency of the Royal Institution and this: "offer raised the spectre of the much-feared second exclusion."[68] The result of this crisis was both his refusal of the Presidency and resignation of his Eldership.

1.3.8 Chronic Illness

The years 1839 to 1845 were a fallow period in terms of research. It is believed that one reason for the lack of scientific activity relative to the 1830's was the start of progressive illness in 1831 (another reason was his duties as Elder). Faraday's symptoms included depression, giddiness, forgetfulness, and a great aversion to company. From early days Faraday suffered from poor memory, dizziness and headaches. This is one reason given for Faraday recording events in a meticulous manner.

---

[64] Faraday to Christian Friedrich Schoenbein, April 12 1844, in James, Frank A. J. L; 1993; letter 1575.
[65] Cantor, Geoffery; 1991: 275.
[66] Cantor, Geoffery; 1991: 276.
[67] Cantor, Geoffery; 1991: 277.
[68] Cantor, Geoffery; 1991: 279.

The year 1836 saw Faraday suffering great pain in his right knee. His doctor, Benjamin Brodie, advised complete rest, so Faraday went to the Isle of Wight to recover. But 1838 to 1840 was a period of sustained illness. The giddiness and malaise were not thrown off until 1845, though from 1839 Faraday experienced severe and almost unrelenting headaches, and he suffered his first nervous breakdown in 1839. A sustained loss of memory is thought to have been the consequence. In the Summer of 1841 Faraday and Sarah travelled to Switzerland where Faraday tried to regain his health. The winter of 1849 to 1850 saw Faraday suffer from an almost continuous sore throat. The cause was found to be his teeth, for which he received dental treatment in 1850.

There has been much speculation about the cause of Faraday's Illness, with no consensus or decision having been reached. It is probably safe to say that a combination of social, psychological, and chemical factors conspired to destroy Faraday's health. If we look broadly over Faraday's life we know that he hovered on the brink in 1840, 1850, and 1864, as the result of crises precipitated by exclusion from the Sandemanian church. He was accused of plagiarism twice, in 1821 and 1832, and was ostracized. In terms of his work Faraday was pushing against a very conservative and entrenched science; he did experience distaste for his views and perspectives, and his researches demanded tremendous concentration. Isolation may have been a stressor, in terms of the nature of and idealism inherent in his work, and possibly in terms of social class. Chronic chemical poisoning may be attributed to Faraday's illness for he did work with a lot of mercury, as well as handling other chemicals such as benzene based anilines, hydrochloric acid, nitrogen tetrachloride, and nitrous oxide.[69]

1.3.9 The Final Years

In 1855 Faraday ceased working due to declining mental powers, though he continued as a lecturer for another six years. He refused the presidency of the Royal Society, when it was offered to him in 1857, and that of the Royal Institution in 1864. Faraday submitted his final paper for publication on April 16, 1860. This year also saw him returned to the role as Elder of the Sandemanian

[69] Cantor, Geoffery; 1991: 281.

church, which he subsequently resigned in 1862. Two years before his death Faraday resigned his position as Superintendent of the Royal Institution and left Trinity House.

Faraday spent a lot of his time during his last years at the house in Hampton Court bestowed on him by Queen Victoria. The years 1865 to 1867 saw a slide into senility. Faraday died on August 25, 1867, a pensioner of Queen Victoria, in his chair in his study at Hampton Court. The funeral was described by Faraday's niece, Jane Barnard, as "a 'strictly private and plain' affair",[70] at which the scientific community was requested by the family to stay away.

1.4 Faraday and Language

Faraday was very passionate about his experimental work. It was his life. His use of imagination in science was revolutionary;[71] what was especially important for Faraday was that: "experiment [was] a check on the imagination, not [a] substitute for it."[72] The apotheosis of the reign of the inductive style in scientific writing came about in the mid-1800's.[73] Faraday was in no little way responsible for the passing of the inductive style. His use of imagination, his innovative use of language at a time when scientific language was highly standardized,[74] and his heretical argumentative or 'dialectic' writing style[75] - "action and reaction" as Faraday called it[76] - all inveighed against the inductivism.[77] Faraday was writing his notes (and *Experimental Researches in Electricity*) in the style of a diary; "his observations are used as arguments".[78]

The importance of Faraday's writings to my research is that Faraday was a highly articulate 'hands-on' scientist, whose use of scientific language was

[70] Jane Barnard to Bence Jones, 3 September 1867, in Jones, Bence Henry; 1870(b): 482, in Forgan, Sophie; 1985: 51.
[71] Tyndall, John; 1870.
[72] Agassi, Joseph; 1971: 126.
[73] Agassi, Joseph; 1971.
[74] Agassi, Joseph; 1971: 117.
[75] Faraday was accused of plagiarism in 1821 and 1834, both of which is thought to have contributed to his development of a new style of scientific writing (Agassi, Joseph; 1971).
[76] Agassi, Joseph; 1971: 125.
[77] Faraday gave inductivism one final chance in 1834 in a paper on static electricity after a four-year period of absence from research due to depression and illness (Agassi, Joseph; 1971).
[78] Agassi, Joseph; 1971: 125.

innovative, and whose writing style was no less innovative and mould-breaking. We create the realities through the linguistic practices we perform and live within. These practices include how we place words together and how we signal and alter meanings. It is no different in science; one of our created and contingent realities. All ideas and concepts have histories and are given form through language. This research will try to demonstrate how Faraday created his discovery of electromagnetism through the way he communicated his ideas and thoughts in language and writing.

1.5 Studies in Discourse Analysis and Their Relevance to my Research

I shall now discuss some relevant work done in the fields of analysis of scientific discourse and quantitative discourse analysis by Badalamenti,[79] Bolton and Roberts,[80] Gilbert and Mulkay,[81] Gooding,[82] Holmes,[83] Shapin,[84] and Zahar.[85]

1.5.1 Badalamenti and Rules of Language Use

In 1994 Badalamenti *et al* published a quantitative-mathematical study of word usage. This study was designed to determine whether there are mathematical laws of word usage and whether this: "reflects the existence of deep structures in the human mind and brain." (p46). Thus the main aim was to study the temporal properties of word usage and test the hypothesis that the emergence of new words is a Poisson process (a model for the occurrence of random events). The authors believed that the progression of words in a poem constitutes a type of time series.

Their source material consisted of primary (seven poems and one prose segment) and secondary material (two poems, opening pages of a scientific paper, and an "emotionally charged dialogue").[86] Each text was converted into a

---

[79] Badalamenti, Anthony F. *et al*; 1994: 46-71.
[80] Bolton and Roberts; 1995: 295-302.
[81] G, Gilbert Nigel and Mulkay, Michael; 1984: 105-125.
[82] Gooding, David C.; 1990; 1992: 65-112.
[83] Holmes, Frederick L.; 1981: 60-70; 1984: 131-142; 1985; 1986: 19-35; 1987: 220-235.
[84] Shapin, Steve; 1984: 125-130.
[85] Zahar, Alexander; 1995.
[86] Primary sources: Lord Byron: *She Walks In Beauty*; Shakespeare: *Sonnett VXIII*; William Henry Earnest: *Invictus*; William Wordsworth: *Sonnett XIV*; Edgar Allan Poe: *Annabel Lee*; Samuel Taylor Coleridge: *Kubla Khan*, or, *A Vision In A Dream, A Fragment*; Lewis Carroll:

word sequence. The first occurrence of each word was assigned a number (integer), and its subsequent occurrences replaced by this number. A change in tense was only scored differently if it involved a change in meaning. For example, Frost's use of "had" in line 5 of stanza 2 and "has" in the last line of his poem (p53). Thus integers were mapped to word position. The distinctiveness of meaning was considered a priority and was preserved in three ways:

i) the mapping of words to numbers
ii) homonyms were scored according to usage, for example, love as a verb and a noun
iii) compound words scored as single words.

The principal measures used are the waiting times between new words and complexity (informational entropy[87]) of word usage. The number of words and the number of distinct words were counted, and their ratios calculated. Three models were applied to the sequence of numbers:
i) statistical: word length, number of distinct words, the ratio of distinct words to word length, visit mean/sigma ("the average number of times a word is used and the standard deviation of it use"), and revisit mean/sigma ("the average number of times a word is reused, given that it is used more than once, and the standard deviation of its use") (p53).
ii) mathematical: the Shannon entropy and regression analysis.[88]
iii) stochastic: the Kolmogorov Smirnoff test.

Histograms were constructed for the 'waiting times': "the number of words between the current new word and the next new word. Thus the position measures "time"." (p53). The Kolmogrov-Simirnoff criterion was used to estimate the rate constant lambda (a measure of the probability that a next word

---

Alice's Adventures In Wonderland, Chapter 1, Down The Rabbit Hole. Seconday sources: Emily Dickinson: *A Bird Came Down The Walk*; Julia Ward Howe; *Battle Hymn of the Republic*; opening pages of a scientific paper; "an entire emotionally charged monologue" (p58).

[87] The Shannon definition of entropy was used: $E(n) = -\sum_{i=1}^{i=2400} p_i \ln(p_i)$ Where: "p_i is the probability of event i (here the ith distinct word) at time n (word position n into the text. E(n) denotes the estimate of entropy at word n into the work." (67-68).

[88] The statistical package BMDP statistical package was used to compute histograms and regressions.

will be a new word). The use of Shannon entropy is a measure of how: “does the *variety* of the words selected by an author - the use of both new words and repetitions of previously used words - grow with the unfolding of the poem?” (p53).

The author’s findings were fourfold as follows:
1) Statistical: correlations exist between the word length and the number of distinct words.
2) Stochastic: the histograms of waiting times for the appearance of new words represented a negative exponential density, indicative of a Poisson process.
3) Mathematical: entropy regressions reveal a logarithmic growth of the complexity of word usage.

Thus, the overall conclusion was that the generation of new words for each poet is a Poisson process and entropy grows logarithmically, but these findings are independent. Thus in human creativity there is a ‘psychobiological’ urge for the innovative use of new words.

Two conclusions in particular seem very useful to add to the pool of approaches for text analysis. The first is that the introduction of new words slows as a work progresses. I would add that bursts of new words appear not only at the beginning of a text but also when a new concept or idea is introduced by the author. The second conclusion is that there is a ‘psychobiological’ urge to use a fresh word. Thus the longer it was since a new word was used, the greater the probability that the author’s next word will be a new word. This is presented as a measure of creative urge (the need to use a new word or phrase, a feeling which increases in urgency the longer a new word is avoided) by Badalamenti *et al*, who concluded that the creative urge is greater in poetry than in prose.

A general difference with my current text analysis research is the importance attached to social, historical, and cultural issues in language use, and more specifically, in the development of scientific communication and ideas. In contradistinction to Badalamenti, the wider contextual circumstances of discourse - the cultural, sociological, and historical settings are considered when experiments are being designed and when findings are interpreted.

This research studies the emergence or progression of words in poetry. This approach is something my current research is interested in with regard to scientific texts. The methods used to explore this emergence were to: i) calculate the "waiting time"[89] between new words and ii) answer the question: 'how does the *variety* of the words chosen by an author grow with the unfolding of a text.' The measure used here was Shannon entropy. It would be useful to incorporate into my work the first method used to characterize the emergence of new words. In particular, exploring the issue of how certain words appear and disappear over time, and also how some words come to fall into certain fields or categories. One theme my research is concerned with is the linguistic characteristics of scientist's reconstructions as their writings move from the personal into the public idiom. I believe that the non-linear and personal reconstructions would demonstrate the least waiting time, with the longest waiting time being for the reconstructions in the public idiom.

Two other measures Badalamenti *et al* use are the *mean visit time* ('the average number of times a word was used') and the *mean revisit time* ('the average number of times a reused word was reused'). I would hypothesize that these averages would be highest in a public scientific text or if Faraday was discussing a scientific matter in a letter. The former measure is similar to the CLARITY's compression score function, which calculates the degree of ellipsis as a percentage of paragraph size.

It would also be interesting to test Badalamenti *et al's* conclusion regarding the measuring of the creative urge. As my research investigates scientific language use in personal and public idioms this test could be applied to Faraday's personal writings, such as his correspondence, and, for example, his published researches. Similarly this measure could be used to test Faraday's creative urge when he discusses a topic, for example, electromagnetism, in both personal and public writings.

Finally, in terms of corpus size, a study of the unfolding of the author's word usage over time would be more interesting and revealing, that is, a large corpus

89 Waiting time is: "the number of words between the current new word and the next new. Thus position measures "time". (p53).

of the writer's work over a period of years as is the case with the Faraday material.

A criticism of Badalamenti's work is that it seems to consider language as a closed physical system. The author draws on Chomsky's theory of language as an abstract system divorced from wider communicative and social interaction.[90] But there is no such thing as a "completely homogenous speech-community".[91] It seems that this work does not rebut the traditional positivistic notions of knowledge and language which assume that: "the objective world can be known directly and this knowledge remains separate from language and the theories articulated in language."[92] This is unfortunate when text analysis and computational linguistics are increasingly looking at ways to analyze natural language directly.[93] All communication and language exists in a social and cultural context in which they create knowledge, and this context directly influences the creation of that knowledge and the development of technical discourse.

1.5.2 Zahar and Constructivism

I now want to look at research that uses the same constructivist philosophy of language that I do. This is a study by Zahar on language in science the primary aim of which was to refute the *dualist* model of language and support the *constructivist* conception of language.[94] The former view has dominated mainstream, analytical philosophers of science and the AI community who see language as a linear series of narratives, sequential in construction. Here 'language' and the 'world' are seen as separate, because language is perceived to only describe a world of pre-given objects or facts. This is assigning a referential function to language, where: "meaningful statements are those which describe a state of affairs in the world."[95] Dualists maintain a relatively static theory of language development. They tend only to study monographs or

---

[90] Fowler, Roger; 1991: 25-32.
[91] Fowler, Roger; 1991: 27.
[92] Dombrowski, Paul M.; 1994: 7.
[93] See, for example, Rank Xerox's Language Tools Web site: http://www.rxrc.xerox.com/research/mltt/toolhome.html; accessed January 14th, 1997.
[94] Zahar, Alexander; 1995.
[95] Widdowson, Peter and Selden, Raman; 1993: 148.

textbooks in which the active processes and skilled aspects of theoretical and experimental work are not conveyed.

Languages are not representations of the world; language does not describe "the way the world is" nor can it seen as "fitting the facts".[96] The contingency and opaqueness of language is argued by philosophers such as Davidson, Goodman, Putman, and Wittgenstein, and is found in the work of philosophers of science such as Hesse and Kuhn. Similarly, it must be argued that words cannot exist in isolation, they are not separate linguistics units of meaning, instead words: "constitute a special medium of intercollective communication", and are coloured as they conform:

> "to a given thought style,[97] a character which changes during their passage from one collective to the next, always [undergoing] a certain change in their meaning as they circulate intercollectively."[98]

The constructivist theory of language is potentially very useful for text or literary-analysis. It informs us that you cannot just passively lift meaning and knowledge from texts. This is the case whether human-based or computer-based methods of text analysis are employed. Texts do not contain unassailable objective facts, which affect the reader in an unmediated sense, *nor* are readers able to make what they will of a text. Willard McCarty makes the point that for almost all of us: "a text is thoroughly assimilated to our shared and private mental worlds as we read". [99] When analyzing meaning of text we must bring intertextuality to bear. That is, a text is not isolated from the world but is a part of a wider human, cultural, and linguistic process. This is no different when considering the activities of science.

Language is seen as a shared, creative tool through which the world has existence. In terms of science, language is but one of the 'many activities and skills associated with experimentation'.[100] Thus, language is part of experimentation: a means by which 'scientists communally make sense of their

[96] Rorty, Richard; 1989: 20.
[97] "the readiness for directed perception and appropriate assimilation of what has been perceived." Fleck, Ludwig; 1979 [1935]; 142.
[98] Fleck, Ludwig; 1979 [1935]: 109.
[99] McCarty, Willard, http://www.kcl.ac.uk/kis/schools/hum/ruhc/wlm/essays/litmech.html; November 1995; accessed January 14th, 1997.
[100] Zahar, Alexander; 139.

experimental experience.'[101] Scientific practices use verbal behaviour or language to construct a 'real' world of objects; this 'real' world is situated as separate from language by logico-positivists.[102] The view of language applied in my research emphasizes non-linearity, that language is slippery and indeterminate, and, as Wittgenstein stressed, the meaning of a word is sought in its use, and language is not simply a matter of rule following:

> "One cannot guess how a word functions. One has to *look at* its use and learn from its use."[103]

I believe that Wittgenstein's view helps us with constructivism because it supports the premise that language and words do not exist separate from the world. If we want to understand a meaning of a word we need to understand how the word is used by the author within the wider linguistic, interpersonal, and social environments. In his book *Textual Power*, Robert Scholes talks about little social and cultural 'earthquakes' that disrupt and alter perceptions of meaning and methods of interpretation.[104] This highlights the important issue of how meanings are embedded in social and cultural practices, which are themselves not givens but constructed and maintained - or allowed to atrophy - through representation and use.

The use of verbal behaviour reaches its zenith in the peer-reviewed scientific paper. This is where all pre-verbal behaviours are filtered out and knowledge is ossified in language to be presented as 'facts'. Thus we see the dualist disjunction of 'language' and 'world'; the paper exists as the culmination of scientific investigation and a repository of certain knowledge. If one took this view then research would be aimed at reconstructing the investigative pathways *lying behind* published papers. My perspective views the scientific paper as a *part* of experimental practice and a literary effort.

Let me briefly enlarge on the latter part of this last point. Literature studies analyse texts as 'literature' proper, that is, for its form, content, style, and meaning.[105] Scientific texts are not passive repositories of facts, they are not the

---

[101] Zahar, Alexander; 1995; 2.
[102] for example Collins, Harry M; 1985, Gooding, David C.; 1990, Lynch, Michael; 1985.
[103] Wittgenstein, Ludwig; 1967 (third edition): 109[e] (remark 340).
[104] Scholes, Robert E.; 1986.
[105] Holmes, Frederick L.; 1987: 220.

endpoint of a linear process of objective reasoning, they do not just represent experimental skills and results. Scientific texts are writings, they are a creative achievement and cultural artefact; language use and thinking are integrated, and are also dynamic features of experimental practice. Scientific writing is a genre of literature. Like other forms of literature, not only has scientific writing form, content, style, meaning, and affect on its audience but these features are interconnected.

Such a perspective on scientific writing is a useful window through which to view the construction of ideas and concepts. Not only are the four features of literature interconnected as an integral whole, but this whole is constituted by - and partially constitutes - a mosaic of other texts, thoughts, ideology, and intended audiences. It is the same with ideas and concepts. They always have a pre-history and are links in the cyclical process of creation and re-creation. Despite how they seem to be represented in scientific papers, ideas and concepts are not static objects frozen in language.

I believe the constructivist model provides an important framework for interpreting the perceptions and articulations of new ideas and concepts. My first point is that ideas and concepts are products, and not static objects frozen in language. Ideas are not composed of *a priori* concepts, rather they are fluid and open to revision. They do not exist exclusively in one's mind; they are not a private activity or investment. Instead they are tied up with preconceptions and are the outcome of interpersonal, linguistic, and social processes.

Textual evidence for this emerging property of ideas and concepts is change both in words themselves (including the abandonment of words) and in the use of words. As the scientist thinks about an idea or concept the thoughts are expressed in language, in the written word. This act of using language and solidifying thought in writing affects the content of the idea or concept. Change in words and word use will be seen across private and public writings, and whether the language is being used to argue, persuade, or describe, for example.

The constructivist approach teaches us that an idea or knowledge is not found solely in the text. To understand how an idea is being framed and articulated we must look at the social, cultural, and linguistic communities within which the idea, and any accompanying text, is embedded and in interaction. The constructivist views language and reality as intertwined. Similarly the writing of papers is an active part of a cyclical process of scientific investigation and not a discrete end-point of scientific enquiry. The writings feed back into thoughts and ideas about other experimental work.

This research will take this constructivism a stage further. By tracing the development of concepts and language change within the 'context of discovery' and along the private-public continuum it should also be possible to reconstruct discovery narratives in terms of tracing the history or biography of an idea or concept as it evolves outside of the laboratory or lecture theatre. Here we have the belief that scientific facts, objects, and theory go beyond what has been empirically and experimentally proven. This allows for various processes of persuasion, such as emotions, group dynamics, economic changes, and cognitive factors: "[f]eeling, will, and intellect all function together as an indivisible unit."[106] Thus not all procedures are apparent; Hegel believed that theories conceal as they reveal. This research views experimentation as a skilled performance where practical measures and theory construction are intertwined.

My research is interested in the functions of language and Zahar presents a classification of language functions based on the five 'inter-connected-spaces' Gooding uses to accommodate his maps of cognitive action,[107] which I wish to analyse. The purpose of Zahar's classification is to:

> "elucidate interconnecting and overlapping functions of language associated with Gooding's two 'non-linear' levels of reconstruct-ion".[108]

These two 'non-linear' levels are the 'cognitive' and the 'demonstrative' from a set of six kinds of reconstructive activities of language.[109] For Gooding reconstruction is what:

---

[106] Fleck, Ludwig; 1979 [1935]: 94.
[107] Gooding, David C.; 1990: 16. The five 'spaces' are: 'mental', 'computational', 'social', 'physical (laboratory)', 'and 'rhetorical'/'literary'.
[108] Zahar, Alexander; 1995: 142.
[109] Gooding, David C.; 1990: 7.

> "is needed to produce an account ordered enough to enable further action, the communication of what is going on and the redefinition of problems."[110]

The two non-linear levels of reconstruction reflect the earliest stages of experimental activity, before the (inter)personal, social, and cultural processes which are integral to experimental work are filtered out of a scientist's recounting as:

> "he strive[s] to build a world conforming to his chosen concepts and obeying his universal laws."[111]

What we get from Zahar's classification is the total immersion of events in language. This is quite antithetical to the ideas of the separation of language and the world, and of the referential nature of language. Instead language is seen as a creative tool in the construction of phenomena, of allowing the phenomena to be rendered intelligible, communicated to, and shared by others. Interpretations of experimental work are reinterpreted through language to fit with the theory, and to be seen as the product of abstract hypotheses or theories. This is very relevant to my research. Phenomena, ideas, concepts are not things-in-themselves, they are contingent social and human products which emerge and change shape as we interact with language.

1.5.3 Language as Abling and Disabling

Before leaving the discussion of constructivism I want to address a use of language which seems to be missing from the constructivist philosophy of language and Zahar's classification, but implicit in Gooding's six types of reconstruction of scientific accounting. Language not only enables but it also disables and constrains; language, like theory for Hegel, conceals as much as it reveals. The use of language is never neutral, disinterested, and it is situational. Meanings slip and slide; they are always shaped around particular social groups, agendas, and beliefs: "meaning is infused with forms of power."[112] Language can be used to promote or discourage egalitarianism in communication. For example, language can enable description, communication, persuasion, criticism, understanding, ordering, comparison, ostension (pointing

[110] Gooding, David C.; 1990: 6.
[111] Goodman, Nelson; 1976: 15.
[112] Thompson, John B.; 1984: 254.

and labelling), and self-knowledge. On the other hand, language use can obscure meaning, distort communication, and reify. We never participate in a discourse where everyone is aware of everyone else's perspectives and interpretations.[113] Language is fallible and never absolute: either as a medium of freedom or as a medium of enslavement. In 'traditional' or 'ordinary' language the context is provided by the social and linguistic community, by traditions. In scientific-technical language the context is shaped by abstractions; the individual speaker is shunned. The more a language loses its context and content the closer it gets to being propaganda.[114]

1.5.4 Holmes: Unpublished Writings and Thinking

Another researcher whose work on scientific writing and creativity informs my research is Frederick Holmes, who illustrates, using the example of Lavoisier's writings about his experiments on respiration, that the process of writing is an invaluable tool for working through and modifying ideas.[115] There is a dynamic and intimate relationship between thinking and language use and change. Holmes talks about the: "clarifying effects of scientific writing on....scientific thought."[116] This is one important reason for studying personal and public scientific writings. In the personal writings - notebooks and letters - we can recover and reconstruct the investigative pathways and revisions as ideas and concepts become visible in their public, reified form; the messy thought processes, epistemic cul-de sacs, and errors hidden from view. The selection of the portions of inquiries and ideas to construct self-contained concepts and neatly packaged solutions is a highly creative and human process.

Holmes talks about the usefulness and problems of using private writings to reconstruct historical accounts of scientific investigations. He points out that different documents can give different perspectives on the nature of the investigation. Unpublished writings, such as correspondence with colleagues, can reveal what a scientist was thinking, though a complete picture is not usually possible, because, as Holmes writes:

---

[113] Habermas, Jurgen; 1990.
[114] Lovekin, David; 1991.
[115] Holmes, Frederick L.; 1987: 220-235.
[116] Holmes, Frederick L.; 1987: 225.

> "letters are usually too infrequent to fill in all of the critical steps along the trail, and they tend to capture only those selective aspects of the problem about which the investigator feels a need to consult other texts."[117]

Direct access to a scientist's thoughts are less likely for laboratory notebooks, Holmes states:

> "for they most often do not give the reasons why he performed the particular operations recorded; but they do typically provide a steady, full record of the investigation itself, from which the historian has a good chance to infer the rationale that guided it."[118]

The important point to bear in mind is that historical records are very likely to be incomplete. Thus when studying Faraday's work on electromagnetism a qualitative-interpretative 'piecing together' across the broad plain of a variety of private documents, correspondence written for publication, and published work is an important context of my analysis.

Holmes also makes the very interesting assertion that creativity is not found just at one point but is found infused throughout the work of a scientist. This is another very good reason for studying the construction of ideas and how they are communicated in language across the private and published writings of a scientist. According to Holmes:

> "A great scientific investigation is not captured in a small set of great moments. It is an organic, growing, slowly changing movement, a network of intertwined problems which themselves develop, change their relationships with one another and their relative importance within the complex."[119]

Private and unpublished writings may well: "reveal the finer structure of the creative process",[120] and enable the reconstruction of the interaction between thought, language, and action. Gruber described a paradox that is to be borne in mind when studying Faraday's language use and the emergence of his ideas:

> "The stream of thought is incredibly swift, but the emergence of and solidification of new ideas is a relatively slow process. Creative thinking is often treated as an isolated act, but if instead it is treated as a growth process it may be easier to understand why progress is slow."[121]

---

[117] Holmes, Frederick L.; 1984: 138.
[118] Holmes, Frederick L.; 1984: 139.
[119] Holmes, Frederick Lawrence; 1985: xx.
[120] Holmes, Frederick Lawrence; 1981: 64.
[121] Gruber, Howard E; 1974: 114, quoted in Holmes, Frederick Lawrence; 1981: 66.

1.5.5 Gilbert and Mulkay: Centrality of Experiments

In the journal *Isis* (1984) Gilbert and Mulkay[122] describe how scientists report on the history of their field and explain theory choice either by depicting key experiments as the main vehicle for scientific change, or in terms of the interactions between social, cultural, and psychological contingencies. The experiments are used as a powerful means by which to persuade the experimenter, and other's in the same field, of the veracity of their current beliefs. By studying the private and public/published writings of a scientist it should be possible to recontextualize the role of experiments. That is, to recover the personal and social contingencies and linguistic practices, and how these have or have not influenced the development of a concept or idea.

In the same edition of *Isis* Shapin, commenting on Gilbert and Mulkay, reminds us, as Holmes (1984) does, that:

> "[s]cientists do talk in different ways in different contexts....[t]alk is not just an account of behaviour and belief; it is itself behaviour that varies according to audience and purpose."[123]

The importance of this point for my research is that language is not a structure that passively describes action and events, but rather language, and its use, varies with the context, is part of the context, and affects the shape of the context.

1.5.6 Bolton and Roberts: Scientists' Writing Styles

Another piece of research that I wish to discuss briefly is by Bolton and Roberts,[124] who carried out statistical analysis using sentence-length distribution and the χ^2 test on seven letters and seven articles to Einstein from Born. The aim was to investigate the hypothesis that scientists write scientific prose and literature (letters) in different styles. Their results found this to be the case. My thoughts are that I am not convinced that Bolton and Roberts' distinction is entirely useful. The belief embodied in my research is that scientists have many 'styles' of writing, which are shaped by the use of

[122] Gilbert, G. Nigel and Mulkay, Michael; 1984: 105-125.
[123] Shapin, Steve; 1984: 126.
[124] Bolton, H.C. and Roberts, Alan; 1995: 295-302.

language. Bolton and Roberts are equating 'scientific prose' with research articles, whereas my research unpacks the notion of 'scientific prose' to show that it take several different forms of personal and published writing. Also a concept, such as electromagnetism, is written about differently depending on, *inter alia*, the language style and language use.

1.6 Critique of the Quantitative Analysis of Language

This research uses qualitative-quantitative text analysis to investigate the implications of language change and use in Faraday's private and public writings. Quine asserts that:

> "[I]n a scientific theory even a whole sentence is ordinarily too short a text to serve as an independent vehicle of meaning. It will not have its separable bundle of observable or testable consequences. A reasonably inclusive body of scientific theory, taken as a whole, will indeed have such consequences."[125]

Critics of quantitative studies of language argue that such a reductionist venture causes a divorce between meaning and form; creating a duality. Furthermore, that meaning in language is always vague and undefinable, and can only be approached through knowledge of the background information; all meaning is relational: "sentences are connected not only with the theoretical side of behavior but also with the emotional, volitional, and other factors."[126] Van Peer argues that:

> "In counting words as words only, one may easily overlook the fact that such words will have very different meanings and connotations in different literary works, or in the hands of different authors."[127]

Discussing Quine's argument in *Two Dogmas of Empiricism*, Putnam writes that:

> "Individual sentences are meaningful in the sense of making a systematic contribution to the functioning of the whole language: they don't having "meaning" in the form of isolatable objects, properties, or processes which are associated with them individually and which determine individual assertibility conditions."[128]

---

[124] Quine, Williard van Orman; 1981: 70.
[126] Fodor, Jerry A. and Katz, Jerrold J.; 1964: 412.
[127] Peer, Willie van.; 1991: 303.
[128] Putnam Hilary; 1990: 278.

Even observation sentences - "something that we can depend on other witnesses to agree to at the time of the event or situation described"[129] - are not necessarily meaningful, for:

> "[a] witness might of course forget and give divergent testimony later, or might fail to notice a feature at the time until it was pointed out."[130]

All sentences lack determinate meaning, so we 'should talk about interpretation and not about meaning'. It is Quine's belief that no general rules can be elicited to determine a context of assertibility, or not, for a sentence. [131] Assertibilty, Putnam proclaims: "to the extent that it is rational, is pragmatic and depends on the entire context."[132]

1.7 A Note On The Qualitative and Quantitative Dimensions of Research

Following on from the previous section, the type of analysis employed in this research combines the quantitative dimension with the qualitative. This means that the results obtained by quantitative methods become meaningful by being placed in a qualitative context, that is, qualitative interpretation requires knowledge of the text or of the landmarks events in the text. Thus the social and human elements of knowledge construction and language use is never lost sight of; it is present throughout the analysis and interpretation of results. For example, the qualitative work by Geoffery Cantor on Sandemanianism informs what my quantitative data means in relation to the influence of Faraday's religious views on his science and his writing.

Before leaving this area I want briefly to look at what Karl Kroeber thought about the qualitative being allowed to 'speak', writing that:

> "When we speak of a novelist's choice of words we are referring to something different from mere language, just as when we speak of language we refer to something different from mere sound. Just as language is made out of sounds, so literature is made out of language, but to reduce the "higher" system to elements of the "lower" material of which it is composed is to obscure the difference we wish to define. Quantitative analysis tends to make this fatal reduction; by discriminating parts it blurs the qualitative relationship of part to whole which in fact defines the essential nature of the part."[133]

---

[129] Quine, Williard van Orman; 1970; 23.
[130] Quine, Williard van Orman; 1970: 23.
[131] Quine, Williard, van Orman; 1960.
[132] Putnam, Hilary; 1990:208.
[133] Kroeber, Karl; 1969: 206-207.

The argument put forward by Kroeber is very important and highly persuasive. The thesis expounded in the last sentence of the quote can be refuted by 'inducing the context'. That is: "if we wish to grapple with the intentions of a writer, then we somehow determine the context that was evoked and present at the time of writing."[134] This process requires the following:

1) a concentration on the text, that is, the 'local context', what preceded and what followed the text,

2) a knowledge of: "what kinds of ideas and information are in the head of the writer at the time of writing." To achieve this we draw upon biographical information, an understanding of the culture within which the scientist and/or writer laboured, the person's social stratum, and their role(s) in that culture they inhabited.[135]

This means that in studying a set of texts by Faraday we would do the following steps to ensure that we relate text/context:

1) be knowledgeable about the text being analyzed,

2) construct a brief biography of Faraday (section 1.3),

3) be familiar with the historical context of a concept such as electromagnetism.

1.8 Overview of Chapter Contents

My thesis has nine chapters in total. **Chapter two** provides a historical context for my research. The first part of which is a history of human-based and computer-supported 'text analysis'. This is followed by an investigation of how the issues of language and meaning are treated in linguistic and literary theories, feminism, science studies, and science. Finally the conclusion discusses what my research has to offer that is different from other work in computer-assisted text analysis. I also discuss what implications: i) the historical context has for my research and ii) my research has for science studies. Next, **chapter three** presents two studies: the use of discourse analysis as a research tool in science studies and the quantitative analysis of text outside of

---

[134] Simon, Herbert; http://shr.stanford.edu:80/shreview/4-1/text/simon4.html; April 1995; accessed September 17, 1997.

[135] Simon, Herbert; http://shr.stanford.edu:80/shreview/4-1/text/simon4.html; April 1995; September 17; 1997.

science studies. The second part of this chapter discusses how some methods identified in the studies are relevant to my research.

Chapter four provides a detailed description of the methods and resources for my research project and is divided into three main parts. The first part details the issues addressed by the research and its aims, and the questions asked and the conjectures thought to be testable. The second part describes the original work to be done in my research, describing my case studies. The final part discusses what revisions will be made response to results of the pilot study and what will count as interesting results. Also what the specifications for the large 'runs' (large scale text analyses) to help evaluate the conjectures will be detailed, and problems encountered following completion of the pilot study, and how they may be countered, will be described.

Chapter five presents my pilot study, in which I first test how effective various textual features, such as compression and modality, are as discriminators of different text types. This is followed by the primary empirical work of my thesis in chapters six to eight, which are preceded by a short introduction. **Chapter six** is a study of the language used in Faraday's correspondence with two scientists and five non-scientists, over a twenty-five year period. I test which textual features are most successful in allowing discriminations between different types of Faraday's writings, for example, between his letters to scientists and those to friends. A test of this chapter's main findings is the aim of **chapter seven**, with three investigations undertaken. The first explores Faraday's early correspondence with friends, and the second his letters to and from non-Sandemanians. The third study involves the analysis of private and published writings of Faraday and Darwin. The linguistic construction of Faraday's discovery of electromagnetism is the concern of **chapter eight**. More specifically, I look how Faraday's language changes in his private and public writings on electromagnetic rotation and induction between 1821 and 1831. I then consider the implications of this for Faraday's 'discovery' of electromagnetism.

Chapter nine contains conclusions, discussions, and interpretation. I evaluate the themes, aims, questions, and theory the research engaged with, and summarize the main findings, saying why they are interesting. The implications of my research are discussed in relation to science studies, analysis of scientific discourse, and textual analysis. This part will also relate back to the work by Badalamenti, Holmes, Zahar, Gooding, *et al* discussed in earlier in this chapter (section 1.5). My results will also be interpreted in relation to my overriding question: 'How is language used as a medium for thinking and communication?'. In this chapter I defend my methods, saying what was useful and unique about them, as well as discussing their limitations. I conclude chapter nine with three sections, the first of which looks at the issue of meaning. The second sections discusses what I have learnt about the relationship between Faraday, language, and religion, and the third, my thoughts on how Faraday himself perceived and used language.

I now move onto chapter two which provides a comprehensive treatment of the history of text analysis. This chapter also places my research in relation to prevailing issues in current computer-assisted text analysis.

2 A History and Theory of Texts

Abstract. The theoretical and historical contexts for my research are provided by this chapter and chapter three. This chapter is written around three themes that are addressed by my research: text analysis, language and science. The first theme is covered by a historical resumé of 'text analysis', encompassing both human-based and the computer-supported aspects. The second theme involves a treatment of language meaning and how language is perceived by a number of linguistic and literary theory disciplines. The third theme investigates the literary properties of scientific discourse, which includes the implications of private and public prose for recovering experimental narratives. Related to these three themes are two empirical studies concerned with how discourse analysis is used as a tool within and outside of science studies presented in chapter three.

2.1 Introduction and Contents Discussion

My research is concerned with recovering language use in the evolution of ideas and concepts, and reconstructing discovery narratives in nineteenth-century scientific practice, principally in the work of Michael Faraday. This chapter, and the next, provide a detailed treatment of the context for my research. The context covers three areas, the first of which is a history of human-based and computer-supported 'text analysis'. This includes a discussion of electronic language analysis research, a resumé of a current computer-supported text analysis projects, found in appendix one, and a breakdown of projects and related software tools in computational linguistic and text analysis in appendix two. The next context involves the meaning and uses of science and language from a variety of perspectives, such as literary studies, feminism, history of science, and scientometrics. Lastly, in this chapter, the conclusion describes what my research has to offer that is different and discusses the implications of the research for science studies.

The third context is the use of discourse analysis as a research tool in science studies and the quantitative analysis of text outside of science studies. This is presented as two empirical studies in **chapter three**, which includes a survey of how some methods identified in the studies are relevant to my research.

The wider aim of my research is to look at how language is used as a medium for thinking and communication. This chapter relates to this aim by describing historically the development of text and language analysis, and how language has been used across linguistic and literary theories, for science studies, and in science.

2.2 A Historical Resumé of 'Text Analysis'

2.2.1 Introduction

This part provides a comprehensive history of text management and analysis from early BC up to the present time. First, there is a discussion of 'text analysis' from early BC up to the 1950s, which looks at 'human' manipulation of written texts, from frequency counts to produce word lists through to literary statistical analysis to discern authorship style. This is followed by a synopsis of computerized text management and analysis systems from the 1950s to the present day. I then discuss the advantages and disadvantages of computerized text management and analysis tools. Lastly, I discuss my research questions and issues in light of the historical treatment of text analysis methodologies, which begins with an overview of issues uppermost in current computer-assisted text analysis. The conclusion that follows draws together the main historical issues and questions in text analysis. This work helps to identify: i) the main issues and questions, historically and currently, being addressed in text analysis, ii) any issues and questions being repeatedly proposed, iii) why particular text analysis methodologies have been used, and not others, if computer-based methods are allowing the investigation of new issues and questions, and v) principal issues and questions to be addressed by this research about science and texts, especially in relation to the work of Michael Faraday and Charles Darwin.

2.2.2 Frequency counts to produce restricted word lists: early BC to the 16th century

The starting point for any type of quantitative textual analysis is frequency counts; the count of the occurrence of linguistic elements. To elaborate on this point DeRocher *et al* explain that:

> "Traditionally the counts have been of words, but have also been of phonemes, morphemes, syllables or idiomatic expressions. The purpose of these counts has usually been to develop a vocabulary of a special type such as of a rare, frequent, useful, or important words with the ultimate objective of developing vocabularies for the teaching and learning of stenography, spelling, or reading in the easiest and most efficient manner possible."[136]

There has survived from the 7th century BC a short Akadian word list from central Mesopotamia.[137] Frequency counts of words can also be traced back to Alexandria (founded in 322 BC), where scholars:

> "distinguished between rare and frequent words of Homeric Greek for the benefit of local students of Literary Greek."[138]

Another example of early word counting is that of the Talmudists categorising the words in the Torah during the tenth century.[139]

DeRocher *et al* (1973) describe how from the fifteenth century in Europe frequency counts were used as an aide to language learning. It was believed that this learning was facilitated by becoming familiar with those words used most often so that they could be learnt first. These lists initially comprised of difficult or useful words, as well as tending to be:

> "restricted for occupational or instructional purposes providing mainly lists such as birds, animals, parts of the body and of occupations."[140]

Concordances[141] (an index of words) have been used as a research tool since the Middle Ages. What is usual is that what is normally called a concordance provides some context surrounding the use of each occurrence of a word. This might throw up particular habits of an author or give clues to how a word is

---

[136] DeRocher, James E. *et al*; Ist July 1972 to 15th May 1973: 1.
[137] Goetz, W. Philip; 1986: 385 (column one).
[138] DeRocher, James *et al*; 1st July 1972 to 15th May 1973: 1.
[139] DeRocher, James *et al*; 1st July 1972 to 15th May 1973: 1.
[140] DeRocher, James *et al*; 1st July 1972 to 15th May 1973: 2.

being used. Early examples of concordances are those by Christopher Barker: the *Concordance to the Bible* in 1300, *The Bible*, published in 1588, and the *A Midsummer's Night Dream: A concordance to the text of the first quarto of 1600.*[142]

2.2.3 Construction of self-contained, basic vocabularies: the 16th and 17th centuries

Early counts were restricted to specialized vocabulary lists. For example, Timothie Bright published in 1588 his *Characterie: An Arte of Shorte, Swifte, and Secrete Writing by Character.* This work was both a first known attempt to produce a phonetically centred type of shorthand and to develop:

> "a self-contained, basic or "island" vocabulary capable of being used to express all necessary concepts with as few words (or symbols) and their variations as possible."[143]

Another example of an early specialized word list was published in 1698 by Johannes Leusden, who counted in the Dutch version of the New Testament 4,956 different words and 1,686 occurring only once or only in one verse.[144]

2.2.4 Dictionary building up to the 18th century

DeRocher *et al* state that the first attempt made to construct a comprehensive dictionary appeared in 1721, with the publication Nathanial Bailey's *Dictionarium Britannicum.*[145] Retracing our steps, the tradition of dictionary construction began in Greece. A lexicon was built by Pamphilus of Alexandria in 1 AD. Other important lexicons made by the Greeks include those compiled by the Atticists[146] in 2 AD, that of Hesychius in 5 AD, and that of Photius in the Middle Ages. In the fourteenth and fifteenth centuries many interlingual

---

[141] According to Joseph Raben: "the root of the word *concordance* is the "heart" that was thought to underlie the overlapping and partly contradicting versions on Jesus' life as recorded in the Gospels." (emphasis orginal). In, Raben, Joseph; 1991: 343.

[142] Barker, Christopher; Concordance to the Bible; 1300, The Bible; translated according to the Ebrew and Greek; 1588/1594 (with a concordance of eighty-two leaves at the end), and A Midsummer's Night Dream: A concordance of the text of the first quarto of 1600; Clarendon Press: Oxford; 1600.

[143] DeRocher, James *et al*; 1st July 1972 to 15th May 1973: 2.

[144] DeRocher, James *et al*; 1st July 1972 to 15th May 1973: 3.

[145] Bailey, Nathanial; 1721.

[146] "The Atticists were compilers of list of words and phrases thought to be accord the usage of the Athenians." In Goetz, W. Philip (Editor-in-Chief); 1986: 385 (column one).

dictionaries appeared, such as the *Promptorious peurorum* ("Storehouse [of words] for Children"), which was brought out by Richard Pynson in 1499.[147]

The first entirely English dictionary was issued in London in 1604, by an Oakham schoolmaster called Robert Cawgrey, and entitled: *A Table Alphabeticall, conteyning and teaching the true writing and understanding of hard usuall English wordes, borrowed from Hebrew, Greeke, Latineor French &c.* The first work to have the word 'dictionary' in its title was *The English Dictionarie: or an Interpreter of hard English Words*, by Henry Cockerham in 1623.[148] The dictionary that became the seminal work was Dr. Samuel Johnson's *A Dictionary of the English Language*, of about 43,500 words. First published in 1756, it was issued in a further three editions. Dr. Johnson's *Dictionary* became the seminal form because it was the first vernacular dictionary to include citations, alphabetic lists, and 'easy' and 'hard' words.[149]

2.2.5 Word counts and frequency of occurrence in the 19th century

One of the largest frequency counts was undertaken by F. W. Kaeding in 1897. Eleven million German words and twenty million syllables collected from fourteen categories of material were counted with the assistance of six thousand assistants. Kaeding found 258,173 unique words, half of which occurred only once. The purpose of this count was to help in the teaching of stenography.[150] DeRocher *et al* state that the importance of this work was that:

> "[I]t firmly established the method of counting large numbers of words from a variety of sources in order to find truly general or representative words, and established frequency of occurrence of a word as the basis for a determination of its linguistic importance or value." [151]

A revision of this study, correcting deficiencies, was published in 1928 by Baynard Morgan.[152]

---

[147] Goetz, W. Philip (Editor-in-Chief); 1986: 385 (column two). This work: "is better known under its later title of *Promptorium parvulorum sive clericorium* ("Storehouse for Children or Clerics") commonly attributed to Geoffrey the Grammarian (Galfridus Grammaticus), a Dominican friar of Normandy, who is thought to have composed it about 1440." (*ibid*).
[148] Goetz, W. Philip (Editor-in-Chief); 1986: 386 (column two).
[149] Green, Jonathon; 1996.
[150] Kaeding, Wilhelm Friedrich; 1897.
[151] DeRocher, James *et al*; 1st July 1972 to 15th May 1973: 4.
[152] Morgan, Baynard Quincy; 1928.

Another use for frequency counting is in: "the teaching of spelling of the real word, as opposed to some representation of it as in shorthand".[153] This was the rationale behind the word count undertaken by Joseph Rice, which appeared in his *Rational Spelling Book*.[154] The reason why frequency counts or word lists help with spelling is that misspellings show up quickly.

2.2.6 Statistical analysis of writing style to identify authorship in the 19th and early twentieth century

Modern textual analysis began with the statistical analysis of writing style. It can be attributed to Augustus de Morgan (Professor of mathematics in London) (1806-1871). In 1851 de Morgan expostulated, in a letter to a Cambridge friend, the Reverend W. Heald, that the controversy over the authorship of St. Paul's Epistles (the fourteen letters from Romans to Hebrews) may be resolved through a count of the number of letters in the words - the word-length - of the corpus.[155] Thus such counts were used for resolving a controversy over authorship for the first time. De Morgan proclaimed that:

> "If scholars knew the law of averages as well as mathematicians, it would be easy to raise a few hundred pounds to try this experiment on a grand scale...I should expect to find that one man writing on two different subjects agree more nearly with himself than two different men writing on the same subject. Some of these days spurious writings will be detected by this test. Mind, I told you so."[156]

De Morgan's letter, writes Morton:

> "[I]s noteworthy for containing so many of the basic principles of stylometry, the use and description of samples, the disregarding of the meaning of words and the concentration on their occurrences."[157]

The thesis that word-length could a deciding factor in the study of literary style was first empirically investigated by an American geophysicist called Thomas Corwin Mendenhall. In his 1887 paper in the journal, *Science*, Mendenhall writes that he had reflected at various times over a period of five or six years on de Morgan's belief that: "the identity of an author of a book, a poem, or a play" might be identified by the average length of words in the composition.[158] This

[153] Rice, Mayer Joseph; 1898: 5.
[154] Rice, Mayer Joseph; 1898.
[155] Kenny, Anthony; 1982: 1.
[156] Lord, R. D.; 1958: 282.
[157] Morton, Andrew Queen; 1978: 166.
[158] Mendenhall, Thomas C.; Friday, March 11, 1887:213.

suggestion of de Morgan's possibly, Mendenhall believes, came from his *Budget of Paradoxes.*[159] Whereas Lord claims that the reference is more likely to be found in *Memoir of Augustus de Morgan* by his wife Sophie, and published in 1882.[160]

Mendenhall expresses his dissatisfaction with this method and is attracted to what we today call the frequency distribution of words of different lengths: a method for which spectroscopic analysis of a beam of light was the inspiration. He writes, describing his new method, that:

> "So certain and uniform are the results of this analysis, that the appearance of a particular spectrum is indisputable evidence of the presence of the element to which it belongs. In a manner very similar, it is proposed to analyze a composition by forming what may be called a 'word-spectrum' or 'characteristic curve', which shall be a graphic representation of an arrangement of words according to their length and to the relative frequency of their occurrence."[161]

Kenny elucidates how Mendenhall's method purported to improve upon de Morgan's method:

> "De Morgan seems to have thought that the average word-length by itself might be an indication of authorship; by Mendenhall's studies showed that texts with the same average word-length might possess different spectra...It was by comparing the whole spectrum of word-length preferences that Mendenhall hoped to offer a scientific solution to disputes about authorship."[162]

This study involved graphical representation of the construction of word-spectra (frequency distribution of word-length) for various authors. Mendenhall used one, five, and ten thousand word samples from, amongst others, *Oliver Twist*, *David Copperfield*, *Vanity Fair*, Mill's *Political Economy*, and the first five thousand five hundred words of Caesar's *Commentaries* furnished by a friend. One interesting result of this study was that:

> "It was soon discovered that among writers of English the three-letter word occurred much more frequently than any other."[163]

In his concluding remarks Mendenhall is quite convinced of the validity of his method in settling disputes about authorship:

---

[159] de Morgan, Augustus; 1859.
[160] de Morgan, Sophie; 1882.
[161] Mendenhall, Thomas C; 1887: 214.
[162] Kenny, Anthony; 1982: 2.
[163] Mendenhall, Thomas C.; December 1901: 99.

"If striking differences are found between the curves of known and suspected compositions of any writer, the evidence against identity of authorship would be quite conclusive."[164]

Mendenhall's method and claims did not go unchallenged. Three letters to the same journal followed Mendenhall's paper.[165] The letter writer A. B. M suggests that a sentence-length distribution would be a more revealing test in terms of an author's style:

"[I] have lately thought some instructive results might be obtained from examining sentences with regard to length, as measured by the number of words. Length of sentences is a matter in which psychological peculiarities which it might be instructive to inquire into."[166]

After applying Mendenhall's method to three hundred sentences of Carlyle's *Sartor Resartus*, H. A. Parker is somewhat dismissive of the usefulness of the technique, writing that his investigation:

"[G]oes to show, if it does not prove, that for detective purposes the method is valueless."[167]

A more ambitious project in finding a solution to authorship identity was a study by Mendenhall published in *The Popular Science Monthly* in 1901.[168] This work was enabled by a grant donated by a Boston philanthropist, Mr Augustus Heminway, which Mendenhall used to employ two secretaries, a Mrs Richard Mitchell and a Miss Amy C. Whitman, of Worcester, Massachusetts, and construct a counting machine.[169] Thus:

"by which a registration of a word of any given number of letters was made by touching a button marked with that number. One of the counters, with book in hand, called off 'five', 'two,' etc., as rapidly as possible, counting the letters in each word carefully and taking the words in their consecutive order, the other registering, as called, by pressing the proper buttons."[170]

This work involved an examination of the Shakespeare and Francis Bacon controversy, that is, Bacon at the time was perceived to have authored the Shakespearean literature:

[164] Mendenhall, Thomas C.; 1887: 246.
[165] M, A. B.; March 1889: 226, Parker, H. A.; March 29, 1889: 246, and M.; 1889: 269.
[166] M, A. B.; March 1989: 226.
[167] Parker, H. A.; March 29, 1889: 246.
[168] Mendenhall, Thomas C.; December 1901: 97-105.
[169] Kenny, Anthony; 1982: 2, and Mendenhall, Thomas C.; 1901: 101-102.
[170] Mendenhall, Thomas C.; 1901: 102.

"The principal hypothesis under scrutiny in this work was based upon the prevailing view that no unlettered Stratfordian could have written the plays attributed to Shakespeare. As it was fashionable to assume, Francis Bacon *must* have been the author".[171] (emphasis original).

A total of 400,000 words of Shakespeare and 200,000 words of Bacon were analyzed, as well as writings of, *inter alia*, Jonson, Milton, Marlowe, and Goldsmith. Mendenhall found that Shakespeare's 'characteristic curve' was very consistent across his prose and poetry, and two features of his writings became evident:

"Shakespeare's vocabulary consisted of words whose average length was a trifle below four letters, less than that of any writer of English before studied; and his word of greatest frequency was the *four-letter* word, a thing never met with before." (emphasis is original).[172]

The characteristic curves of Shakespeare and Bacon differed considerably: Bacon could not have authored writings ordinarily attributed to Shakespeare, Mendenhall concluded. This did not entirely satisfy Heminway who, quoted by Mendenhall, said robustly: "the question still remains, who did?" One further revelation arose from this study. This concerned Marlowe, for quite a different scenario arose with the comparison of Shakespeare and Marlowe's characteristic curves:

"[I]t was discovered that in the characteristic curves of his plays Christopher Marlowe agrees with Shakespeare about as well as Shakespeare agrees with himself".[173]

Towards the end of the nineteenth century mean sentence-length was the focus of analysis for literary analysts. An example of this work was undertaken by Professor Sherman, who, convinced that the influence of classical scholarship had waned, purported that liberty from the stifling effect of this literary tradition was reflected in a trend embracing shorter sentence length.[174] That is:

"The decline in the use of complex (compound) sentences" [and]..."the movement towards sentential simplification".[175]

Sherman initially attempted to ground his thesis in a study of the average sentence-length of several authors. He claimed that a sample (of three-hundred sentences):

"[W]ill generally reveal the sentence-rhythm of any writer who has achieved a style."[176]

[171] Bailey, Richard, W; 1969: 218.
[172] Mendenhall, Thomas C.; 1901: 102.
[173] Mendenhall, Thomas C.; 1901: 105.
[174] Sherman, L. A.; 1888: 119-130, and Sherman, L. A.; 1892: 337-366.
[175] Sherman, L. A.; 1888: 129.
[176] Sherman, L. A.; 1888: 130.

I do not know how Sherman arrived at a figure of three-hundred but he seems to be claiming that a writer can be identified by a single style. Whereas I argue that a writer has many styles, which will vary with language function - arguing, reflecting, or persuasion, for example. As well as with respect to the intended audience or the genre of writing.

To conclude, according to Morton, for the identification of authorship:

> "[A] habit must comply with three conditions: first, it must be a habit apparent in a choice which frequently confronts all authors; second, it must be a habit which can be shown to be unaffected by changes in subject matter, by passing of periods of time, by reasonable differences in literary form and all other possible influences which might affect the habit."[177]

2.2.7 Statistical Stylistic Studies: The 1930's to The 1950's

In this section I shall primarily consider the contribution of three central academics, that is, George Zipf, G. Udny Yule, and C. B. Williams.

One of the best known attempts to ascertain a statistical law for linguistic data was undertaken by G. K. Zipf, who occasioned the observation of the inverse relationship between word-length and frequency.[178] Though Zipf's work followed the original observations made by Estoup[179] and Wills,[180] he was the first to thoroughly investigate this relationship, which for Zipf meant that:

> "a list of word-types from a large body of text ranked in order of frequency will show a *constant* decrease in the frequency with which they appear (with the exception of the most common words."[181] (emphasis original).

One particular criticism of Zipf's 'law' is described by thus:

> "[R]ank and frequency are of necessity lawfully related not by observation but by definition."

Word-length distribution, following the work of Mendenhall, was used by Brinegar in a study comparing the writings of Mark Twain and a number of unidentified authors.[182]

---

[177] Morton, Andrew Queen; 1978: 98.
[178] Zipf, George Kingsley; 1935, and Zipf, George Kingsley; 1949.
[179] Estoup, J. B.; 1916.
[180] Wills, J. C.; 1922.
[181] Bailey, Richard W.; 1969: 220.
[182] Brinegar, Claude S.; 1963: 85-96

In 1944 the statistician G. Udny Yule published *The Statistical Study of Literary Vocabulary*, a study of the frequency distribution of sentence lengths of Greek authors. According to Morton this work was flawed because of an error in the calculation of the standard errors of the sentence length distribution.[183] This analysis was continued and improved upon by W. C. Wake in 1947 and 1957.

It was in this work that Yule devised 'Characteristic K', an instrument for measuring vocabulary richness from the frequency-distribution of data. The original 'Characteristic' was expressed thus:

$$S_1 = S(f_x X), S_2 = S(f_x X^2),$$
$$K = 10{,}000 \frac{S_2 - S_1}{S_1^2}$$

Where S_1 and S_2 are the samples, and f_x is the number of words occurring X times, and to avoid the inconvenience of small decimals the expression multiplied by 10,000.[184]

Bennett attests that:

> "What the characteristic measures is the repetitiveness of the vocabulary of a given literary work, which is certainly a significant stylistic trait. We have all noticed how one writer will hammer away at certain words and employ them again and again, where another writer may scrupulously avoid exact repetition of the same words...[and] Yule's characteristic *is independent of the length* of the literary work."[185] (emphasis original).

Thus, as Herdan describes, the 'Characteristic':

> "[S]hows that a particular style is characterized by a constant relation between uniformity and diversity in the number of repetitions of the items of vocabulary."[186]

The original formulation of Yule's Characteristic K was:

> "based on the assumption that that the occurrence of a given word is based on chance and can be regarded as a Poisson distribution."[187]

This notion of randomness was criticised by Ross[188] and lead to a revision of the Characteristic, thus defined by Herdan as:

> "[T]he coefficient of variation of a mean, or the relative fluctuation of a mean."[189]

---

[183] Morton, Andrew Queen; 1987: 208.
[184] Yule, Udny G.; 1944: 44-53.
[185] Bennett, Paul E.; 1969: 30 and 32.
[186] Herdan, Gustav; 1958: 269.
[187] Bailey, Richard W.; 1969: 223.
[188] Ross, A. S. C.; 1950: 19-59.
[189] Herdan, Gustav; 1955: 333.

Another author whose work is prominent in the field of literary statistics is C. B. Williams, who, as a biological statistician, noticed that several of the frequency distributions facilitative in the study of animal populations were salient to the word 'populations' in the writings of authors. Williams proposed in his 1940 paper that sentence-length distribution could be more suitably measured using the log distribution. This was because:

> "If the log distribution is normal we can infer that the extent to which the process of writing is likely at any level proportional to the length of the sentence."[190]

2.2.8 Conclusion and Summary

In summary, the quantitative manipulation of texts has a long and complex history, which started with the construction of word lists and lexicons in classical Greece. Then through to the first attempts to build dictionaries in the late Middle Ages, with the nascence of statistical analysis of textual features in the late nineteenth-century. The use of this statistical analysis in stylistics and authorship-attribution studies took hold in the first half of this century. My research is not concerned with stylistics nor authorship-attribution but instead with meaning and the role of language in writing, thinking, and the construction of ideas.

One main drawback of manual analysis was the limitation in terms of size of corpus. This was one problem the introduction of computer-supported analysis could sweep aside, and it is to computer-supported text analysis we turn next.

2.2.9 Computer-based Analysis of Text and Language Processing: Introduction

This section encompasses, first, a discussion of the initial divergence between, on the one hand, humanities and literary studies and on the other, computational and statistical methods. Then there is a digest of computer-based text analysis software, followed by a consideration of some of the advantages and disadvantages of electronic texts and computerized text analysis.

---

[190] Williams, C. B.; 1940: 360-361.

2.2.10 Computational Linguistics and Humanities Computing

Machine-based language processing was an inchoate field in the 1950's and the early 1960's. In fact, the first ever computer-based project on a humanities text - a concordance of the works of St Thomas Aquinas (*Index Thomisticus*)- was begun in 1949 by Father Roberto Busa, and the first printed volume produced in 1974.[191] Computer-aided concordance was first achieved in 1951, and this initiated the electronic processing, that is: "the acquisition, encoding, and analysis", of large corpora of texts.[192] From the mid-1960's there was a shift to complete linguistic analysis, thus the field of computational linguistics focused on sentences rather than text. Computational linguistics, describes Ide is:

> "concerned with the treatment of natural language in computational terms. Quite often, but not necessarily always, the goal of work in this field is to enable computers to understand, generate, and translate natural languages, which would in turn enable machines to understand and respond to user enquires and commands which are expressed in natural languages rather than an artificial computer language...To give a slightly more detailed understanding of what is involved in computational linguistics, consider that the treatment of natural language must involve the following:

- *phonology*
- *morphology* (finding root forms for words in inflected or derived forms)
- *syntax* (determining the constituent grammatical pieces of sentences)
- *semantics* (determining the relations among elements of a sentence in terms of their meaning)
- *pragmatics* (determining which pieces of general word knowledge affect the meaning of a sentence, and how)".[193]

Through the 1960's and 1970's what transpired was a move to analyse smaller textual units, primarily sentences. According to Renear:

> "a number of computer scientists and software designers came independently to the conclusion that the best way to design text processing systems was to base them on the view that there are certain features of texts - such as titles, chapters, paragraphs, lists and the like - which are fundamental and salient and that all processing of texts (editing, formatting, analysis, etc) should be implemented indirectly through these features." [194]

[191] Busa, Roberta S. J.; 1974.
[192] Ide, Nancy and Walker, Donald; 1993: 327.
[193] Ide, Nancy; 1991: 7-9.
[194] Renear, Allen; 1992: 221-248 (222).

According to Ide and Walker, the research in the 1970's and 1980's:

> "was restricted to language drawn from limited domains, using very small, hand-coded lexicons".[195]

In the late 1980's a convergence of the methodologies and tasks of computational linguistics and humanities computing developed, with the computational linguists becoming involved in work characteristic of the humanists. Computational linguists are increasingly resorting to statistical methods, such as factor analysis and cluster analysis, in order to garner information about general characteristics of language use from large corpora ("a body or collection of linguistic data for use in scholarship and research").[196] For according to Dolezel and Bailey quantitative stylistics was in wretched state because of:

> "the widespread ignorance of statistical methods characteristic of most linguistic or literary scholars."[197]

More recently, Ide and Walker proclaim, computational linguists are applying concordance analysis, for example, to similar tasks as the computational humanists, such as, "part of text tagging", "collocation studies", "parallel text alignment", and "term extraction".[198] Ide describes how humanist scholars and computational linguists have reached an impasse, whereby the humanists' reliance on human-based analysis is reaching its analytic nadir, and computational linguistics are suffering from a lack of knowledge about how languages are used by people. Thus they sought to surmount their difficulties by borrowing and sharing methods and results:

> "Recently, both humanists and computational linguistics found themselves up against a wall. Humanists have found that they cannot go further in analysing things like style and theme without deeper information about syntax and semantics. Computational linguists have found linguistic theories predict the possible, but they need information about the probable and characteristic properties of language in order to make more progress in handling language by machine. So, humanities people are becoming interested in some of the methods and results computational linguists have been working on the past few decades, while the computational linguists are beginning to apply statistical methods to large corpora in order to gather information about general properties of language use, and have begun to use tools like concordance and word lists. The methods they are

---

[195] Ide, Nancy and Walker, Donald; 1993: 328.
[196] Anderson, James M. (Ed.); 1991: 73.
[197] Bailey, Richard, W.; 1969: 219.

> using and the resources they are applying to them are those of humanities computing. The main difference is that humanists continue to be concerned primarily with remarkable language [literary language] and computational linguistics are interested, for the time being, in unremarkable language [straightforward language *sans* metaphor or irony, for example] ."[199]

The sharing of resources has been one primary reason for the activity spent on integrating humanities computing and computational linguists. These resources include the Text Encoding Initiative, the ACL Data Collection Initiative, the Network of European Research Corpora, the Consortium for Lexical Research, and the Linguistic Data Consortium."[200]

This picture is not as unblemished as it may seem for Bruce laments the failure of humanities computing to integrate: "the key theoretical elements of contemporary text and discourse theory."[201] Humanities computing, Bruce claims, have overly concentrated upon individual micro-analyses of discourse, for traditional scholars have had to work with small corpora. The concentration upon micro-analysis has not allowed computer-aided researchers to exploit the potential of computers to facilitate systematic macro-textual analysis. An advantage of macro-analysis is that, whether one is studying an author's style or how Faraday constructed his ideas and concepts in language, the larger the corpus the more likely it is patterns will be exposed. Though saying this, I do not hold with a micro-macro analysis distinction. My research embraces a more holistic approach, whereby language use, ideas, writing, and the interpersonal and social environments are all inextricably interconnected.

2.2.11 The Evolution of Text-Management and Analysis Software

One of the first computer-based data analysis systems was WORDS (1965), which was a content-analytic program.[202] During the 1970's the number of text processing programs, which were not generally designed for the analysis of text

---

[198] Ide, Nancy and Walker, Donald; 1993: 328.
[199] Ide, Nancy; 1991: 8.
[200] Ide, Nancy and Walker, Donald; 1993: 330.
[201] Bruce, Donald; 1993: 357-364.
[202] Iker, H. P. and Harway, N. I; 1965: 173-183; *idem*; 1968: 134-154; Iker, H. P.; 1974: 430-438; Jandt, F. E.; 1972: 25-31.

content, burgeoned. These included *Bravo,*[203] *CLAS,*[204] *DPS,*[205] *PUB,*[206] *Scribe,*[207] *TEX,*[208] *Texture,*[209] and *TROFF.*[210]

Another software package was CLOC, on which work began at the University of Birmingham in 1973. Automatic parsing programs for stylistic analysis developed in the 1970's included EYEBALL and its updated version OXEYE.[211] This software performed collocation analysis and generated concordances and word listings on natural language texts, and grew out of a dissatisfaction with COCOA, the software generally available for this type of analysis.[212] In the early 1980's examples of such software are *BORIS,*[213] *PARTS,*[214] and the *Writer's Workbench*™.[215]

Come the 1980's computer software is being used for more diverse and complex quantitative analysis of texts. Towards the end of this decade and into the next computerized co-word analysis was being undertaken with software such as LEXIMAPPE and CANDIDE.[216] Snelgrove in a 1990 paper employs a software package called STRAP (**STR**utural **A**nalysis **P**rogram) to analyse the structure of narrative texts.[217] This software was constructed: "as a general-purpose text utility with special application to the close analysis of literary texts."[218]

---

[203] Lampson, Butler; 1978.
[204] A "Computerized Language Analysis System", designed to: "perform standard statistical tests on natural language texts as well as print a concordance or index of all the words of that text or a specified subset of these words." Borden, George, A and James, Watts J.; January 1971; 129-141.
[205] Sibbald, Kern E.; 1976.
[206] Tesler, Larry; September 1972.
[207] Reid, Brian K.; 1978.
[208] Knuth, Donald E.; November 1978.
[209] Gorlicjk, M. *et al*; 1975.
[210] Ossanna, J. F.; 1977.
[211] Ross, D and Aitken, A. J. *et al*; 1973: 85-99, Ross, D.; 1974: 94-108, Ross, D. Jr and Rasche, R. H.; 1972: 213-221.
[212] Reed, Alan and Schonfelder, J. Laurie; 1979: 59.
[213] Lehnert, Wendy G. *et al*; 1983: 15-62, Germain, Ellen; 1992: 30-35.
[214] Cherry, L. L.; 1980.
[215] Cherry, L.; *Writing Tools*; 1982; 100-105; Frase, Lawrence T. *et al*; November 1981: 21-24; Kiefer, Kathleen E and Smith, Charles R.; 1983: 201-214; Sterkei, Karen S; 1986: 43-61; MacDonald, Nina H. *et al*; 1982: 105-110; Sims Brenda R. and DiMaaggio, Donna; Winter 1990: 61-68.
[216] for example, Callon, M. and Courtial, J. P.; 1989, Courtial, J. P.; 1989: 527-534, Callon, M. *et al*; 1991: 155-205, Courtial, J. P. *et al*; 1994: 173-192, and Tijssen, R. J. W. and van Raan, A. F. J.; 1989: 285-297.
[217] Snelgrove, Teresa; 1990: 221-225.

To bring us almost up to date, in 1995 Eric Johnson wrote and released a computer program, *WORDS*, designed to count the number of "running words in a text", the "number of unique word forms", and "to list the number of occurrences of each unique form."[219] Even more up-to-date, in April 1997 version 4.0 of INTEXT text analysis software was released.[220] This software package was originally developed for computer aided content analysis. The search patterns can include words, parts of words, word sequences, and sequences of words. Intext can perform, for example, compilation of lists of words, (sorted by alphabet or frequency, ascending or descending), KWICs (key word in context), SITs (search patterns in text unit), list of word permutations, and control of multiple search patterns.

This brief discussion of the development of text manipulation and analysis software is supplemented by a detailed breakdown of projects and related software tools in computational linguistic and text analysis in appendix two. The next theme to be addressed will be the advantages and disadvantages of electronic technology in scholarship.

2.2.12 Advantages and Disadvantages of Computerized Text Management and Analysis Tools

This section will discuss some of the issues arising from scholars employing electronic technologies with respect to texts.

i) Advantages

Electronic texts and on-line textual databases can allow greater access to texts and wider dissemination of ideas. In this way knowledge-boundaries can become more permeable, easier to surmount, and allow researchers to germinate their ideas with a wider pool of viewpoints and tangents: "The greater access", Ruhleder concludes, "can balance out the relationship between poorer

---

[218] Snelgrove; 1990: 222.
[219] Johnson, Eric; Spring 1991: 8-17.
[220] Klein, Harald; http://ifsws.soziologie.uni-jena.de:80/home/klein/intexte.html; September 12, 1997; accessed September 18, 1997.

and richer departments, between scholars at the beginning of their careers and those at the end."[221]

A similar advantage of electronic text technologies is the ability to overcome the seemingly unbridgeable gulf between users and non-users of a language, thus enabling texts in an unknown language to be analyzed relatively effectively. A specific instance of this is the *Perseus Project*. This project was commenced in the Summer of 1987, with the early work undertaken at the Boston University Centre for Remote Sensing, and entails the creation of an electronic database - "Interactive Sources and Studies on Ancient Greece" - to include many aspects of classical Greek culture. The software allows the user to see the original Greek and its English translation side-by-side on the computer screen. This collection, yearly updated, contains:

> "the equivalent of 40 volumes of source material, several hundred encyclopaedia articles, a 35,000 word Greek-English lexicon, hundreds of maps, plans, and drawings, roughly 15,000 still color images, and 15 minutes of motion video."[222]

Further to this, an electronic milieu, Crane explains, not only:

> "allows authors to use written words but also to turn to still images, sound, or motion video when these are better suited to their purposes."[223]

Concomitant with easier and faster access to texts the sheer quantity of material can become more user-friendly and less daunting. This facilitates deeper investigation of both the original text and correlative material, the entry to which is, for example, is attained by 'clicking on' 'links' in hypertext and Internet documents. Raben observes that:

> "The ability to search...with the range, speed and precision that the computer makes possible will encourage greater exploitation of extant resources and diminish the tendency to ignore the past simply because there is so much of it."[224]

As Stubbs points out computers "help considerably with the criterion of comprehensiveness of coverage." Furthermore, that:

> "when correctly instructed, computers make it more difficult to overlook inconvenient instances, and to that extent move towards descriptive neutrality."[225]

---

[221] Ruhleder, Karen; Winter 1995: 5.
[222] Crane, Gregory; June 1991: 294.
[223] Crane, Gregory; June 1991: 302.
[224] Raben, Joseph; 1991: 347.
[225] Stubbs, Michael; 1996: 154.

Also wider access to texts may loosen an institution's ability to control the use and interpretations of a text, may enhance a reader's creative ability to extract meanings from a text, so undermining the appeals to 'traditional' authority.

A final advantage worth considering is that a statement can often become 'truthful', authoritative, and archetypal through one being able to: "muster on the spot the largest number of well aligned and faithful allies".[226] The process of enrolling others is crucial for empirical experience to become constituted as 'matters of fact'.[227]

This scenario can be facilitated in terms of access to a statement or document; an 'inscription'. Boyle purported that these 'matters of fact' could acquire certainty and validation through the multiplying of witnesses; through: "the aggregation of individuals' *beliefs*" (emphasis original). [228] For such:

> "Matters of fact were the outcome of the process of having an empirical experience, warranting it to oneself and assuring others that grounds for their belief were adequate."[229]

A similar process of enrolment takes place if the inscriptions are "immutable mobiles", which have the properties of being not only mobile, but 'immutable', 'presentable', 'readable', and 'combinable'.[230]

ii) Disadvantages

One disadvantage of the accessibility afforded by electronic 'multimedia' is a possible diminishment of scholarly authority:

> "material that it previously took years to glean is available immediately; consequently, far less prestige is invested in the possession of such "knowledge".[231]

But access is not all. It is only the beginning of research. The material still needs to be read, assimilated, and interpreted, the most human of work.

Another consequence of the electronic environment is a redefinition of the notion of 'text', whereby boundaries between author, text and user are

---

226 Latour, Bruno; 1990: 23.
227 Shapin, Steve and Schaffer, Simon; 1985.
228 Shapin, Steve and Schaffer, Simon; 1985: 25.
229 Shapin, Steve and Schaffer, Simon; 1985: 25.
230 Latour, Bruno; 1990: 26.
231 Martindale, Charles and Stoddart, Simon; 1994: 305.

disturbed, if not dissolved. For me what is important about this redefinition of the notion of 'text' is that the author is no longer seen as the sole and absolute producer of the text and its meaning, nor as the only authority for its interpretation.[232] The author is overthrown as the centred subject. Instead different readers reconstruct the text through their own variant interpretations of the meaning of the text. Thus, alluding to Barthes, Selden and Widdowson write that:

> "The worst sin a writer can commit is to pretend that language is a natural, transparent medium through which the reader grasps a solid and unified 'truth' or 'reality'."[233]

Furthermore readers are:

> "are free to open and close the texts' signifying process without respect for the signified. They are free to take their pleasure of the text, to follow at will the defiles of the signifier as it slips and slides evading the grasp of the signified. Readers are also sites of languages' empire, but they are free to connect the text with systems of meaning and ignore the author's 'intention'."[234]

In peer-reviewed, published scientific papers we find the positivistic use of rhetoric, whereby language is used to keep meaning and interpretative authority with the author. Thus presenting the text as a transparent medium to some unadulterated 'truth'; language is being used to manufacture certainties.

A cautionary riposte to the easier access to texts afforded by electronic technology is discussed by Ruhleder thus:

> "[T]he view is broadened in that one can "see" more texts from where one stands, but also shallower, as critical information is decoupled from the tool [electronic database] at hand."[235]

I think the problem here can be one of meaning. The electronic text can have the effect of distancing the text from the socio-cultural conditions of the text's production, thus richness of meaning is lost. Frederic Jameson makes a related point:

> "Unfortunately, no society has ever been quite so mystified in quite so many ways as our own, saturated as it is with messages and information, the very vehicle for mystification."[236]

---

[232] Selden and Widdowson; 1993.
[233] Selden and Widdowson; 1993: 131.
[234] Selden and Widdowson; 1993: 132.
[235] Ruhleder, Karen; Winter 1995: 7.
[236] Jameson, Frederic; 1981: 60-61.

The issue of electronic media distancing the writer and reader from the text is also addressed by Bolter who explains that an "unusual feature" of electronic writing is that it is:

> "not directly accessible to either the writer or to the reader. The bits of the text are simply not on a human scale. Electronic technology removes or abstracts the writer and reader from the text. If you hold a magnetic tape or optical disk up to the light, you will not see text at all. . . .In the electronic medium several layers of sophisticated technology must intervene between the writer or reader and the coded text. There are so many levels of deferral that the reader or writer is hard put to identify the text at all: is it on the screen, in the transistor memory, or on the disk?"[237]

What must be avoided is the trap of textual-technological determinism, and the "autonomous" model of literacy.[238] This model states that meaning can be lifted solely from the text. The complexities and vagaries of human creative work and social life, and their relationship, cannot be garnered from the words on the page. In records of experimental practice there is inevitably incompleteness.[239] This can be the result of gaps, where there are no written accounts for a particular day or even for a whole month. Or it can be due to the omission of what Polanyi described as tacit knowledge:[240]

> "the skills, techniques, assumptions of which practitioners were either unaware, or which, by their nature, could not be recorded in writing or drawings."[241]

Leydesdorff describes this difficulty in trying to understand or piece together the complexity of scientific practice :

> "[E]ach unit of analysis in science studies is a composite of, among other things, cognitions, texts, and scientists. These building blocks are different in their nature: one can not reduce scientists to their cognition or a text to it(s) authors, nor can a cognition be equated with the language in which it is expressed. Observable units of analysis, however, are complex; complex units may be taken as objects in quite different types of theories, since the aggregation and organization of the composite may refer to different systems and their development over time."[242]

---

[237] Bolter, David J.; 1990: 42-43.
[238] Hirsch, Eric; 1991: 305; Finnegan, R; 1988; Street, B; 1984.
[239] Gooding, David C.; 1989 (a): 64.
[240] Polanyi, Michael; 1964.
[241] Gooding, David C.; 1989 (a): 64.
[242] Leydesdorff, Loet; 1995: 12-13.

iii) Janus

When we look below the surface the notion of a Cartesian-like dichotomy of advantages separated from disadvantages of electronic text analysis does not stand up to scrutiny. The reality is more *Janus*-faced in character. This point is demonstrated by Coulthard who writes that the use of computing tools allow hypotheses to be tested over large corpora, and for results to be returned swiftly:

> "[T]he opportunity is the freedom to speculate and get fairly quick feedback from the computers about the accuracy and potential of the speculations. Far from restricting the theorist, computers will actually encourage hunch-playing and speculation at the creative stage."[243]

What is the *quid pro quo*? What compromises this rosy picture? According to Coulthard this advantage of computer tools can only be attained through crafting more precise and robust statements and intentions:

> "[W]hich will add pressure to move linguistics towards scientific rigour."[244]

The uncertain and ambivalent nature of computer-supported literature analysis is played by Mark Olsen, who states that the failure of computerized literary analysis to impact significantly on the discipline:

> "is traced to a concentration on how a text achieves its literary effect by the examination of subtle semantic or grammatical structures in single texts or the works of individual authors. Computer systems have proven to be very poorly suited to such refined analysis of complex language.[245]

Though with a reorientation of theoretical models underpinning computer-supported textual work:

> "Computer-aided methods are thus capable of opening up new areas of study, which can potentially transform the way in which literature is studied."[246]

2.2.13 Current Computer-assisted Linguistics and Text Analysis Centres and Research Projects

There are a great number of University departments and independent research institutions around the world concerned with computational linguistics, discourse analysis, textual analysis, natural language processing, and language

---

[243] Sinclair, John McH.; 1994: 19.
[244] Sinclair, John McH.; 1994: 19.
[245] Olsen, Mark; http://tuna.uchicago.edu/homes/mark/Signs.html; 1993; September 18, 1997.
[246] Olsen, Mark; http://tuna.uchicago.edu/homes/mark/Signs.html; 1993; September 18, 1997.

technology. Principal centres are located at Universities of Edinburgh, Oxford, Pennsylvania, Toronto, Princeton, and Rutgers, amongst others. A full description of several centres and institutions concerned with electronic texts and machine-based language analysis is located in appendix two.

2.2.14 Conclusion and Summary

In summary, the use of electronic databases and computational analysis tools is associated with a wide variety of advantages and disadvantages. We need to remember that the acquisition and use of any technology is never neutral and apolitical, and the consequences of its 'impact' are never fully transparent in advance:

> "The evolution of a technology is thus the function of a complex set of technical, social, economic, and political factors."[247]

Furthermore, although electronic technology may allow easier and wider access to databases and texts, their acquisition is only ever a first step. Computational tools are necessary to analyse the material and human skills to interpret the results. Thus access and the computational tools are not a panacea, for their usefulness depends on how results are interpreted and shaped by the research agenda and intended audience.

2.2.15 Research Questions and Issues in Light of the Historical Overview of Text Analysis

i) Overview of Current Issues Being Addressed In Computer-Assisted Text Analysis

Historically, text analysis, has been limited, *inter alia*, by the amount of written material it has been humanly possible to analyse. Modern text analysis, over the past century, has generally focused on authorship attribution and stylistics. The explosion in computer technology since the 1960's has allowed far more complex analysis not only on texts, but also on language. The interactive text retrieval and analysis programs available today are quite removed from the mechanical study of style and authorship we have seen in the articles by

[247] Wajcman, Judy; 1991: 23.

Mendenhall. This must be qualified on two counts. First, the advent of computerized applications has brought its own attendant problems. Early projects in the humanities were hindered by limitations in hardware and software, such as lack of storage space, a problem remedied in the 1970's when magnetic tape storage was replaced by disk storage, allowing random, rather than sequential, data access.[248] Second:

> "Essentially we are still able to only search text by specifying strings of characters, possibly linked by Boolean operators, whereas most users are interested in concepts, themes and the like. String searches cannot effectively disambiguate homographic forms, for example, 'bank' as in money bank as opposed to 'bank' of the river, or the verb 'bank' (used of an airplane)".[249]

Today researchers in computational linguistics are attending to large-scale corpora and natural language understanding systems, after years spent working with rule-based systems. Applied computational linguistics focuses on the practical outcomes of modelling human language use. There are also computer applications and software available which attempt to perform morphological and syntactic analysis, though success has so far proved limited.[250] The design of such software for linguistic analysis is undertaken by, for example, the Multi-Lingual Language Technology team at the Rank Xerox Research Centre.[251] One of their projects is called 'Finite-State Technology', which involves the design of morphological analyzers for various languages.[252]

Thus computer-based methods are allowing the investigation of new issues and questions, and assisting the development of new tools to analyse text. Though it is also the case that computer technology has breathed new life into humanistic tools much older than computers. For example, concordances, which they have been around since the middle ages, and: "are suitable to a number of different purposes, not only philological ones."[253] Also, returning to the work by Mendenhall (section 2.2.7), we saw how he employed rudimentary statistics in

---

[248] Sutton, Brett; 1994 quoted in, Hockey, Susan *et al*; 1996: 58.
[249] Sutton, Brett; 1994 quoted in, Hockey, Susan *et al*; 1996: 64.
[250] Sutton, Brett; 1994 quoted in, Hockey, Susan *et al*; 1996: 64.
[251] MLLT Home Page; http://www.rxrc.xerox.com/research/mltt/home.html; 12 September 1997; accessed September 23, 1997.
[252] Finite-State Technology Home Page; http://www.rxrc.xerox.com/research/mltt/fsHead/home.html; May 8, 1997, accessed September 23, 1997.
[253] Aarseth, Espen; http://www.hf-fak.uib.no/hi/espen/HI.html; September 24, 1996; accessed September 21, 1997.

his authorship studies (stylistics). For over three decades computational stylistics have been in use, particularly in the field of authorship attribution.

To sum up, currently in text analysis, with the application of computer technology and use of electronic texts, we have arrived at a particular juncture. Computerized tools are, on the one hand, allowing the investigation of new issues and questions, such as those concerned with morphology and meaning. On the other hand, this technology is being applied to older concerns, pre-dating computers, such as the building of concordances and stylistics.

ii) Issues for Text Analysis and Scientific Writings Central to my Research

On a general level meaning is an issue being addressed by my research. I will explore how amenable analysis of meaning is to quantitative and computer-assisted text analysis. Moving away from the general to the more specific, one key issue for my research is how far can textual features and indicators be used to investigate the role of language use and change in the development of ideas; as an indicator of thinking. For this purpose relations between private and public writings are regarded as existing on a continuum. A related issue concerns the reconstruction of investigative pathways scientists tread to reach the conclusions published in papers; from exploratory work through to the public discourse. Here language is perceived as a process and: "as one among many *activities* associated with experiments".[254] Language development and use, for example, in terms of electromagnetism in chapter eight, is followed from the earliest conception in personal writings through to the 'final product' in the public arena, in the form of a published article.

iii) Questions for Text Analysis and Scientific Writings

The first question this research will aim to answer is: 'Is it possible to use a set of textual features to recover a scientist's thinking in the development of an idea or theory?' Another question addressed is: 'What type of text analysis is required to test the conjecture that scientific writing is a form of literature?' A third question asks: 'How far is it possible to recover experimental practice or

[254] Zahar, Alexander; 1995: 139.

discovery narratives using the analysis of private scientific documents analyzing language use and change?'

2.3 Language and Meaning as Issues for Linguistics, Science Studies, and for Science

2.3.1 Abstract

The purpose of part three is to place the constructivist view of language within the wider context of linguistics and literary theories, of science studies, and feminist critiques of language and science. This includes a look at the role of language in scientific discourse. Science writing is viewed as a form of literature, constructed as a social and human process, with multiple interpretations, some sign-posted, some hidden away, all with an intent to communicate, if allowed:

> "Readers, particularly "trained" readers, are programmed to read a text as if it were a code. In order to discover the literal meaning behind the text, readers interpret the messages offered by the text. To some extent, all texts have "solutions" or "meanings," albeit, not usually only one solution or meaning, which readers believe they can discover and interpret, as if the text has a key which automatically unlocks the door to the text's true significance."[255]

2.3.2 Introduction

Mainstream, analytical philosophers of science and the AI community view language as a linear series of narratives, sequential in construction. This is also a relatively static theory of language development. They tend only to study monographs or textbooks which do not convey the procedural, skilled aspects of theoretical and experimental work. This can lead to an underestimate of the amount of qualitative, constructive work that enables quantitative data.

The view of language applied in this research emphasizes non-linearity, that language is slippery and indeterminate. Furthermore for this critique of

[255] Privett, Ronna; http://pegasus.acs.ttu.edu/~ykfrp/pynchon.htm; December 1st, 1996; accessed September 19, 1997.

language, the so-called 'new rhetoric',[256] centred on the work of Kenneth Burke,[257] views human discourse in terms of starting:

> "with the assumption that all discourse is somehow addressed to an audience, either real or imagined, and takes as its responsibility the elucidation of how speakers or writers adapt to and shape their audiences."[258]

The fallible and corrigible nature of scientists work is depicted in the imagery of Neurath:

> "We are like sailors who have to rebuild the ship on the open sea, without ever being able to dismantle it in dry-dock and reconstruct it from the best components."[259]

This part of the chapter places the constructivist, post-positivistic philosophy of language in my research within a tripartite context of: i) linguistic and literary theories, ii) science studies, that is, feminist, scientometric, and post-humanist perspectives, and iii) science, where I elaborate upon my view that science writing is a genre of literature. This also incorporates the belief that an analysis of private and public prose can reveal the non-sequential, constructed, and fallible character of scientific writing, thus of scientific knowledge itself.

2.3.3 Perspectives on Language for Linguistic and Literary Theories: Structuralism, Post-Structuralism, Post-Modernism, and Feminist Standpoints

Structuralism is the 'scientific' and ahistorical attempt to disclose all the rules and systems which form the framework underpinning all human and social behaviour. The foundation stone of structural linguistics were the insights into language of Ferdinand de Saussure who recognized that the study of language was also synchronic. It was this opposition of a synchronic (unhistoric and abstract) to a diachronic view of language, the distinction between diachrony and synchrony, which shaped and inspired modern linguistics, both being widely accepted, and criticised, for example, by the Prague Linguistics School. In Saussurean linguistics, Fowler attests:

> "language is an autonomous system, self-contained, self-regulating and quite arbitrary in its genesis and its relations with non-linguistic world".[260]

[256] Simons, H. W.; 1990, and Aviva, Freedman and Medway, Peter (eds.); 1994: 3-4.
[257] Burke, Kenneth; 1950.
[258] Halloran, Michael S.; 1978: 79.
[259] Neurath, Otto; 1983: 92.
[260] Fowler, Roger; 1991: 26.

Similarly for Wales:

> "Saussure's opposition between the abstract or systematic *langue* and the concrete individualistic *parole* belies an understanding of language as a living, diversified construct between living, diversified speakers; language as an 'utterance' in a social, and historical, context".[261]

For Saussure language is not a: "word-heap gradually accumulated over time",[262] but instead operates as a system of signs (one among many), and that all words, and all communicating objects, are signs. A sign has two aspects: the *signifier* and the *signified*:

> "The signified is not a thing but the notion of a thing, what comes into the mind of the speaker or hearer when the appropriate signifier is uttered. The signifier thus constitutes the material aspect of language: in the case of the spoken language a signifier is any meaningful sound which is uttered or heard, in the case of the written language it is a meaningful mark inscribed on the page. "[263]

The perspective in my research does not see language as a closed, autonomous system. Rather language, as well as communication, is seen as a human and cultural product and bound up in a complex of human experiences, and interpersonal, social and historical contexts. It is also a social, creative tool; a third-party - the intended audience of our discourse - is always present when we use language.

The study of signs instigated by Saussure gave way, later in the century, to semiology in the West European countries, especially in France, and semiotics in the USA.

The Formalist movement saw its conception in Russian Formalism. The Moscow Linguistic Circle was founded in 1915, whose leading members included Roman Jakobson and Petr Bogatyrev (who later were founding members of the Prague Linguistics Circle), and Opojaz (the acronym for 'The Society for the Study of Poetic Language') in 1916.[264]

The Russian Formalist group sought to establish literary theory on a 'scientific' footing. As van Dijk puts it these scholars: "all seem highly preoccupied with the scientific status of the *forms* of language, literature, discourse or other semiotic

---

[261] Wales, Kathleen; 1991: 178.
[262] Selden, Raman and Widdowson, Peter; 1993: 104.
[263] Sturrock, John (Ed.); 1979: 6.

artifacts."[265] But there were various phases of Russian Formalism, and Peter Steiner has constructed a tripartite metaphor-based system for capturing the historical evolution of Russian Formalism,[266] which Selden and Widdowson describe thus:

> "The model of the 'machine' governs the first phase which sees literary criticism as a sort of mechanics and the text as a heap of devices. The second is an 'organic' phase which sees literary texts as fully functioning 'organisms' of interrelated parts. The third phase adopts the metaphor of 'system' and tries to understand literary texts as the products of the entire literary system and even of the metasystem of interacting literary and non-literary systems."[267]

The early Formalists (the first phase), particularly Shklovsky, Tomashevsky, and Eikhenbaum, were concerned with an explicit distinction between form and content, and they:

> considered that human 'content' (emotions, ideas and 'reality' in general) possessed no literary significance in itself, but merely provided a context for the functioning of literary 'devices'...[also aiming] to outline models and hypotheses (in a scientific spirit) to explain how aesthetic effects are produced by literary devices, and how the 'literary' is distinguished from and related to the 'extra-literary'."[268]

In my research the 'human content' is very important when considering the role of language in communication and in the development of ideas. I believe there are no binary oppositions such as form/content, literary/extra-literary, or language/world.

The later and most prevalently 'structuralist' Russian Formalism was embodied in the work of Jakobson and Tynyanov, whose enunciation's were known as the 'Jakobson-Tynyanov theses' (1928), which:

> "reject a mechanical formalism and attempt to reach beyond a narrow literary perspective by trying to define the relationship between the literary 'series' (system) and other 'historical series'. The way a literary system develops historically cannot be understood...without understanding the way in which other systems impinge on it and partly determine its evolutionary path. On the other hand...we must attend to the 'immanent laws' of the literary system itself if we are to understand correctly the correlation of the systems."[269]

---

[264] Selden, Raman and Widdowson, Peter; 1993:28.
[265] van Dijk, Teun A.; 1985: 2.
[266] Steiner, Peter; 1984.
[267] Selden, Raman and Widdowson, Peter; 1993: 28.
[268] Selden, Raman and Widdowson, Peter; 1993: 27-28.
[269] Selden, Raman and Widdowson, Peter; 1993: 42.

Formalism and the 'structural' approach were continued and developed by the Prague Linguistic Circle. This school was initiated in 1926, and held discussions in Prague up to 1945, by a number of Czech, French, and Russian linguists.[270] The inspiration for the Circle's work were the principles of Saussure, and its aim was to investigate *la langue*[271] as a "functional system".[272] The program of the Prague Circle was entitled: "Methodological Problems Stemming From The Conception Of Language *{la langue}* As A System And The Significance Of This Conception For Slavic Languages".[273] More specifically, the Prague linguistics applied Saussurean principles to a study of phonological theory, such that:

> "Prague scholars did not treat the phoneme as a mere class of sounds or as a transcription device, but as a complex phonological unit realized by the sounds of speech."[274]

The Czech linguistics of the Prague Circle emphasized that the study of phonological theory must not be separated from the wider social and cultural context.

Russian Formalism tended to remove the socio-historical context from language, thus arising in the later period of Formalism was The Bakhtin School, whose primary exponents were Mikhail Bakhtin, Pavel Medvedev, and Valentin Voloshinov. They were concerned with language as a human and social construction, not an abstract system; the relationship between language, social and historical meanings, ideologies, and personal knowledge is complex and inseparable. Thus for the Bakhtin School:

> "'words' are active, dynamic social signs, capable of taking different meanings and connotations for different social classes in different social and historical situations."[275]

The philosophy running through my research follows the Bakhtinian principles. That is to say, to hope to understand how Faraday is thinking and using language we need to situate the textual data in the wider social and historical contexts.

---

[270] The Czech linguists: V. Mathesius, B. Havránek, J. Mukarovský, B. Trnka, J. Vachek, and M. Weingart. The French linguistics: L. Bruo, L. Tesnière, J. Vendryes, E. Benveniste, and A. Martinet, and the Russian linguistics: R. Jabokson, P. Bogatyrev, and N.S. Trubetskoi. Kristeva, Julia; 1989: 225-226.

[271] Defined by The Prague Circle as: "goal-orientated means of expression" in Steiner, Peter (Ed.); 1982: 5.

[272] Kristeva, Julia; 1989: 226.

[273] Kristeva, Julia;1989: 226.

[274] Robins, R. H.; 224-225.

[275] Selden, Raman and Widdowson, Peter; 1992: 38.

It was during the 1960s that structuralism gave way to post-structuralism, associated with thinkers such as Barthes, Derrida, Foucault, Kristeva, and Lacan. Whereas structuralism views language as a medium for reflecting a pre-given reality and as an impersonal system, poststructuralism employs the conception of 'language-in-use'. That is, as Snelden and Widdowson write:

> "Language cannot neatly be dissociated from social living; it is always contaminated, interleaved, opaquely coloured by layers of semantic deposits resulting from the endless processes of human struggle and interaction."[276]

Postmodern thought perceives language as: "open, discontinuous, improvisational, indeterminate, or aelotory", with an absent centre and embodying an "ontological uncertainty".[277]

A central theme for feminist critiques of language is the 'silence or absence of female voices'. Certain linguistic genres are typically associated with women. Cameron gives as examples of these genres: "'gossip', storytelling, private letters, and diaries".[278] These forms of languages tend to be associated with the private or domestic sphere. Those genres affecting greatest social currency - "religious ceremony, political rhetoric, legal discourse, science, poetry" - women's voices are generally absent from; silent. Women:

> "often are explicitly *prevented* from speaking, either by social taboos and restriction or by the more genteel tyrannies of custom and practice."[279]

2.3.4 Discussion

In the course of looking at primary stages in the development of linguistic and literary theories this century we have seen perspectives on language change quite radically. At the outset the Saussurean view of language as an autonomous, self-contained system reflecting a pre-given world of objects dominated. Language as impersonal and a system separate from the world was sustained by Russian Formalism and the Prague Circle linguists. We arrive at a conception of language shaped by constructivism, post-structuralism, and contemporary feminist literary theories. Language is an organic, dynamic tool, constructed through use, and in which we come to represent a reality.

---

[276] Selden, Raman and Widdowson, Peter; 1993: 127.
[277] Selden, Raman and Widdowson, Peter; 1993: 177-178.
[278] Cameron, Deborah (Ed.); 1990: 3- 4
[279] Cameron, Deborah (Ed.); 1990: 3- 4.

Scientists, just as we all do, use language - rhetoric - to create 'truths' and persuade the reader of the veracity of the statements made; certainty can only come about if consensus is achieved and then only provisionally.

2.3.5 Linguistic Discourse Analysis

I will close this look at the linguistic and literary context for my research by talking about linguistic discourse analysis and its relevance to my project. More specifically I shall examine two pioneering studies in linguistic discourse analysis, one by Harris (1952),[280] and the other by Mitchell (1957),[281] which can be seen as prominent representatives of two traditions - the Bloomfieldian and the Firthian.

The article by Harris is set within the Bloomfieldian framework, whereby Bloomfield asserted that linguists should not be concerned with meaning, but instead with form and substance,[282] and it embarks upon an: "analysis of connected speech (and writing)".[283] The separation between meaning and personal knowledge, with the suppression of the latter, is central to Harris's method, which depends:

> "only on the occurrence of morphemes as distinguishable elements; it does not depend upon the analyst's knowledge of the particular meaning of each morpheme."

The discourse analysis of Harris is concerned with suprasentential structure as the object of analysis:

> "Language does not occur in stray words or sentences, but in connected discourse - from one-word utterance to a ten volume work, from a monolog to a Union Square argument. Arbitrary conglomerations of sentences are indeed of no interest except as a check on grammatical description; and it is not surprising that we cannot find interdependence among the sentences of such an aggregate."[284]

Traditionally linguistics had emphasised the minimal units of language, and made no provision to study beyond sentences.

---

280 Harris, Zellig S.; 1952: 1-30.
281 Mitchell, T. F.; 1957: 31-71.
282 Coulthard, Malcolm; 1985.
283 Harris, Zellig S.; 1952: 1.
284 Harris, Zellig S.; 1952: 3.

In contrast, Mitchell's analysis is motivated by the Firthian tradition of the intimate relationship between language and meaning. The London School was closely associated with the work, primarily in phonological theory and semantic theory, of John Firth, who, from 1944 to 1956, was the first British Professor of General Linguistics.[285] Coulthard writes that:

> "For Firth language was only meaningful in its context of situation; he asserted that the descriptive process must begin with the collection of a set of contextually defined homogeneous texts and the aim of description is to explain how the sentences or utterances are meaningful in their contexts."[286]

It is the context in which language is located, including its historical evolution and cognitive representation, which lends meaning and significance to language. As Stubbs expounds:

> "language, action, and knowledge are inseparable...communication is impossible without shared knowledge and assumptions between speakers and hearer."[287]

This is quite the antithesis of the classical type of structuralism whereby individuals are understood through an underlying system; it is the systems which construct meaning and not the individual. We can see how such a system of thought requires a static, stable, and 'transparent' reality to bear fruit. This can be put into perspective through a contrast with a poststructuralist view on reading and meaning, where the human-meaning interaction is indeterminable, interminable, chameleon-like, and inherently unstable:

> "reading itself has become a signifying practice. Rather than the classical bourgeois readable *(lisible)* text, which reduces the reader to a passive consumer, the writable *(scriptable)* text is liberating. It forces active participation in the production of literature itself. The reader always remains linguistically (i.e. semiotically) constituted within the broader domain of intertextuality and within a cultural context. Consequently, reading, though indefinitely variable, is never free of presuppositions, and remains caught in the web of meaning, an unanalyzable whole. Textual indecidability becomes the norm."[288]

The Firthian tradition is upheld by my study of scientific language. This perspective underpins the need to analyse Faraday's language use and the

---

[285] Bright, William (Editor in Chief); 1992: 170-172.
[286] Coulthard, Malcolm; 1985: 1.
[287] Stubbs, Michael; 1983: 1.
[288] Bright, Wiliam (Editor in Chief); 1992: 409 (column one).

development of his concepts and ideas across his personal and published writings, as well as within the biographical, social, and historical settings.

An issue to bear in mind is that of avoiding being seen to endorse textual and cultural determinism. That is to say, texts do not contain unassailable objective facts, which affect the reader in an unmediated sense, but neither are readers able to make what they will of a text. First, there are grammatical and stylistic features of a text which 'favour' certain interpretations. Then there are literary and social conventions which readers live within. Frawley points out that:

> "The knowledge does not exist in the text; texts are not containers; there would be no knowledge unless the text were read. Knowledge is knowing, and knowing is the activity of engaging the text."[289]

What varies is the degree to which different texts are objective or 'robust' and are pre-interpreted. The structure and features of a text - the typeface, the use of the active and passive voice, the sequence of events, tropes, claims to impartiality, for example, as well as the level of social sanctions supporting a canon of authorized and 'preferred' interpretations, help construct particular readings of it. Texts are judged through 'horizons of expectations', as Jauss put it.[290] Duchan points out that:

> "discourse is designed to fit the intended audience. Linguistic devices have been identified that signal a listener or reader how to understand the narratives. Intensity markers such as "really" tell the audience what is important...words and phrases such as "so," "then," "anyway," "by the way" can be used to mark degree of continuity or discontinuity in the discourse".[291]

Similarity, reading and interpretation are as much a political act as writing for which: "an individual's choice of words is a form of selected action or behaviour"[292], and reading-interpretation: "is always a politically-interested act of persuasion."[293] Readers cannot escape indexicality and interpretative flexibility, though they are an active constituent of the context and shape of its impact. Analysis of prose, Coupland argues, should:

[289] Frawley, William; 1987: 55.
[290] Jauss, Hans Robert.; Brighton; 1982.
[291] Duncan, John; 1994: 4.
[292] Rabon, Don; 1994: 16.
[293] Mailloux, Steven; 1995: 127.

> "[F]ocus not on the text isolate but on the text as composed and...as performed...with the pre- and post-textual processes."[294]

To conclude this section with a final point about the shared space of the textual milieu and the socio-political world. An author's work, whether a poem, a scientific monograph, or a piece of literary criticism cannot be separated out from the economic, social, political, emotional, and gendered contexts. This has been a primary contention of Marxist, Frankfurt School (Critical Theory), Post-Structuralist, and Feminist positions.[295] Thus it is always very difficult, if not impossible, to fully describe an author's intentions at the time of writing, especially as the meaning of writing changes in content and quantity as writing builds on writing.

2.3.6 Language and Science Studies

i) Feminist Science Studies

Meaning is made, it is corrigible and never fully transparent. It is always created by particular social groups and around certain agendas and beliefs. There is no communication situation where all participants are fully aware of each other's perspectives and interpretations.[296] Ginzberg purports that: [w]oman *are* trapped in an androcentric[297] world....one in which language and meaning have been constructed around androcentric goals and enterprises."[298] (emphasis original) The authority and robustness of scientific knowledge, *inter alia*, is conferred by a particular definition of objectivity. An objectivity premised on reducing nature to separate objects, where the scientists relationship to these 'objects' is filtered out, a process at its apotheosis in the recounting of experimental practice in published science papers. This illusion, Hubbard believes, is reinforced by scientific prose:

> "because it implicitly denies the relevance of time, place, social context, authorship, and personal responsibility."[299]

---

[294] Coupland, Nikolas (Ed.); 1988:12.
[295] Pease, E. Donald; 1995: 111.
[296] Habermas, Jurgen; 1990.
[297] *andro* Greek for male.
[298] Ginzberg, Ruth; 1989: 81.
[299] Hubbard, Ruth; 1989: 125.

Hilary Rose has stated that the gendering and racializing of science is present in laboratories practices and embedded in knowledge.[300] Gender is a cultural construct rather a biological given, thus is acquired through social and cultural practices. Jordanova claims:

> "As embodiments of such practices, the natural and bio-medical sciences fully participate in the forging of gender, and are themselves informed by it. As fields with a privileged relationship to nature, they play a major role in explaining and disseminating gender as a naturalized category."[301]

While gender is only one piece in the complex jigsaw puzzle of science studies, it:

> "may well prove to be a powerful tool for opening up a more critical awareness of our own forms of being and knowing, as well as for interpreting the past."[302]

ii) Scientometrics

The field of scientometrics involves research concerned with the quantitative features, dynamics, and characteristics of science and technology, such as calculations of publication distributions, co-citation mapping of research specialities, co-author and co-term analyses. Scientometrics is closely related to communication studies, thus it also investigates the creation and dissemination of scholarly or authoritative information. Our interest lies with scientific discourse, which, for Leydesdorff:

> "are...debates about the methods appropriate to warrant the evidence adduced. Scientific methods are embedded in scientific practices; they are not 'meta', 'but 'epi' to the sciences."[303]

An unpacking of scientific developments to reveal the underlying networks through analysis of co-occurrences and co-absences is seen as a fruitful approach by, amongst others, Hesse[304] and Callon *et al*[305]. Leydesdorff can see drawbacks with this approach. One such drawback is that document structure analysis in terms of co-words would succeed best in restricted document sets (selected from a wider database) rather than in a natural language context.[306]

---

[300] Rose, Hilary; 1994.
[301] Jordanova, Ludmilla; 1993: 482.
[302] Jordanove, Ludmilla; 1993: 483.
[303] Leydesdorff, Loet; 1995: 28.
[304] Hesse, Mary; 1980.
[305] Callon, Michel *et al*; 1983: 191-235.
[306] Leydesdorff, Loet; 1995: 38-39.

iii) Post-Humanism: Human and Non-Human Agency

Human imagination and creative behaviour are an outcome of the process of interconnectedness - a fluid interweaving dance of the human, the material, the corporeal, the technological, and of embodiment; creating rhythms of growth and change. All are bound together and given meaning through metaphors and rituals. Human assumptions of species superiority are a tired and outmoded concept of life. Western science since the sixteenth century has played on this concept. This is the public face of science. If we zoom in closer and closer at the detail of the practices of science we experience science as just part of the 'fluid interweaving dance' of life:

> "We can no longer place the assemblages on a quantitative scale measuring how close or far they are from the plane of consistency. There are different types of abstract machines that overlap in their operations and qualify the assemblages...Every abstract machine is linked to other abstract machines, not only because they are inseparably political, economic, scientific, artistic, ecological, cosmic -- perceptive, affective, active, thinking, physical, and semiotic -- but because their various types are as intertwined as their operations are convergent. Mechanosphere."[307]

This research supports the view that texts and language are not passive receptacles of single meanings but, rather, they are sites of agency. Language is the medium in which science develops. Scientific practices are constructed through the interaction of numerous human and non-human agents. Pickering in his *Mangle of Practice* describes this intermingling of agencies:

> "Scientists are human agents in a field of material agency[308] which they struggle to capture in machines. Further, human and material agency are reciprocally and emergently intertwined in this struggle. Their contours emerge in the temporality of practice and are definitional of and sustain one another."[309]

In traditional sociology of scientific knowledge the human subject is decentred and one can have human and nonhuman agency but not both: "Traditional science studies is asymmetric about agency, recognising genuine agency only in the human realm but not in nature, which is typically regarded as inert matter,

---

[307] Deleuze, Gilles and Guattari, Felix; www.scsn.net/~efolley/; accessed September 19, 1997.
[308] Pickering, Andrew (1995; 6) when discussing material agency describes how: "[t]he world is continually *doing things*, things that bear upon us not as observation statements upon disembodied intellects but as forces upon material beings." This agency: "comes at us from outside the human realm and....cannot be reduced to anything within that realm."
[309] Pickering, Andrew; 1995; 21.

passively awaiting representation."[310] Pickering, as with actor-network theory, perceives of scientific practice as a field of human and non-human actors (the latter includes skills, social relations, machines and instruments, as well as scientific facts and theories), where human intentionality is no longer privileged, so that the distinctions between the human and material factors are subverted, creating a 'posthumanist' space:

> "a space in which the human actors are still there but now inextricably entangled with the non-human, no longer at the center of the action and calling the shots."[311]

Interpreting language as a site of agency makes us analyze it differently because, first, we cannot see meaning as an homogenous entity able to be lifted directly from the text. Instead meaning is plural and changes with language use, the intended audience, and the interpersonal and social environments. Second, language does not have a passive referential or representational relationship with the world. Rather the world has existence in and through language and communication; all human behaviour is language-embedded. Also language is a human and social product, thus its existence is brought about by being used. What this means for my research is that Faraday is not using language to describe his experiments and his ideas, rather language-use - the change in words and their use - has a crucial shaping effect on what comes to labelled as experiment *x* or concept *y*.

2.3.7 Language and Scientific Prose

As this research is concerned with how new language, ideas, and concepts arises within science this section will deal specifically with the issues of language and meaning for science. The illusion of linearity and sequentiality in scientific discourse will be discussed first. The next section deliberates on the 'ordinary' nature of scientific language and meaning. This is followed by a treatment of the issue of 'scientific English', continuing with a look at whether science writing is formulaic or can take numerous forms, and argues that science writing is another literary genre. Finally, I look at the issue of reconstructing experimental narratives using private and public writings.

---

[310] Pickering, Andrew; gaia.lis.uiuc.edu/leep3/stim/pickcyb.html; 1995; accessed September 19, 1997.
[311] Pickering, Andrew; 1995; 26.

Before proceeding I want to clarify what is meant by 'science'. This research embodies the post-Kuhnian, constructivist conception of science. That is, science as skilled labour, employing tacit knowledge, craft skills, and practical skills. Public, 'objective' knowledge emerges out of creativity, which is contingent, personal, and idiosyncratic. What is rejected is the 'traditional' Whig and 'great man' history of science. In the case of Faraday, he is credited with the discovery that certain substances when dissolved in water could conduct an electric current, concluding that atoms were electrically charged and he called those charged particles, ions. Faraday also formulated the law of electromagnetic induction, whereby a changing magnetic field induces an electric field. But he did not work alone in the basement laboratory of the Royal Institution, being surrounded by assistants, servants, and instrument makers. This science is in the business of construction, construal, and persuasion. Woolgar states that:

> "Scientists are not engaged in the passive description of pre-existing facts in the world, but are actively engaged in formulating or constructing the character of the world.[312]

This point about the collaborative and socially-embedded nature of scientific work is important as it applies to the notion of 'private' discourse used in this research. We want to avoid exaggerating the independence of private and personal actions from the encompassing cultural and interpersonal processes. For:

> "Even the most private deliberations participate in conceptual, practical, methodological and other resources, on which all scientists rely in order to think, work, communicate and defend their results."[313]

Thus we are concerned with 'science' as a human activity, its practices inseparable from wider cultural issues, such as gender, politics, and economics, as well affect and cognitive attributes. This can be evidenced in the life and work of Faraday. Faraday constructed new experimental practice, he was an important part of the European wide movement which opened out construction of new theory in electro-magnetism, set into motion by Oersted in 1820, and invented the concept of a field of force. Faraday was also a member of: "a very

---

[312] Woolgar, Steve; 1988: 87.
[313] Gooding, David C.; 1989 (b): 423.

small and despised sect of Christians", the Sandemanians.[314] Cantor argues that:

> "Faraday's science bore the imprint of his religion in many different ways....Faraday did not classify science (except for the politics of science) as a mundane activity. Instead he conceived science to be closely akin to true Christianity, since order and peace were to be found in both. Moreover, in engaging in science he was pursuing a quasi-religious activity, since he was discovering the way God had constructed and ordered the Universe."[315]

This opinion is shared by Habgood in an address to the Royal Institution in 1991:

> "in Faraday it has become increasingly obvious that the two worlds in which he lived, the world of the laboratory and the world of a small exclusive sect, were connected."[316]

We have a picture of a science which is about creativity and skills. Then we look at the public, published accounts of discovery and findings, and a process of 'creative deskilling' has occurred. The original concepts have had their rough edges removed, content has been lost, and standardization of facts is one consequence.

We can take the ideas of non-linearity and non-sequentiality a stage further. It is the reader who assembles and reinterprets the text, the boundaries between reader and text disintegrates, and the text is not necessarily perceived as an otherness.[317] This could be said of private documents and non-scientific literature. To borrow from Baudrillard, the absence of otherness can lead to:

> "an artificial dramaturgy stimulating and dramatizing the absence of the other; in this dramaturgy, the subject becomes interactive, a candidate for all possible connections and combinations."[318]

Whereas a scientific monograph is constructed around a sequential model of a text and a linear model of language. In this case the number of connections and combinations to be navigated and the space for interpretative flexibility are minimized, so the science text appears to speak univocally; the style of scientific reasoning works to diminish capriciousness.[319] The result is that the

[314] Faraday, Michael in a letter to Ada, Countess of Loveleace, 24 October 1944, in Jones, Bence Henry; 2; 1870: 191.
[315] Cantor, Geoffery; 1991; 10 and 286.
[316] Habgood, John, Archibishop of York; 1992: 1.
[317] Paul; Christiane; August 1995.
[318] Paul; Christiane; August 1995: 265.
[319] Hacking, Ian; 1990.

myth of 'science as paradigm' and scientists uncovering unequivocal, objective facts about nature is perpetuated. The illusion of certainty and final answers in the world is maintained. It is as if the scientific image of the world, as Kant and Quine maintain, could only contain real 'knowledge': "[I]t is within science itself...that reality is to be identified and described."[320]

2.3.8 The Meaning and Role of Language in Science

The seventeenth-century witnessed the practitioners of the 'new' science cast disparaging overtures at rhetoric, which was deprecated as mere verbal dressing, whereas scientific prose was to aspire to what Thomas Sprat referred to as:

> "the primitive purity and shortness, when men deliver'd so many things, almost in an equal number of words."[321]

Sprat insisted upon an anti-Ciceronian 'plain style' for scientific writing; a "mathematical plainness", encapsulating his "formulaic one-word/one-thing ideal".[322]

Language, in opposition to this static and mathematical conception put forward by Sprat, is a dynamic process, enabling and disabling social action, and is a technology used in the construction of knowledge. This view contrasts with the more orthodox, analytical position:

> "the scientific idea of language, where language is seen as a formal and *passive* structure - syntax, semantics, phonetics - a structure that is passively available to, but does not define or organize, the human beings who use it....[t]hus science must presume that scientific language can somehow be used to study the structure of language *from the outside*, as though scientific language is detached from and external to language."[323]

The rapid growth in science in the eighteenth and nineteenth centuries required new words for newly described objects or phenomena. Events and hypotheses need to be converted into facts through being expressed in fact-affirming language.[324] Language confers legitimation on scientific knowledge, where:

---

[320] Quine, Williard van Orman; 1981: 21.
[321] Jones, R. F.; 1971: 63.
[322] Walters, Frank D.; 1993: 239-258.
[323] Wright, Will; 1992: 16, 19.
[324] Latour, Bruno and Woolgar, Steve; 1979: chapter three.

"language is seen as providing privileged explanatory access to true reality".[325] The human activity of science attempts to construct language that has a precise and univocal meaning. This was the task when science was undergoing a rapid growth in knowledge which required a burgeoning lexicography.[326] Natural processes came to be able to be seen only through physical experiments; a reality was defined that was detached from the human beings experiencing it. The scientist through language comes to understand knowledge as the asocial, ahistorical mathematical representation of reality.[327] I argue that scientists do not use language to furnish pristine insights in the characteristics of nature, but rather scientists produce 'metaphoric redescriptions' of nature. Scientific discourse does not constitute a separate, special language, but instead scientific language borrows from and feeds into ordinary, everyday discourses. As Sellars claims:

> "the scientific enterprise is the flowering of a dimension of discourse which already exists in what historians call the "prescientific stage", and that failure to understand this type of discourse "writ large" - in science - may lead, indeed has often led to a failure to appreciate its role in "ordinary usage", and, as a result, to a failure to understand the full logic of even the most fundamental, the "simplest" empirical terms."[328]

The view of language in this chapter rejects viewing scientific language through the lens of the 'representational' philosophy of language, whereby language is seen as a passive vehicle for communicating ideas connected with immutable objects in the world. Rather language is a tool used to construct knowledge; to render the culturally invisible, empirically visible. As Golinski puts it:

> "language [is] not a tool to serve individual persuasive purposes but a common resource by which shared knowledge [is] constructed and bonds of trust reinforced."[329]

2.3.9 Scientific English and Text

Scientific texts constitute finely crafted case-studies of textual or linguistic determinism and reductionism. The rhetorical devices and stylistic structure employed by the genre of scientific discourse aim to 'wash out' background information, the labour involved in experiments, subtext, and temporality. This is

[325] Wright, Will; 1992: 60.
[326] Gooding, David C.; 1990: chapter one.
[327] Wright, Will: 1992.
[328] Sellars, Wilfred; 1956: 302.
[329] Golinski, Jan; 1992: 250.

a view of text as a watertight receptacle of given, prearticulated certainties, undisturbed by 'extra-linguistic considerations, and being consumed without fear of indigestion by passive readers. Not a view shared by this research.

Modern science seems to put us in the Nietzschean position of having to philosophize and garner knowledge without foundations. Fawley talks of:

> "the fallacy of containedness...[and that] texts have no insides or outsides as such, either. Where is the inside of the text which contains all of the meaning? Meaning is put into texts and sent across to readers who extract it felicitously. Texts, like words and sentences, are activities."[330]

It is through these texts that the 'sciences' are taking up doctrinaire positions about the structure of the universe. Texts create the distance necessary for the reification of knowledge to take place, and for active subjects to be made into objects, fit for transparent analysis; the "concretization of truth" takes place.[331] Halliday makes the point about 'scientific English', whereby this genre or 'register':

> "[H]as to reconcile the need to create new knowledge with the need to restrict access to that knowledge (that is, make access to it conditional on participating in the power structures and value systems within which it is located and defined)."[332]

Once a 'fact' is sewn into a scientific paper, the characteristics of modern science culture render it behoven upon the scientist to cover their tracks in the sand; deny that the fact was ever tentative. The refereed scientific text is an excellent means to achieve the 'black-boxing' of facts.

The genre of scientific English embodies and manifests a Bretchian conception of the typical relationship between text and audience. A relationship in which they are deemed to be separate and estranged from each other; they are alienated. This is quite opposite to the notion of intertexuality, which, according to Barthes:

> "requires an attempt to abolish (or at least lessen) the distance between writing and reading, not by intensifying the reader's projection into the work, but by linking the two together in a signifying process."[333]

---

330 Fawley, William; 1987: 132.
331 Fawley, William; 1987.
332 Halliday, Michael A. K.; 1994: 137.
333 Barthes, Roland; 1979: 79.

At this point I wish to introduce another important role of the genre of scientific language found in published papers, that of being a 'certainty-producing technology'.[334] This idea can be useful as we analyse certain words or a concept as they are traced from private to public scientific documents. Thus the question asked is: 'How is certainty being linguistically or textually created as we move along the private-public continuum?'

2.3.10 Is Scientific Writing A Form of Literature?

I will now look at two different perspectives on science writing. One is concerned with science prose conforming to a template, and the other perspective views science writings as able to take a multitude of forms.

There is a received belief that scientific papers are written in a uniform style, according to a style-sheet or formula even, for:

> "science is a place where rules abound...there are the more or less explicit practical rules and guidelines for experimentation, analysis of data and writing up results [and an]...adherence to generalized norms [as]...scientists try to negotiate a more or less stable scientific reality".[335]

Thus the 'imprint' of the author(s) is not part of the text; no variance of authorship style is discernible. Bazerman claims that:

> "[T]o write science is commonly thought not to write at all, just simply to record the facts...the popular belief of this past century is that science is a transparent transmitter of natural facts."[336]

It may be the desire of (some) scientists that their papers and monographs are perceived as heterogeneous, semantically closed, and neutral entities, for this can, *inter alia*, perpetuate the mysteriousness and authority of scientific knowledge. This is an example of 'writing degree zero',[337] where the illusion of the semantic content as free of social and political import is sustained:

> "the degree zero testifies to the power held by any system of signs, of creating meaning 'out of nothing'."[338]

---

[334] Thompson, James D.; 1967.
[335] Potter, Jonathan and Margaret, Wetherall; 1987: 65.
[336] Bazerman, Charles; 1988: 14.
[337] Barthes, Roland; 1967.
[338] Barthes, Roland; 1967: 138.

Within this *Weltanschauung* literary writing serves to communicate emotions and the responses of individuals to the world, whereas scientific prose communicates facts and cause-effect relationships. To put this differently, literary texts are a 'process', whereas scientists are concerned with consensus and product. It could be claimed that scientific written discourse is not writing in the literary and rhetoric sense, but acts as a form of reporting.

A counter-argument to the received wisdom that science writing is all of a kind states that there exist many kinds of, and purposes for, writing. The instructions for the assembly of a child's toy, the score of Beethoven's Symphony Number 6, Newton's *Opticks*, the *Encyclopaedia Britannica*, the ingredients on food packages, and the listings in telephone directories, are all forms of writing: writing as the translation of thoughts into written symbols for dissemination and reconstruction in the minds and hearts of readers. Also just as there are many forms of writing, scientific prose appears in many guises. For instance, Stephen Jay Gould writes in one style in *It's a Wonderful Life* and another style in *Ontogeny and Phylogeny*. There are many scientific authors who have very individual and 'literary' writing styles, such as Charles Darwin, Freud, Douglas Hofstadter, Daniel Dennett, Oliver Sacks, and Isaac Asimov. In the journal *Science and Public Affairs* White proclaims: "Science writing is the new literary art form".[339] This point I would not necessary agree with. There is no such entity as 'science writing' *per se*. Instead there are many different types of 'science writing', which vary with the topic, the intended audience, and the language use.

2.3.11 Private and Public Prose: The Recovery and Variability of Experimental Practice

A principal argument of this research is that the published account of a scientist's experimental labours and the resultant discoveries, whether in the form of a lecture or as a refereed paper, do not constitute a fully transparent record, nor is it all that is required to understand the 'nature' of scientific work. Rather, scientific papers: "are generally viewed only as vehicles for conveying information."[340] Thus we attempt: "to expose the pathway by which the scientist

[339] White, Harry; Winter 1995: 7.
[340] Holmes, Frederick L.; 1987: 220.

reached the conclusions embodied in the published texts."[341] That is, to recover the contingencies and uncertainties consummate with innovation and creativity.[342] For this research: "texts are read, not as static arguments, but as part of the dynamic process of research activity."[343] Evidence will be provided by using the analysis of private and public documents to:

1) Follow an idea or theory through its various stages of development so as to try to cast:

> "light on the dynamic processes of cognitive growth on the individual level...and [the] group levels of scientific practice."[344]

2) Compare descriptions of practice with its final form in public accounts. This is important because public accounts of experimental practices reflect the reified, completed artefact, with the skills, craft, defeats and ingenuities of actual practice distilled out.[345] Eisenberg compares the process of scientific creativity to painting:

> "If a work of art could be completely translated into words so after the painting had been described in words, there was no unaccounted for residue, then one could write about painting and no longer have to paint. There is something in the painting which defies translation into words....The painting is a direct description of what it can never fully be."[346]

2.3.12 Conclusion

In conclusion, science prose is a form of writing: "the evolution of choices".[347] For it employs narrative conventions, rhetorical devices, and literary *tropes*, that are: "designed to inform, but also to enforce the authority and objectivity of its results and its conclusions".[348] This is not a contradistinction to the assertion that language and the 'real world' are separate entities, and are treated as such in published scientific writing. It is not the case that the science text constitutes an 'ideal' language, through which objective and unquestionable knowledge of the world is attainable. Language is being used, as with all other forms of writing, to not only to say certain things to a particular audience, but also to avoid saying or revealing other things. It is also used as a tool to manufacture

[341] Holmes, Frederick L.; 1981: 61-62.
[342] Gooding, David C.; 1990: 168-170.
[343] Bazerman, Charles; 1985: 3.
[344] Rudwick, Martin J .S.; 1982: 186.
[345] Gooding, David C. 1989 (a): 63-82.
[346] Krebs, Hans A. and Shelley, Julian A. (Eds.); 1975: 24.
[347] Bazerman, Charles; 1988: 13.
[348] Doorman, S. J.; 1980: 187.

consensus on what is 'truth', or the dominant discourse, at a distance from the laboratory.

2.4 Final Conclusions

This chapter has provided the broad historical and linguistic contexts for my research. I have tried to chart two main trends. First, that during this century the role of language has shifted away from being perceived as a 'mirror of nature' and a mechanical medium for the expression of pre-constructed ideas. Instead we see language conceived of as a 'shared narrative' or dialogue that preserves the complexity of human behaviour with its interplay of interpersonal and social contexts, thought processes, and contingent happenings.[349] Second, that text analysis in the second half of this century is increasing used as tool to study the uses of language and language meaning.

The implications of my research for science studies are as follows:
1) The quantitative analysis of scientific language has up to now not received a lot of attention.
2) The use of computational tools to study the linguistic development of ideas in science is a relatively untapped area of research.
3) Investigation of how language-use affects thinking and concept-construction in science is also hardly encroached on by science studies.
4) My research places emphasis on a holistic approach to scientific practice, that is, it is not actually the case that the peer-reviewed paper is the final outcome of a linear pathway of scientific discovery. Rather, writing and language-use are integral parts of experimental processes, for writing and communicating help shape further ideas and experiments, a process which may span the whole of a scientist's career.

Computer-supported text analysis and computational linguistics are well established disciplines as we have seen. What has this search got that is different?
1) Modern text analysis from the 1960's has been increasingly interested in meaning, though such research sits alongside work on authorship-attribution,

[349] Hatch, J. A. and Wisniewski, R.; 1995.

for example, in Shakespeare studies. Meaning, in terms of the interplay between language use and the construction of ideas, is a key issue addressed by this research.

2) At the time of writing, this research is unique in analyzing a large corpus of scientific writing, encompassing private and public documents, using a computer-assisted research tool.

3) Research – with quantitative data being interpreted within a qualitative environment - is undertaken on scientific and non-scientific prose that spans across an author's writings in terms of both private and public (published) writings.

3) The private and public dimensions are not taken as separate entities, with published accounts and facts appearing as 'final' products, denuded of any history or psychological and social preparations. Instead a continuum of writings is perceived. Thus the idea is of the unfolding of an idea or nomenclature, characterized by non-linearity and non-sequentiality.

4) Computer-based text analysis tends to start with the quantitative analysis. My research begins with qualitative preparatory analysis to give wider meaning to the quantitative results. This takes the shape of a biography of Faraday in chapter one and two studies of the use of discourse analysis in science studies and in the humanities, which is the concern of my next chapter.

3 Discourse Analysis as a Tool for Science Studies and for the Humanities: Two Case Studies

Abstract: This chapter completes the context for my research began in chapter two. Here I explore the use of discourse analysis through two case studies: i) discourse analysis as a research tool in science studies, and ii) quantitative analysis of text outside of science studies. The first study investigates the use of three categories of discourse analysis (quantitative, qualitative, and qualitative-quantitative) in the journal *Social Studies of Science* from its inception in 1971 (as *Science Studies*) up to May 1995. We found that the number of papers using discourse analysis, and the number of qualitative and qualitative-quantitative papers has risen quite steeply since 1986, with a concomitant dip in the number of quantitative methodologies used. Quantitative methodologies in science studies are very little in evidence at present, except in scientometrics. The prevalence of quantitative text analysis in fifty-eight humanities publications (1990 to 1995) is the concern of the second case study. The two case-studies were not designed for direct comparison.

3.1 Science Studies and Discourse Analysis

3.1.1 Introduction

The invention of the *Science Citation Index* in the late 1960's lead to references and footnotes quickly acquiring the status of valuable empirical data, with citation, co-citation, and other bibliometric analyses the inevitable outcome.[350] This case study investigates the relation between discourse analysis and science studies by searching the journal *Social Studies of Science* (formally *Science Studies* up to 1974) from 1971 to May 1995 for scholarly papers using discourse analysis. The analysis is categorised as either quantitative, qualitative, or quantitative-qualitative. It is believed that qualitative-interpretative methodology is in the ascendancy, and that the use of computer-supported discourse analysis in science studies is generally absent.

3.1.2 Methodology

The articles published in *Science Studies* and *Social Studies of Sciences* from the inaugural edition in January 1971 until the May 1995 volume were studied to uncover the number of articles based on analyse of discourse. The articles had to be based upon empirical analysis, where the sample consisted of some type of scientific or science-based written discourse, thus both biology monographs and science-centred patents were admissible.

The three types of analysis in the articles were delineated as follows:
1) 'quantitative-bibliometric' (statistical, bibliometric, or computer-supported),
2)'qualitative-interpretative' (non-numerical, interpretative) or,
3) 'quantitative-qualitative' (aspects of both).

The method of deciding which papers to be included was simply a matter of reading through all the numbers of *Science Studies* and *Social Studies of Science* from 1971 to May 1995.

3.1.3 Results

In the four volumes of *Science Studies* (1971-1974) there were sixty-seven scholarly articles, out of which five papers were 'quantitative-bibliometric' and five were 'qualitative-interpretative' (7.46%). From 1975 to 1990 there have been 410 papers in *Social Studies of Science* (hereafter SSS), and twenty-five (6.10%) of these have used quantitative-bibliometric methodologies. In the same period there have been twelve (2.93%) 'qualitative-interpretative' articles, and three (0.73%) that were 'quantitative-qualitative'.

Since 1980 there have been twenty papers published in SSS employing 'quantitative-bibliometric' methodologies, nineteen were 'qualitative-interpretative', and six were 'quantitative-qualitative'. This includes the only two examples of co-word analysis.[351] To complete this picture, from 1971 until

---

[350] MacRoberts and MacRoberts; 1986: 151.
[351] Law and Courtail; 1989, Whittaker; 1989.

February 1995 there has been a total of three uses of co-word analysis. The third being found in a paper by Leydesdorff in 1991.[352]

From 1990 until the present time there have been eighteen scholarly articles (out of 120; 15%) published in *Social Studies of Science* concerned with discourse/textual-based analysis using very different methodologies, including 'co-occurrence of words',[353] 'socio-technical analysis',[354] and 'genre analysis'.[355] It was also found that there has been seven (38.89%) papers either wholly or partly quantitative, and the other eleven (61.11%) papers qualitative, since 1990.

The first discourse-based qualitative study was published in May 1982. Out of the first twenty-two articles, all quantitative, nineteen (86.36%) are written up using either citation or co-citation analysis, and for two of the remaining three papers cluster and factor analysis is the methodology. From May 1982 until 1990 fifteen of the twenty-eight (53.57%) articles are qualitative, that is, employing a form of discourse or textual analysis. Citation analysis remains the predominant quantitative tool during this period, with data being analysed using this method for ten of the thirteen (76.92%) quantitative papers. Over the next five years (1991-1995) there are eighteen papers of which three (16.67%) are quantitative (co-word analysis, citation analysis, and relationship patterns). From 1982 until 1990 there were four quantitative-qualitative articles (14.29%), and five such (27.78%) articles during the last period.

Table 3.1 (and charts 3.1 to 3.5) details the number and percentage of papers in *Science Studies* and *Social Studies of Science* (SSS) from 1971 until May 1995 of those articles either quantitative, qualitative, or qualitative-quantitative papers. The ratios given are in relation to the total number of articles over five year periods (for rows three to five the upper percentage concerns the total number of papers and the lower percentage the number of 'discourse analysis' papers for each time period). The number of: i) total papers, ii) discourse analysis (D.A) papers, iii) quantitative papers, iv) qualitative papers, and v)

[352] Leydesdorff, Loet; August 1991: 75-110.
[353] Callon, Micheal; 1991.
[354] Latour, Bruno; 1992.
[355] Swales, John; 1994.

quantitative-qualitative papers for the five-years periods are displayed in graphs one to five.

Table 3.1. The Percentage of 'quantitative-bibliometric' (q-b), 'qualitative-interpretative' (q-i), and 'quantitative-qualitative' (q-q) articles in *Science Studies* and *Social Studies of Science* from 1971 until 1995.

	Science Studies: 1971-1974	SSS: 1975-1980	SSS: 1981-1985	SSS: 1986-1990	SSS: 1991-5/1995	Total
Number of papers	67	135	134	141	120	597
D.A papers	5 (7.46%)	13 (9.63%)	16 (11.94%)	13 (9.23%)	18 (15.00%)	65 (10.72%)
q-b papers	5 (100%) (7.46%)	13 (100%) (8.89%)	6 (37.50%) (4.48%)	7 (53.85%) (4.96%)	4 (28.57%) (3.33%)	35 (56.67%) (5.69%)
q-i papers	0	0	8 (50.00%) (5.97%)	5 (38.46%) (3.55%)	10 (55.55%) (8.33%)	23 (35.94%) (3.85%)
q-q papers	0	0	2 (12.50%) (1.49%)	1 (7.69%) (0.71%)	4 (22.22%) (3.33%)	7 (10.94%) (1.17%)

Chart 3.1 Total Number of Papers for Five Year Periods

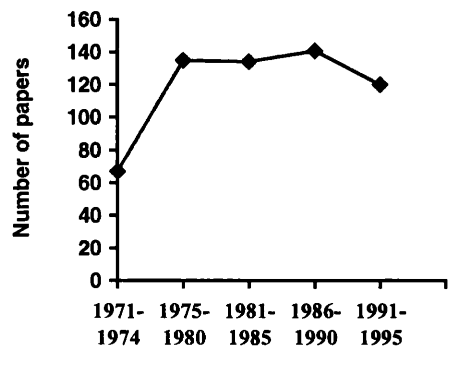


Chart 3.2 Number of Discourse Analysis Papers for Five Year Periods

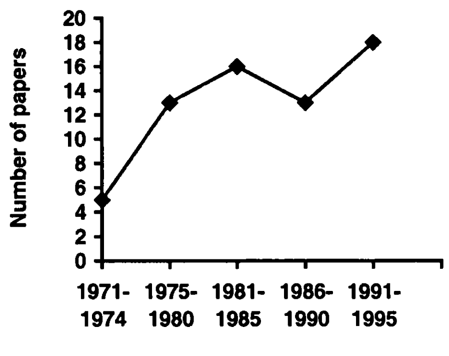


Chart 3.3 Number of Quantitiative Papers for Five Year Periods

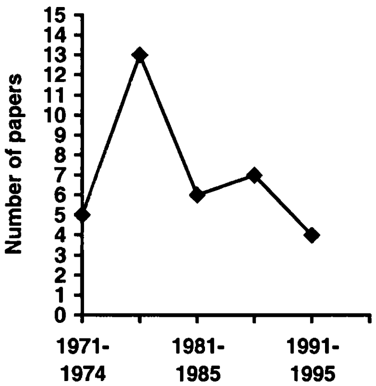


Chart 3.4 Number of Qualitiative Papers for Five Year Periods

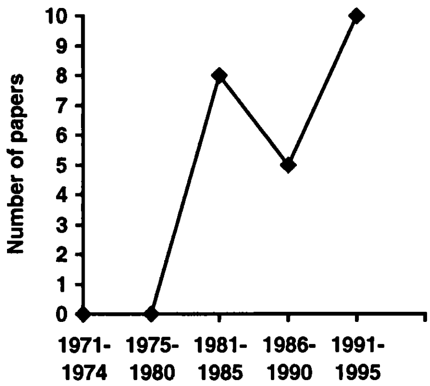
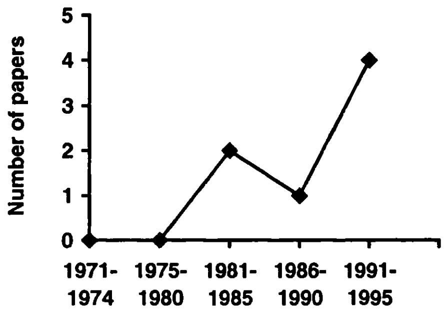


Chart 3.5 Number of Quantitative-Qualitiative Papers for Five Year Periods



What we can see is that from 1971 to 1980 all the discourse analysis articles were quantitative (five and thirteen respectively). Then over the next five years the percentage of quantitative papers fell from 100% to less than 40% (6 out of 16), with now half of the papers being qualitative. The percentage of quantitative articles increased in the next period (1986-1990), with seven (58.33%) quantitative and five (41.67%) qualitative, out of a total of thirteen. This fell to under 30% for the present period, during which time more than half of the papers used qualitative methods. Papers using a mixture of quantitative and qualitative methodologies were first published in 1984 and 1985. For the present period four out of eighteen articles (22.22%) are of this type.

The total number of papers created a plateau from 1975 to 1990 (chart 3.1), whilst the number of discourse analysis papers has shown a steady increase, except for a dip during the period 1986 to 1990. We can clearly see the decline in quantitative methods after a peak in the period 1975-1980 (chart 3.2). The ascendancy in the use of qualitative discourse analysis papers from 1981 is very evident in chart 3.3.

3.1.4 Conclusions

The waning in the use of quantitative-bibliometric analysis was accompanied by the use of qualitative discourse analysis in the early nineteen-eighties.[356] Traditional sociological analysis was perceived as inadequate by not fully appreciating that data can be interpreted in a myriad of ways, and scientists discourses do not just tell one story, but rather have *plurivocity*.[357] It is argued by Gilbert and Mulkay that:

> "Most sociological analyses are dominated by the authorial voice of the sociologist. Participants are allowed to speak through the author's text only when they appear to endorse his story. Most sociological reports are, in this case, univocal. We believe that this form of presentation grossly misrepresents the participants' discourse."[358]

Significantly, among all of these scholarly papers concerned with quantitative and qualitative aspects of bibliometric and discourse\textual analysis there is very little in-depth statistical work and certainly no computer-supported analysis.

Thus in conclusion, there does seem to exist a noticeable reliance on qualitative-interpretative analysis among science studies scholars. The primary exception to this observation is the quantitative analysis of science and technology in scientometrics. Computer-based and statistical-quantitative analysis is far more prevalent among humanities and literary-linguistic researchers, which will be shown in the second case-study.

---

[356] For example, Latour, Bruno and Woolgar, Steve; 1979; Mulkay, Michael; 1981: 163-172; Yearley, Steven; *Textual Persuasion*; Philosophy of the Social Sciences; VII; 1981: 409-435; Mulkay, M., Potter, J. and Yearley, S.; 1982: 171-204; Mulkay, Michael and Gilbert, G Nigel; 1982: 309-319; Gilbert, G Nigel and Mulkay, Michael; 1984; Mulkay, Michael; 1984: 531-549; Mulkay, Michael; 1984: 265-283; Shapin, Steve; 1984: 125-128; Shapin, Steven; November 1984: 481-520; Mulkay, Michael; The Word and the World; 1985; Tibbetts, Paul and Johnson, Patricia; 1985: 739-749; Fuhrman, Ellesworth R. and Oehler, Kay; 1986: 293-307; Woolgar, Steve; 1986: 309-317.

[357] Mulkay, Michael, Potter, Jonathan, and Yearley, Steve; 1982, G. Nigel Gilbert and Mulkay, Michael; 1984.

[358] G. Nigel Gilbert, and Mulkay, Michael; 1984: 15.

3.2 Quantitative Text Analysis Outside Science Studies

3.2.1 Introduction

Using the papers published in *Social Studies of Studies* as a representative study of science studies research it was found that quantitative methodologies are little used, and computer-based analysis not at all. Aspects of my research straddle science studies and the humanities, thus I will analyse papers from humanities journals to investigate the prevalence of quantitative text analysis methods.

3.2.2 Methodology

Fifty-eight journals, for the years 1990 to 1995, were searched for articles using different quantitative techniques for the analysis of text. The journals were chosen through key-word searches in electronic bibliographic databases, the Internet, and manual searching of the reference section in journal articles. Key-words used, singularly and in combination included: text(ual), computer(ized), computing, science, scientific, literature, literary, humanities, discourse, genre, linguistics, language, and communication. The criteria employed for the selection of articles was twofold:

1) the article had to be an empirical research paper,
2) the sample analyzed had to comprise (a) written text(s).

The aim was to identify a representative example of different types of quantitative analysis for written texts. The titles of the journals searched, the number of articles per journal, and the number and year of the selected articles per journal are located in appendix three. Thirty-five articles were selected from twenty-one of the fifty-eight searched journals. Six articles come from 1990, 1991, 1993, and 1994, seven articles from 1992, and four from 1995. The articles were chosen by 'manually' checking the method of analysis for each paper.

3.2.3 Findings

The number and percentage of selected articles for each of the twenty-one journals are displayed in table 3.2.

Table 3.2. The Number and Percentage of the Selected Papers per Journal.

Number and Percentage	Journal Title
7 (20.00%)	Scientometrics
2 (5.71%)	Journal of Pragmatics
2 (5.71%)	Applied Linguistics
2 (5.71%)	Discourse Processes
2 (5.71%)	Literary and Linguistic Computing
2 (5.71%)	Computers & the Humanities
2 (5.71%)	Language Sciences
2 (5.71%)	Information Processing and Management
2 (5.71%)	Written Communication
1 (2.86%)	Journal of Technical Writing and Communication
1 (2.86%)	The Technical Writing Teacher
1 (2.86%)	Science in Context
1 (2.86%)	English for Specific Purposes
1 (2.86%)	Language
1 (2.86%)	Knowledge Acquisition
1 (2.86%)	Journal of the Proceedings of the Annual Conference of Cognitive Science
1 (2.86%)	Behavioral Science
1 (2.86%)	Poetics
1 (2.86%)	Notes and Records of the Royal Society of London
1 (2.86%)	TEXT Technology
1 (2.86%)	Journal of the American Society For Information Science

It can be seen that one journal (*Scientometrics*) contained the greatest number of articles, and that just under 46.00% of the journals (12) contain two articles each. Twelve of the twenty-one journals (57.14%) contain just one article.

The material analyzed in the articles was then classified by eighteen genres. For thirteen of the articles scientific discourse was analyzed. This genre accounts for more than three times the number of articles in the next genre, literary novels, found in four articles. Fourteen of the genres contain two articles each: 40.00% of the thirty-five articles. Twelve of the eighteen genres (66.67%) represent one article each, accounting for 34.32% of the articles. The results are displayed in table 3.3 and charts 3.6 and 3.7.

Table 3.3. The Number and Percentage of Genres for the Thirty-Five Selected Articles.

Number and Percentage	Sample Genre
13 (37.14%)	Scientific
4 (11.43%)	Literature (novels)
2 (5.71%)	Literature (poetry)
2 (5.71%)	English composition
1 (2.86%)	Literature (notebooks)
1 (2.86%)	Literature (various)
1 (2.86%)	Literature (personal correspondence)
1 (2.86%)	Medicine
1 (2.86%)	Student technical writing course material
1 (2.86%)	Science and engineering
1 (2.86%)	Scientometrics
1 (2.86%)	written and spoken registers
1 (2.86%)	University science textbooks
1 (2.86%)	Written advertisements
1 (2.86%)	Social science articles
1 (2.86%)	Family history literature
1 (2.86%)	'Science, medicine, economics, psychology, philosophy, and law'
1 (2.86%)	Computer manuals (UNIX)

Chart 3.6 The Percentage of Articles for Each Genre.

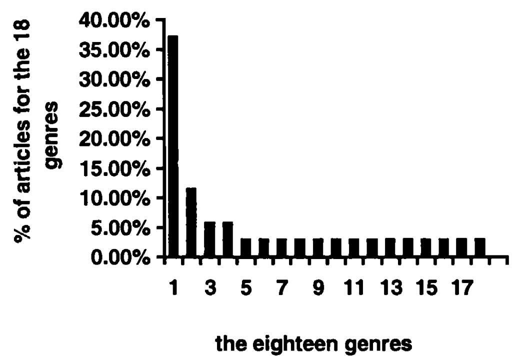
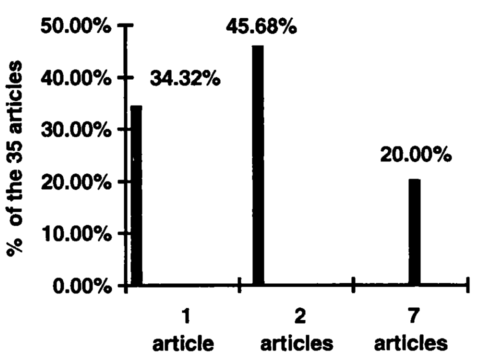


Chart 3.7 The Percentage of the Selected Articles in the Twenty-One Journals.



Genres: 1) Scientific 2) Literature (novels) 3) Literature (poetry) 4) English composition 5) Literature (notebooks) 6) Literature (various) 7) Literature (personal correspondence), 8) Medicine 9) Student technical writing course material 10) Science and engineering 11) Scientometrics 12) 17 written and spoken registers 13) University science textbooks 14) Written advertisements 15) Social Science articles 16) Family history literature 17) 'Science, medicine, economics, psychology, philosophy, and law' 18) Computer manuals (Unix).

Table 3.4 displays the types of quantitative methods used in the thirty-five articles. We do see a significant use of computer-supported analysis (mostly computerized text analysis software).

Table 3.4 Number and Percentage of Articles Using Different Text Analysis Methods.

Methodology	Number and Percentage of the 35 Articles
statistical analysis	11
computer-supported	8
co-word analysis	2
content analysis (stylistics)	7
content analysis (genre analysis)	1
content analysis (contrastive discourse analysis)	1
content analysis (semantic analysis)	1
content analysis (map analysis)	1
content analysis (main clause analysis)	1
word-frequency analysis	1

3.2.4 Conclusions

The first study investigated the use of qualitative and quantitative analysis methods in one of the leading academic journals in science studies, from 1971 to 1995, and found that qualitative methods are currently most prevalent. Then I investigated thirty-five articles using different methods of quantitative discourse analysis selected from twenty-one humanities journals, from an original trawl of fifty-eight publications. A plethora of quantitative discourse analysis research tools were seen to used, something not seen in the issues of *Science Studies*. From a knowledge of other journals for the social, historical, and philosophical studies of science the dearth of papers using quantitative/computer-supported methodologies exists across the science studies community. The obvious exception to this is found in the field of scientometrics and its eponymous journal. This is not too startling when you consider that the discipline of scientometrics is concerned with the quantitative analysis of science and technology communication, developments, and performance.

3.3 Review of Eighteen of the Thirty-Five Humanities Articles

3.3.1 Introduction

Part three presents a survey of different methods used in eighteen of the thirty-five papers selected in the second case-study in part two. I shall discuss any relevance for or bearing on my own research, or pertinent issues raised.

The first method is that of quantitative analysis of tagged texts using software called STRAP (STRuctural Analysis Program) used by Snelgrove to study the structure of narrative texts and chart reader response.[359] This tool is used to:

> "objectively analysis those elements in the text that are intended to make us think and feel."[360]

Specific reactions to a text are identified by computer-generated patterns. The issue of the relationship between the reader (the audience) and the text is something that is explored in my research. One point to be made here is that the author of a text is simultaneously a writer, reader, and interpreter of the text. The text is not a neutral space for the non-mediated description of 'facts'.

Callon *et al* use co-word analysis to study interactions between academic and technological (applied) research.[361] That is, they measure the relative intensity of the occurrence of keywords which describes the contents of documents. Co-occurrence of words is a method I may employ in my research. It may be revealing, for example, to know which words Faraday is combining with keywords such as 'electromagnetism', and to what degree.

A computerized-statistical technique called 'Vocabulary-Management Profiles' was used by Youmans to investigate the distribution of new and repeated

---

[359] Snelgrove, Teresa; 1990: 221-225.
[360] Snelgrove, Teresa; 1990: 221.
[361] Callon, Micheal *et al*; 1991: 155-205.

vocabulary.[362] The total vocabulary (the number of types) is plotted against the total number of words (tokens):

> "For the first few words of a normal discourse, every new token is also a new vocabulary word; initially, then, the number of types equals the number tokens. After the first repeated word, however, the number of tokens exceeds the number of types, and this difference increases with each repetition…it seems plausible to predict that new topics in essays, new episodes in stories, and the like should coincide with bursts of new vocabulary...an upturn in the curve signals an increase in new vocabulary at the end of the interval, whereas a downturn signals an increase in repetitions. The peaks and valleys on these curves prove to be surprisingly successful in signalling the ebb and flow of information in texts".[363]

Bursts of new vocabulary was said to reflect innovation, and repeated vocabulary to reflect discourse coherence. The distribution of vocabulary is of interest to my research. I would certainly agree that where there is a run of new vocabulary then innovation, a new idea or concept, may well be developing or being stated.

An investigation of the dimensions of discourse complexity using confirmatory factor analysis by Biber found that written registers (press reportage and science fiction texts, for example) exhibit greater differences in terms of the types and extent of discourse complexity than spoken registers (such as telephone conversations and planned speeches).[364] In terms of my research the equivalent of the 'written registers' would be public/published scientific documents, and the private scientific texts the 'spoken registers'. My belief is that the private scientific texts would exhibit variable sentence length, greater compression, and shorter waiting time thus greater complexity than published material.

Miall researched identification and analysis of collocates of emotion words in Coleridge's notebooks.[365] The mapping of changes in collocation were grounded in a premise, namely that:

> "A collocate that occurs more frequently than expected within a set span of the target word is likely to be playing an important role in the usage of that word."[366]

---

[362] Youmans, Gilbert; 1991: 763-789.
[363] Youmans, Gilbert; 1991: 765.
[364] Biber, Douglas; 1992: 133-163.
[365] Miall, David S.; 1992: 1-12.
[366] Miall, David S.; 1992: 11.

This study comprised two phases. In phase one of Miall's research the occurrence of all words collocating with emotion words - *feel*, *feels*, *feeling*, *feelings*, *felt*, *emotion*, *emotions*, *passion*, *passions*, and *passionate* - in the notebooks were extracted. The collocation span was five words either side of an emotion word. Those words occurring but once were eliminated, and a *z*-score computed from the remaining list of words. The aim of phase two was to demonstrate the extent to which Coleridge's vocabulary altered over the period covered by the notebooks. The collocation span was lengthened to fifteen words (the mean span counted was nine words) either side of a target word. This was to avoid missing links within Coleridge's: "distinctive, interconnected manner of thinking."[367] Biber's method involved the use of change-point analysis and two collocation programmes, COLLOC and COLLREAD. One particularly interesting aspect of Biber's research is the attempt, at least partially, to understand Coleridge's manner of thinking. How Faraday used language as a medium for thinking and communication is a central issue of my research.

Content analysis, with manual scanning of related articles in molecular genetics and linguistics, was used by Myers in order to, first, compare the introductory sentences, in which fact assertions are found, of science discourses with other contexts. Secondly, to investigate how the textual features account for such claims.[368] It would be interesting to study the introductory sentences (and paragraphs) of both private and public scientific texts. I suspect fact assertiveness, modality, short sentences, and non-human actants would be more prevalent in published documents.

A study of the grammatical subjects in main clauses from thirty-six research articles was done by Godsen through the identification of four domains: the participant domain, the discourse domain, the hypothesized and objectivized domain, and the real world domain.[369] Examples of discourse for these domains are: 'our viewpoint', 'it is concluded', 'there was evidence', and 'it was found' respectively. It was concluded that these domains form a continuum, with the greatest degree of writer visibility at the 'participant domain' end and the least writer visibility at the real world domain, at which: "there is a greater focus on

[367] Miall, David S.; 1992: 5.
[368] Myers, Greg; 1992: 295-313.

research-based, i.e. real-world physical and mental entities and activities."[370] In terms of my research the 'participant domain' end, where writer visibility is greatest, would be in letters and laboratory notebooks. This would be evidenced by use of the first-person pronoun, relative little use of modality, reference to other people and to emotions, and errors, for example. What Godsen calls the 'real world' domain would manifest in papers, treatises, and textbooks. This is where the experimentalist and author is virtually filtered out, replaced by abstract prose, modality terms, non-human agents, and jargon.

Using computer-assisted textual analysis, incorporating Michael Gregory's Communication Linguistic model,[371] Matsuba analyzes Shakespeare's sonnet number CXXX.[372] This model incorporates the thesis that, in resonance with post-structuralism, language is effected by a variety of non-linguistic factors, such as social class or education, which influence the reading and interpretation of a text.[373] The aim is to try, using computer tools, to get beyond the following situation:

> "Most computer scientists [in AI] approached the relationship between language and meaning with the view that there is a single "correct" meaning to any language act. They assume for the most part that all users of a language share the same semiologies and knowledge."[374]

There is never just one voice or meaning in a discourse or interpretation of an idea. One dominant meaning or interpretation may be attempted to be manufactured into a text, but with a little analysis other voices or meanings can be unearthed. This research will try to demonstrate that by looking at how an

---

[369] Godsen, Hugh; 1993: 56-75.
[370] Godsen, Hugh; 1993: 62.
[371] Gregory, Michael; 1982.
[372] My mistress' eyes are nothing like the sun
Coral is far more red than a her lips' red
If snow be white, why then her breasts are dun
If hairs be wires, black wires grow on her head:
I have seen roses damasked, red and white,
But no such Roses see I in her cheeks,
And in some perfumes is there more delight
Than in the breath that from my mistress reeks.
I love to hear her speak, yet well I know
That music hath a far more pleasing sound,
I grant I never saw a goddess go;
My mistress when she walks treads on the ground
And yet by heaven I think my love as rare
As any she belied with false compare.
[373] Matsuba, Naoyuki Stephen; 1993; 331-340.

idea or concept is presented in language in different genres, at different stages of accounting and reconstruction, there is no one voice of 'scientific authority' existing throughout.

Schils and de Haan studied alteration in sentence length in a large corpus.[375] They found that literary texts had a higher degree of alteration in sentence length than the non-literary texts. I think that sentence length varies across more situations than alluded to by Schils and de Haan. That is to say, sentence length may vary with the genre of writing, whether in a letter or a peer-reviewed paper. Also if the author is using language, for example, to argue, persuade, or for description.

Carley investigated and compared the relative benefits for using content analysis and map analysis (which has its theoretical basis in an understanding of human cognition) for extracting and analyzing culture from texts.[376] Carely concludes that map analysis not only:

> "enables the researcher to locate the rhetoric of change and the extent to which different concepts are used [but] takes the researcher a step further and enables the analysis of meaning."[377]

The analysis of meaning is an issue central to this research. What has to be borne in mind is that meaning is a very slippery concept. A great number of personal, cultural, and historical factors shape meanings, such as the intended audience and the social and professional standing of the author. It cannot just be plucked out from amongst the words on a page.

An investigation of lexical density was a key aspect of Kopple's research on scientific discourse.[378] Lexical words are the nouns, adjectives, main verbs, and adverbs *vis-à-vis* grammatical words - the pronouns, determiners, auxiliary verbs, conjunctions, and prepositions. Lexical density is calculated in terms of the number of lexical words per unembedded clauses. Scientific text is a representative of the attic or synoptic style (*vis-à-vis* the doric or dynamic style), where:

---

[374] Matsuba, Naoyuki Stephen; 1993: 336.
[375] Schils, Erik and de Haan, Pieter; 1993: 20-26.
[376] Carley, Kathleen; 1994: 291-312.
[377] Carely, Kathleen; 1994: 309.
[378] Kopple, William J. Vande; 1994: 534-564.

> "the world is a world of things, rather than one of happening; of product, rather than of process; of being rather than becoming"[379]

Thus "people can freeze what they write about" and "take it in as a whole."[380] The synoptic style of writing is characterised by 'a high degree of lexical density and relatively simple sentence structure'. The scientific texts analyzed in my research include personal writings and peer-reviewed papers. Thus when we follow the development and articulation of an idea we see the move from the attic or synoptic style to the doric or dynamic style, with lexical density increasing.

In 1995 Johnson wrote a computer program called WORDS to perform three tasks:

1) "to count the number of running words in a text",
2) "to count the number of unique word forms",
3) "to list the number of occurrences of each unique form."[381]

My research is concerned with meaning and not with stylometrics, and meaning cannot not be lifted from a text, nor does it reside in statistics. Thus the measures advocated by Johnson would need to used to supplement qualitative analysis and the interpersonal, social, and historical contexts the author is writing within.

3.4 Conclusions

We have seen how quantitative methods for analysis of discourse are relatively more prevalent in disciplines outside of science studies. My research thus goes against the grain of one current trend in science studies: the prevalent use of qualitative methods. I use a computerized tool to scan my texts and the quantitative results are then interpreted within a wider qualitative framework. One advantage of a computerized tool is that far larger amounts of texts can be analyzed. An idea or concept will not appear at any one moment but be infused, growing and developing, over time and throughout a scientist's work.[382] Thus the larger the corpus of material investigated the more possible it should be to

---

[379] Halliday, M. A. K; 1987: 146-147.
[380] Halliday, M. A. K.; 1989: 97.
[381] Johnson, Eric; Spring 1995: 8-17 (8).
[382] Holmes, Frederick Lawrence; 1985.

understand the process of scientific discovery. This is an issue I shall return to in later chapters.

Now that I have provided a biographical, historical, and theoretical context for my research over chapters one to three the next chapter will describe the characteristics of my empirical work.

4 The Nature of the Research Project

The world little knows how many of the thoughts and theories which have passed through the mind of the scientific investigator have been crushed in silence and secrecy by his own severe criticism and adverse examination; that in the most successful instances not a tenth of the suggestions, the hopes, the wishes, the preliminary conclusions have been realized.
--Michael Faraday

Abstract: The aim of this chapter is to present a clear and comprehensive account of the aims, methods, empirical work, and resources involved in the research project. The overall aim of the project is to use computational methods to study the relationship between language and the construction of ideas in science. This will involve analyzing and comparing private and public (published) writings of, primarily, Michael Faraday over a period of about thirty years. Some analysis of writings of Charles Darwin will also be undertaken.

4.1 Introduction

This chapter is a detailed discussion of my methods and resources for undertaking my case studies presented in chapters five through eight. Each of my main studies have been written up to include a specific methods section, thus this chapter provides a detailed overview of how the case studies will be carried out in three sections. The first of which discuss the aims, conjectures, and questions addressed by my research. This is followed by a description of my main studies, and then the final section describes the methods, textual features, and computational tool used in the case-studies. Additionally this section discusses revisions to be made and expected results in light of my pilot study, and problems encountered in the pilot study and how they may be remedied.

4.2 Aims and Conjectures to be Tested, and Questions Addressed

This text-analysis research is part of a larger project concerned with source constrained modelling of the cognitive and social processes of science.[383] This approach to modelling utilizes: "a variety of constraints on the behaviour of a model."[384] One such constraint is text-based, that is, it depends on the contents of the source files. The long-term aim of the project is to:

> "develop a more context-sensitive approach to modeling cognitive and social aspects of complex processes, as these are described by historians and sociologists on the basis of their studies of sources such as correspondence, publications and other documentary records."[385]

This project is part of a larger research program which is:

> "aimed at establishing the usefulness of functional modeling, conducted via graphical methods, to the analysis, and understanding of complex historical and social processes."[386]

A primary aim of my research is to identify the best features of texts to use to 'constrain researches initiated by running models'.[387] This is to be achieved by making a different kind of detailed analysis of written/published records of the development of certain scientific discoveries. A related aim is to develop and evaluate computational methods of text analysis as a means of investigating the social and cognitive processes at work in the construction of scientific knowledge in a variety of documents, including diaries, correspondence, and published sources.

Concomitant to this, the research will also test the conjecture that an author's writing style will ossify as she/he progresses through their career.[388] Put another way, an author's 'stylistic diversity' will diminish:[389] "after an early peak of diversity since his [her] use of language should crystallize as certain words and patterns become increasingly preferred."[390]

---

[383] Undertaken by David C. Gooding and Tom R. Addis funded by the JCI and ESRC.
[384] Gooding, David C and Addis, Tom R.; 1995: 1.
[385] Gooding, David C and Addis, Tom R.; April 1996 (a): 2.
[386] Gooding, David C and Addis, Tom R.; April 1996 (a): 11.
[387] Gooding, David C and Addis, Tom R.; April 1996 (a): 9.
[388] Brainerd, Barron; 1979: 6.
[389] Joose, Martin; 1962.
[390] Tallentire, D.R.; 1976: 315.

Two principal questions are attended to by this research. The first is whether it is possible to investigate the question: 'how do new expressions and ideas arise in language?', using quantitative methods of text analysis. That is, the pathway from private writings to scientific theories or concepts reified in public/published accounts can be revealed. This is not at all straight forward because it is postulated that all language is second hand and that every time we form a sentence, we are reassembling used words, cutting and pasting.[391] Similarly it was Roland Barthes who:

> "[A]rgued that writers only have the power to mix already existing writings, to reassemble or redeploy them; writers cannot use writing to 'express' themselves, but only to draw upon that immense dictionary of language and culture which is 'always already written'".[392]

Furthermore that:

> "[A]ny text is an intertext; other texts are present in it, at varying levels, in more or less recognisable forms: the texts of the previous and surrounding culture. Any text is a new tissue of past citations. Bits of code, formulae, rhythmic models, fragments of social languages, etc. pass into the text and are redistributed within it, for there is always language before and around the text."[393]

The second question is whether it is possible to: i) expose the differences in laboratory activity and the experimental recounting in published papers,[394] and ii) describe and reconstruct the human activity of scientific work through the study of the micro-history contained in private writings. These documents:

> "ordinarily follow along close enough behind the activity itself to provide a track still 'fresh' enough to trace the daily actions of those who have left their imprint there."[395]

To address these two questions we need to answer two further questions:

i) How is it possible to use a set of stylistic or textual features to recover a scientist's thinking in the development of a new idea, theory, or nomenclature?

ii) How far is it possible to recover experimental practice or discovery narratives using the analysis of private scientific documents to investigate language use and change?

---

[391] Byatt, A. S; 1996.
[392] Selden, Raman and Widdowson, Peter; 1993: 104.
[393] Barthes, Roland; 1981: 39.
[394] Gooding, David C; 1989 (a): 63-82.
[395] Holmes, Frederick L.; 1981: 65.

4.3 Case Studies

The initial exploratory working testing the computational tool and testing textual features was done in the pilot-study. This study, found in chapter five, investigated the use of compression and modality in the first twenty paragraphs of a variety of private and published texts of Faraday and Charles Darwin. The main empirical work is made up of five studies over three chapters. Chapter six comprises one case-study, the main findings of which are then tested by the three smaller studies written up in chapter seven. Lastly, chapter eight contains the fifth case-study. I now wish to describe my three main case-studies, from which I can decide which features and indicators are most and least useful, and construct normal ranges, baselines, and typical values for different writings of Faraday, which would be of use to other researchers. All of the wordlists used in the studies are found in appendix four.

i) A Study of Modality and Compression in Faraday's Letters to Scientists and Non-Scientists

One of the main findings of my pilot study was that discrimination between texts was enabled most by successfully scanning for compression and 'empirical positive' modality. This case-study attempts to validate these findings through a large-scale investigation of the language used in Faraday's correspondence with two scientists and five non-scientists, over a twenty-five year period. The discriminations I want to make are between early and late Faraday letters, 'experimental' and 'non-experimental' topics in letters, and letters to scientists and non-scientists.

ii) Testing of the Main Findings of the First Case-Study

In this case-study three smaller studies are presented which test the primary findings of the first main study. More specifically, I want to explore how the language use in Faraday's letters to scientists relates to his early letters to friends: does Faraday's language use change from his early letters to friends to his more 'formal' letters to scientists? Also am I able to discriminate between these two sets of correspondence? Secondly, I investigate whether the measures of compression and modality can discriminate between notebook and published writings of Faraday and Charles Darwin. Lastly, I test to what extent Faraday's writing style can be attributed to him being a Sandemanian by analyzing letters to Faraday from three non-Sandemanian scientists.

iii) The Role of Language in the Construction and Reconstruction of Faraday's 'Discovery' of Electromagnetism

Third case-study will involve comparative work analyzing the articulation and linguistic construction of the 'discovery' of electromagnetic rotation in 1821 and that of electromagnetic induction in 1831 in Faraday's private and public writings. This investigation will enable the comparison of successive stages in the process of writing about and communicating new ideas and knowledge. The analysis will be supplemented by the inclusion of the letters Faraday wrote dealing with electromagnetism between 1821 and 1831.

4.4 Methods

My case studies measure a number of textual features to try and meet their aims and objectives. In this section I first discuss the textual features and then describe my computational tool, methods, and resources. This is followed by discussion of revisions to be made in light of my pilot study, what will count as interesting results in my main studies, and problems encountered in my pilot study and the remedies.

4.4.1 The Features of Texts Best Suited to Achieving the Objectives

The textual features I will use are defined as indicators of discrimination between different types of texts, and are of two different kinds. Lists of keywords are indicators of the content of a text, these include modal terms and conjunctives. The other type of textual feature used include the degree of ellipsis (compression) and paragraph length.

More specifically, the features of texts I am considering measuring are: 1) the degree of ellipsis and 2) paragraph length. As well as indicators such as: 1) structured (nested) keyword list searches (for modal words, 'technical',[396] and 'grammar' lists, such as conjunctives and adverbs), 2) the construction of lexicons from words which differentiate texts to be used to generate a frequency distribution of the introduction of new words in any text, and 3) the waiting time (the distance (number of words) between new words or key-words).

Ellipsis

The notion of ellipsis and compression scores that I am using derives from the work of David Gooding.

Grammatical function words (for example: a, the , an , and , not , but , is , are) occur frequently in almost all English sentences. The most commonly used, thus repeated, 'nuts and bolts' words are articles and connectives, such as prepositions and conjunctives, and they are most likely to be omitted from private writing, for example, a diary entry or a non-published account. The number of words can be counted, from which is subtracted the number of repetitions, this latter count is then subtracted from the first, and the difference is returned as a percentage of the text. This can be used as an indicator of how compressed a text is. The smaller the difference between the two counts - the fewer the repeated expressions - the more compressed the discourse, which is more likely to be private writing, than a public narrative written with all grammatical complexity, and thus demonstrates a lower compression score.[397]

---

[396] Two examples of such lists are:
1) chem_analysis
"very" "care" "can" "find" "contain" "adulteration" "impur" "pure" "accura" "experiment" "trial" "examin" "sample" "opinion" "view" "substance" "compos" "analys" "weigh" "contain" "grain" "ounce" "proportion" "minute" "estimate"
2) new_discovery
"experiment" "examin" "expts" "supported" "adduced" "demonstr" "beautiful" "lovely" "show" "apparatus" "arrang" "instrument" "new" "ascertain" "verify" "establish" "result" "shew" "law" "describe" "repeat" "confirm" "proof" "expect" "anticipate" "made" "make" "sensible" "notice" "obtain" "illustrate" "act" "curious" "attention" "assumption" "suppos" "convinc" "phenomen"
[397] Gooding, David C ; April 1996: 2.

Thus to measure compression in a text the function scan_%diff_wrd converts a paragraph to words, counts them, removes all repetitions, counts the words again, and subtracts this second total from the first, to produce a difference. This difference is returned as a percentage of paragraph size. In the more compressed (private) writing, such as notes or rough drafts, we could expect some articles, conjunctives, and connectives to be omitted.[398] As an example, we have a personal note containing 120 words, of which 6 are repeated words, thus the number of expressions fall to 114, that is, 5 percent of the expressions are repeated; a compression score of 5 percent. On the other hand a paragraph of a published paper may contain 80 words, of which 20 are repeated, giving a compression score of 25 percent. Thus the lower the compression score, the smaller the number of repeated words, and the more the text is said to be compressed.

A high 'compression score' (expressed as a percentage) equals use of a lot of repeated words, and therefore low compression. A lot of repeated words actually means the text is enlarged or extended, and this is what the percentage indicates. Therefore referring to this score as a 'compression score' is misleading for it is, in reality, a 'repetition score'. A high 'repetition score' means the text has a high percentage of repeated words. For consistency I shall use terminology that is used within the larger project, and therefore the expression 'compression score' will be used throughout.

Modality

The structured wordlists include modal terms, parts-of-speech, and terms relating to experimental work. When a text is scanned using a wordlist the terms are expressed as a percentage of paragraph size. I have constructed seven sets of modal terms, which range from those terms expressing the most robust truth-asserting modality ('empirical positive') and to those which signify the most uncertainty ('contingent negative). The modal word-lists are found in appendix four.

[398] Gooding, David C ; April 1996: 5-6.

Finally in this section matrix 4.1 below maps out the relationship between distinctions, discriminations, and wordlists for my main case studies.

Matrix 4.1 Relationship Between Distinctions, Discriminations, and Wordlists for the Main Studies.

Case-study	Distinctions	Discrimination	Lists for each discrimination
1	degree of (truth) modality	between: i) letters dealing with experimental and non-experimental topics, ii) letters from scientists and non-scientists	all seven modal wordlists
2	degree of (truth) modality	between: i) early letters to friends and later letters to scientists, ii) private and public material of Faraday and Darwin, iii) letters written by Faraday and non-Sandemanians	'empirical positive' and 'contingent negative' wordlists
3	degree of cohersion; reliance of text on results and 'discovery'	between: i) private and published texts, ii) early (1821) and later (1831) writings, on electromagnetism	conjunctives; 'any_experiment', 'bench_words', and 'new_discovery' wordlists

4.4.2 Computational Tool and Methods Used To Measure The Textual Features

The computational methods consist of numerous text search and analysis programs executed as functions in a computer language called FAITH.[399] I had not previously worked with text analysis software before using the FAITH text search and analysis programs in CLARITY. The search and analysis functions belong to a suite of programs written in the functional language, FAITH. They are found in the FAITH code file TEX_AN.ddb, which is loaded by dragging it onto the Clarity icon. The programs are designed for use on flat/ASCII format text files. The source files have to be in the same directory as CLARITY. These

[399] Addis, T. R. and Addis, J. J. T.; Clarity; http://www.sis.port.ac.uk/research/clarity/index.html; February 26, 1997; accessed March 17, 1998.

programs have a graphical user interface, CLARITY, and were developed as part of a larger research project on modelling scientific discovery, principally to:

- "enable discovery programs to access text resources such as scientific writings of experimental (especially letters and laboratory manuscripts), so as to provide guidance for searches and support searches through large corpora by scholars designing the models and
- to support editorial work on a large-scale study of the collected works of scientists."[400]

The current version of CLARITY for Macintosh is 3.6.8, and version 4.7.3 for PC (Windows '95 and NT) is available.[401]

There are significant differences between the approach embodied in the programs and that of computational linguistics and humanities computing. These include the functions being designed to allow investigation of texts that have been left relatively untreated, and enabling:

> "the use of numerical values which can be assigned to any number of fields, either as weights or as thresholds."[402]

This approach has three major advantages:

> 1) "it supports the analysis of text that has not been indexed or structured according to a data model",
> 2) "it provides far greater flexibility than is possible with Boolean searches", and
> 3) "it supports the training of software to recognize types of text according to classifications designed by user."[403]

Texts are prepared in digital format from different source materials. The first step in the preparation of the Faraday texts involved them being transposed into digital form by OCR scanning during 1995 and early 1996. My job then was to proof-read the electronic texts against photocopies of the originals. Faraday material digitized include his *Experimental Researches in Electricity Volume 2* (pages 127-158), and *Diary Volume 1* (1821) (pages 45-117). Three of Charles Darwin's notebooks - 'M', 'N', and 'Old & Useless' - were also scanned in and proof-read. Then we apply and modify existing search and analysis functions in

[400] Gooding, David C. and Addis, Tom R.; April 1996 (a): 3.
[401] http://www.sis.port.ac.uk/research/clarity/index.html; accessed March 17, 1998.
[402] Gooding, David C. and Addis, Tom R.; April 1996 (b): 6.
[403] Gooding, David C.; 9 February 1996: 4.

the FAITH/CLARITY package. The numerical data acquired can be inserted in graphical, statistical, and word-processing software for further analysis.
The basic method is to identify features whose presence or whose change in numerical value may indicate a quality or process I am interested in. Although word-frequency counts for example, are the raw material for analysis, we are looking for well-defined sets of features and for well-defined changes in features.

The proposed method of data analysis includes:
i) Data visualization, using MS Graph (or MS Excel) for the pilot study, and Minitab 10.5 *Xtra* for large-scale analyses.
ii) Statistical analysis is done also using Mintitab 10.5 *Xtra*.

My data in the main case studies will be presented in two main forms: i) as numerical results, that is, the (truncated) mean[404] and the minimum (smallest non-zero) and maximum values, and ii) as (connected) histograms. My research is not attempting an overtly statistical analysis of Faraday's writings, rather I am more interested in being able to discriminate between different texts - private and published, as well as over time. Thus I chose to use the mean so that I had a relatively straightforward value with which to compare texts and different uses of textual features. The minimum and maximum values will allow me to build normal ranges for Faraday's writings. One refinement to the methods used in main studies will be to generate connected histograms rather than the scatter plots used in my pilot study. This is so that I can more easily construct shapes for Faraday's use of language, in tandem with my numerical results. The shapes can then be readily compared and distinct uses of language easily visualized.

4.4.3 Resources

This work draws on three main resources:

[404] The truncated mean is where the smallest 5% and largest 5% of the values are removed, and the remaining values are averaged. A 'neater' set of raw scores is acquired; where anomalous large scores are removed.

1) Powerful and versatile text search and analysis programmes. FAITH is the functional programming language, and CLARITY is the graphical interface to FAITH.[405]

2) Text stored as documents in electronic (machine-readable) form.

3) A primary research database of texts of Faraday extending over approximately thirty years. This database also contains several writings of Charles Darwin: the first two chapters of the *On The Origin of the Species* and *The Voyage of the Beagle*, and his 'M' and 'N' notebooks.

4.4.4 Revisions Made in Response to the Results of the Pilot Study

It was found in the pilot study that Darwin's 'N' notebook contains the greatest percentage of both empirical and contingent modal words. My expectation had been that the public writings would demonstrate the most authoritative and truth-affirming modality ('empirical positive' modality). Compression was greatest in Faraday's *Diary* and in Darwin's *Beagle* and 'N' notebook. These compression results are as expected, except for Darwin's *Beagle*, because the most commonly used, thus repeated, 'nuts and bolts' words, such as connectives, are most likely to be omitted from private narrative. The pilot study was also designed to find which textual features were most and least able to allow discrimination between 'private' and 'public' texts. I found that compression and 'empirical positive' modality were most able to allow discrimination between private and public texts. Whereas the 'Gilbert and Mulkay' wordlists, as well as the list of contractions and 'contingent negative' modality proved the least successful. In light of these results only compression and the 'empirical positive' and 'contingent negative' modal wordlists will be used for the main case-studies.

[405]The development of computer modelling software was initiated by Dr. David C. Gooding and Professor Tom R. Addis in 1991, using the researches of Michael Faraday as an application. This work was funded from 1991 to 1993 by the Joint Council Initiative on Cognitive Science and Human-Computer Interaction, and continues with ESRC funding (1994 to 1996; Gooding, David C. and Addis, Tom R.; **MRC**, SPG # 9107137 and Gooding, David C. and Addis, Tom R.; **ESRC**, R000 23 52 86). The graphical interface, CLARITY, is written in C for MPW, and will run on any Macintosh with a minimum of 8 RAM. A PC version is also being developed. The interface enables computational methods to be continually developed in response to new research questions.

4.4.5 What Will Count As Interesting or Significant Results?

As my research will study Faraday's writings over time, and given Faraday's robust use of 'empirical positive' modality in my pilot study, a significant result would be if a strong use of this modality was found across Faraday's private and published writings, and over time. Though it would also be noteworthy if this modality allowed us to consistently, over time, discriminate Faraday's published texts from his private ones. Similarly with compression, a textual feature more associated with the narrative of private discourse, it would be interesting if it was found that compression was consistently well represented in Faraday's texts. This would mean that compression is one of the best features of texts to use to 'constrain researches initiated by running models',[406] and that the software was sensitive enough to consistently capture this feature in Faraday's private writings over time. Furthermore, for the study of electromagnetism it would be significant if we captured distinctly different uses of the textual features measured for the private and published texts for both 1821 and 1831. This would tell us that Faraday did not write about electromagnetism in any one way, but used certain textual features to construct and reconstruct electromagnetism depending on the audience.

In terms of the histograms it would be interesting to find unique profiles that allow discrimination between or identification of, for example, letters Faraday wrote to scientists and non-scientists, or his published and notebook writings on electromagnetism. Again this would indicate how successful these features are in terms of being captured in Faraday's different writings. Also this would provide further - visual - evidence for how differently Faraday wrote in his private and public writings. The use of quantitative (numerical) and qualitative (visual) results. Lastly, histograms which are generated from samples with less than ten data points will be omitted.

[406] Gooding, David C and Addis, Tom R.; 1996 (a): 9.

4.4.6 The Specifications for the Large Scale Text Analyses To Enable the Evaluation of the Conjectures

Tables 4.1 to 4.3 below give the word length and number of paragraphs for the texts analyzed for each of the chapters (six to eight) dealing with the main case-studies.

Table 4.1 The Texts and Their Number of Words and Paragraphs for Chapter Six.

	Chapter 6	
Correspondence	words	paragraphs
Faraday to Herschel (1825-48)	6,159	290
Herschel to Faraday (1825-46)	6,495	261
Faraday to Whewell (1831-48)	10,635	331
Whewell to Faraday (1833-48)	8,644	272
Faraday to Barlow (1839-48)	2,396	117
Barlow to Faraday (1846)	805	6
Faraday to Herbert (1841-48)	24,381	401
Herbert to Faraday (1844-48)	2,968	142
Faraday to Magrath (1825-46)	6,036	203
Magrath to Faraday (1831-36)	398	22
Faraday to Phillips (1838-45)	3,949	118
Phillips to Faraday (1838-45)	1,142	55
Faraday to Sarah (1819-46)	4,272	92

Table 4.2 The Texts and Their Number of Words and Paragraphs for Chapter Seven.

	Chapter 7	
	words	paragraphs
Darwin's *Origin* (ch. 1 and 2)	16,581	94
Darwin's 'M' notebook	13,727	232
Darwin's 'N' notebook	8,353	142
Schoenbein's letters to Faraday (experimental topic)	32, 487	366
Schoenbein's letters to Faraday (non-experimental topic)	7,249	159
Faraday's letters to T. Huxtable	1,436	25
Faraday's letters to B. Abbott	54,905	728
Faraday's letters to R. Phillips	2,896	105

Table 4.3 The Texts and Their Number of Words and Paragraphs for Chapter Eight.

	Chapter 8	
	words	paragraphs
Diary (1821; pp49-63)	4,135	75
ERE2 (pp147-158)	4,702	26
Electro-magnetic Motions Paper	13,603	88
Sketch (all)	18,093	144
Sketch (part 1)	2,509	19
Sketch (part 2)	8,364	54
Sketch (part 3)	7,220	71
Diary (1831; pp367-430)	24, 180	678
ERE1	42, 522	501
'electro-magnetism' letters (1821-1831)	8,222	159

4.4.7 Expected Results

In matrix 4.2 below I detail the expected presence of the wordlists in each of the genre of writings analyzed, based on the results of my pilot study.

Matrix 4.2 Expected Presence of Modality and Compression in Each Type of Writing.

Word-list	Notebook	Letters	Sketch	Book	Research papers, monographs
‘empirical positive’ modalwords	hardly any	some	some	many	many
‘empirical negative’ modalwords	hardly any	many	some	many	many
‘empirical’ inclusive	hardly any	many	some	many	many
‘contingent’ positive modalwords	many	some	some	hardly any	hardly any/none
‘contingent’ negative modalwords	many	some	some	hardly any/none	hardly any/none
‘contingent’ inclusive	many	some	some	hardly any/none	hardly any/none
contractions	some	none	none	none	none
compression	most prevalent	some	hardly any	hardly any	hardly any

4.4.8 Problems Encountered

Having completed the pilot study I have come up against three problems:

1) The software - and the methods - are new and continually under development.

2) Once the documents are captured electronically time is required to proof-read the material.

3) Not all textual features are able to be captured at all or with little sensitivity.

Additionally I have been conscious of the fact that during my research a collection of private and public writings will not provide a entire record of experimental activity, a point Holmes warns us about.[407] This is because historical documents will be incomplete (see section 1.1).

4.4.9 Remedies

My use of CLARITY/FAITH text analysis programs in my main studies will be intensive, thus I expect my interpretation skills to become more intuitive. The proof-reading is certainly time-consuming, but there are two advantages. First it helps me acquire a general feel for Faraday's writing style across different genres, and second, this reading will add to my knowledge of Faraday biographically and in terms of the era he was working and writing in. I think the correspondence will prove especially useful here. The sensitivity of my textual features in terms of enabling discrimination between private and published texts was tested on small amounts of texts, but my main studies will comprise very substantial volumes of writings over about thirty years (see tables 4.1 to 4.3). This will allow a much wider textual landscape for any patterns to emerge. Furthermore, due to the flexibility and modular nature of the programs software problems can usually be corrected or software updated quickly.

My first use of the computational tool to assist in the analysis of compression and modality in Faraday's writings was my pilot study which is written up in the next chapter.

[407] Holmes, Frederick L.; 1984.

5 Pilot Study

Abstract. This study investigates the use of compression and modal words in personal and private writings of Michael Faraday and Charles Darwin. The work draws on Gilbert and Mulkay's analysis of scientific discourse using two linguistic repertoires: the empirical and the contingent.[408] This research project is interested in how effective the textual features measured are of acting as indicators able to discriminate between private and published texts. Thus this pilot study is an initial test of how sensitive compression and the modal wordlists are as potential indicators. The key measures used are compression, lists of modal words and lists of words and phrases from Gilbert & Mulkay's analysis of scientific discourse. The findings indicate that Darwin's 'N' notebook contain the greatest percentage of both empirical and contingent modal words. Compression was highest in Faraday's *Diary* and in Darwin's *Origin*. Two features were most able to allow discrimination between 'private' and 'public' texts, namely compression and 'empirical positive' modality.

5.1 Introduction

This study has two main aims. The first is to provide an initial test of how sensitive compression and modal wordlists are at allowing discriminations between different types of texts. The second is to demonstrate that a broader classification of scientific language use than Gilbert and Mulkay's dual repertoire (1984) can be used to capture the complexities and variability of scientific discourse. Their dual repertoire describes two different types of social accounting: the empiricist repertoire and the contingent repertoire. These constitute two different ways of accounting for or talking about how science works.

The use of the words 'contingent' and 'empirical' in my modal wordlists does not imply a direct correspondence to Gilbert and Mulkay's dual repertoire. Their use of 'contingent' is in terms of 'that which is dependent upon or varies with'. The

[408]Gilbert, G. Nigel and Mulkay, Michael; 1984.

contingent repertoire thus refers to how talk is used to situate scientific practices as contingent or dependent upon 'non-scientific' factors, such as personality, status, and gossip. Whereas the empirical repertoire accounts for or defines scientific work in relation to 'experimental' factors. In this repertoire non-human actors are given agency, modality is deliberately removed, and the actions of the author are presented as not relevant to the findings. A scientist's own position and work is presented in the empirical/experimental idiom, and errors and the position of other scientists in the contingent.

The aims of my study diverge from Gilbert and Mulkay's in that I am specifically interested in the words used in scientific writings and how these are employed to express degrees of certainty. For example, modality is equivalent to terms such as 'maybe or 'could be', and terms in the empirical modality include 'are' and 'must be'. I am therefore using 'contingent' in the way it has come to be specifically used in the sociology of science to mean language or ideas that are uncertain or indeterminate, in contrast with those that are presented as empirical, factual, and immutable. My use of the word 'empirical' describes this latter style of language. One of the significant conclusions of my pilot-study was the rejection of the Gilbert and Mulkay 'contingent' and 'empirical' wordlists (see appendix four), since these words did not allow discrimination between different types of texts.

Gilbert and Mulkay use material gleaned from informal interviews with scientists, as well as published papers. My equivalent private/personal scientific discourse consists of letters and notebooks. I am using measurement of compression and the wordlists as indicators to try and discriminate between private and published writings of Faraday and Darwin. Thus this study tests whether the personal writings have more of the informal or contingent modal terms than the published works. It is expected that the published writings would demonstrate more modality because the sciences use language to construct truth statements for their audiences. In this way consensual certainties are construed and knowledge is legitimated as 'factual'. Also it is conjectured that compression is more prevalent in the private writings because the number of repeated words, such as 'and', 'of', 'at' should be fewer, due to less concern with formal grammatical sense compared to material for publication.

5.2 Data

The data consists of twenty paragraphs from the beginning of seven types of writing: two private and two public writings by Faraday and one private and two public by Darwin:

1) Michael Faraday:
i) Private: *Diary 1* (laboratory notebook), *Letters Volume 1*
ii) Public: *Electro-Magnetic Motions 2* (*ERE2*), *Historical Sketch of Electro-Magnetism* (*Sketch*)

2) Charles Darwin:
i) Private: 'N' Notebook
ii) Public: *The Origin of the Species* (Chapter 1), *Descent (or Origin) of Man* (Part 1, Chapter 1)

5.3 Method: Data Preparation

The scans followed the general procedure below for unsorted data:

1) Preparation of Files for Analysis: Selecting text from a source file:

- Open Word and a new file.
- Open the source file.
- Use the mouse to 'block' the first twenty paragraphs.
- Use the copy and paste functions to place the twenty paragraphs in the new file.
- Save this file in plain text format.
- Open the new file to ensure that it contains the text you required.
- Save the file in ASCII format and then close it.

2) Analysis of Text:

- Ensure the source file is closed and in the Clarity directory.
- Scan the text with the 'single_list_%score2_table' function, choosing a paragraph size of ten words (#10 in the example of a 'Query' below). This function scans the file, generates a data file, and converts this to a *.tab* data file containing a tabbed list stored in the Data Tables directory. The tabbed list can be imported directly into a graphics/statistics package:

```
QUERY>    single_list_%score2_table    #10    [(modalwords    "all    inclusive")]
":Resource_Files:F_Diary:FaradayD1_20paras" "FD1_incl.tmp"
['
' "Scanning source file" '
']['
' "Scan complete.  Converting results to table" '
']['
' "Results in tabular data (.tab) file. " '
' ]True malloc/free/realloc: clobbered space detected
```

- Open the files to check that a column of scores has been written to the file and remove any extra paragraph marks by doing a search and replace. Close the file. The converted (table) file looks like this:

2.531646
9.333333
11.904762
0.000000
0.000000
2.777778
5.769231
3.389831
10.00000

Place an extra paragraph mark at the beginning to ensure that the data starts at column one when pasted into a statistical package.

3) Importing Data into Minitab v.10:

- Launch MS Graph.
- Open a new worksheet.
- Start a recording session.
- Import the data using 'copy and paste'.
- Label each column.

5.4 Method: Data Analysis

The first twenty paragraphs of each of the texts were selected and pasted into a new MS Word document. These texts were then proof-read and spell-checked, creating seven text files. The methodology used compression and word lists calculated as a percentage of paragraph size. I constructed seven lists of modal terms, and three further lists of words and phrases were gleaned from reading Gilbert & Mulkay's *Opening Pandora's Box*. The modal word lists were designated either 'contingent', that is, the words expressed doubt or error, or 'empirical', conveying certainty or conviction. The lists of words are found in appendix four. The 'empirical positive' modality represents the most robust form of conviction and truth-assertion. These terms perpetuate the myth of 'taken-for-granted' knowledge and act as enthymemes, statements which exclude the expression of key assumptions which ground conclusions. At the opposite end the weakest or most 'contingent' form of modality is the 'contingent negative'.

The specific discriminations I wanted to make are set out in table 5.1. I am using measurement of compression and the wordlists to try and discriminate between Faraday's private writings (laboratory notebook and letters) and his published texts (*ERE 2* and the *Sketch*). In the case of Darwin's writings the attempted discrimination is between his 'N' notebooks and the two books (*Origin* and *Descent*).

Table 5.1 Discriminations To Be Attempted, Indicators Used, and Texts Scanned.

	Private v public	private v public	private v public
compression	Faraday: *Diary 1* v *ERE 2* and *Sketch*	Faraday: Letter V1 v *ERE 2* and *Sketch*	Darwin: 'N' notebook v *Origin* and *Descent*
seven modalities	Faraday: *Diary 1* v *ERE 2* and *Sketch*	Faraday: Letter V1 v *ERE 2* and *Sketch*	Darwin: 'N' notebook v *Origin* and *Descent*
'Gilbert & Mulkay' wordlists	Faraday: *Diary 1* v *ERE 2* and *Sketch*	Faraday: Letter V1 v *ERE 2* and *Sketch*	Darwin: 'N' notebook v *Origin* and *Descent*

The most interesting results of the scans are presented numerically in tables, with the (truncated) mean and the (non-zero) minimum and maximum scores displayed. As I was also interested in whether the text features measured would produce 'trademark' patterns, where there is clustering of data or where an author seems to use a significant percentage of particular modal words, some results are also shown as cluster plots.

5.5 Results

It was found that the most interesting results were produced from scanning for compression and 'empirical positive' modality. The list of contractions and the 'Gilbert and Mulkay' terms were present only to a very small degree or not at all. The main results are found in numerical form in tables 5.2 and 5.3 (page 127). It is realised that analysing only twenty paragraphs may not have been enough to allow significant patterns or clusters to emerge.

Table 5.2 Results of Compression and 'Empirical Positive' Modality (ep) Scans for Faraday and Darwin's Private and Published Texts.

	compression - mean	compression - minimum	compression - maximum	ep - mean	ep - minimum	ep – maximum
Faraday						
Diary 1	14.11	4.55	33.33	3.17	1.69	10.00
Sketch	37.49	11.43	59.73	3.50	1.41	6.80
ERE2	38.34	24.49	59.84	3.06	1.32	6.74
Letters VI1	30.70	7.41	46.58	2.52	0.59	7.69
Darwin						
Origin	41.24	22.11	55.43	2.56	0.49	7.73
Descent	15.57	3.70	38.35	1.75	1.69	5.08
'N' notebook	16.38	3.45	28.16	3.66	1.37	15.00

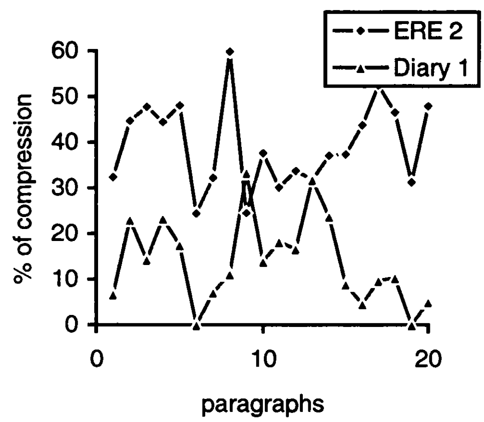
5.5.1 Compression

The first thing to note is that Faraday's texts have a similar mean compression except for the *Diary*. Even though Faraday's letters are a personal discourse, they are written to a public audience, and, as such, there is full attention given to use of grammar. His laboratory notebook is the only 'private' writing, that is, there is less concern over the use of articles and connectives, thus less repeated words, and so more compression. Thus the measurement of compression has allowed discrimination between Faraday's private and public writings to be made.

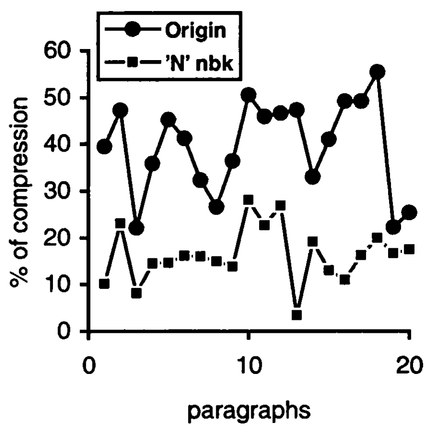
In the case of Darwin it is the paragraphs from the *Origin of the Species* which demonstrates compression consistent with a published narrative. What is interesting is that the degree of compression for Darwin's other book, *Descent*, is similar to the 'N' notebook. Compression has again acted as an indicator of a private or public text, and allowed discrimination between these two genres, though only between the *Origin* and the 'N' notebook. In terms of its compression it has been written as if it was a 'private' discourse.

The difference in compression scores between the private and public texts of Faraday and Darwin are clearly visualized in graphs 5.1 and 5.2 respectively. There is virtually no overlap in the scatter plot patterns for each graph.

Graph 5.1 Degree of Compression for Faraday's *Diary 1* and *ERE2*



Graph 5.2 Degree of Compression for Darwin's *Origin* and 'N' notebook



5.5.2 Modal wordlists and 'Gilbert and Mulkay' Terms

The main results are displayed numerically in table 5.3 (page 127) and in scatter plots. The 'all inclusive' wordlist produced percentages generally below twelve percent for Faraday and Darwin. This is except for Darwin's 'N' notebook whose paragraph eleven contains thirty percent (Graph 5.3). The highest percentage for the empirical positive modal words was fifteen. This was for Darwin's 'N' Notebook, again paragraph eleven (Graph 5.4). Darwin's 'N' notebook produced the highest percentage for empirical negative modal words, but even these were largely less than one percent. The percentage of 'empirical' modality terms for Faraday's texts is uniformly low; no greater than eight percentage. The highest percentages are for the 'empirical positive' modality in Faraday's *ERE2* and *Sketch* respectively (Graph's 5.5 and 5.6) and Darwin's *Origin* (Graph 5.7).

The percentage of contingent positive modal words is poorly represented by both Faraday and Darwin. Of the seven texts only Darwin's *Origin* exhibits any notable degree of these terms (Graph 5.8). For Faraday it is his *Sketch* which warrants attention only, with three paragraphs containing words for this word list. The contingent negative modal list fairs worse still, with only the paragraphs in Darwin's *Origin* containing any notable amount of these terms, even then the percentages did rise above 0.8.

Darwin's 'N' notebook has greatest percentage of terms from the empirical inclusive list (Graph 5.9). This text contains twenty percent of these terms. Again Faraday's modal terms do not rise above ten percent except for paragraph three in the *Diary*. Apart from Darwin's notebook in only two other texts are the empirical inclusive modal terms found to any noteworthy degree: Darwin's *Origins* (Graph 5.10), and Faraday's *Sketch* (Graph 5.11). The analysis using the contractions word list produced no significant results for neither Faraday nor Darwin (zero percentages were commonplace).

Lastly the Gilbert & Mulkay modal lists. Faraday's more formal, empirical texts contain small percentages of the empirical words and phrases. This is more so with Faraday's *ERE 2* and Darwin's 'N' notebook. The *Descent of Man* has two or three paragraphs at the end where the percentage rises to a maximum of approximately six percent (Graph 5.12).

Table 5.3 The Mean, Truncated Mean, and the Minimum and Maximum Values for the Wordlists Scatter Plots.

modal words	Truncated Mean	Minimum Value	Maximum Value
all inclusive (1)	**5.52**	**2.13**	**30.00**
empirical positive (1, 2, 4, 5)	**3.66/3.89/4.45/2.56**	**1.37/0.43/1.49/1.31**	**15.00/7.52/6.74/6.91**
contingent positive (2)	**0.22**	**0.22**	**2.33**
empirical inclusive (1, 2, 5)	**2.77/3.78/4.17**	**1.37/1.70/1.52**	**20.00/7.27/7.93**
G & M empirical (3)	**1.17**	**0.46**	**6.09**

1 = Darwin's *'N' Notebook*

2 = Darwin's *Origin of the Species*

3 = Darwin's *Descent of Man*

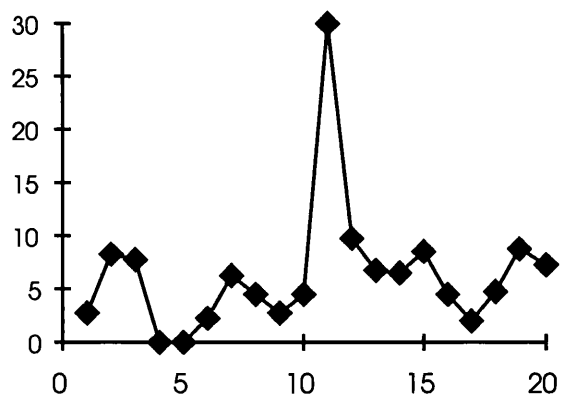
4 = Faraday's *Experimental Researches in Electricity 2*

5 = Faraday's *Sketch*

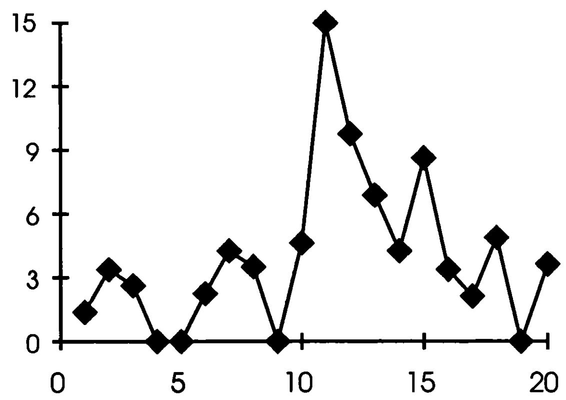
5.5.3 Scatter Plots for Wordlist Results

The x-axis denotes the paragraph number and the y-axis displays the percentage of words.

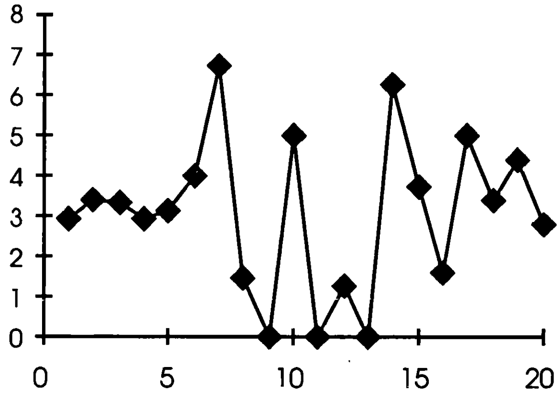
Graph 5.3 Modal words (all inclusive) for Darwin's *'N' notebook*



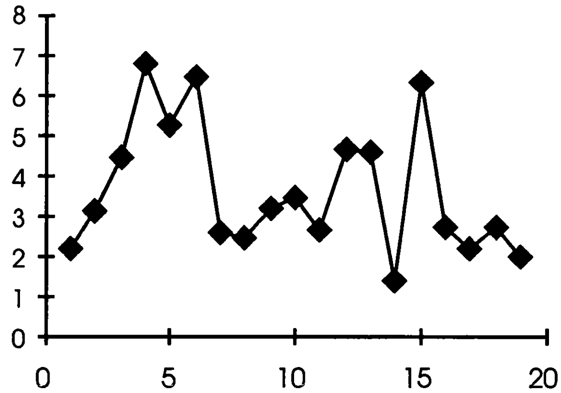
Graph 5.4 Modal words (empirical positive) for Darwin's *'N' notebook*



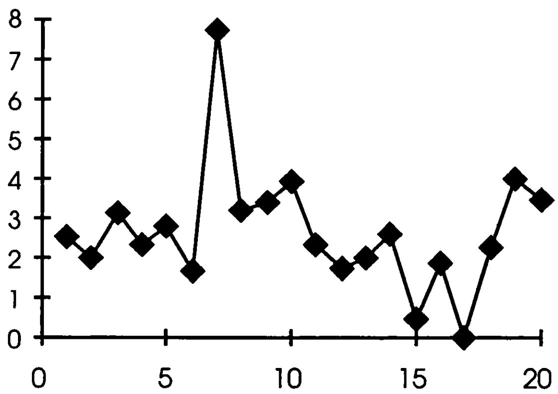
Graph 5.5 Modal words (empirical positive) for Faraday's *ERE2*



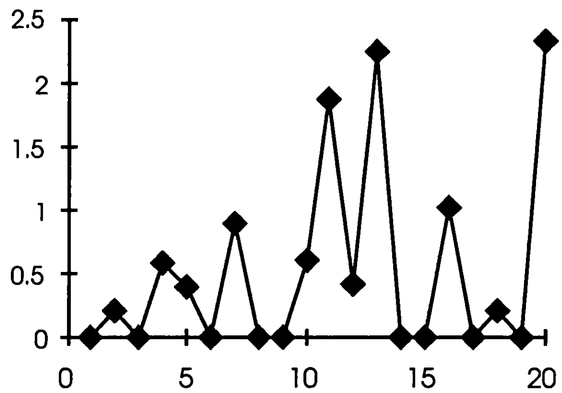
Graph 5.6 Modal words (empirical positive) for Faraday's *Sketch*



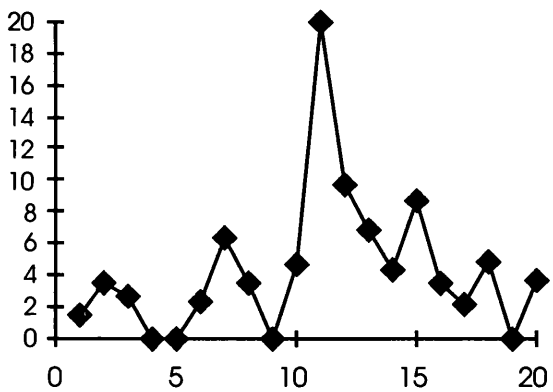
Graph 5.7 Modal words (empirical positive) for Darwin's *Origin of the Species* (chapter 1)



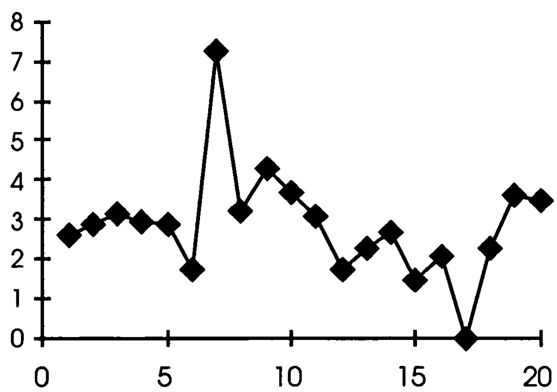
Graph 5.8 Modal words (contingent positive) Darwin *Origin of the Species* (chapter 1)



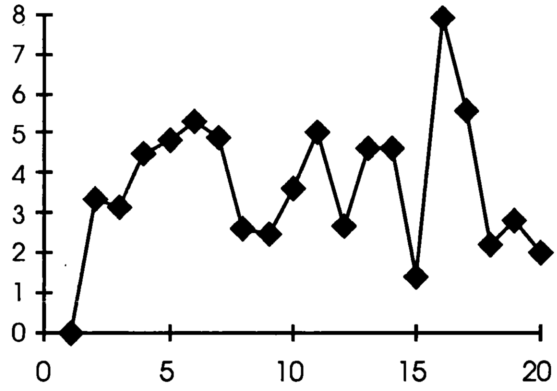
Graph 5.9 Modal words (empirical inclusive) for Darwin's *'N' notebook*



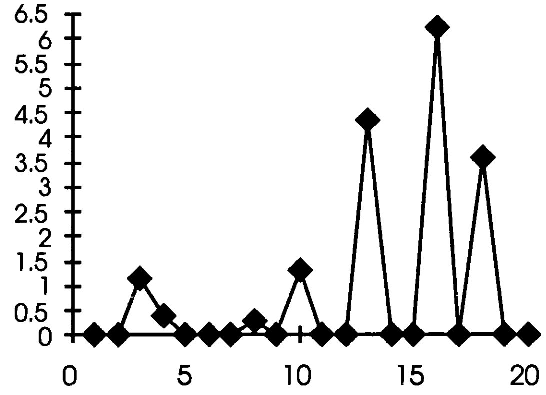
Graph 5.10 Modal words (empirical inclusive) for Darwin's *Origins of the Species* (chapter 1)



Graph 5.11 Modal words (empirical inclusive) for Faraday's *Sketch*



Graph 5.12 Modal words (G & M empirical) Darwin's *Descent of Man*



5.6 Conclusions

To conclude, there are two noteworthy findings. It is the twenty paragraphs of Darwin's *N notebooks* that quite consistently contain the highest percentage of both empirical and contingent modal terms. This is not the case with the Gilbert and Mulkay's word lists, which did not enable discrimination. Faraday's *Sketch* contains the highest percentages of these terms, and only Darwin's *Origin* displays any significant degree of their contingent terms and phrases. It is also worthy noting that Faraday's paragraphs consistently score low percentages for all the modal word lists. Thus we find that compression and the modalities, 'empirical positive' and 'contingent negative', are most able to act as indicators of the discriminations I wanted to make.

The work by Gilbert and Mulkay suggests that when accounts of experiments and findings enter the public sphere language is used that avoids contamination by those human, contingent, or social processes that are an integral part of experimental practices. This, albeit small-scale, study demonstrates that things are not that simple. Language use is complex and is shaped by such factors as the intended audience, the subject matter, peer support or pressure, and the stage a scientist is at in their career.

In more general terms I have found that the most useful indicators are measurement of compression and 'empirical positive' modality. Measurement of compression meant I was able to make the discriminations laid out in table 5.1 between Faraday's private and public texts. Furthermore, Faraday's letters produced mean and maximum compression scores closer to that of the published writings. So even though the scores for the letters and the published material were different this difference was not as stark as that between Faraday's *Diary 1* and the published texts. Thus I conclude that the significant discrimination for Faraday, in terms of compression, is between his laboratory notebook and the published writings: the *Sketch* and *Electro-Magnetic Motions VI 2*.

When we look at the degree of 'empirical positive' modality we find that Faraday's *Diary 1* has one of the higher mean scores, comparable to the published texts. Thus measurement of 'empirical positive' modality has not so readily enabled discrimination between Faraday's private and public writings.

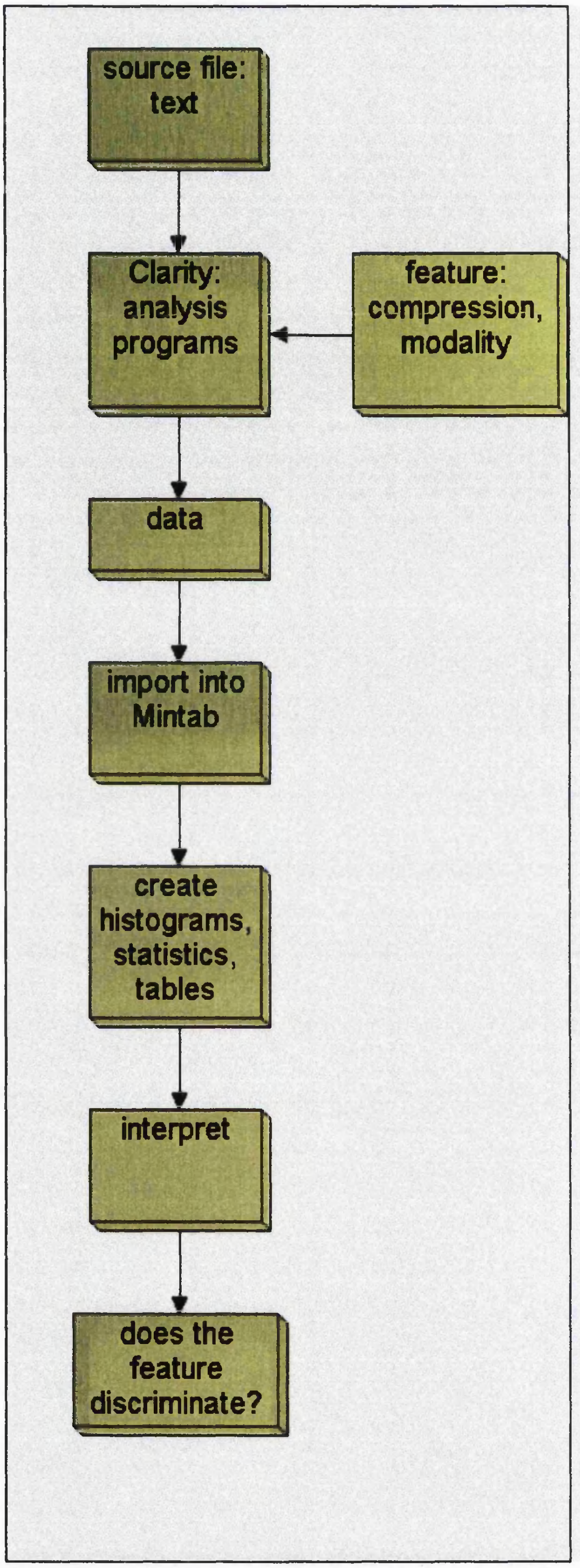
The measurement of compression and 'empirical positive' modality were also effective in making the discriminations in table 5.1 for Darwin's texts, though they were not as straight forward. The private/public discrimination when measuring mean compression is between the 'N' notebook and *Origin.* The mean compression score for *Descent* is similar to the notebook. The 'empirical positive' modality allowed discrimination between the 'N' notebook and the two published works. Taking on board the conclusions and main results of my pilot study I now move onto the chapters dealing with the main case studies, which begin with a brief introduction.

Introduction to Chapters Six, Seven and Eight

Abstract: The three chapters which follow comprise my main empirical studies. Chapter six tests the sensitivity of the features - compression and the seven modality lists - as indicators of the discriminations I want to make, that is, between: i) 'formal' (dealing with experimental matters) and 'informal' (dealing with non-experimental matters) letters, written to scientists and non-scientists, ii) 'early' and 'late' letters, and iii) letters to friends and to scientists. In the seventh chapter I test the main findings of chapter six. I investigate the sensitivity of the compression and the modalities, 'empirical positive', and 'contingent negative' as indicators of further discriminations I want to make, between: i) private and public texts, ii) early letters to friends and later letters to scientists, and iii) letters to and from Sandemanians/Non-Sandemanians. This chapter also continues the investigation of the sensitivity of word-frequency based text analysis. Chapter eight focuses on Faraday's private and public writings on his most important work - electromagnetism. I analyze the language use over the ten-year period - 1821 to 1831 - which saw the blossoming of his work in both electromagnetic induction and rotation.

On a general level what I am doing in chapters six, seven and eight is described in flowchart 1 on the next page.

Flowchart 1. The General Method for Testing What Features Enable Discriminations.



Finally, the computer-assisted analysis of the texts will follow certain procedures, an example of which is laid out below:

1) Preparation of files for analysis: Selecting text from a source file using the function file_cut01.[409]

- Ensure the source file is closed, saved as a text file, and is in the same directory as CLARITY.
- Apply the file_cut01 function to the source file, as in the following example:

QUERY> file_cut01 "FLET_V2(525-1333).txt" "FLet1_Hersch(end)" ["Faraday to John Frederick Herschel" "END"]

This function selects and copies text from a source file between two markers (the co-respondents and either END of EDITOR'S NOTES(S)) writing the text to new file.

- Open the new file to ensure the it contains the text you required.
- Save the file in ASCII format and then close it.

For this experiment the letters between Faraday and co-respondents, John Herschel and William Whewell were selected from volumes one, two, and three of electronic versions of Frank James' The Correspondence of Michael Faraday'.[410] Twelve files were created: Faraday to and from Herschel for each of the three volumes, and the same with Whewell. The files were saved in MS Word as text files.

- Open each file to further select the letters for a given time period.

These files were opened in MS Word letters and divided up into five year periods: 1825-1829, 1830-1834, 1835-1839, 1840-1844, and 1845-1849. This procedure resulted in seventeen files.

2) Scanning of texts.

a) Comparison of the ellipsis (compression) scores.

- Ensure the file is closed.
- Decide on the size of paragraphs to be selected for scanning.

The compression scores are expressed as a percentage of paragraph size.

- Scan the text using the ellipsis-measuring function:

---

[409] This function reads a source file and copies the text to another file from between two specified markers.

QUERY> scan_%diff_wrd_tls_only2[411] #10[412] ":Resource_Files:Cut_letters:F_Her-scf/non-scf:F_Her/1844-49/scf" "F_Her/1844-49/scf.tmp"

- Open the files to check that a column of scores has been written to the file. Close the file.
- Convert the file to a tabbed data format for pasting into a graphics/statistical package:

QUERY> lfile_to_tfile[413] "F_Her(1844-49/scf).tmp" "F_Her(1844-49/scf).tmp.tab" "F_Her(1844-49/scf).tmp.tab"

- Open the file and remove any extra paragraph marks by doing a search and replace.

An extra paragraph mark at the beginning ensures that the data starts at column one when pasted into a graphics/statistical package.

b) Modalword frequencies as percentage of paragraph size.

The same initial steps are followed as for compression, then scan the text using a single 'prog' function:

QUERY> single_list_%scores2_table[414] #10 [(modalwords "empirical positive")] ":Resource_Files:Cut_letters:F_Her-scf/non-scf:F_Her/1844-49/scf" "F_Her/1844-49/scf.tmp"

4) Importing Data into Minitab

- Launch Minitab.
- Open a New Worksheet.
- Start a recording session.
- Import the data using 'copy and paste' starting with column two.
- Label each column.

I now turn to the first of my main case studies. Chapter six explores the use of modality and compression in correspondence between Faraday and scientists and non-scientists.

[410] Published by the Institute of Electrical Engineers; 1991, 1993, and 1996.
[411] This function calculates measures compression as a percentage of paragraph size.
[412] This is the minimum paragrpah size (number of words) to be scanned.
[413] This function converts data to table-form to faciltate import into a statistics package.
[414] This function scans the source file, and converts the data file in to a .tab dat file. This file contained a tabbed list which can be imported directly into a statistics package.

6 Faraday's Language of Facts and Purification of Words: A Study of Modality and Compression

Abstract: The research described in this chapter is in part an investigation of the sensitivity of word-frequency (frequency of percentage scores) based text analysis. What I do is study the language used in Faraday's correspondence with two scientists and five non-scientists, over a twenty-five year period. More precisely, I look at compression (ellipsis) and the use of modality as potential indicators of discriminations I want to make. These discriminations are between early and late Faraday letters, 'experimental' and 'non-experimental' letter topics, and letters to scientists and non-scientists. Also, an interest in whether the tables and specific indicators show any pattern or trend is pursued. The main finding is that Faraday demonstrates a quite consistent use of a 'strong' and confident truth modality across his letters to scientists and non-scientists. Also though when the subject matter is non-experimental the compression scores are generally greater, in Faraday's later letters to the scientists the 'experimental' topic tends to produces the higher compression scores.

6.1 Introduction

The study described in this chapter, and the others in chapters seven and eight, are in part an investigation of the sensitivity of word-frequency (that is, frequencies of percentage scores per paragraph) based methods of text analysis. Using this type of text analysis I am looking at particular textual features as indicators of discriminations I want to make. In this chapter the interest lies with the degree of compression and with modality. These features are investigated in order to try make three discriminations, that is, between: i) letters dealing with 'experimental' and 'non-experimental' topics, ii) early and late Faraday letters, and iii) letters between scientists and non-scientists.

6.2 Modality

I chose to study modal words because they are used to highlight modality, an interpersonal practice in the use of language. Interpersonal linguistic practices: "always have some statement to make, and often work by implied propositions or presuppositions."[415] What I am looking at is the commitment to the truth and certainty of a statement or proposition, which varies in robustness and confidence, from deep conviction:

> "This Water *is* excellent in taste and smell, and contains only a small portion of Saline matter".[416]

To far less certainty:

> "If I had known the circumstances as you describe them, I should....have probably not spoken at all....though I might have presumed to lay my view before you in private conversation."[417]

Traditional or positivistic linguistics perceives language primarily as a vehicle for the conveying of thoughts, ideas, and facts about the world. The relationship between thought and language is assumed to be fixed and unchanging. As has been realized by Wittgenstein,[418] Vygotsky,[419] Austin,[420] Searle,[421] and other anti-foundationalists, there is no one-to-one-link between word and referent. Language is also a site of agency, a practice: "a mode of action."[422] So when we speak or write, we are also doing something through this action.

When modality is used with great strength we are seeing the linguistic construction of truth statements. For Carnap the language of science means "the use of language for making assertions".[423] In published papers this is robust, truth assertions, as, I believe, evidenced by a high degree of the 'empirical positive' modality. This is a positivistic linking of thoughts, language, and physical action. That is, an idea or thought is expressed as a statement of

---

415 Fowler, Roger; 1991: 85.
416 Faraday to Jacob Herbert; January 30th, 1843.
417 Faraday to Edward Solly; 9 February 1832.
418 Wittgenstein, Ludwig; 1967.
419 Vygotsky, Lev, S.; 1988
420 Austin, J. L;1962.
421 Searle, J. R.; 1969.
422 Fowler, Roger; 1991: 87.
423 Carnap, Rudolf; 1937: 3

fact or truth both in language, say in a laboratory notebook, and in physical action, in terms of an experiment. Two things are being constructed here. One is a linear pathway of internal thoughts and ideas to external speech or writing (that is, communication) and physical action. The other construction is the non-necessity or irrelevance of the mediated nature of the relationship between thought, words, and action; a mediation through temporary, linguistic, emotional, and social factors. This latter construction is situating communication more as a biological phenomenon.

6.3 Compression

The second aspect of language investigated is the degree of compression (ellipsis). The compression function:

> "converts a paragraph to words (a list or string), counts them, reduces the list to a set (i.e. removes all repetitions), counts the latter and subtracts the second total from the first, to give a difference [returned] as a percentage of paragraph size".[424]

Compression was originally designed to capture variance in the grammatical complexity of successful drafts of a text or writings on a particular theme from a private narrative to the public account, for example, electromagnetism. In private writing, for example, in a diary entry, we could expect the omission of prepositions, articles and conjunctives. Thus the text would be more compressed.[425] This indicator can be used in this chapter because, for instance, I would expect, on the above criteria, the letters dealing with 'experimental' topics to show less compression than those predominantly discussing 'non-experimental' matters. One role of language here is as an *aide-de-memoire*. Also thoughts and ideas do not arrive complete or fully articulatable. It is only after we enact the thought, or manifest it in speech or writing, that we are able to give it the social and cultural context it needs to have meaning, and to see how ideas may interlink and lead off on other tangents.

[424] Gooding, David C.; January 1995: 61.
[425] Gooding, David C.; January 1995: 63.

6.4. Methodology

6.4.1 Source Material

The scientists chosen were John Frederick Herschel[426] and William Whewell,[427] not least because a sizeable number of letters could be collated over three volumes of Faraday's correspondence. Five non-scientists with whom Faraday co-responded were chosen:[428] Sarah Faraday,[429] Edward Magrath,[430] Samuel Phillipps,[431] John Barlow,[432] and Jacob Herbert.[433] The letters for both sets of correspondence covered the years 1825 to 1849. As I was looking at the correspondence over time I first divided the letters up into periods of five-years: 1825-29, 1830-34, 1835-39, 1840-44, and 1845-49. Letters were not available for every time-period for all the co-respondents,[434] and the first letters from Faraday to Sarah spanned an earlier period: 1819-24. The first and last five-year periods of Faraday's correspondence with Herschel and Whewell were divided up according to whether paragraphs primarily dealt with an 'experimental' or an 'non-experimental' topic. An example of an 'experimental' topic is when Faraday discusses glass experiments with Herschel or crystal experiments with Whewell. Examples of 'non-experimental' topics are Faraday's mention of eye inflammation to Herschel and his requests for nomenclature from Whewell. The same classification was applied to the letters to and from non-scientists. If it was not the case that letters dealt with one of these two themes predominantly then the label 'mixed' was applied. Tables 1 and 2 detail the number of letters for the scientist and non-scientist groups.

426 John Frederick Wiliam Herschel (1792-1871) was involved in astronomy, chemistry, and physics. Secretary of the Royal Society, 1824-1827.
427 William Whewell (1794-1866) was primarily an historian and philosopher.
428 My choices were greatly assisted by the suggestions of Frank James.
429 Sarah Faraday (neé Bernard) (1800 - 287). Married Faraday on June 12th, 1821.
430 Edward Magrath (1791 - 1861). Secretary of the Athenaeum Club, 1824-1855.
431 Samuel March Phillipps (1780-1862). Permanent Under Secretary of State at the Home Office, 1827-1848.
432 John Barlow (1798-1869). Secretary of the Lecture Committee at the Royal Institution, from1841, and Secretary of the Royal Institution, 1843-1860.
433 Jacob Herbert (1788-1867). Secretary of Trinity House, 1824-1856.
434 There were no letters available between Faraday and Whewell for 1825-1829. Nor from Herschel to Faraday for the period 1840-1844.

Table 6.1 The Number of Letters for Each Five-year Period of Faraday's Correspondence with Herschel and Whewell.

	Faraday to Herschel	Herschel to Faraday	Faraday to Whewell	Whewell to Faraday
1825-1829	12	10	-	-
1830-1834	1	1	7	5
1835-1839	2	1	5	7
1840-1844	1	-	3	2
1845-1849	7	4	11	9
total	23	16	26	23

Table 6.2 The Number of Letters and Primary Topics For Faraday's Correspondence With the Non-Scientists.

Co-respondents	Five-year period	number of letters	topics
Faraday to Sarah	1819-24	10	non-experimental
Faraday to Sarah	1835-39	1	non-experimental
Faraday to Sarah	1835-39	1	non-experimental
Faraday to Phillipps	1835-39	1	non-experimental
Faraday to Phillipps	1840-44	5	non-experimental
Faraday to Phillipps	1845-49	4	non-experimental
Phillipps to Faraday	1835-39	2	non-experimental
Phillipps to Faraday	1840-44	6	non-experimental
Phillipps to Faraday	1845-49	1	non-experimental
Faraday to Herbert	1840-44	13	experimental
Faraday to Herbert	1845-49	18	experimental
Herbert to Faraday	1844-49	4	experimental
Herbert to Faraday	1845-49	19	experimental
Faraday to Barlow	1835-39	1	non-experimental
Faraday to Barlow	1840-44	10	non-experimental
Faraday to Barlow	1845-49	10	non-experimental
Barlow to Faraday	1845-49	1	non-experimental
Faraday to Magrath	1825-29	6	non-experimental
Faraday to Magrath	1830-34	5	mixed
Faraday to Magrath	1835-39	13	non-experimental
Faraday to Magrath	1840-44	5	non-experimental
Faraday to Magrath	1845-49	1	non-experimental
Magrath to Faraday	1830-34	1	experimental
Magrath to Faraday	1835-39	1	non-experimental

6.4.2 Analysis

I built seven lists of modality words ranging from expression of great certainty and truth to that of doubt and uncertainty. The modal words lists used are described in appendix four. The most 'objective' modalities, especially 'empirical positive', depicted by the use of words such as 'is', 'will', and 'are not', 'never' (found in my 'empirical positive' and 'empirical negative' wordlists) demonstrate the greatest commitment to presupposition, certainty, truth statements, and a linear, non-mediated relationship between thought, language, and action.

My data will be analyzed in two forms: i) as numerical results, that is, the (truncated) mean[435] and the minimum (smallest non-zero) and maximum values, and ii) as (connected) histograms. The numbers, the modal word and compression scores, are expressed as percentages of paragraph size, and it is this that the mean, and minimum and maximum, values correspond to.

My texts thus divided up into seven groups, each analyzed for modality and compression:

i) Faraday to/from Herschel: undifferentiated,

ii) Faraday to/from Herschel: 'experimental' topics,

iii) Faraday to/from Herschel: 'non-experimental' topics,

iv) Faraday to/from Whewell,

v) Faraday to/from Whewell: 'experimental' topics,

vi) Faraday to/from Whewell: 'non-experimental' topics,

vii) Faraday to non-scientists (the five non-scientists were studied separately).

The source material was analyzed using text analysis programs written by David Gooding and Tom Addis in FAITH.[436] The data was converted into histograms and statistics gained using Minitab 10 Xtra. My findings were then interpreted and evaluated in terms of how well the chosen language features

[435] The truncated mean is where the smallest 5% and largest 5% of the values are removed, and the remaining values are averaged. A 'neater' set of raw scores is acquired; where anomolous large scores are removed.

[436] FAITH is the functional programming language for which the graphical programming interface called CLARITY. Developed by Tom R. Addis and Jan Townsend in collaboration with David C. Gooding and Simon Gray.

allow discrimination between: i) Faraday's early and late letters, ii) letters dealing with 'experimental' and 'non-experimental' matters. The files in this experiment were analysed with ten words chosen as the default paragraph size.

6.4.3 Empirical Work Undertaken

This empirical work has four parts. The first part is my main study that examines Faraday's letters to scientists and non-scientists, which is written up in this chapter. The main results of this chapter are tested through three further studies, which form chapter seven. The second study acts to check the compression scores and modality in the earliest available letters of Faraday to his friends - Benjamin Abbott, Thomas Huxtable, and Richard Phillips. This time period covers the nascence of Faraday's scientific career. Thus in these letters I expected Faraday's language to be characterized by relatively high compression, for the precise and certain writing of an established scientist would not yet have emerged, and that he was writing to friends. Also the use of the more 'contingent' modalities I would suggest would be more in evidence for the same reasons. The next study tests how well measurement of compression and modality can discriminate between private and public writings of scientists using texts by Faraday and Charles Darwin. This is an additional discrimination to those featured in the main study. Finally, the fourth study looks at the prevalence of compression and 'empirical positive' and 'contingent negative' modalities in letters to Faraday from non-Sandemanians.

The primary interest of this study was not a concern with statistical analysis. Instead, I was interested in shapes and patterns; specifically, to what degree the shapes were consistent over time, and what this could tell us about the author's style when discussing scientific and non-scientific matters. It must also be noted that we are not concerned with word frequencies, but with frequencies of percentage scores representing the degree of modal words and compression in each paragraph of a text.

6.4.4 Expected Results and Predictions

My expected results fall into three areas: 1) the ‘empirical positive’ modality would be: a) greater when Faraday is discussing an ‘experimental’ topic with another scientist, as well as with a non-scientist, b) more evident during 1845-49, c) seen to increase over time. This would reflect Faraday's evolving status as a scientist, from a novice, still finding his own way, to a scientist with an international reputation; a ‘grand, old man’ of science, 2) mean compression would be greatest in Faraday's: a) letters to non-scientists discussing ‘non-experimental’ topics, and b) letters for the early periods.

In table 6.3 I describe six categories of prediction for compression and the modalities, ‘empirical positive’ and ‘contingent negative’ (the predictions are relative). I shall compare these with my actual results.

Table 6.3 Predicted Results for Mean Scores for Compression and ‘Empirical Positive’ and ‘Contingent Negative’ Modalities.

	1825-29	1845-49	‘Experi-mental’ Letters	‘Non-experimental’ Letters	Letters to Scientists	Letters to Non-scientists
Ellipsis	**High**	**Low**	**Low**	**High**	**Low**	**High**
EP Modality	**Low**	**High**	**High**	**Low**	**High**	**Low**
CN Modality	**High**	**Low**	**Low**	**High**	**Low**	**High**

The distinctions tested for using compression and modalities as potential indicators across a variety of genres of texts in this chapter are summarized in table 6.4 I return to this table in my ‘discussion’ section to examine how my results relate to the design in the table.

Table 6.4 Relationship Between Distinctions, Features, and Source Texts.

	early v late letters	'experimental' v 'non-experimental' letters	letters to scientists v to non-scientists
compression	Faraday to scientists and non-scientists 1825-1849	Faraday to scientists and non-scientists 1825-1849	Faraday to scientists and non-scientists 1825-1849
seven modalities	Faraday to scientists and non-scientists 1825-1849	Faraday to scientists and non-scientists 1825-1849	Faraday to scientists and non-scientists 1825-1849

6.5 Findings

6.5.1 Introduction

My analysis of my main findings will be done in two parts. The first part will concentrate on the numerical results of my scans - truncated mean, the minimum (lowest non-zero score) and maximum values, for the correspondence between the scientists and the non-scientists. These results are presented as tables which are found in appendix five. The second part will deal with the histograms; those samples with less than ten data points will be omitted. Following on from which will be the interpretation of my findings, and conclusions and discussion.

6.5.2 Numerical Results

I shall look at the modal word results for Faraday's correspondence with Whewell and Herschel for the undifferentiated, 'experimental', and 'non-experimental' groups, followed by the results for the correspondence with the non-scientists. Next I analyse the results of the compression scans in the same order.

i) Modality Results for Correspondence Between Faraday and Whewell and Herschel: Undifferentiated

In the majority of cases Faraday's letters for 1840-1844 to Whewell show the largest (truncated) mean of modalwords (table 6.5). This situation does not arise with the two modalities which express the most informality or the least certainty. My belief was that the highest degree of modality would reside with the 'empirical inclusive' and 'empirical positive' wordlists. In the case of Faraday's letters to Whewell the results do not show such a simple scenario. These letters have the highest (truncated) mean scores for the 'empirical positive' wordlists, but these scores are closely followed by those for the 'contingent negative' modal words, which give the highest maximum values. The wordlists where the truth modality is greatest - the 'empirical inclusive' and 'empirical positive' - Faraday shows the lowest and highest scores for the first and last five-year period respectively.

Similar features emerge from Whewell's letters to Faraday (table 6.6). The 'empirical positive' modality demonstrates the highest (truncated) mean scores followed by the 'contingent negative' scores. This modality again contains the largest maximum values.

In Faraday's letters to Herschel the 'empirical positive' modality has the higher (truncated) mean scores (table 6.7). This time these scores for the 'contingent negative' modality are amongst the lowest. The largest maximum values are for the 'empirical positive' modality, though we do not see an increase in values over time. With Herschel's letters to Faraday (table 6.8) we see again that the 'empirical positive' modality displays the highest mean scores. Though this time the 'contingent negative' is the second most used modality in terms of mean scores. These two modalities have the largest maximum values for 1825-29 and 1845-49, with the 'contingent negative' values prevailing.

ii) Modality Results for Correspondence Between Faraday and Whewell and Herschel: 'Experimental' and 'Non-Experimental' Topics

For Faraday's letters to Herschel (table 6.9), when discussing experimental topics, it seems that the 'contingent negative' modality produces the highest mean score and maximum value. This is followed by the mean score for the 'empirical positive' modality. Both of these results are for 1845-49. The maximum values for this period for the 'empirical inclusive', empirical negative', and 'contingent negative' modalities are greater than for 1825-29.

In Herschel's letters to Faraday (table 6.10) we have almost a reversal of the results found Faraday's letters to Herschel, with the largest mean scores for the 'empirical positive' modality, followed by the 'contingent negative' results. Two further results to note are, first, that the maximum values for Herschel's letters to Faraday are generally higher than those in Faraday's letters. Secondly, for each modality the values are highest for 1825-29; with the largest values being for the 'empirical positive' and 'contingent negative' modalities.

Moving onto letters concerned with non-experimental matters, we see a feature of Faraday's letters noted in his letters discussing experimental topics. The 'contingent negative' and 'empirical positive' modalities have the two highest mean scores (table 6.11). Though this is for 1825-29, as opposed to 1845-49 in the case of the 'experimental' letters. We see that in both sets of Herschel's letters to Faraday, the 'empirical positive' modality has higher mean scores but not maximum values than the 'contingent negative' modality (table 6.12). Virtually the opposite result from Faraday's letters to Herschel.

The mean scores for the 'empirical positive' modality are higher than for the 'contingent negative' in Faraday's letters to Whewell concerned with experimental topics (table 6.13). The largest maximum values are for the 1845-49 periods. When Whewell writes to Faraday on experimental matters during 1845-49, again, the largest mean scores are for the 'contingent negative' and the 'empirical positive' modalities (table 6.14).

In his letters to Whewell, discussing non-experimental issues (table 6.15), Faraday, as with the letters to Herschel dealing with the same genre of topic, is using more of the 'contingent negative' than 'empirical positive' modality. We also find that the maximum values are generally higher for the earlier time period. Again this is consistent with Faradays letters to Herschel. Whewell's letters to Faraday, relating to non-experimental matters (table 6.16), like Herschel's letters to Faraday, show higher mean scores for the 'empirical positive' than the 'contingent negative' modality, with the latter producing larger maximum values.

iii) Compression Scan Results for Correspondence Between Faraday and Whewell and Herschel: Undifferentiated

I shall now move onto analyzing the results for the compression scans. In Faraday's letters to Herschel (table 6.17) the mean scores for 1840-44 and 1845-49 are higher than for 1825-29, and the maximal value occurs during 1845-49. Herschel demonstrates the highest mean compression for 1830-34 in his letters to Faraday. While the largest maximum value is found, marginally, during 1845-49. The letters from Faraday to Whewell (table 6.18) present the greatest mean compression for 1835-39, and the most maximal value for 1845-49. With Whewell's letters to Faraday the highest mean compression score and maximum value are for 1830-34.

iv) Compression Results for Correspondence Between Faraday and Whewell and Herschel: 'Experimental' and 'Non-Experimental' Topics

What we find is that in the case of Faraday's letters to Herschel and Whewell the mean compression is greatest for 1845-49, for both experimental topics and non-experimental topics (table 6.19). The minimum and maximum values are highest for both topics and the latter time period, except for the 1825-29 letters dealing with non-experimental matters, with Faraday's letters to Herschel. Whereas Faraday's letters to Whewell demonstrate more compression for the earlier time period for both topic genres. The letters for experimental concerns have higher minimum and maximum values. Herschel's letters to Faraday, dealing with experimental issues, almost show the most mean compression and the highest maximum values (table 6.20). The mean compression is greater for the earlier time periods for both types of topic. With Whewell's letters to Faraday the mean compression, and the minimum and maximum values are highest for the non-experimental letters for 1830-34.

v) Modality Results for Correspondence Between Faraday and the Non-Scientists

We now look at the numerical results for Faraday's correspondence with the non-scientists. First, Faraday's correspondence with Phillipps, where the topics covered where primarily non-experimental. The mean scores for Faraday's letters to Phillipps (table 6.21) are higher for the 'empirical positive' than the 'contingent negative' modality, whose maximum scores are generally the largest of these modalities. In both cases the maximum value, more or less, increases over time. In the case of Phillips' letters to Faraday (table 6.22) the lowest and the highest mean scores are for the 'empirical positive' and 'contingent negative' modalities respectively during 1845-49.

The matters discussed in the correspondence between Faraday and Herbert are experimental, and we see that the highest mean scores in Faraday's letters are for 1845-49 for all the modalities (table 6.23). Though the most maximal score tends to be for the earlier of the two periods, 1840-44. Herbert's letters to Faraday also deal primarily with experimental topics, and what is unusual about the results, is that the 'contingent negative' modality does not appear for 1840-44, though does show the most maximal score for 1845-49 (table 6.24).

Faraday's letters to Sarah dealt with personal matters, and a few were love letters (table 6.25). The 'empirical positive' modality has the largest mean scores, with the highest maximum values being for 1835-39 for the 'empirical positive' and 'contingent negative' modalities.

Faraday's letters to Magrath mainly deal with non-experimental topics and we once again find that the mean scores and maximum values are highest for the 'empirical positive' and 'contingent negative' modalities (table 6.26). The largest maximum value being for 1825-29 for both these modalities. The lowest mean compression for the 'empirical positive' modality and the highest compression for the 'contingent negative' modality are during 1845-49. In Magrath's correspondence with Faraday the largest maximum value is for the 'empirical positive' modality during 1845-49 (table 6.27).

Faraday's letters to Barlow deal mainly with non-experimental issues (table 6.28), where the 'empirical positive' and 'contingent negative' modalities produce the largest mean compressions for 1845-49. Also these two modalities have the most maximal values, again during 1845-49. There were only letters from Barlow to Faraday available for 1845-49 (table 6.29). These results reveal that the highest mean compression and maximum value were for the 'empirical positive' modality.

vi) Compression Scan Results for Correspondence Between Faraday and the Non-Scientists

We now return to consider results for compression scans time for Faraday's correspondence with the non-scientists. In Faraday's letters to Magrath the mean compression is highest for 1830-34, and during 1840-44 we find the maximum value. These two results are largest for 1835-39 when Magrath writes to Faraday. The correspondence between Faraday and Magrath discusses non-experimental issues, except for Magrath's letters to Faraday for 1830-34 (table 6.30).

The mean compression is highest for 1840-44 and 1845-49 in Faraday's letters to Barlow and his letters to Faraday respectively. Faraday has the highest maximal value for his 1845-49 letters (table 6.31). The mean compression is at its highest for both sets of the correspondence between Faraday's and Herbert for 1845-49, as is the maximum score (table 6.32). With Faraday's letters to Sarah the mean compression is highest for 1835-39, and the maximal value for 1819-24 (table 6.33).

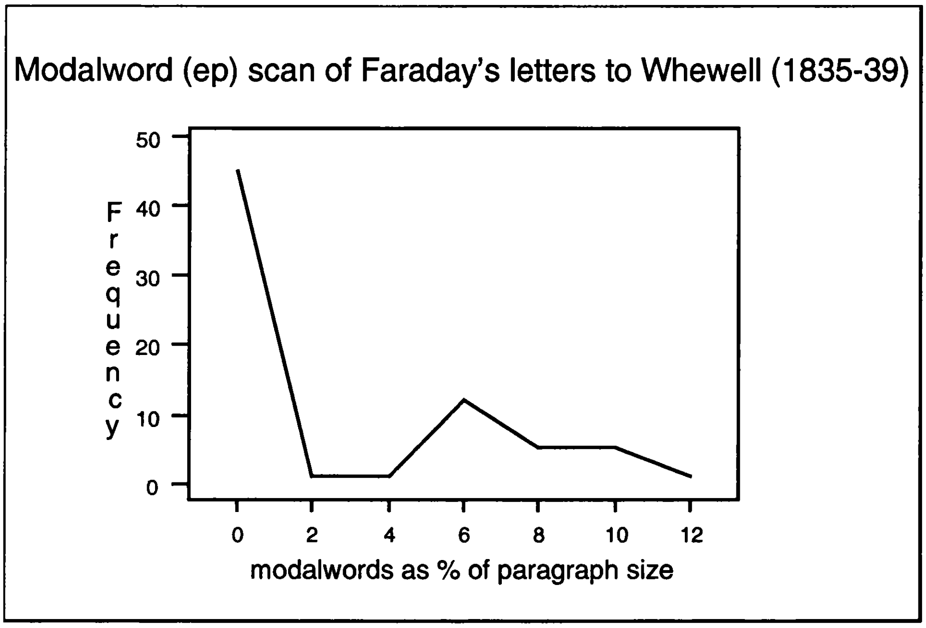
A familiar theme is played out when we consider Faraday's letters to Phillips. The mean compression and maximum score are largest for 1845-49. The three periods covering Phillipps' letters to Faraday (1835-39 to 1845-49) exhibit similar scores for both mean compression and the maximum value (table 6.34).

6.5.3 Histograms

i) Modality Histograms For The Correspondence Between Faraday and Herschel and Whewell: Undifferentiated

The histograms produced by Faraday's correspondence with Herschel and Whewell for each modality show little variation in shape. These histograms all had similar shapes to that seen in histogram 6.5[437] (70 paragraphs).

Histogram 6.5



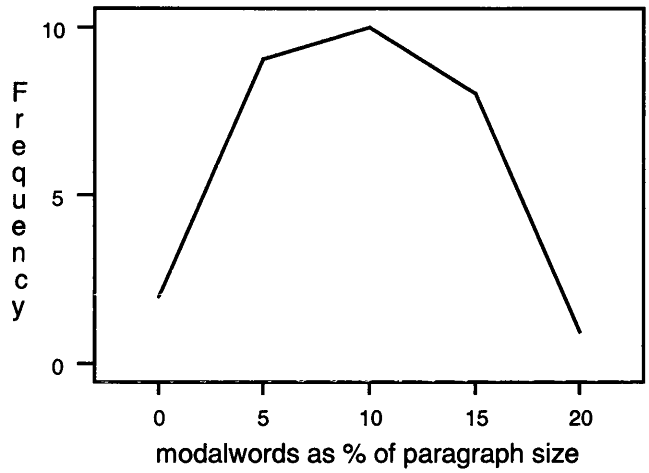
ii) Modality Histograms For The Correspondence Between Faraday and Herschel and Whewell: 'Experimental' and 'Non-Experimental'

With Faraday's letters to Whewell concerning 'experimental' topics a particular shape is found for the early (1825-1829) and late (1845-1849) periods for the 'empirical positive' modality (see histogram 6.13.1; 30 paragraphs). Though the early and late time periods for the 'contingent negative' modality demonstrate quite different shapes for the same group of letters (see histogram 6.13.2; 30 paragraphs).

[437] Each histogram corresponds to a table in appendix 5. For example, histogram 5.5 belongs to data table 6.5, and histogram 6.9.**2** means that it is the **second** histogram to correspond with data table 6.9.

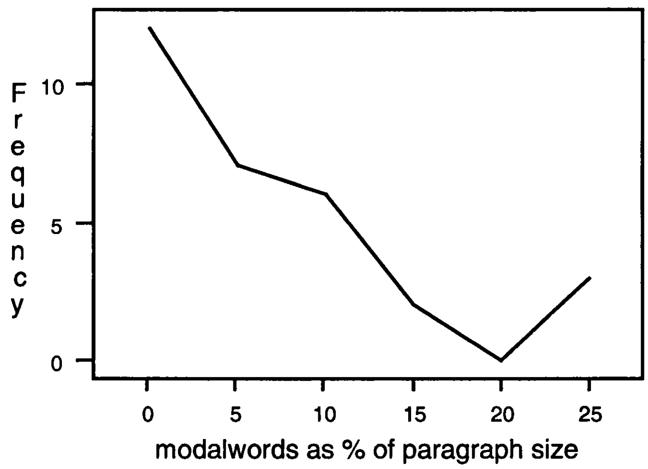
Histogram 6.13.1

Modalword (ep) scan of Faraday's letters to Whewell (1845-49)
Experimental topics.



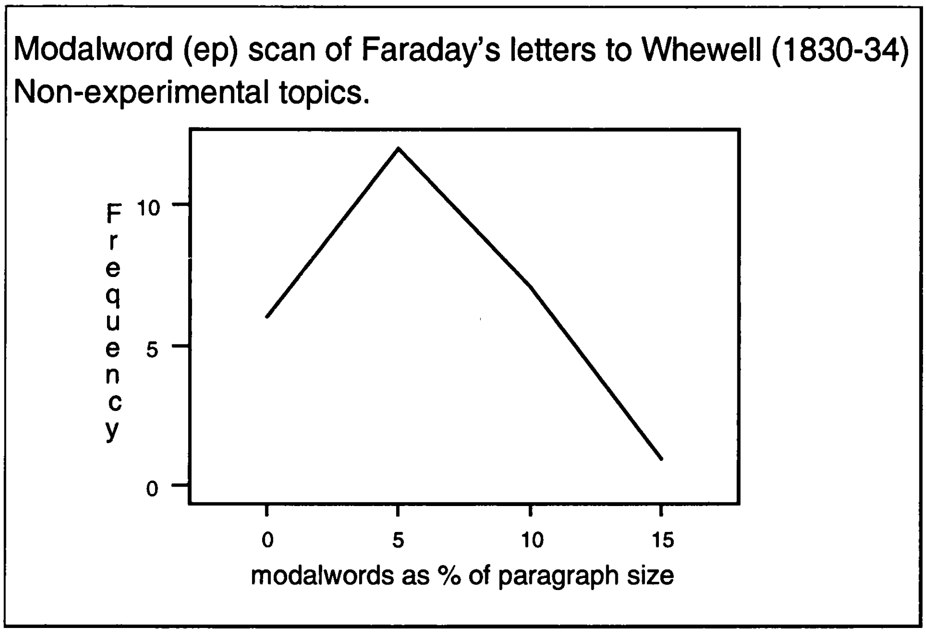
Histogram 6.13.2

Modalword (cn) scan of Faraday's letters to Whewell (1845-49)
Experimental topics.

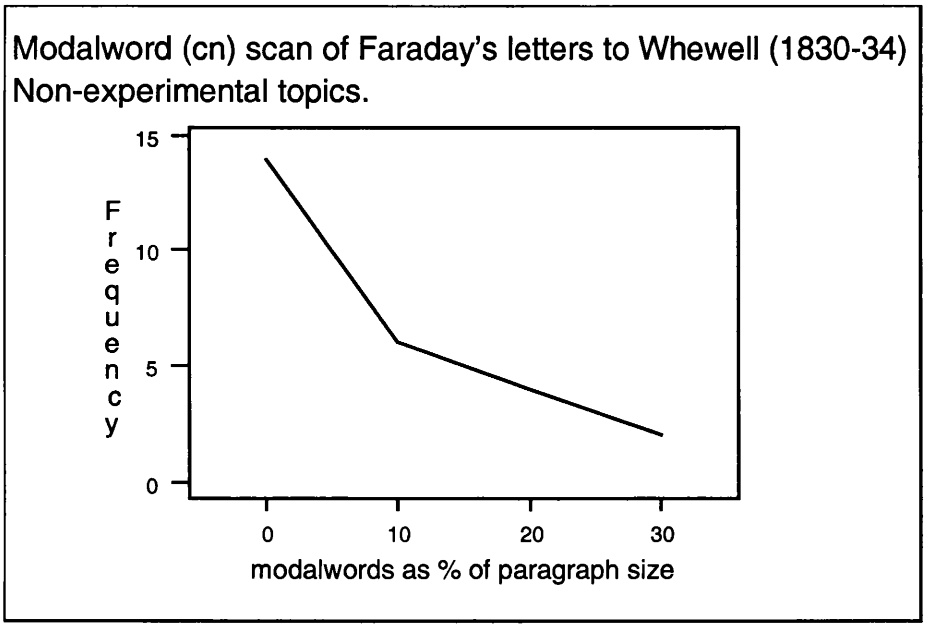


When we consider the letters written by Whewell to Faraday discussing 'experimental' issues the histograms for the earlier and later periods, which was one of the discriminations I wanted to make, these letters are quite different for the 'contingent negative' modality. The letters written by Faraday to Whewell, discussing 'non-experimental' matters, tend to display a similar histogram shape over time for the three 'empirical' modalities. An example of this shape is histogram 6.15.1 (26 paragraphs). The other modality producing significant results - 'contingent negative' - tended to have a histogram shape that was quite different from this group of letters (see histogram 6.15.2; 26 paragraphs).

Histogram 6.15.1



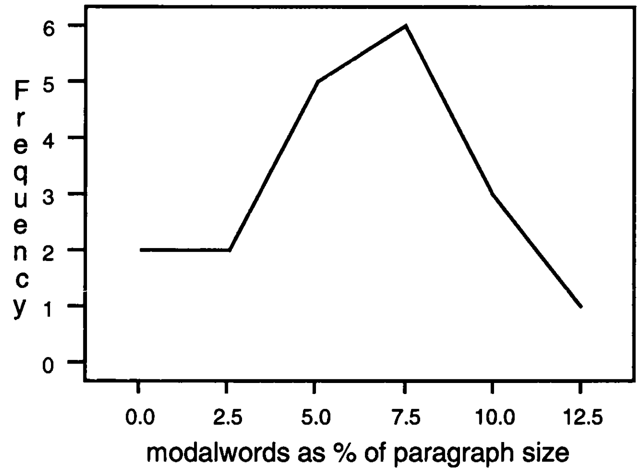
Histogram 6.15.2



Whewell's letters to Faraday for the same topics generally produced, over time, very similar histograms for the 'empirical inclusive', 'empirical negative', and 'contingent negative' modalities to those produced by Faraday's letters (see histograms 6.16.1 and 6.16.2; 19 paragraphs each).

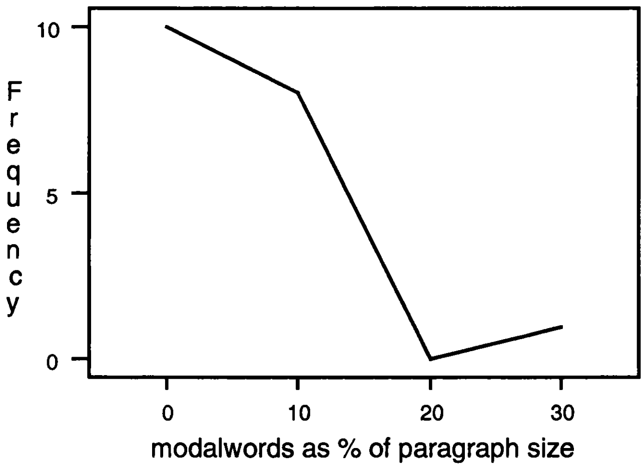
Histogram 6.16.1

Modalword (ep) scan of Whewell's letters to Faraday (1830-34)
Non-experimental topics.



Histogram 6.16.2

Modalword (cn) scan of Whewell's letters to Faraday (1830-34)
Non-experimental topics.

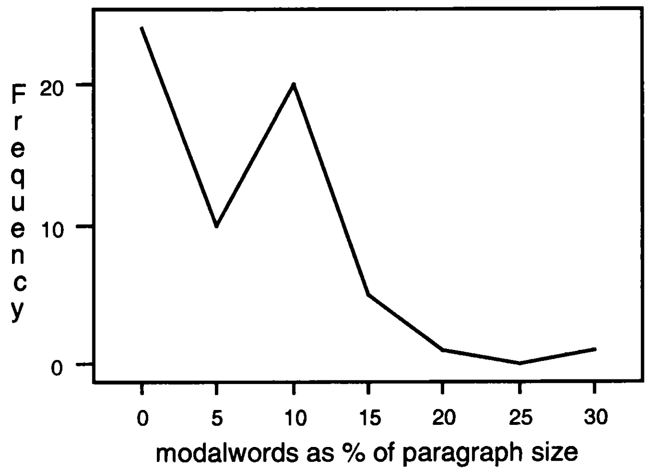


One interesting finding about the histogram shapes for the 'empirical positive' and 'contingent negative' modalities, when Faraday is discussing 'experimental' matters in his letters to Herschel, is that these shapes are quite the opposite of those for the 'non-experimental' letters. What could this tell us about the content or meaning? I think we have a case for having made a discrimination between 'experimental' and 'non-experimental' letters. Faraday is using different degrees of these modalities as part of his linguistic construction of certainty and meaning, as well as how persuasive he is being. Meaning within a text with a high degree of 'empirical positive' modality is constructed to be far less open to interpretation, and thus perceived as neutral, authoritative, and factual.

The histograms produced for the 'empirical positive' modality by Herschel's 'experimental' letters to Faraday for 1825-29 and 1845-49 are quite different (see histograms 6.10.1 and 6.10.2; 61 and 14 paragraphs respectively).

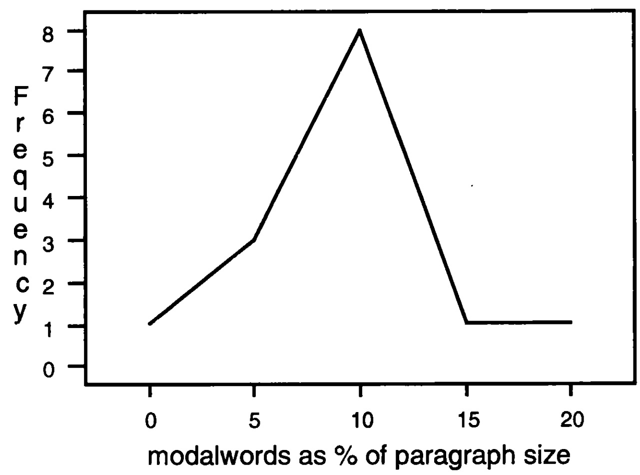
Histogram 6.10.1

Modalword (ep) scan of Herschel's letters to Faraday (1825-29) Experimental topics.



Histogram 6.10.2

Modalword (ep) scan of Herschel's letters to Faraday (1845-49) Experimental topics.

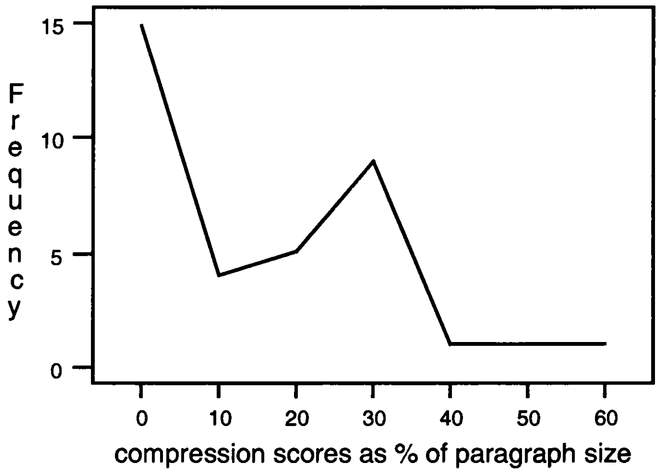


iii) Compression Histograms for Faraday's Correspondence With Herschel and Whewell: Undifferentiated

Three of Faraday's histograms for compression scores for his letters with Herschel are not dissimilar (for example histogram 6.17; 36 paragraphs). One histogram for the Faraday to Whewell correspondence stands out. This is for the period 1835 - 1839, where we see a peak at approximately twenty percent compression, which exceeds the frequency of zero compression scores (see histogram 6.18; 35 paragraphs).

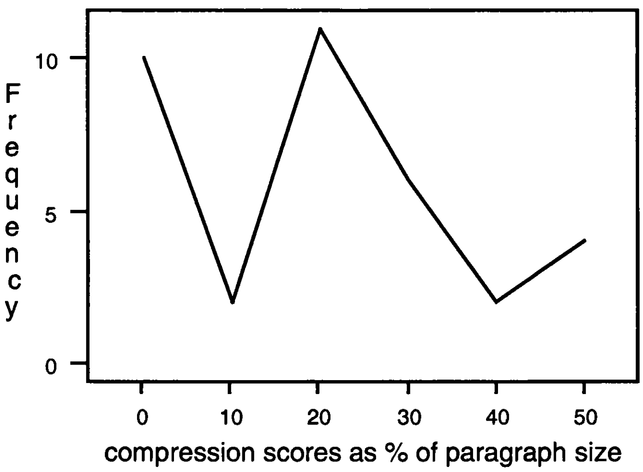
Histogram 6.17

Compression scan for Faraday's letters to Herschel (1845-1849)



Histogram 6.18

Compression scan for Faraday's letters to Whewell (1835-39)

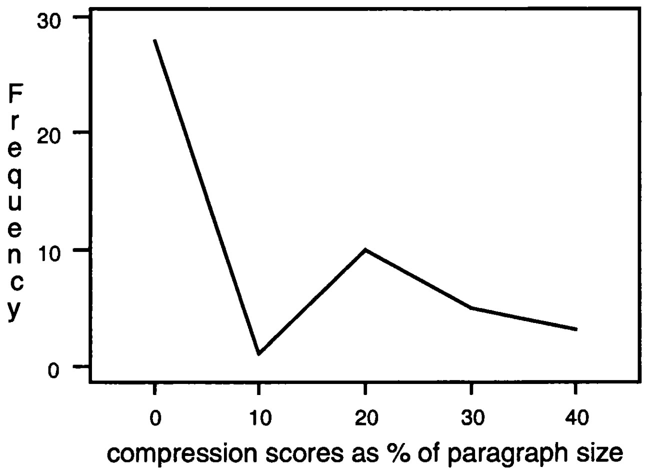


iv) Compression Scans for Faraday's Correspondence With Herschel and Whewell: 'Experimental' and 'Non-Experimental' Topics

The histogram shapes for the 'experimental' and 'non-experimental' letters for Faraday's 1825-1829 correspondence with Herschel are quite different (see histograms 6.19.1 and 6.19.2; 47 and 11 paragraphs respectively).

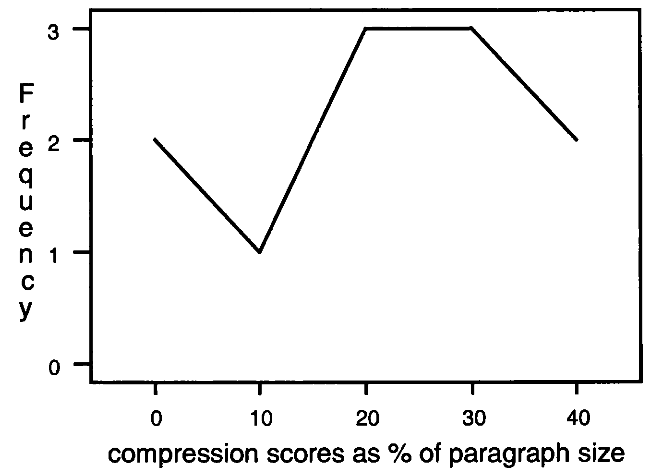
Histogram 6.19.1

Compression scan for Faraday's letters to Herschel (1825-29)
Experimental topics



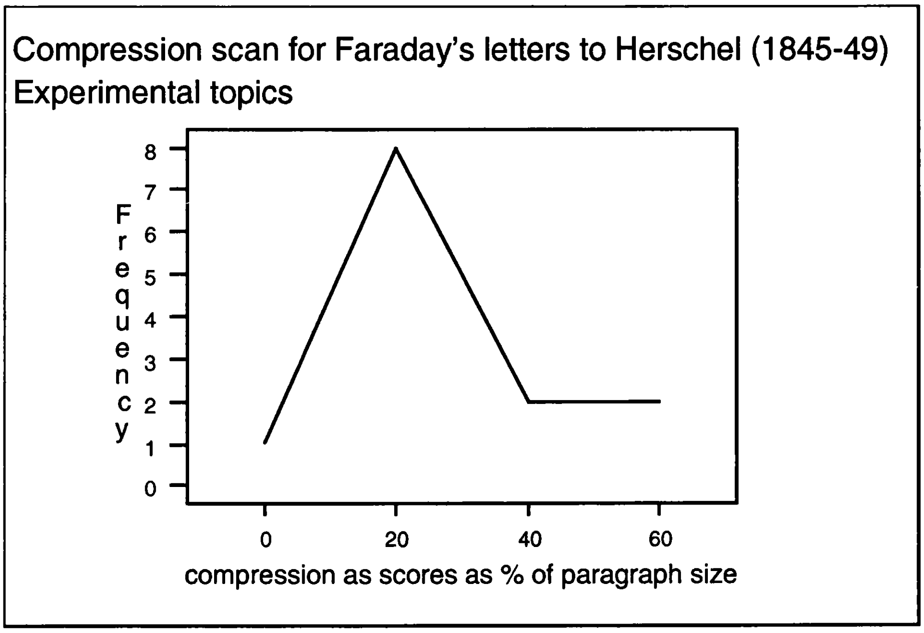
Histogram 6.19.2

Compression scan for Faraday's letters to Herschel (1825-29)
Non-experimental topics

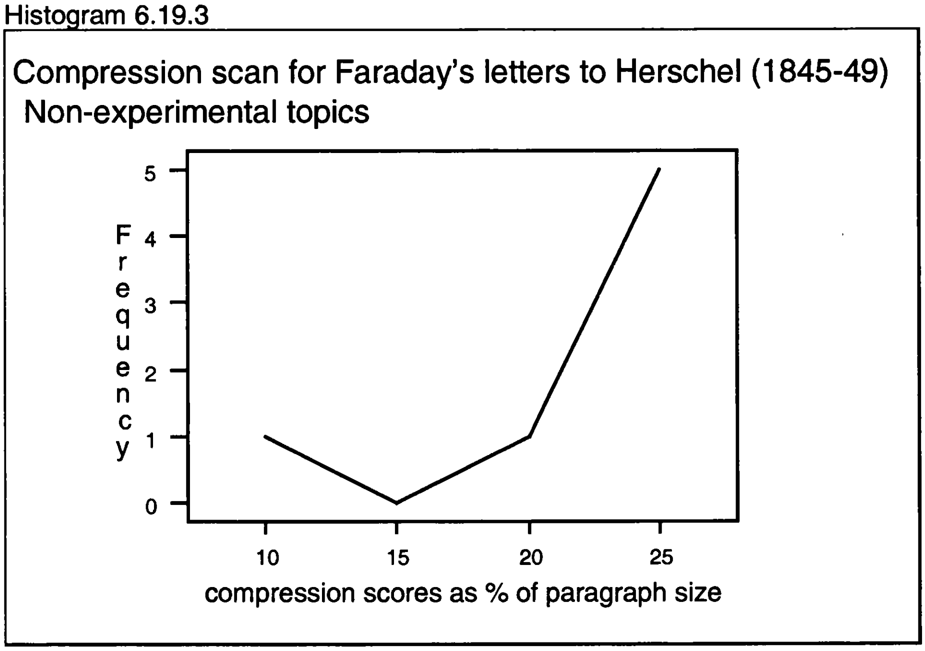


For the 1845-1849 period the 'experimental' and 'non-experimental' correspondence also produce quite different histogram shapes (see histograms 6.19.3 and 6.19.4; 13 and 7 paragraphs respectively).

Histogram 6.19.3



Histogram 6.19.3

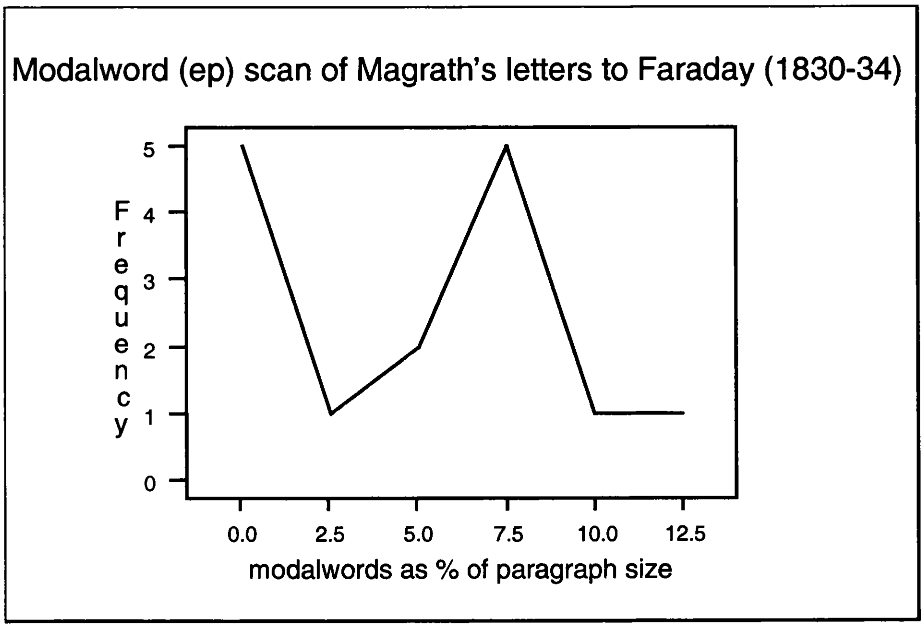


This difference in early and later histogram shapes continues for Faraday's 'experimental' and 'non-experimental' letters to Whewell. Also the 'experimental' letters of Faraday's early correspondence with Whewell have a very similar shape to the 'non-experimental' paragraphs for Faraday's 1845 to 1849 correspondence with Herschel. When we examine the histograms for Herschel and Whewell's letters to Faraday we do not see these differences for the early and later time periods.

v) Modality Histograms For Faraday's Correspondence With The Non-Scientists

The histograms for the letters to and from Faraday and Phillips show little in the way of notable differences in shape. There is little to note in terms of histograms for Faraday and Herbert's correspondence. The histograms for Faraday's letters to Sarah are also generally unremarkable, with a majority of scores of low frequency lying along the x-axis. This is generally the case for the letters between Faraday and Magrath, and Barlow, except for Magrath's letters to Faraday for 'empirical positive' modalities for 1830-34, where there is a sharp peak near the end of the graph (see histogram 6.26; 15 paragraphs each). This lends further support to the 'empirical positive' modality as a relatively sensitive indicator.

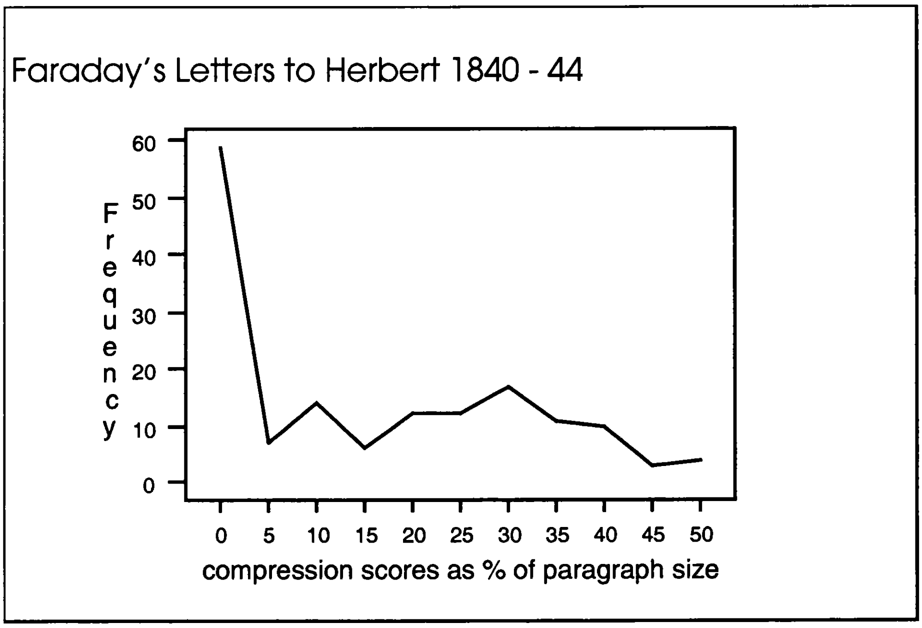
Histogram 6.26



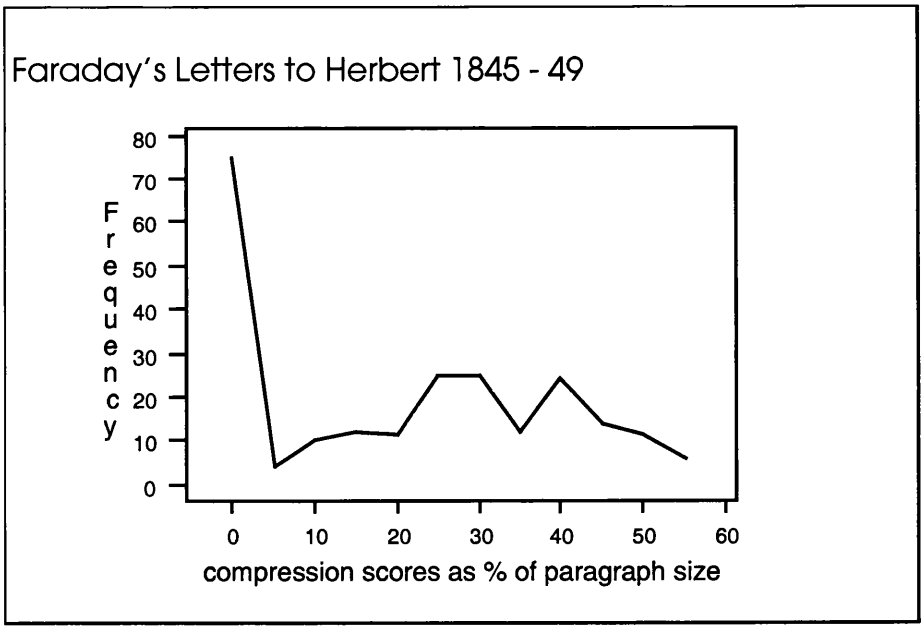
vi) Compression Histograms For Faraday's Correspondence With The Non-Scientists

If we first consider the histograms for the compression scans of the letters between Faraday and Phillips what is striking is the relative consistency, that is, the majority of these histograms demonstrate very abrupt and angular shapes. A situation of reasonable similar histogram shapes is also found for the correspondence between Faraday and Herbert with the very undulating pattern seen in histograms 6.32.1 and 6.32.2 (155 and 229 paragraphs respectively).

Histogram 6.32.1

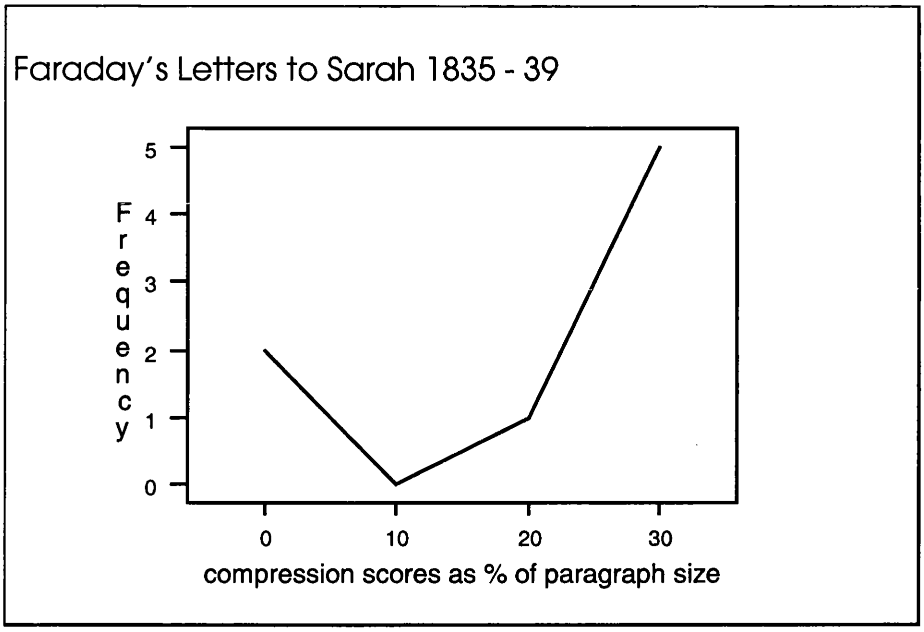


Histogram 6.32.2



The histograms for Faraday's letters to Sarah for 1835-39 and 1845-49 share the same shape (see histogram 6.33; 13 paragraphs). This shape is quite the opposite of the histogram for Faraday's letters to Sarah of 1819-24.

Histogram 6.33



The indicators produced histograms for Faraday's correspondence with Magrath and Barlow that were unremarkable, thus any discriminations were not possible.

6.6 Interpretation of Findings

I shall examine the findings from my analysis of modality and compression for the correspondence between the scientists first, followed by the histograms produced. Then a similar path will be followed with the results from Faraday's correspondence with the non-scientists.

For all of Faraday's correspondence between Herschel and Whewell, where the letters were not subdivided, modality showed the highest mean scores. This is what I had expected because the 'empirical positive' modality represents the greatest commitment to truth statements and certainty, thus language would be especially didactic and fact-ascertaining. This modality relates to Gilbert and Mulkay's empirical repertoire.

What was unexpected was the predominance of the 'contingent negative' modality, This most 'uncertain' and 'informal' of modalities had the second highest mean scores, apart from Faraday's letters to Herschel, where the means for this modality were among the lowest. This set of letters produced the most maximal value for the 'empirical positive' modality, but at all other times this achievement went to the 'contingent negative' modality. It was only in Faraday's letters to Whewell did we see the 'empirical positive' modality increase over time, which was not my initial expectation.

When we consider the results for 'experimental' and 'non-experimental' letters the dominance of the 'empirical positive' and 'contingent negative' modalities is striking. Faraday's letters for the two genres to Herschel and his 'non-experimental' letters to Whewell all exhibit this trait. The largest mean scores and maximum values in Faraday's 'experimental' letters to both scientists occur during 1845-49. This is generally not the case with his 'non-experimental' letters where the highest scores reside in the earlier years of Faraday's career. Interestingly the maximum values for the 'empirical inclusive', 'empirical positive' and 'contingent negative' modalities when Faraday is writing to Herschel on 'experimental' concerns are highest for 1845-49. This genre of letters to Faraday from Herschel show their highest mean scores across the modalities for 1825-29. I would tentatively suggest that one explanation for this pattern is that because Faraday is a junior or unknown scientific quantity in 1825-29 his lowly status is being constructed and displayed in the scientific pecking order through his writing and use of truth modality. That is, only an established scientist of some reputation can construct and exhibit certainty and statements of truth and fact in their writings. Though this may also reflect the cautious and conservative side to Faraday's nature, from the 1840's his investigations were cloaked no longer in the language of orthodox theory, and he went on the offensive and used persuasive, heretical arguments. This was after 1834 when Faraday was accused of plagiarism for a second time (the first time was 1821), which is thought to have contributed to his development of a new post-inductive style of scientific writing.[438]

[438] Agassi, Joseph; 1971.

If we now consider the results of the compression scans for the undifferentiated letters, the mean score and maximum value when Faraday writes to Herschel are highest for 1845-49. Also his letters to Whewell for these results are highest from 1835. When we examine Faraday's 'experimental' and 'non-experimental' letters to Herschel and Whewell the mean compression is highest in both genres, as is the minimum and maximum values in the Herschel letters, for 1845-49. Also these values are maximal in Faraday's 'experimental' letters to Whewell. This is a somewhat more complex picture than painted by my expectations. Faraday seems to use not a little compression in his letters of 1845-49 to both scientists.

Faraday's letters to Phillips, which discuss the deleterious effects of the polluted London air on the Raphael cartoons upon their removal from Hampton Court and matters arising from the Coroner's inquest into the explosion at the Haswell Collieries, show most use of the 'empirical positive' modality. Though the maximal value is presented by the 'contingent negative' modality. All the modalities have their largest mean score for 1845-49 when Faraday writes to Herbert on 'experimental' matters. Whereas the 'empirical positive' modality is most in evidence in Faraday's letters to Sarah. This modality has the largest mean and maximum value in the letters written to Magrath by Faraday, again 'non-experimental' topics are primarily discussed. A scenario repeated in the correspondence between Faraday and Barlow. With the letters between Faraday's and the scientists it is the 'contingent modality' that predominates, but in the correspondence to and from Faraday and the non-scientists it is the 'empirical modality' that we see most of. It was found that Faraday's correspondence with the non-scientists, except for the letters to Magrath and Sarah, show the highest mean compression score and maximum value for 1845-49.

I shall now move onto looking at the histograms, of which we saw little variation in shape for the undifferentiated letters between Faraday and Herschel and Whewell. More interesting results arose from Faraday's letters to Whewell involving 'experimental' topics, where the histograms for the 'empirical positive' modality have a similar shape over time, whereas those for the 'contingent modality' differ between the early and later time periods. This result is repeated with Whewell's letters to Faraday. When Faraday writes to Whewell on 'non-experimental' matters the 'empirical' modalities produce similar histograms. It seems as if as the 'empirical modality' is not only central to Faraday's writing but its use is quite consistent over the years, something not really seen with the 'contingent modality'. More evidence for this is that the histograms over time for the 'empirical positive' modality for Faraday's letters to Herschel, discussing 'non-experimental' issues, are similar to those for his letters to Whewell. The compression histograms for the 'experimental' letters of Faraday's early correspondence with Whewell have a very similar shape to the 'non-experimental' letters from Faraday to Herschel for 1845-1849. Also the histogram patterns for the 'experimental' and 'non-experimental' letters for Faraday's 1825-1829 correspondence with Herschel are similar. Furthermore, with Faraday's correspondence with Herbert the only histograms which are notable are those produced for the 'empirical positive' modality. These letters involve 'experimental' topics.

6.7 Are there Any Discernible Patterns or Trends Over Time?

One aim of this study was to try and discern any patterns or trends over time. With this in mind, I have, first, summarized the (truncated) mean scores for what seems the most sensitive indicators, that is, compression and the modalities: 'empirical positive' and 'contingent negative' in tables 6.5 to 6.7. These tables give the scores for Faraday's letters to scientists and non-scientists for each applicable five-year period.

What we observe is that the mean compression scores for all three indicators are highest for 1845-49, except in the case of Whewell for 1840-44. A very similar picture is found for Faraday's letters to the non-scientists. I believe this to be a reflection of Faraday's status as a ground-breaking scientist, granted much respect and reverence. The 1830's were his most productive decade in terms of research, and his research position became increasingly independent due to rising income. Though the scenario is not as transparent as this. The 1840's was a decade characterized by a lack of scientific activity relative to the

1830's, which can primarily be accounted for by a combination of Faraday's Eldership in the Sandemanian Church and his increasing illness. Another important point is that through the 1840's (and 1850's) Faraday ceased to cloak his investigations in the language of orthodox theory. Instead he went on the offensive and used more robustly persuasive arguments. I think the latter point is most telling of Faraday's predominant use of 'empirical positive' modality through the 1840's.

Table 6.5 Summary of Mean Compression Scores of Faraday's Letters Over Time.

	1819-24	1825-29	1830-34	1835-39	1840-44	1845-49
Herschel		8.34	17.14	8.95	13.75	13.37
Whewell			13.18	19.58	8.34	16.42
Sarah	13.62			18.53		18.22
Magrath		11.44	12.24	6.32	9.21	6.90
Phillips				15.81	4.66	24.81
Barlow				5.47	7.09	11.55
Herbert					14.76	20.26

Table 6.6 Summary of Mean 'Empirical Positive' Modality Scores of Faraday's Letters Over Time.

	1819-24	1825-29	1830-34	1835-39	1840-44	1845-49
Herschel		0.89	0.46	1.50	1.27	1.69
Whewell			1.70	2.10	1.55	2.64
Sarah	2.10			2.77		2.60
Magrath		2.26	2.51	1.93	2.49	1.65
Phillips				4.07	1.48	4.67
Barlow				1.38	1.14	1.91
Herbert					3.34	4.75

Table 6.7 Summary of Mean 'Contingent Negative' Modality Scores Over Time.

	1819-24	1825-29	1830-34	1835-39	1840-44	1845-49
Herschel		0.00	0.00	0.00	0.00	0.03
Whewell			0.95	1.40	0.61	0.94
Sarah	0.96			2.64		0.51
Magrath		0.00		0.00	0.00	0.26
Phillips				3.13	0.15	2.24
Barlow				2.30	0.60	1.71
Herbert					0.80	1.66

6.8 What Became of My Expected Results and Predictions?

In Faraday's letters to scientists it turned out not to be the case that the correspondence discussing 'experimental' topics consistently had the highest scores for the 'empirical positive' modality (table 6.8). There is not a very large pool of figures available to compare 'experimental' letters with 'non-experimental' letters for the non-scientists. It is only in the letters between Herbert and Faraday that the primary topic is 'experimental'. Though we see that the mean 'empirical positive' modality is the highest for all the non-scientists (table 6.6). Additionally, the 'empirical positive' modality was found to be highest for the 1845-49 period, for the scientists and the non-scientists, except for Faraday's letters to Magrath (table 6.6). Faraday's letters to Herbert ('experimental') produced some of the highest mean compression scores, going against my expectation, that letters dealing with experimental concerns would have less compression (section 4.4.1). Lastly, it transpired that, generally, the mean compression was highest during the 1845-49 periods for Faraday's letter to the non-scientists (table 6.9). Then with Faraday's correspondence to Herschel compression is greater for 1845-49, but 1825-29 in the case of Whewell (table 6.8).

The compression is higher in the 'experimental' letters than the 'non-experimental' letters for both Herschel and Whewell, except for the 1825-1829 letters to Whewell (table 6.8). This goes against my prediction in table 6.3, which was based on the assumption that less concern over the use of conjunctives and articles in more informal writings would translate as higher compression. Compression is also higher in the later letters to Herschel but not in the case of Whewell (table 6.9). Faraday was far more of an established and renowned scientist during 1844 to 1849 than during the earlier period. For example, in 1845 Faraday 'discovered' magneto-optical effect and diamagnetism, this lead him to formulate the field theory of electromagnetism.

To conclude, Faraday's writings on matters 'experimental' would have been robustly truth-asserting and unequivocal in tone, thus a higher usage of 'empirical positive' modality expected. At least this is what was predicted, but generally this turns out not be the case. Also there is not a pattern of higher 'contingent negative' modality in the 'non-experimental' letters. The 'empirical positive' modality was not lower in the Faraday's letters to the non-scientists as I originally expected.

Table 6.8 Summary of Mean Compression Scores for Early and Late 'Experimental' and 'Non-Experimental'.

	Compression		EP Modality		CN Modality	
	1825-29	1845-49	1825-29	1845-49	1825-29	1845-49
Herschel (x)	9.33	27.31	7.00	0.02	1.08	9.17
Herschel (n-x)	21.76	22.41	8.72	5.58	9.25	7.72
Whewell (x) [439]	43.37	30.28	7.87	9.41	6.92	5.79
Whewell (n-x)[440]	23.64	21.40	5.67	0.00	6.73	3.71

[439] The first period for Faraday's 'experimental' letters to Whewell is 1835-1839.
[440] The first period for Faraday's 'non-experimental' letters to Whewell is 1830-1834.

Table 6.9. Summary of Mean Scores for 'Empirical Positive' and 'Contingent Negative' Modalities, and Compression For Faraday's Letters to the Non-Scientists.[441]

	1819-24	1825-29	1830-34	1835-39	1840-44	1845-49
Phillips						
EP				4.07	1.48	4.67
CN				3.13	0.15	2.24
Ellipsis				15.81	4.66	24.81
Herbert						
EP					3.34	4.75
CN					0.80	1.66
Ellipsis					14.76	20.26
Sarah						
EP	2.01			2.77		2.6
CN	0.97			2.64		0.51
Ellipsis	13.62				18.53	18.22
Magrath						
EP		2.26	2.51	1.93	2.49	1.65
CN		1.64	1.27	0.82	1.54	5.49
Ellipsis		11.44	12.24	6.32	9.21	6.90
Barlow						
EP				1.38	1.14	1.91
CN				2.30	0.60	1.71
Ellipsis				5.47	7.09	11.55

6.9 Conclusions and Wider Discussion

I conclude that I was able to make the discriminations I wanted and the features most able to indicate these discriminations were compression and 'empirical positive' modality. The 'contingent negative' modality was almost, but not quite, as effective as an indicator. The results seem to indicate that Faraday's thinking and writing style is very systematic, precise, and methodical. I do not think this is seen with such consistency for Herschel and Whewell from my analysis.

Faraday's writing has less compression when he is dealing with an 'experimental' matter. Especially with Faraday's letters to the scientists for later

[441] It is only the correspondence between Faraday and Herbert which contained topics that were primarily 'experimental'.

in his scientific career. This feature I would say is the consequence of a very economical and careful use of words, and thinking that is tuned to dealing systematically with one task. Such a way of thinking in and about life is consistent with Faraday's membership of the Sandemanian church. It is as if Faraday is trying to give very precise form to his thoughts in writing. We have seen how use of the most confident and truth-ascribing modality - the 'empirical positive' - is used quite consistently across 'experimental' and 'non-experimental' topics, and with scientists and non-scientists alike, and over time. Faraday was writing his notes in a diary style; "his observations are used as arguments".[442] I suspect that he was doing the same with his letters from the 1840's (or even from the late 1830's), when he shed the language of orthodox, inductivist theory, and started to use a more argumentative, pointed language. Scientific language at the time was rounded on for being obscure and inaccessible, and Faraday was definitely aiming for great clarity in language, in communication of thoughts and ideas. He was a pioneer in this respect.

We have seen the relative success of three features as indicators of the discriminations described in table 6.1, that is, compression and 'empirical positive' and 'contingent negative' modalities. The significance of the success of the two modalities I think lies in giving empirical substance to a remark by Helmoltz in 1881:

> "His [Faraday] principal aim was to express in his new conceptions only facts, with the least use of hypothetical substances or forces. This really was an advance in general scientific method, destined to purify science from the last remnants of metaphysics."[443]

This notion of 'purifying science' ties in neatly with what I see as a primary aim of Faraday's work, that is, to produce a 'language of facts'.[444] Through the considered and deliberate application of words Faraday was attempting an act of 'linguistic cleansing'. Thus hopefully:

> "By purifying ideas and words the link between experiment and language was complete and facts could be expressed and communicated without distortion by the linguistic medium."[445]

[442] Agassi, Joseph; 1971: 125.
[443] Helmholtz, Hermann von; 1881: 277.
[444] Cantor, Geoffery; 1991: 213.
[445] Cantor, Geoffery; 1991: 213.

The irony of this is that to try to 'purify' communication and language is to proactively engage with words which means contamination by ideology. Any use of language is ideological because words are not used arbitrarily, divorced for the world, but come with cultural baggage, value orientations, and presumptions. Thus far from 'conceiving to make ideas clear and distinct'[446] Faraday engaged in an imaginary and linguistic construction of reality, probably not a little aided and abetted by beliefs inherent to Sandemanianism:

> "To act as a true, moral Christian was his overriding concern, and therefore his views and behaviour were natural extensions of his [Faraday] Sandemanianism. To be a scientist did not threaten his religious persona; rather science provided a relatively safe area in which he could practice his Sandemanianism....[his] science was predicated on strong beliefs about what the physical world is like, how God constructed it, and how he was to understand nature and its laws"[447]

I think the results of this study bode well as part of a wider investigation of the sensitivity of word-frequency (frequency of percentage scores) based text analysis. I have conducted a comprehensive study of language use in Faraday's letters and found that three features (compression and 'empirical positive' and 'contingent negative' modalities) acted as sensitive indicators of the discriminations I wanted to make. I think this investigation and method are valuable contributions to the repertoire of text analysis methods and to the history of quantitative text analysis in science studies. I shall discuss further in my final chapter how this approach differs from 'traditional' quantitative text analysis and what this means for science studies.

Finally, I think this work highlights an important issue raised, inter alia, by Gooding, Gruber, Holmes and Shapin, and one that is central to the constructivist philosophy of language; that is, there is no simple extrapolation from thought to verbal communication. Numerous factors - temporal, biological, cultural, and economic, for example - are continuously mediating the relationship between word and the world, between mind and language. All we can do is fine-tune our tools and skills, including patience, to try and help us understand what was going on in Faraday's correspondence with Herschel and

[446] Faraday to Benjamin Abbott, July 12 1812.
[447] Cantor, Geoffery; 1991: 294.

Whewell by studying aspects of his language use. To say that I have tried to achieve more would be vainglorious.

In this chapter I have tested the sensitivity of features as indicators of discriminations I wanted to make. In the next chapter further testing and evaluation of the main results and discriminating features - compression, and ‘empirical positive’ and ‘contingent negative’ modalities, will be done.

7 Testing of the Main Findings of Chapter Six

Abstract. This chapter contains three small studies to test the main findings of chapter six which were: i) Faraday's strong and consistent use of the 'empirical positive' modality, ii) the discriminatory power of compression, and the modalities, 'empirical positive' and 'contingent negative'. This chapter continues the investigation of the sensitivity of word-frequency based text analysis, through an investigation of early letters from Faraday to friends, private and published text's of Darwin, and letters to Faraday from non-Sandemanians. It was found that the discriminatory power of compression and 'empirical positive' modality revealed in chapter six is played out in this chapter.

7.1 Introduction

Chapter six described a large-scale study of the use of compression and modality in Faraday's letters to scientists and non-scientists over a twenty-five year period. The primary finding was that Faraday demonstrates a quite consistent use of a 'strong' and confident truth modality ('empirical positive') across his letters to scientists and non-scientists. Also when the subject matter is non-experimental the compression scores are generally greater. It was also found that measurement of compression and 'empirical positive' modality' were most sensitive as indicators of discriminations I wanted to make. In this chapter I describe three smaller studies to test the main findings of the main case study in the previous chapter.

7.2 Questions

The questions this chapter asks are:

i) Does Faraday's predominant use of 'empirical positive' in his letters to scientists exist in early letters to friends?

ii) Does measuring compression and the 'contingent negative' and 'empirical positive' modalities allow discrimination between Faraday's letters to scientists and to his friends?

iii) Can the measures of compression and modality discriminate between the private and public writings of scientists? Here private and public texts of Faraday and Charles Darwin are compared.
iv) Can Faraday's robust truth-asserting writing style be attributed at all to him being a Sandemanian? For this the letters to Faraday from three non-Sandemanian scientists are analyzed.

7.3 Methodology

7.3.1 Studies and Source Material

The first study investigates Faraday's earliest available letters to three friends: Benjamin Abbott,[448] Richard Phillips,[449] and T. Huxtable.[450] This study set out to test whether Faraday's predominant use of the 'empirical positive' modality exists in his early letters to friends. As well as test the use of the 'empirical negative' modality and measurement of compression to discriminate between letters to friends and to scientists. The letters addressed a mixture of 'experimental' and 'non-experimental' subjects. The aim of the next study was to test further whether the measure of compression and modality can enable discrimination, this time, between the private and public writings of scientists. For this purpose I will compare writings of Faraday and Darwin, which comprise Faraday's *Diary* and the corresponding published paper for 1821,[451] which both deal with electromagnetism. For Darwin I have chosen his 'M' and 'N' notebooks, and chapters one and two of *The Voyages of the Beagle* and *On the Origins of the Species.* In the notebooks Darwin is struggling to give shape to the materialist philosophy of biology.[452]

The final study analyzes letters to Faraday from three non-Sandemanian scientists: Christian Schoenbein,[453] John Herschel, and William Whewell. I was interested in whether the predominant use of 'empirical positive' modality in

[448] A clerk in the City, but became a teacher in 1822.
[449] Chemist and curator of the Museum of Practical Geology.
[450] Medical student and member of the City Philosophical Society.
[451] Faraday, Michael; 1821: 74-96.
[452] Gruber E. Howard; 1981: xiv.
[453] Professor of Physics and Chemistry at the University of Basle.

Faraday's letters was present in letters to him from non-Sandemanians. This study compares the letters for 'experimental' and 'non-experimental' topics.[454]

The relationship between the features I use for the discriminations and the texts analyzed is summarized in table 7.1 on the next page.

[454] There is not enough letters from Faraday to Schoenbein for analysis.

Table 7.1 Relationship Between Distinctions, Features, and Source Texts.

	early letters to friends v later letters to scientists	private v public writings	letters from Faraday v from non-Sandemanians
compression	Faraday writing to B. Abbott, (1812-24), T. Huxtable (1811-23), and R. Phillips (1820-30) v to Herschel (1825-49) and Whewell (1830-49)	Faraday: diary and paper for 1821 on electro-magnetism. Darwin: 'M' & 'N' notebooks; chapters 1 & 2 of *Origin* and *Beagle*.	Faraday writing to Herschel (1825-49) and Whewell (1830-49) v Schoenbein (1837-47), Herschel, and Whewell writing to Faraday
'empirical positive' and 'contingent negative' modalities	Faraday writing to B. Abbott, (1812-24), T. Huxtable (1811-23), and R. Phillips (1820-30) v to Herschel (1825-49) and Whewell (1830-49)	Faraday: diary and paper for 1821 on electro-magnetism. Darwin: 'M' & 'N' notebooks; chapters 1 & 2 of Origin and Beagle.	Faraday writing to Herschel (1825-49) and Whewell (1830-49) v Schoenbein (1837-47), Herschel, and Whewell writing to Faraday

Table 7.2 Summary of the Descriptive Statistics for Faraday's Letters to Scientists and Friends.

	Dates	Mean (ep)	Min (ep)	Max (ep)	Mean (cn)	Min (cn)	Max (cn)	Mean (ellipsis)	Min (ellipsis)	Max (ellipsis)
Herschel	1825-1849	2.00	3.62	11.86	0.97	4.94	24.57	12.31	9.70	37.00
Whewell	1830-1849	1.16	2.53	9.43	0.01	2.58	16.90	14.38	8.80	47.70
R. Phillips	1820-1830	2.08	1.09	19.30	1.11	0.98	35.71	10.19	4.76	42.11
T. Huxtable	1811-1823	3.34	3.70	20.00	1.36	3.57	12.04	21.84	7.41	40.64
B. Abbott	1812-1824	2.40	1.19	30.00	1.31	0.34	35.71	24.45	3.33	61.19

7.4 Results

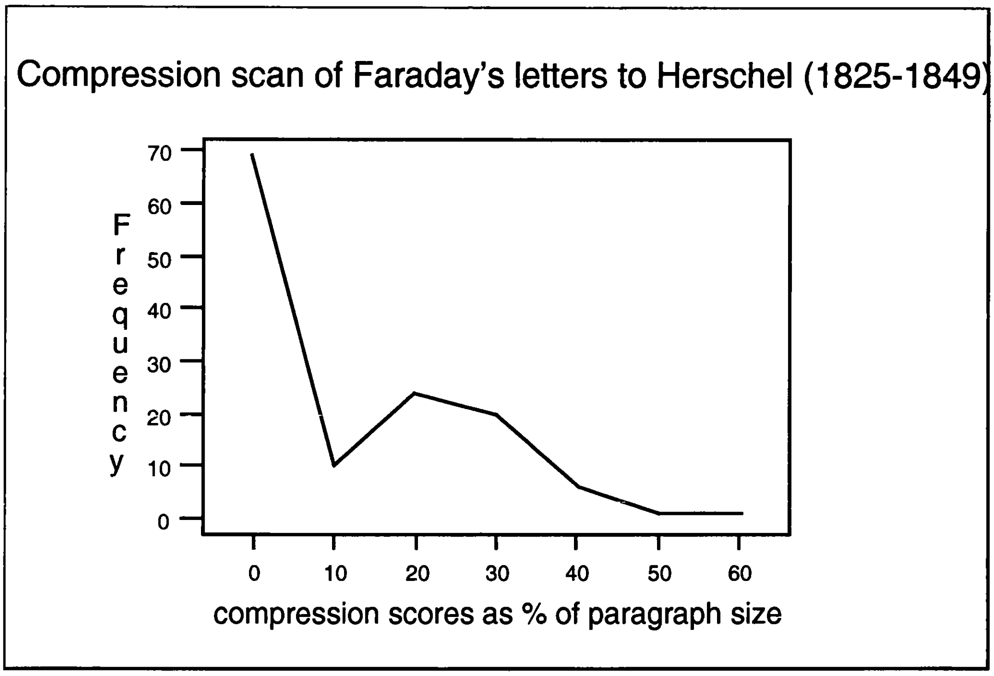
7.4.1 Study One: Faraday's Early Letters To Friends

This study provides a check on my findings for the modalword and compression scans for Faraday's letters to non-scientists. The results are found in table 7.2 on the previous page.

What we find is that the mean and maximum scores for both the measured modalities are higher for Faraday's early letters to friends. This goes against the grain of what I expected. In the later letters, a higher degree of 'empirical positive' modality could have been expected because Faraday would have been writing to other scientists in his capacity as an increasingly established and revered scientist. Thus such truth-asserting language would be more to the fore. The degree of compression is as high or higher in the letters to friends as well. This could be explained by the use of a more 'crafted', careful, and formal prose-style in his later letters to scientists.

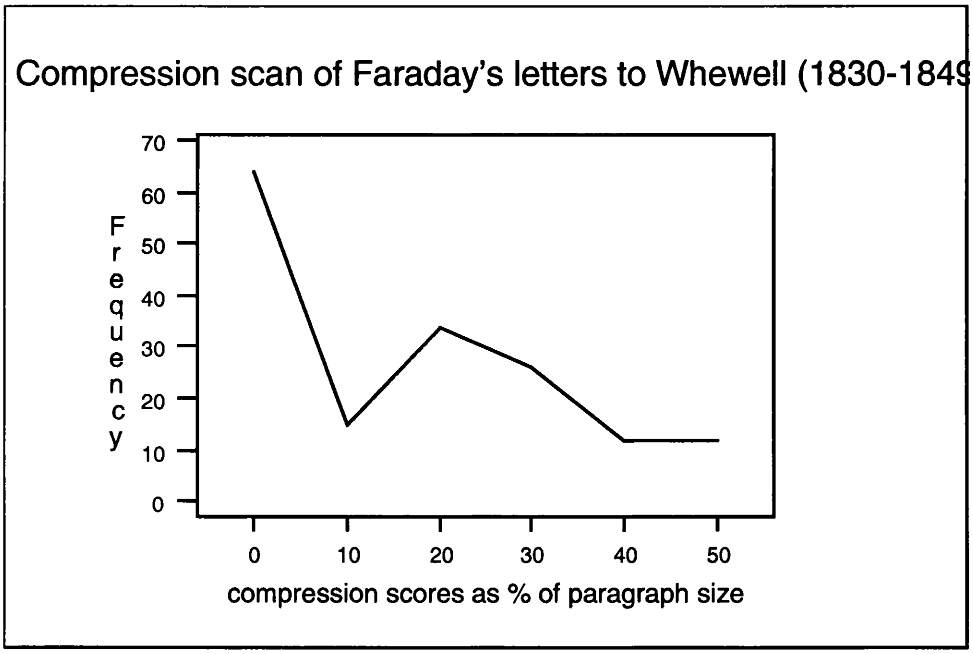
The modal word scans produce histogram shapes with little variation. The three histograms below are the typical shapes for Faraday's early letters to Abbott, Phillips, and Huxtable. The compression histograms for the letters to Herschel (131 paragraphs) and Whewell (163 paragraphs) are quite similar (histograms 7.1 and 7.2), and quite different for those of Abbott (483 paragraphs), Phillips (71 paragraphs), and Huxtable (17 paragraphs).

Histogram 7.1



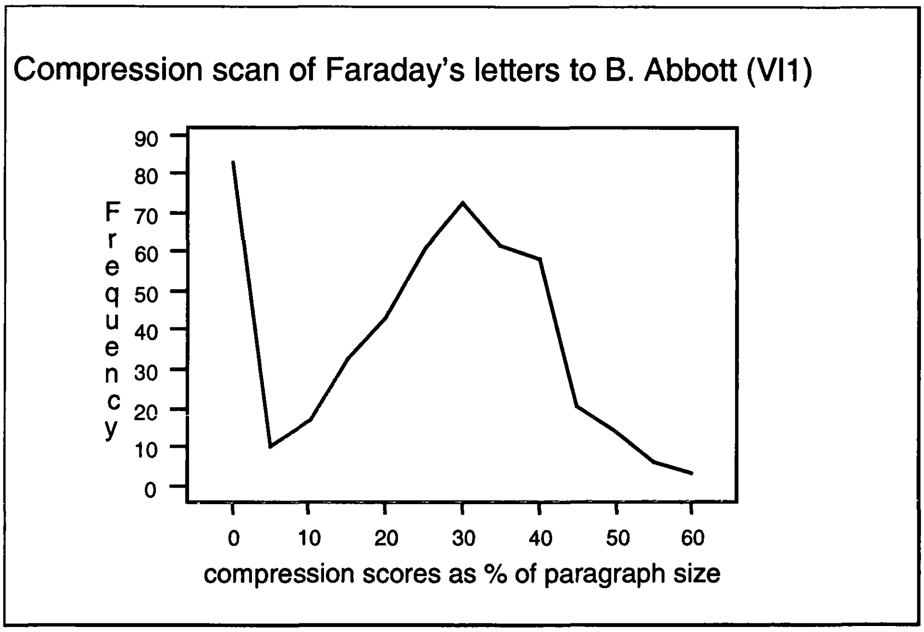
Compression scan of Faraday's letters to Herschel (1825-1849)

Histogram 7.2

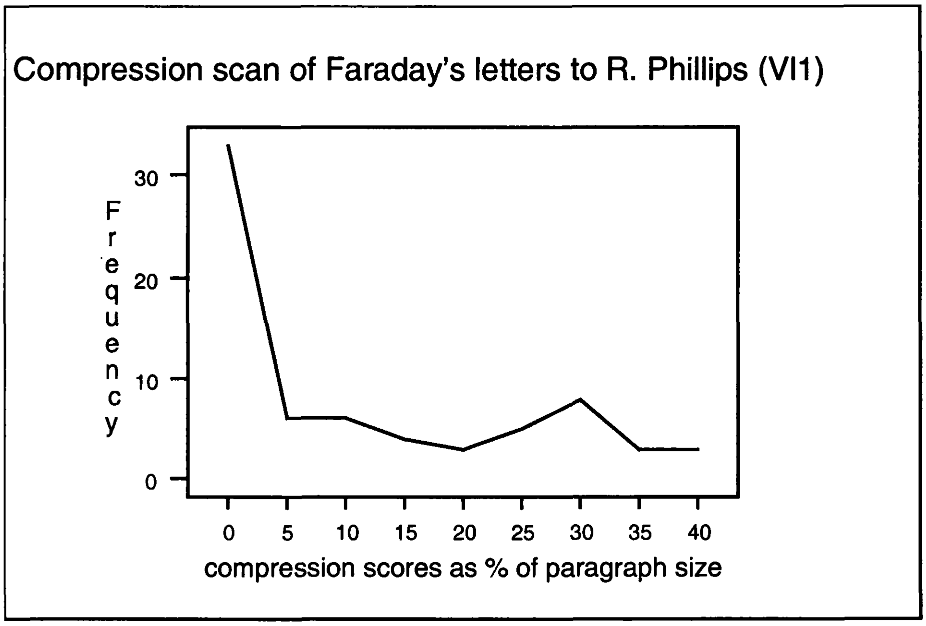


Compression scan of Faraday's letters to Whewell (1830-1849

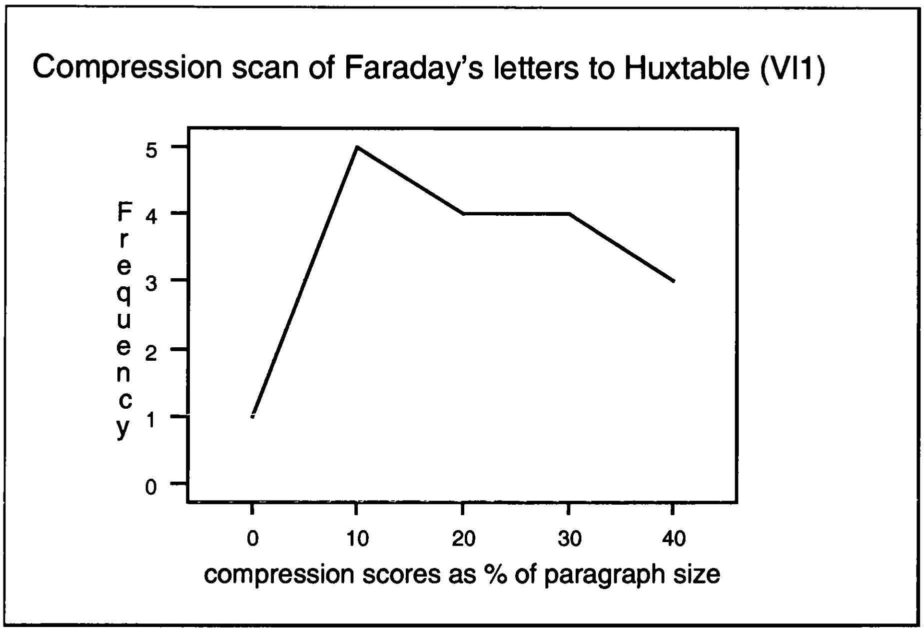
Histogram 7.3



Histogram 7.4



Histogram 7.5



7.4.2 Study Two: Private and Public Writings of Faraday and Darwin Compared

In this study I am interested in whether these features allow discrimination between the private and public writings of scientists, in this case, Faraday and Charles Darwin. The results are summarized in tables 7.3 and 7.4 below.

When Faraday's private and public electromagnetism writings were compared the public text had the largest mean for both modalities and for compression. I would have expected the *Diary* to return the highest scores for the 'contingent' modalities and the compression being the more informal and 'contingent' of the two texts. In the *Diary* Faraday would be immersed in his experimental work on electromagnetism; engaged in 'science in the making'. He would not yet be linguistically constructing a 'clean', linear experimental pathway. Nor would he be presenting truth statements and certainties. This would appear in the published, peer-reviewed paper.

The public text also returned the highest maximum scores for the 'empirical' modalities, while the private writing produced the highest scores for the 'contingent' modalities. This was expected, because for Faraday's work on electro-magnetism to be accepted by his scientific peers an unambiguous, repeatable experimental method it would need to be whittled out of the 'messy' laboratory toil. Faraday would also need to present facts and certainties as emanating smoothly from his method. Thus his language would be steeped in confident, affirming use of truth modality.

The statistical results are reflected in the histograms for Faraday's private and public electromagnetism writings where different profiles are produced for the 'empirical positive' modality and for compression.

When we turn to Darwin's writings we find that the mean scores are slightly higher for Darwin's 'M' notebook than his 'N' notebook. The situation is reversed for the 'contingent' modalities and compression. Maximum scores are greater in the 'M' notebook for all the modalities, but not for compression. The 'empirical positive' modality histogram shape for Darwin's 'M' and 'N' notebooks are quite different. The chapters from Darwin's Origin have higher means for both modality and compression than those from the Beagle, though the histograms show that the modalities are very similar for both these texts.

When we compare the private and public writings for Darwin we find that the mean scores for both public texts are higher than for the notebooks for the modalities, whereas compression is greatest in the notebooks. A similar scenario is found with the maximum scores for the modalities and compression.

In terms of histogram shapes the 'empirical positive' and 'contingent negative' modalities and compression allowed discrimination between Faraday's private and public writings on electro-magnetism. We can see this with the histograms representing compression for Faraday's *Diary* (histogram 7.6; 86 paragraphs) and the corresponding paper (histogram 7.7; 83 paragraphs) below. For Darwin's writings a similar picture is acquired. Histograms 7.8 to 7.12 below are the profiles for compression and 'empirical positive' modality for the 'M' (239 paragraphs) and 'N' notebooks (161 paragraphs). The similarity between the chapters of the *Beagle* (209 paragraphs) and the *Origin* (166 paragraphs) in terms of compression is seen in histograms 7.11 and 7.12 below.

Testing of the Main Findings of Chapter Six

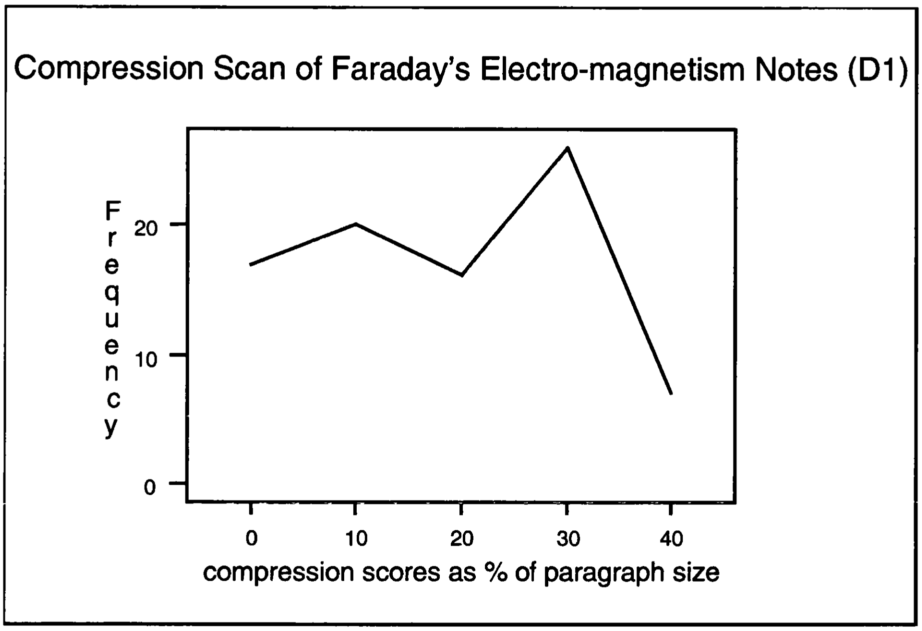
Table 7.3 Statistical Results for Faraday and Darwin's Private Texts.

Diary 1821	ep	cn	compression
mean	3.13	0.89	17.94
min. score	1.32	0.48	3.57
max. score	35.71	33.33	43.82
'M' notebook			
mean	7.56	4.89	14.33
min. score	1.32	0.48	3.57
max. score	38.36	53.85	40.56
N' notebook			
mean	6.32	4.76	14.05
min. score	2.22	0.73	2.86
max. score	22.22	45.46	43.86

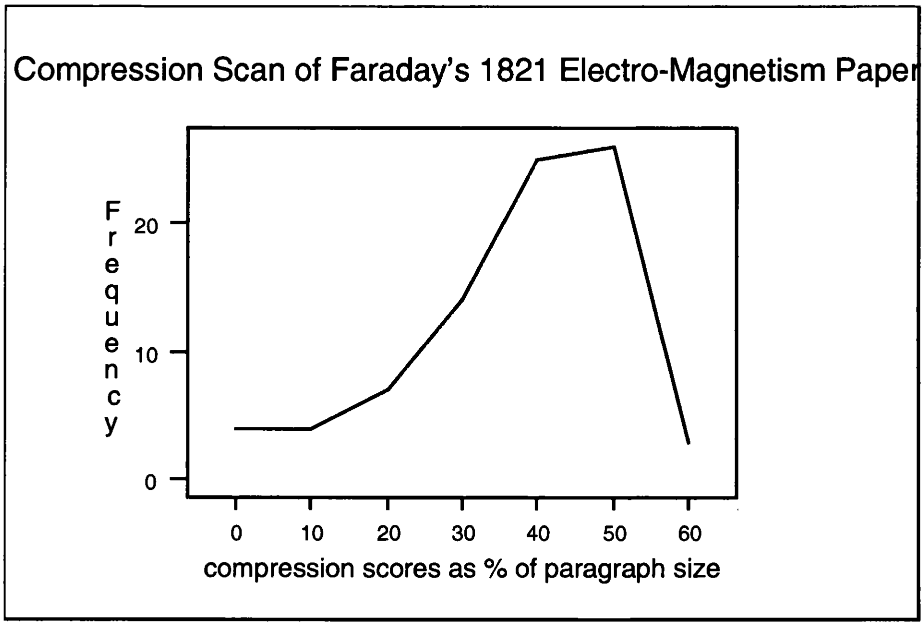
Table 7.4 Statistical Results for Faraday and Darwin's Public Texts

1821 paper	ep	cn	compression
mean	7.98	2.54	38.35
min. score	3.12	0.20	7.69
max. score	18.18	45.46	60.20
Beagle, ch 1 & 2			
mean	2.71	0.75	27.81
min. score	3.17	0.35	6.25
max. score	25.00	15.28	45.86
Origin ch 1 & 2			
mean	3.78	1.74	38.55
min. score	3.90	0.25	8.33
maxi. score	15.12	15.34	55.66

Histogram 7.6

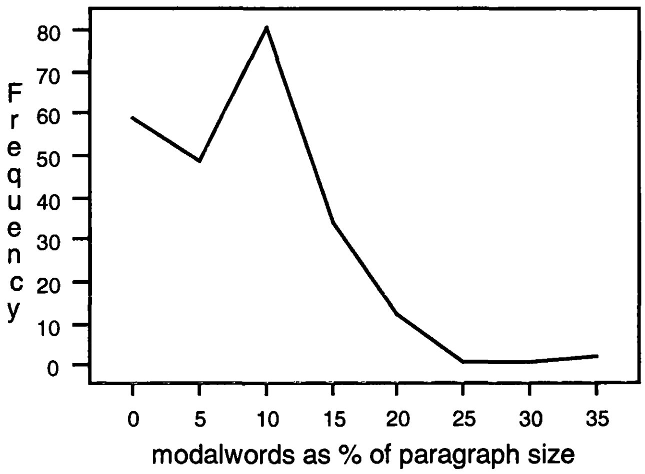


Histogram 7.7



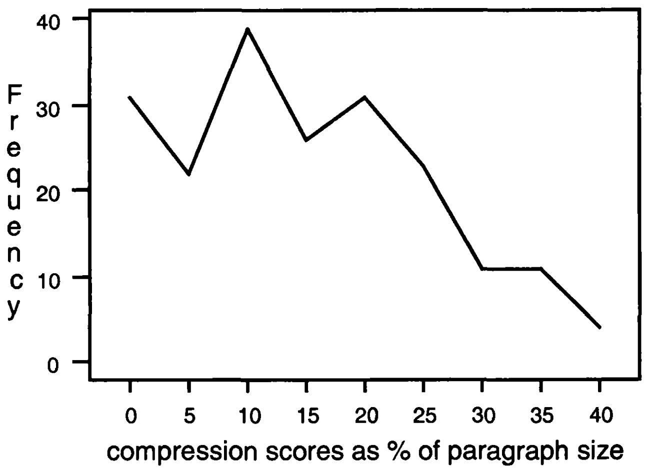
Histogram 7.8

Modalword (ep) Scan of Darwin's 'M' notebook

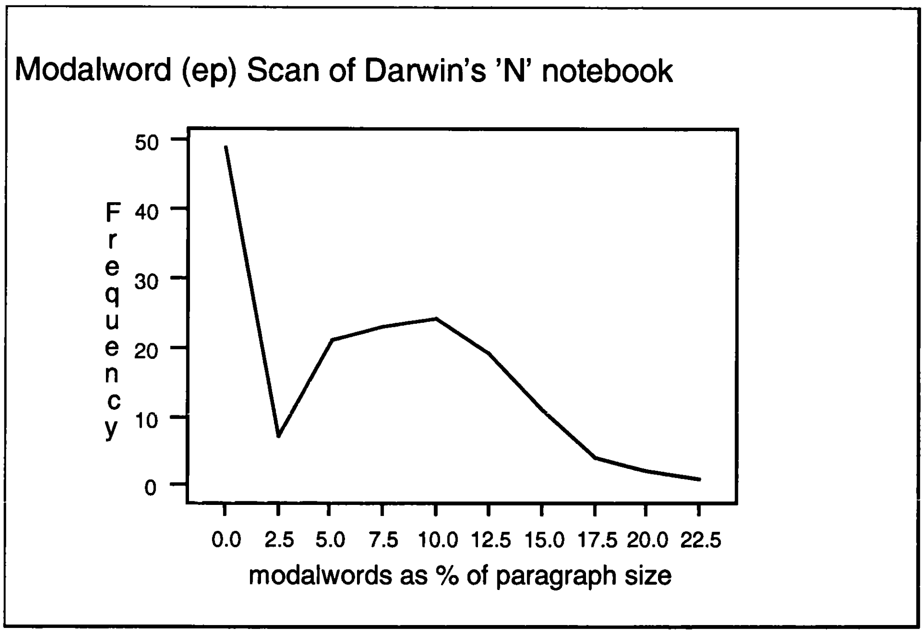


Histogram 7.9

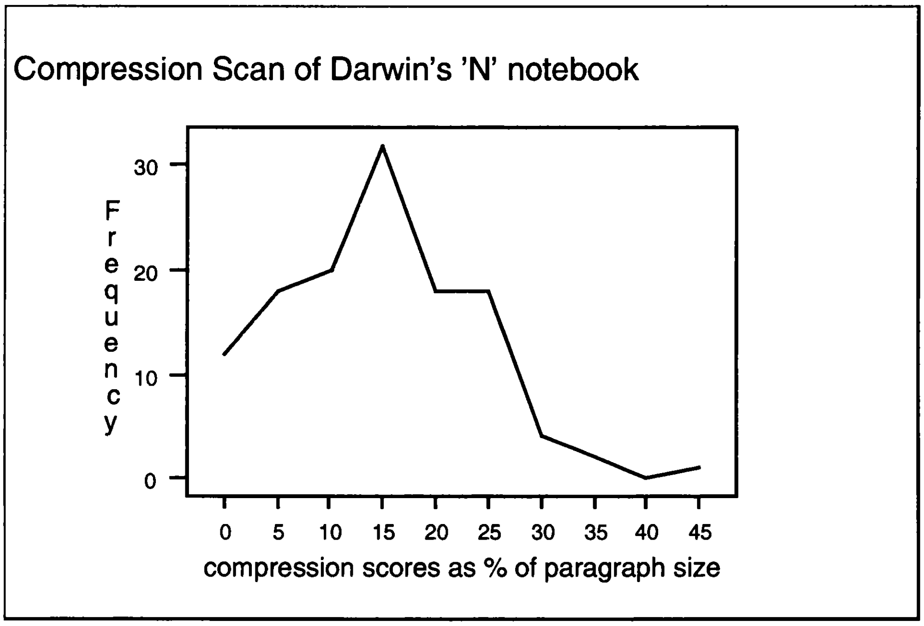
Compression Scan of Darwin's 'M' notebook



Histogram 7.10

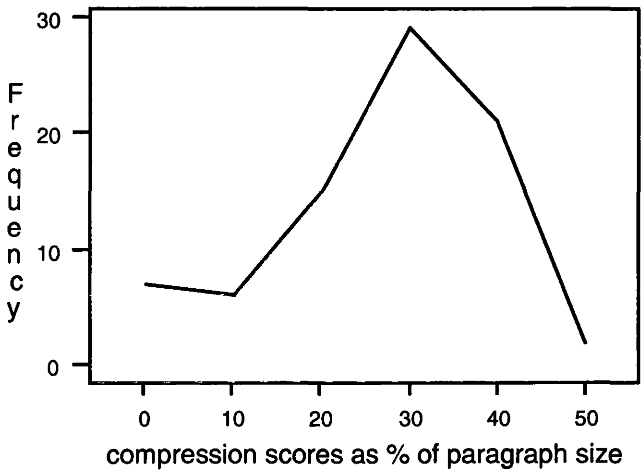


Histogram 7.11



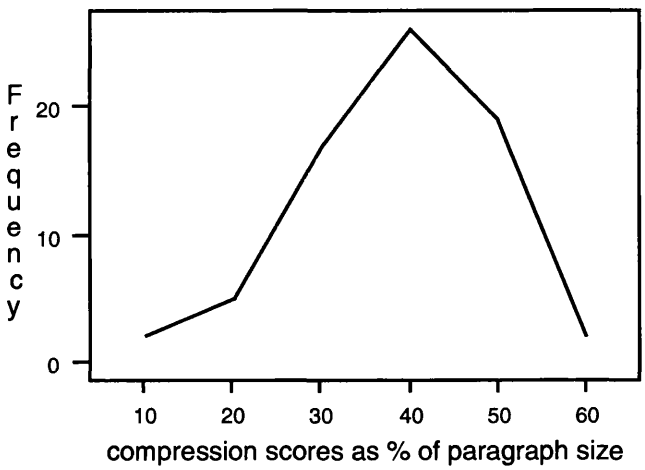
Histogram 7.12

Compression Scan of Darwin's 'Beagle' (ch 1 & 2)



Histogram 7.13

Compression Scan of Darwin's 'Origin' (ch 1 & 2)



7.4.3 Study Three: Letters to Faraday From Non-Sandemanians

It was found that the mean of both 'empirical positive' and 'contingent negative' modalities are almost all higher in the letters written by Faraday for both 'experimental' and 'non-experimental' subject matters. The mean compression scores as almost all higher, thus less compression, in the letters sent by Faraday. Though Schoenbein's letters to Faraday produce a relatively high mean compression score. The maximum scores tend to be higher in those letters written to Faraday. The histogram profiles for Schoenbein's letters to Faraday were generally similar.

The results are summarized in table 7.5, with those in the first two parts of the table for the letters written by Faraday. The last two parts of the table display the results of letters sent to Faraday by Schoenbein, Herschel, and Whewell.

Table 7.5 Statistical Results for the Sandemanian/Non-Sandemanian Correspondence.

	Dates	Mean (ep)	Min (ep)	Max (ep)	Mean (cn)	Min (cn)	Max (cn)	Mean (elp)	Min (elp)	Max (elp)
Exp.										
Faraday -Herschel	1825-1849	4.58	3.48	12.54	5.12	1.93	20.39	18.32	8.19	52.23
Faraday - Whewell	1835-1849	8.64	3.80	15.27	6.36	3.64	17.37	36.82	15.44	32.69
Non-exp.										
Faraday - Herschel	1825-1849	7.15	4.55	12.07	8.49	4.69	29.89	22.09	7.02	52.78
Faraday - Whewell	1830-1849	3.84	3.79	15.96	5.22	1.73	22.76	22.52	12.69	42.82
Exp.										
Schoenbein	1836-1848	4.42	1.37	19.51	3.93	0.26	22.22	20.73	3.33	64.24
Whewell	1835-1849	1.16	6.67	9.43	0.01	0.71	2.60	14.38	29.57	47.70
Herschel	1825-1849	2.08	7.56	19.30	1.11	2.97	35.71	10.19	18.98	42.11
Non-exp.										
Schoenbein	1837-1847	3.94	2.86	20.00	1.04	1.15	31.25	15.58	3.70	48.82
Whewell	1835-1849	3.34	3.68	20.00	1.36	3.80	12.04	21.84	28.63	40.64
Herschel	1825-1849	2.40	8.44	30.00	1.31	4.44	35.71	24.45	21.19	61.19

7.5 Discussion and Conclusions

We found Faraday used a high degree of 'empirical positive' modality in his early letters to friends that was higher than in his later letters to scientists. Also the compression was generally higher in the letters to the scientists. So these two textual features enabled discrimination between these two sets of letters, though the results were the opposite to what I expected. In 1845-49 Faraday was one of the finest scientists of his generation and his correspondence was to other scientists. Thus I surmized that 'empirical positive' modality would be greater than in the early letters to friends, and compression would be less, concomitant with more formal writing by an established man of science (table 6.3).

The corresponding private and public writings on electromagnetism for 1821 provided different profiles for both compression and the modalities. In Darwin's case the modalities were more prevalent in the published book chapters than the notebooks, which demonstrated more compression than in the published works. Also the book chapters and notebooks each shared quite similar profiles for compression and both modalities. Finally, the mean modality scores were almost all lower in the letters from the non-Sandemanians to Faraday, thought the maximum scores were higher, and compression was less in correspondence written by Faraday.

In conclusion, the intended discriminations (table 7.1) have been borne out. The measure of compression and of 'empirical positive' modality are particularly effective as indications of the discriminations I want to make. In the next chapter these measures, and others, such as conjunctives, are tested when I investigate Faraday's language use in his 'discovery' of electromagnetism in his private and published writings.

8 The Role of Language in the Construction and Reconstruction of Faraday's 'Discovery'[455] of Electromagnetism

"To integrate nature is the point of his [humankind's] existence."
J.W. Ritter, *Physics as Art*, 1806

Abstract: In this chapter I turn my attention to Faraday's most significant achievement - his work on electromagnetism. I analyse changes in Faraday's language use in the construction/reconstruction of his 'discovery' of electromagnetism through the comparison of successive drafts of research reports and records from 1821 to 1831. Also I try to discriminate between Faraday's private and public writings, and his early and later writings on electromagnetism. The main findings are that the linguistic construction of electromagnetism is quite different both between private and public texts, and the earlier (1821) and later (1831) writings. The profile of language use in the letters is not very dissimilar from the published material. I also conclude that 'discovery' and concept formation are an organic, growing complex, infused throughout a scientist's work.

8.1 Introduction

This chapter takes a close look at textual features and indicators in Faraday's private and public writings on electromagnetism over the period 1821 to 1832.[456] The period is highly important because Faraday 'discovered' electromagnetic rotation at one end and electromagnetic induction at the later end of this time period. Faraday is seen to be writing about and communicating new knowledge. The aim of this chapter is to compare Faraday's use of language in his private and corresponding published writings on electromagnetic rotation and induction. I also test how sensitive the textual features and indicators are in enabling discrimination between: i) private and public texts on electromagnetism, and ii) early and later writings on this subject.

[455] I have placed 'discovered' (and its various forms) in single quotations because the concept of 'discovery' has been problematized by numerous authors in science studies. See for example, Fleck, Ludwick; 1935, Brannigan, Augustine T.; 1981, Pickering, Andrew, 1984, Nickles, Thomas; 1990.
[456] Faraday's papers on electomagnetism were republished in the *Experimental Researches in Electricity*, 3 vols.; 1839 - 1855.

8.2 Historical and Intellectual Context

Ideas are formed and communicated and understood in languages. All concepts, ideas, and experimentation have a history; a previous body of theoretical and practical work and related language use. Faraday's work on electromagnetism is no exception, and I want to place the empirical work of this chapter within this context. This background section has two parts. The first of which provides a potted history of Faraday's 'discovery' of electromagnetic rotation and induction. The second part looks at electromagnetism on a wider philosophical plane. I do not provide a fine-grained view of Faraday's life for this material is exceptionally well provided for elsewhere.[457]

8.2.1 The Emergence of Electromagnetic Rotation and Induction

The end of the eighteenth century saw the nascence of the science of electricity when Charles de Coulomb succeeded, in the period 1785-1791, in showing that electrostatic forces obey Newton's inverse-square law.[458] In July 1820 the Danish philosopher/theologian, Hans Christian Oersted, discovered the magnetic effects of electric currents; he had demonstrated the linking of electricity with magnetism. Oersted had:

> "accidentally placed a wire carrying a galvanic current parallel to a magnetic needle during a lecture. He observed its deflection and afterwards repeated the experiment and as a result of his investigation the relation between magnetism and electricity was finally established."[459]

This phenomenon was taken up by many members of the European scientific community: "initially to confirm [it] and then increasingly to produce theories to account for it."[460] Ampère produced a theory of electromagnetism based on mathematical principals which he called 'electrodynamics'. This flurry of work also saw Arago and Davy usher in the electromagnet, Nobili's astatic galvanometer, and electromagnetic rotations by Wollaston and Faraday,[461] who was trained as a chemist. Faraday believed that electric currents could be

[457] For example Agassi, Joseph; 1971, Jones, Bence; 1870, Cantor, Geoffery; 1991, Cantor, Geoffery *et al*; 1991, James, Frank; 1991, 1993, 1995, Gooding, David C. and James, Frank (Eds.); 1985, Williams, Pearce L.; 1965.
[458] Taton, René (Ed.); 1965: 100.
[459] Bauer, E; 1965: 185.
[460] Lovie, A. D; 1992: 95.
[461] Ross, Sydney; 1965: 184.

produced by magnetism, whereas Oersted had achieved the opposite, producing magnetism from electric current.[462] In 1822 Faraday made out a list entitled: 'Chemical Notes, Hints, Suggestions, and Objects of Pursuit'. This contained the seeds of many of his future accomplishments; this included the mandate: *Convert magnetism into electricity*.[463]

By the middle of 1821 a huge volume of work had been published on electromagnetism. In an attempt to make sense of this literature, S.M. Phillips, editor of the Annals of Philosophy, commissioned Faraday to review it.[464] To achieve this Faraday had to repeat the primary experiments undertaken by, *inter alia*, Oersted, Arago, and Ampère, as well as closely examine their theoretical perspectives. This work was begun in the Summer of 1821, and came to fruition as the *Historical Sketch of Electromagnetism*. Of which there was three parts, published in volumes 2 and 3 of the *Annal of Philosophy*.[465] What Faraday achieved was to come:

> "to grips, for the first time, with the theory of electric and magnetical action; the result of his own experimental investigations and theoretical reasonings was a startling new phenomenon - electromagnetism rotation - which threatened to destroy all previous theoretical structures."[466]

Faraday's diary is very bare of references to electromagnetism between November 1820 and June 1821. This was a time Faraday was much occupied by other concerns. He was working with Stoddart on experiments on the alloys of steel, as well as working on chlorine and trying to discover its compounds with chlorine. This was also the time of his courtship of Sarah Bernard.[467]

With subsequent work Faraday achieved the effect of a wire, when a current was switched on, rotating around a magnet, as well as the magnet being able to rotate around the wire; electrical current had been converted into continuous movement. This was the advent of the first electric motor:

[462] Seigel, Daniel M.; 1991: 10.
[463] Ross, Sydney; 1991: 102. See Tweney Ryan D. and Gooding, David C. (Eds.); 1991.
[464] Faraday published this account in three parts between October 1821 and November 1822: *Historical Sketch of Electro-Magnetism*, Annals of Philosophy; 18: 195-200, 274-290; 19: 107-121.
[465] Faraday, Michael; October 1821 and November 1822.
[466] William, L. Pearce; 1965: 153.
[467] William, L. Pearce; 1965: 153.

> "the rotatory power of the magnetic force surrounding a current-carrying wire was made obtrusively manifest; the conversion of electricity into mechanical power had been achieved".[468]

Faraday had demonstrated that an electric current could be used as a source of power:

> "he argued that if a magnetic pole can be made to rotate round a current it should be possible to cause a wire carrying current to rotate round a magnetic pole."[469]

The *Historical Sketch* was published anonymously, and was Faraday's first literary foray into electromagnetism. For the eight years prior to its publication Faraday had been Davy's assistant. Faraday achieved his aim of:

> "placing in chronological order the discoveries of a dozen authors dating from latter part of 1820 to the early months of 1821."[470]

For eight days over early September 1821 Faraday conducted his first experimental work on electromagnetism,[471] which produced enough material for Faraday to publish two papers in the October issue of the *Quarterly Journal of Science*.[472] Faraday announced his discovery of electro-magnetic rotation to the world in his paper: *On some new Electro-Magnetic Motions, and on the Theory of Magnetism*, which was published in the *Quarterly Journal of Science* in October 1821.

There was a gap of a decade between the discovery of electromagnetic rotation and that of electromagnetic induction. During this time there are only a dozen entries in Faraday's *Diaries* concerned with electromagnetism. Despite how meagre the sources are, it is believed that three, almost chronological, stages characterized Faraday's ideas that lead to his discovery of electromagnetic induction in 1831, and these stages mirror the path of Faraday's thought between 1821 and 1831. The first stage was further analysis of the theory of Ampère, which subsequently Faraday came to reject. His thoughts were then concentrated on the development of his own ideas about electricity and its passage through conductors. The building of powerful electromagnets by Joseph Henry[473] and the polarity reversal noted by Faraday's friend, Gerritt

---

468 William, L. Pearce; 1965: 157
469 Singer, Charles; 1959: 361.
470 Cantor, Geoffrey; 1991: 228.
471 Cantor, Geoffrey; 1991: 229.
472 Faraday; *On some new Electro-magnetical motions, and on the Theory of Magnetism*; ERE 2: 127-147, and *Electro-magnetic Rotation Apparatus*; *ibid*: 147-158.
473 Professor of Natural Philosophy at the College of New Jersy, Princeton, 1832-1846.

Moll,[474] upon rapid changeover of the electromagnet leads, characterized the third stage of Faraday's work out of which on August 29th 1831 he succeeded in showing that a magnet could induce electricity; electromagnetic induction.[475] Finally, in late November, 1831, he felt he knew enough to announce his findings to the Royal Society. He then wrote a letter to one of his best friends, Richard Phillips, to explain the research program that he had laid out for himself.[476] The paper Faraday read to the Royal Society on November 24 1831, and subsequently published in its *Philosophical Transactions* in 1832 as *Experimental Researches in Electricity*, was one of a long series of papers that were collected together and printed as the *Experimental Researches in Electricity*, which ultimately filled three volumes, with thirty series (1839-1855).

8.2.2 Faraday, German Romantics, and *Naturphilosophie*

Central to Faraday's work on electromagnetism was the idea of the "connectedness of all with all".[477] Revisionist historiography emphasizes that Faraday had different philosophical commitments to begin with, which led him to see new phenomena, to be receptive to different explanations, and to move toward a non-Newtonian theory. According to L. P. Williams, Faraday's discovery of electromagnetic induction, for example, was the outcome of a ten year commitment to the unity of forces, and search for evidence, not its inception.[478]

Traditional histographers point to the role of new experiments on electricity and magnetism, and to Roger Boscovich's *Theory of Natural Philosophy*[479] to explain the shift by Faraday and Davy to a anti-Newtonian stance. Revisionists instead emphasis how *naturphilosophie*, which, *inter alia*, embodied a "dynamic theory of nature", as often contrasted with a 'mechanical, Newtonian theory of nature',[480] informed the philosophical commitments of Oersted and Faraday. Their physical researches being based on vitalist and organismic views promulgated by German biologists of the Romantic period, and thinkers such as

[474] Professor of Physics and Director of the Observatory at Utrecht.
[475] Williams, L. Pearce; 1965: 169-170.
[476] Faraday wrote his letter to Phillips on November 29th, 1831.
[477] Seigel, Daniel M.; 1991: 9.
[478] Williams, L. Pearce.; 1965: 183.
[479] First Edition, 1758. English Translation, 1763.

Goethe, Herder, kant, and Schelling, who opposed the rational mechanism of Newtonian thought. Even before Oersted and Faraday *naturphilosophie* was taken up by Samuel Coleridge, a close friend of Davy.[481] This holistic perspective on scientific research is described by Wise as a:

> "conscious commitment to the unity of all natural powers, to the belief that all physical forces are but manifestations of a fundamental conserved 'force'."[482]

The context of Oersted and Faraday's commitment to *naturphilosophie* and anti-Newtonianism was also a conservative British response to the Godless French Revolution. This last point is important as for Faraday, the religious aspect in his science was expressed and strengthened by his membership of the Sandemanians, a small fundamentalist sect.[483] The reaction to the French Revolution (especially the Terror) also involved a rejection of French materialistic philosophy of matter by British physicists.[484]

Schelling was opposed to the Newtonian view of matter: "as made of hard, impenetrable, inert particles that are acted on by forces external to them."[485] What formed the setting for Schelling's conception of matter was Kant's belief in a "metaphysico-dynamical" conception: "of matter as made of a balance of opposed forces."[486] Thus for Schelling - and Romantics generally - the emphasis was on polarity in nature; nature: "as a balance of opposed forces or tendencies, a balance when disrupted leads to strife and activity."[487] The central themes in the burgeoning knowledge of electricity - attraction and repulsion - were linked to this notion of 'polarity'.[488] This duality of forces was in contradistinction to the push-pull forces of Newtonian mechanics and the Continental School of electromagnetism, premised on the work of Ampère.

---

[480] Gower, Barry; 1973: 320-321.
[481] Knight, David. M.; 1970: 54-75.
[482] Wise, Norton M.; 1990: 347.
[483] Tobey, Ronald C.; http://www.kaiwan.com/~lucknow/horus/guide/ph106.html; accessed September 20, 1997.
[484] Tobey, Ronald C.; http://www.kaiwan.com/~lucknow/horus/guide/ph106.html; accessed September 20, 1997.
[485] Stern, Robert; 1988: x.
[486] Stern, Robert; 1988: xi.
[487] Stern, Robert; 1988: xi.
[488] Porter, Roy and Teich, Mikuláš; 1988: 6.

8.3 Methodology

8.3.1 Source Material

The source materials are three pieces of Faraday's private and seven of his public writings on electromagnetism from 1821 to 1832, which encompass letters, notebooks, and published papers. These writings include the *Historical Sketch in Electromagnetism*, which was published in three parts in volumes two and three of the *Annals of Philosophy*, between October 1821 and November 1822. The third part of the *Sketch* is theoretical, whereas parts one and two are more concerned with describing experiments. Following on from his experiments in September 1821, Faraday published two papers in the October issue of the *Quarterly Journal of Science*:[489] *On some new Electro-Magnetical Motions, and on the Theory of Magnetism*, appearing in volume twelve (pages 74-96),[490] and *Electro-magnetic rotation apparatus*.[491] The corresponding notebook pages are from *Diary* volume one, pages 49 to 63.[492] Turning to electromagnetic induction both the *Diary*[493] and the published[494] forms are available for analysis. Finally, the letters comprise those Faraday wrote between 1821 and 1831 on electromagnetism, where the prose is formal.

Thus the material analyzed comprises the private and public writings for rotation (notebook and two papers) and for induction (notebook and one paper) available for analysis. As well as the whole of the *Sketch* and relevant letters.

8.3.2 Analysis

Each of the texts were scanned for eight features and indicators. These were: compression, conjunctives,[495] paragraph length, modalities ('empirical positive'

[489] Cantor Geoffery; 1991: 229.
[490] Reprinted in *Experimental Researches in Electricity* (3 vols.); vol.2; London; 1844: 127-147.
[491] *Experimental Researches in Electricity*; vol.2: 147-158.
[492] The dates for these diary pages are September 3rd to December 25th, 1821.
[493] August 1831 - June 1832. Pages 367 - 441.
[494] *On the induction of electric currents. On the evolution of electricity from magnetism. On a new electrical condition of matter. On Arago's magnetic phenomena*; Philosophical Transactions; 122: 125-162. Republished in *Experimental Researches in Electricity*; vol.1; 1839.
[495] after and while but or then that because if when since so although before nor than till until unless whereas yet thus meanwhile subsequently at the same time nevertheless therefore by contrast however also like though

and 'contingent negative'), and the word lists: 'bench_words',[496] 'any_experiment',[497] and 'new_discovery' (developed by David Gooding).[498] Compression and the two modalities have been chosen because in chapters six and seven we saw how these features proved very able to allow discriminations between text types to be made. As reasoned in the earlier case studies I expect compression to be more prevalent in the notebooks, than the published papers, in which the use of 'empirical positive' modality would be most robust. The 'bench_words' list contains vocabulary relating to bench-top (experimental) work, which I expect to be used more in Faraday's laboratory notes (*Diary*) than in his published papers. The 'any_experiment' is a list containing words to do with experimentation, and 'new_discovery' a list that would pick out paragraphs dealing with new discoveries, that is, new results. I would expect these two sets of vocabulary to be more prevalent in Faraday's published papers. This is because the linguistic construction and assertion of 'facts' and meaning would be more robust if results were situated as the outcome of discoveries made through experimentation.

Another textual phenomenon I consider in this chapter is conjunction. The use of conjunctives is "a critical device for the cohersion of a text".[499] I believe that the incidence of conjunctives will be greater in Faraday's published work than his private writings for two main reasons. First, in chapter four (section 4.4.1) I discussed how 'nuts and bolts' words, such as conjunctives, are more likely to be omitted from private writing, where cohersion is less important, partially because a readership, beyond the writer, is not expected. Also diary or notebook writing is more a narrative discourse, without an 'external' audience to persuade, thus relying less on argument and logical sentence structure.

---

[496] eye use effect saw see sensible visible touch make made find come out comes out phenomen appear observe work sent send obtain join place push put making tak whil finger hand

[497] very care can find contain adulteration impur pure accura experiment trial examin sample opinion view substance compos analys weigh grain ounce proportion minute estimate expts supported adduced demonstr beautiful lovely show apparatus arrang instrument new ascertain verify establish result shew law describe repeat confirm proof expect anticipate made make sensible notice obtain illustrate act curious attention assumption suppos convinc phenomen eye use effect saw see visible touch come out comes out appear observe work sent send join place push put making tak whil finger hand

[498] experiment examin expts supported adduced demonstr beautiful lovely show apparatus arrang instrument new ascertain verify establish result shew law describe repeat confirm proof expect anticipate made make sensible notice obtain illustrate act curious attention assumption suppos convinc phenomen

Secondly, the discursive space of published science writing is characterized by persuasion and argument for which 'logic orientation' and a high "prevalence of coordination" is crucial.[500] One outcome of the greater use of conjunctives in the more formal, published writing will be longer sentences, and concomitantly, we should see longer paragraphs.

In his writings of the 1820s Faraday did not display the confidence which emerged at the end of this decade.[501] The reason for this not least being the accusation of plagiarising William Wollaston by Sir Humphrey Davy. I would surmise that this could be demonstrated in Faraday's language use. The degree of compression would be greater in his earlier writings of the 1820s. The paragraph lengths would be more consistent at the end of the decade. At this time the use of both the 'empirical positive' (ep) and 'contingent negative' (cn) modalities, as well as the wordlists ('new_discovery', 'any_experiment', 'bench_words'), would be greater.

My results, which will be summarized in tables, will comprise the range of scores from the lowest non-zero scores to the highest, the truncated mean, the standard error (SE). Also I shall run an Analysis of Variance test which displays the confidence level for the mean of each text (mean ± (2 x SE)). This allows me to see whether the range (mean ± (2 x SE)) of the texts I am comparing overlap. If they overlap, then the difference between the means is not statistically significant. Finally, I run off frequency histograms to look for distinct profiles for the private and public texts, as well as early and later writings on electromagnetism. I have overlain the histograms profiles for the published text and the corresponding notebook entries on a single graph. All of the data was gathered using Minitab version 10 *Xtra* for Macintosh. Summarized in table 8.1 are the distinctions, the features measured, and the texts the tests are applied to.

---

[499] Fawley, William; 1987: 96.
[500] Fawley, William; 1987: 98.
[501] Cantor, Geoffrey; 1991: 239-240.

Table 8.1 Relationship Between Distinction, Feature, and Source Texts.

	private v public writings	1821 v 1831 private writings	1821 v 1831 public writings	letters v public writings
compression	electromagnetism texts for 1821 and for 1831 (public includes the *Sketch*)	*Diary* for 1821 and 1831	published papers for 1821 and 1831.	letters (1821-1831) and published papers for 1821 and 1831.
conjunctives	as above	as above	as above	as above
paragraph length	ditto	ditto	ditto	ditto
'ep' and 'cn' modalities	ditto	ditto	ditto	ditto
'new_discovery' words	ditto	ditto	ditto	ditto
'any_experiment' words	ditto	ditto	ditto	ditto
'bench-words'	ditto	ditto	ditto	ditto

8.4 Results: Summary

The results are first summarized in tables 8.2 to 8.7, and then the most striking results are discussed. The Analysis of Variance results are found in appendix six.

Table 8.2 Statistical Results for the *Diary* and Published Papers for 1821.

1821	no. of paragraphs	ellipsis	conjunctives	paragraph length	modality ep	modality cn	'new_discovery'	'any_experiment'	'bench_words'
Diary	**163**								
Mean[502]		17.94	4.61	42.96	3.13	0.89	1.66	2.69	0.80
Min.[503]		3.57	2.94	10	1.32	0.48	1.10	1.70	0.85
Max.		43.82	26.67	214	35.71	33.33	15.39	25.00	9.68
SE		1.35	0.50	4.12	0.43	0.43	0.261	0.37	0.15
Paper[504]	**88**								
Mean		38.35	11.00	156.1	7.98	1.33	3.102	5.43	2.21
Min.		7.69	4.17	12.0	3.12	0.20	0.56	2.03	0.47
Max.		60.20	23.08	517.0	18.18	5.34	13.16	5.43	7.90
SE		1.57	0.44	12.8	0.41	0.70	0.28	0.30	0.17
Paper[505]	**47**								
Mean		42.26	5.04	211.6	3.45	1.11	1.29	2.50	0.86
Min.		7.69	8.14	12.0	4.21	0.30	0.56	2.03	0.47
Max.		55.74	16.67	434.0	13.76	23.26	8.33	9.30	4.84
SE		2.75	0.863	26.4	0.65	0.59	0.31	0.47	0.19

[502] Truncated mean, where the Minitab removes the smallest 5% and the largest 5% of the values, and then averages the remaining values.
[503] The lowest non-zero score.
[504] *On some new Electro-Magnetical Motions, and on the Theory of Magnetism.*
[505] *Electro-magnetic rotation apparatus.*

Table 8.3 Statistical Results for the *Diary* and Published Paper for 1831.

1831	no. of paragraphs	ellipsis	conjunctives	paragraph length	modality ep	modality cn	'new_discovery'	'any_experiment'	'bench_words'
Diary	**719**								
Mean		17.91	7.50	42.11	3.71	1.52	1.39	3.60	1.90
Min.		3.03	1.72	10.00	0.72	0.79	0.62	0.72	0.65
Max.		51.89	37.50	219.0	30.0	54.55	20.83	27.27	20.0
SE		0.49	0.29	1.41	0.175	0.27	0.10	0.18	0.12
Paper	**532**								
Mean		28.07	9.41	90.72	5.87	1.99	2.43	4.67	2.04
Min.		3.85	2.22	10.0	1.31	0.49	0.62	0.72	0.41
Max		58.33	33.33	297.0	30.0	31.25	20.82	23.08	15.38
SE		0.58	2.70	0.20	0.20	0.21	0.13	0.81	0.10

Table 8.4 Statistical Results for the *Sketch*.

Sketch	no. of paragraphs	ellipsis	conjunctives	paragraph length	modality ep	modality cn	'new_discovery'	'any_experiment'	'bench_words'
part 1	**19**								
Mean		36.1	11.33	138.1	11.24	2.80	2.92	4.85	1.58
Min.		11.43	7.41	10.0	6.80	1.14	0.44	1.75	1.31
Max.		59.73	16.20	321.0	36.36	13.89	5.56	9.86	3.57
SE		3.93	1.08	20.10	1.57	0.88	0.430	0.67	0.28
part 2	**54**								
Mean		36.75	11.80	159.0	9.10	3.48	3.61	6.58	2.93
Min.		9.52	6.10	10.0	1.22	0.32	0.52	1.45	0.78
Max		56.29	23.19	337.0	21.57	17.74	29.41	29.41	11.77
SE		1.65	0.52	11.50	0.55	0.66	0.55	0.57	0.28
part 3	**71**								
Mean		34.02	8.99	126.20	6.93	2.77	2.76	4.90	1.89
Min.		8.33	4.76	11	2.38	0.44	1.24	1.92	0.48
Max.		53.33	23.76	313	14.63	30.00	9.84	16.67	8.33
SE		1.67	0.70	10.3	0.55	0.61	0.31	0.47	0.22
all	**144**								
Mean		35.40	10.43	141.86	8.27	3.09	3.01	5.53	2.22
Min.		8.33	4.76	11	1.22	0.31	0.44	1.45	0.48
Max.		59.73	23.76	337	36.36	30.00	29.41	29.14	11.76
SE		1.15	0.43	7.26	0.42	0.41	0.27	0.33	0.16

Table 8.5 Statistical Results for the 1821-1831 Letters.

Letters	no. of paragraphs	ellipsis	conjunctives	paragraph length	modality ep	modality cn	'new_discovery'	'any_experiment'	'bench_words'
(1821-1831)	**114**								
Mean		26.47	6.38	104.3	4.09	2.72	0.87	2.79	1.06
Min.		9.09	3.12	10	3.17	0.32	0.59	1.67	0.59
Max.		57.54	25.93	525	15.63	22.22	7.56	14.29	7.14
SE		1.89	0.68	12.8	0.42	0.51	0.17	0.33	0.16

Table 8.6 Summary of Ranges of Scores for each Text for 1821 and 1831.

	Paragraphs	Ellipsis	Conjunctives	Paragraph Length	EP Modality	CN Modality	'New_ Discovery'	'Any_ Experiment'	'Bench_ Words'
1821 Diary	163	3.57-43.82	2.94-26.67	10.0-214.0	1.32-35.71	0.48-33.33	1.10-15.39	1.70-25.00	0.85-9.68
1821 Paper	88	7.69-60.20	4.17-23.08	12-517	3.12-18.18	0.20-5.34	0.56-13.16	2.03-5.43	0.47-7.90
1821 Paper	47	7.69-55.74	8.14-16.67	12.0-434.0	4.21-13.76	0.30-23.26	0.56-8.33	2.03-9.30	0.47-4.839
1831 Diary	719	3.03-51.89	1.72-37.50	10.0-219.0	0.72-30.0	0.79-54.55	0.62-20.83	0.72-27.27	0.65-20.0
1831 Paper	532	3.85-58.3	2.22-33.3	10.0-297.0	1.31-30.0	0.49-31.25	0.62-20.82	0.72-23.08	0.41-15.38
Sketch 1	19	11.43-59.73	7.41-16.20	10.0-321.0	6.80-36.36	1.14-13.89	0.44-5.56	1.75-9.86	1.31-3.57
Sketch 2	54	9.52-56.29	6.10-23.19	10.0-337.0	1.22-21.57	0.32-17.74	0.52-29.41	1.45-29.41	0.78-11.77
Sketch 3	71	8.33-53.33	4.76-23.76	11.0-313.0	2.38-14.63	0.44-30.0	1.24-9.84	1.92-16.67	0.48-8.33
Sketch (all)	144	8.33-59.73	4.76-23.76	11.0-337.0	1.22-36.36	0.31-30.0	0.44-29.41	1.45-29.14	0.48-11.76
Letters	114	9.09-57.54	3.12-25.93	10.0-525.0	3.17-15.63	0.32-22.22	0.59-7.56	1.67-14.29	0.59-7.14

8.5 Results: Interpretation

8.5.1 Comparison of Faraday's 1821 Private and Public Electromagnetism Writings

The mean compression scores are higher in the notebook than either of the two published papers (table 8.2). This concurs with what I had anticipated. In the private writing - the *Diary* - we have expected more compression as the result of less repetitions, and less use of articles and connectives. This is borne out by the greater incidence of conjunctives (connectives) in both the published papers. The paragraph lengths are also longer in the published papers. In the *Diary* form paragraphs would represent relatively small chunks of thoughts or ideas. The two modalities are more prevalent in the public writings. This would seem to be explained by Faraday writing in a lucid, pure, and precise way, so that unadorned, certain 'facts' shone out from his text; something was framed as categorically this or that. In this way:

> "the link between experiment and language was complete and facts could be expressed and communicated without distortion by the linguistic medium."[506]

Such a use of language would allow Faraday to persuade and mobilize power beyond the confines of his basement laboratory. This "literary technology"[507] is crucial in the process of enrolling and persuading others, and thus for empirical experience to be constituted as 'matters of fact'.[508] The presence of the 'bench_words' vocabulary, contrary to expectation, is greater in the published papers.

The paper that corresponds most closely with Faraday's *Diary* writings on electromagnetism (*new electro-magnetical motions*) demonstrates no overlap in confidence interval levels with the *Diary* (appendix six). These are two quite distinct linguistic constructions of experimental work on electromagnetism.

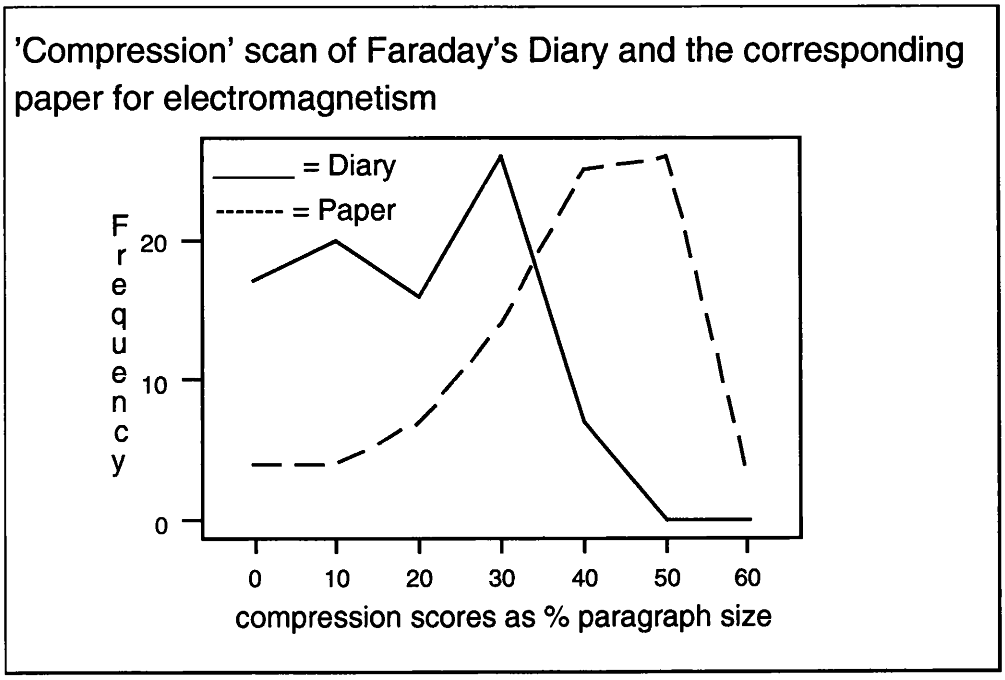
If we now look at the results for Faraday's *Sketch* (tables 8.4) we find that compression is similar across all four of the *Sketch* texts, and once again, compression is highest in the *Diary*. Also the mean score for conjunctives is

---

[506] Cantor, Geoffery; 1991: 214.
[507] Shapin and Schaffer; 1985: 25.
[508] Shapin and Schaffer; 1985.

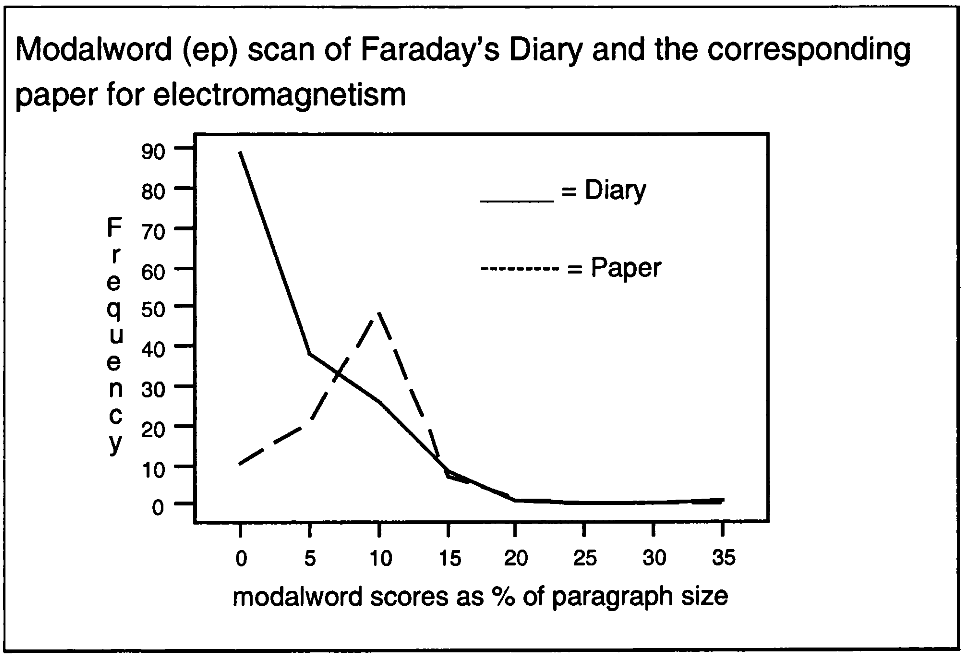
greater in the published writings, though not the maximum scores. This scenario is almost repeated for the two modalities. But the paragraph lengths, on average, are higher for the *Sketch*, as the 'new_discovery', 'any_experiment', and 'bench_words' wordlists. Generally the features and wordlists are more in evidence in the published experimental monographs. In the *Sketch* Faraday is providing a resumé of work done on electromagnetism by other scientists, rather then trying to use language to persuade his audience that his findings are original enough to be accepted as the new default or standard. The *Sketch*, in all its forms, shows little overlap in confidence intervals with the *Diary* (appendix six; graph 1). This again highlights how differently Faraday is using language when writing about electromagnetism in private and public repertoires. Further evidence is provided by the histogram shapes, which are quite different when we compare the *Diary* (163 paragraphs) and the *new electro-magnetical motions* (88 paragraphs) and *Electro-magnetic rotation apparatus* (47 paragraphs) papers.[509] Histograms 8.1 and 8.2 provide examples of this difference.

Histogram 8.1. Comparison of Compression Scores for Faraday's 1821 *Diary* and the Corresponding *new electro-magnetical motions* Paper.



[509] This is also the 1821 paper referred to in histograms 5 and 6.

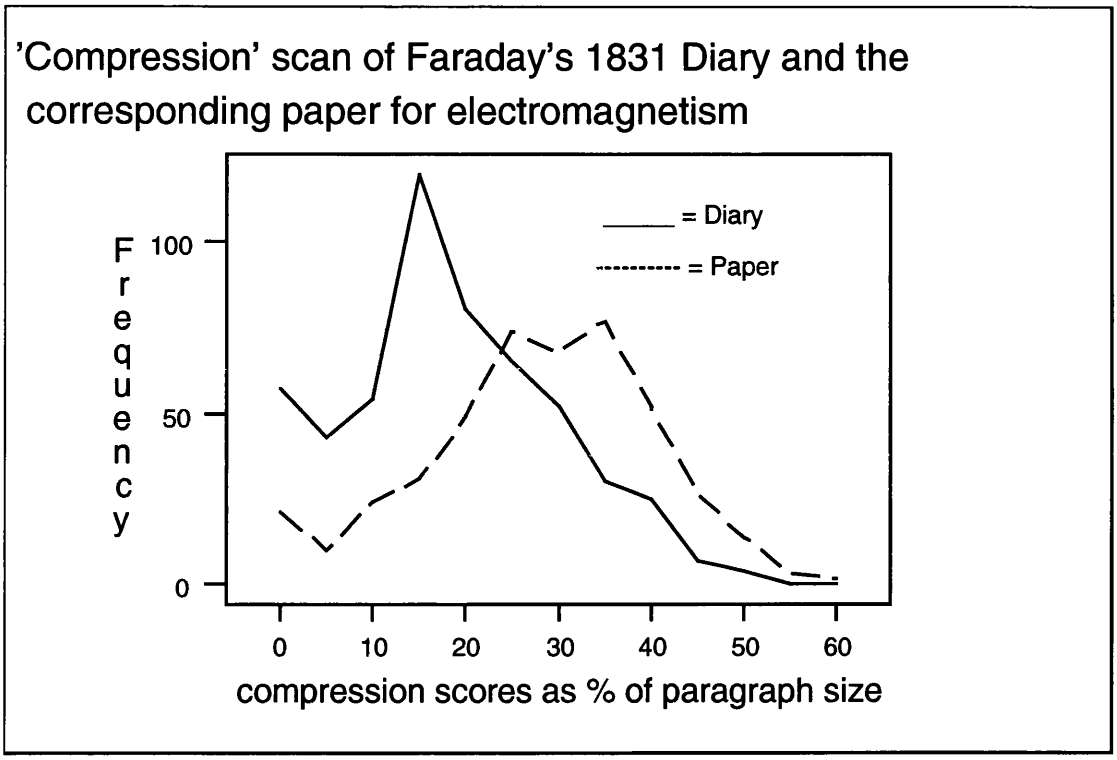
Histogram 8.2. Comparison of 'Empirical Positive' Modality Scores for Faraday's 1821 *Diary* and the Corresponding *Electro-magnetic rotation apparatus* Paper.



8.5.2 Comparison of Faraday's 1831 Private and Public Electromagnetism Writings

As with the 1821 *Diary* and the *new electro-magnetical motions* paper the mean scores for all eight of the textual features and wordlists are higher for the 1831 published paper, than the notebook (table 8.3). The distinctiveness of the *Diary* and the published paper is highlighted by the confidence levels where overlap occurs only for the 'contingent negative' modality and 'bench_words' wordlist (appendix 6, graph 2). This is further evidenced by the different histograms profiles for these two texts, seen, for example, in histogram 8.3 (719 paragraphs for the *Diary* and 532 for the paper).

Histogram 8.3. Comparison of Compression Scores for Faraday's 1831 *Diary* and the Corresponding Paper.



8.5.3 Comparison of the 1821 and the 1831 Private and Public Electromagnetism Writings

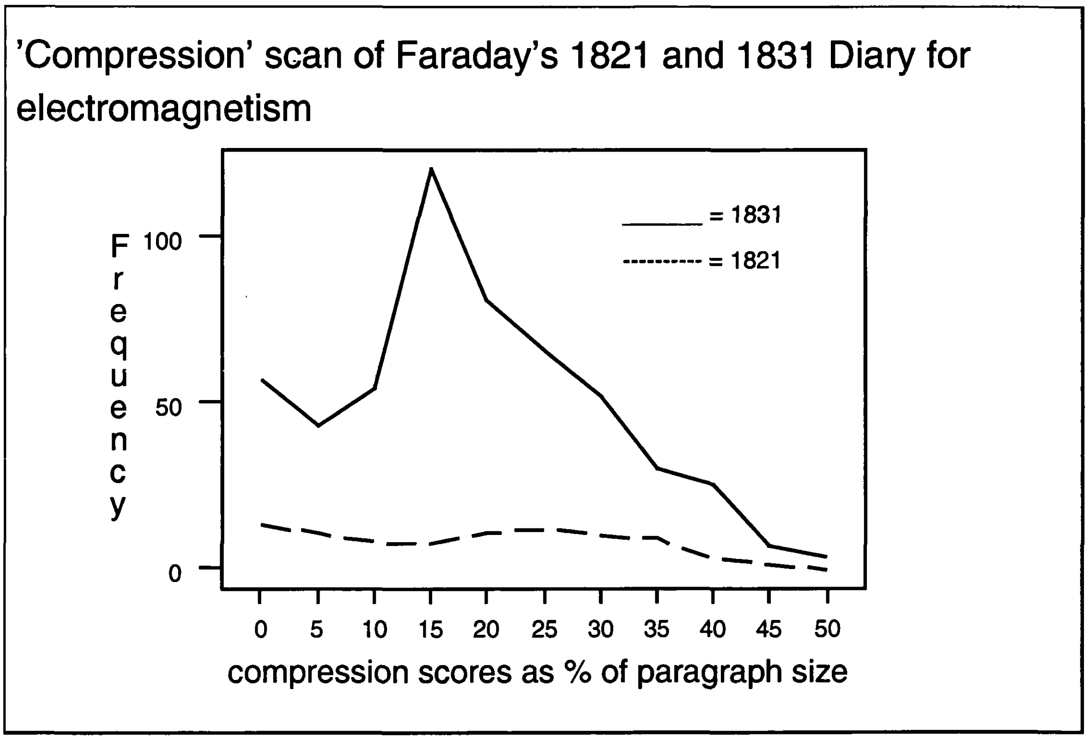
The two laboratory notebooks share fairly similar scores for compression, paragraph length, 'empirical positive' modality, and the 'new_discovery' wordlist (table 8.7 below). This is reflected in the overlap of confidence levels for these features and indicators (appendix six, graph 3). With how systematic and methodical Faraday was in his experimental work I would expect Faraday to demonstrate this degree of consistency. This sort of consistency is only seen between the 1821 *new electro-magnetical motions* and the 1831 papers for the 'empirical negative' modality, and 'new_discovery', 'any_experiment', and 'bench_words' wordlists. Though when we look at the histograms for compression scores for the *Diary* (histogram 8.4; 1821: 163 paragraphs; 1831: 719 paragraphs) the profiles are quite distinct. This is also the case with the similar profiles for the published papers (histogram 8.5; 1821: 88 paragraphs; 1831: 532 paragraphs). This further demonstrates the sensitivity of compression to enable discrimination between different genres of Faraday's writings.

Table 8.7 Statistical Results for the 1821 and 1831 *Diary* for Electromagnetism.

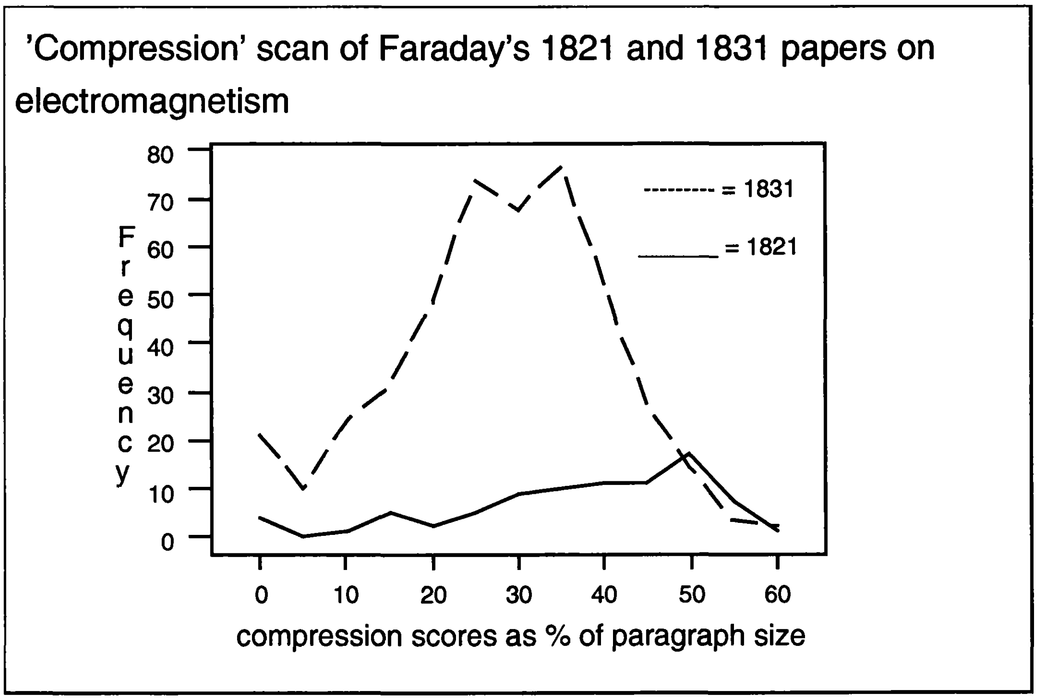
Diary	no. of paragraphs	ellipsis	conjunctives	paragraph length	modality ep	modality cn	'new_discovery'	'any_experiment'	'bench_words'
1821	**163**								
Mean[510]		17.94	4.61	42.96	3.13	0.89	1.66	2.69	0.80
Min.[511]		3.57	2.94	10	1.32	0.48	1.10	1.70	0.85
Max.		43.82	26.67	214	35.71	33.33	15.39	25.00	9.68
SE		1.35	0.50	4.12	0.43	0.43	0.261	0.37	0.15
1831	**719**								
Mean		17.91	7.50	42.11	3.71	1.52	1.39	3.60	1.90
Min.		3.03	1.72	10.00	0.72	0.79	0.62	0.72	0.65
Max.		51.89	37.50	219.0	30.0	54.55	20.83	27.27	20.0
SE		0.49	0.29	1.41	0.175	0.27	0.10	0.18	0.12

[510] Truncated mean, where the Minitab removes the smallest 5% and the largest 5% of the values, and then averages the remaining values.
[511] The lowest non-zero score.

Histogram 8.4 Comparison of Compression Scores for Faraday's 1821 and 1831 *Diary* for Electromagnetism.



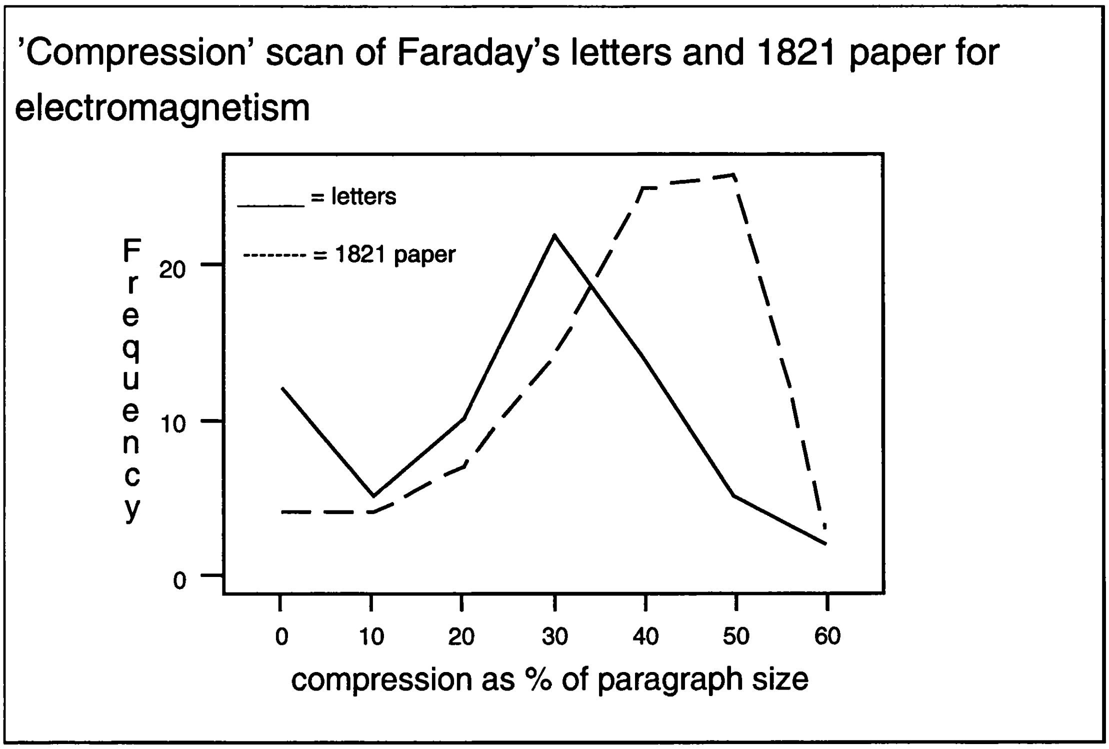
Histogram 8.5 Comparison of Compression Scores for Faraday's 1821 and 1831 Papers on Electromagnetism.



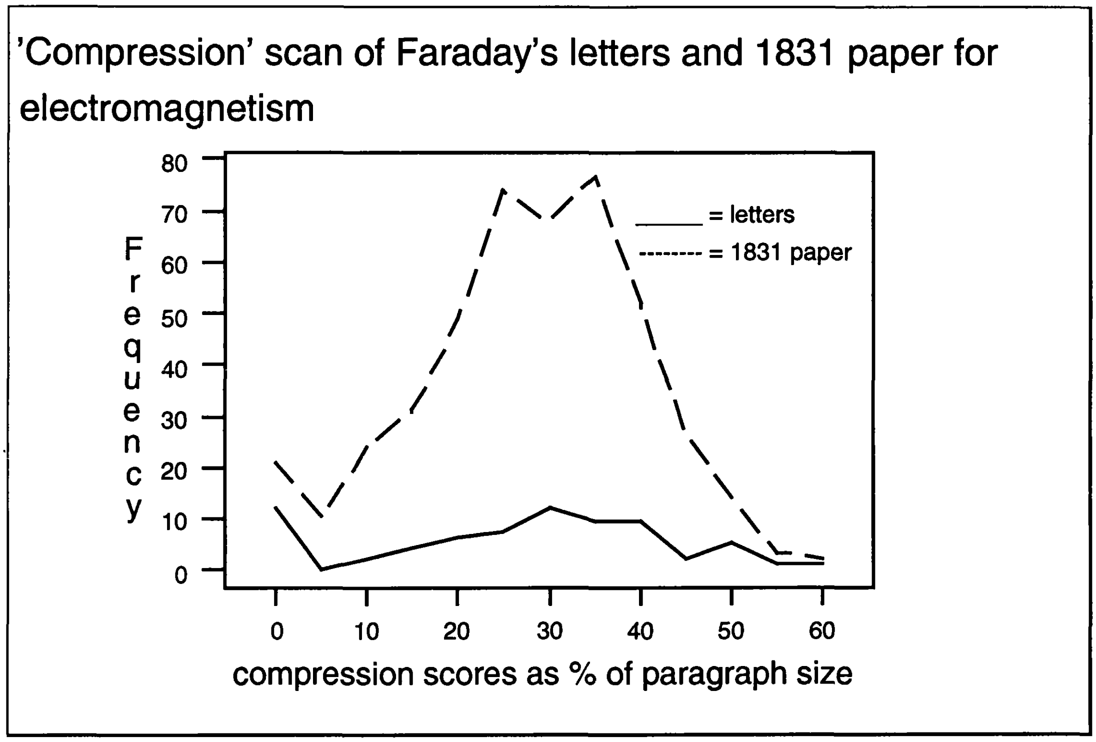
8.5.4 Comparison of Faraday's Letters and the Published Papers (1821 and 1831) on Electromagnetism

The letters produce the least mean compression, and share the lowest maximum compression score (table 8.5) with the second of the 1821 papers (*Electro-magnetic rotation apparatus*). Two of the published papers (including *new electro-magnetical motions*) contain the most conjunctives, which we should expect. These texts also have the highest mean scores for the 'empirical positive' modality and the 'new_discovery', 'any_experiment', and 'bench_words' word-lists. The results for the letters are closer to those of the published papers. This is probably explained by the fact that in these letters Faraday is writing about electromagnetism in formal manner, as he might in a publication. We can see this correspondence with the profiles for compression in histogram 8.6 (letters: 114 paragraphs; 1821 paper 88 paragraphs), and histogram 8.7 (1831 paper: 532 paragraphs).

Histogram 8.6 Comparison of Compression Scores for Faraday's Letters and 1821 Paper on Electromagnetism.



Histogram 8.7 Comparison of Compression Scores for Faraday's Letters and 1831 Paper on Electromagnetism.



8.6 Discussion

We found that the compression and paragraph length scores are very similar for the 1821 and 1831 laboratory notebooks. A greater degree of difference exists for these features for the three published papers. It is the measure of conjunctives, the two modalities, and the 'new_discovery', 'any_experiment', and 'bench_word' wordlists which discriminate most effectively between the 1821 and 1831 notebooks. Whereas with the early and later published writings it is compression, conjunctives, paragraph length, 'empirical positive' modality, and 'any_experiment' wordlist which are the most discriminatory.

For the 1821 writings on electromagnetism all eight features and indicators discriminate between the private and public texts. Though this is most pronounced for Faraday's *On some new electro-magnetical motions, and on the theory of magnetism* paper, which corresponds closest with the *Diary*. This is demonstrated by the statistical measures, the non-overlapping confidence intervals, and the different histogram shapes. With the 1831 texts it is really just the mean compression, conjunctives, paragraph length, and 'empirical positive' modality which are markedly discriminating.

We have seen different usage of our textual features and indicators by Faraday when he is writing about electromagnetism in private and public mediums. Differences are also plainly evident between the 1821 and 1831 material. Faraday is actively using language to construct different accounts of electromagnetism; what is written up in his published papers uses language to create the impression of a definitive, empirically pure account, straight from the laboratory and nature to the printing press.

This chapter has been, amongst other things, concerned with the linguistic construction/reconstruction of Faraday's 'discovery' of electromagnetism. Both the discovery itself and the concept being discovered are not found at one place, but are instead processes. This to say, they are interfused throughout a scientist's work and are situated within a wider context embracing social, historical, intellectual, and linguistic factors. The concept of scientific ideas being a network of organic, slowly developing moments and problems is explored by Holmes.[512]

I looked at the historical and intellectual background earlier in the chapter. To add to this process of construction/reconstruction we see how Faraday uses language to move from an ill-defined, messy set of ideas and thoughts in his laboratory notebooks to placing a concrete, discrete description of a natural phenomenon - electromagnetism - with reified and fixed meaning in the public idiom. For example, the two most 'truth-asserting' modalities are most in evidence in the published writings to ascertain things 'as they are' and greatly reduce scope for interpretative flexibility. This linguistic 'sleight-of-hand' is in keeping with the constructivist/post-structuralist theoretical underpinning of my research, where meanings are not fixed, but semantic closure and fact-assertion are key aims of formal prose by scientists, and other professions, such as law and medicine. This is because writing is the primary means to convince peers, and thus achieve the rewards of this - funding, status, and career advancement, for example.

Faraday, both in this chapter, and in the previous, seems to be striving to avoid what Coleridge argued in *On the Constitution of Church and State*, that:

[512] Holmes, Frederick Lawrence; 1985.

> "there is found even among our educated men a vagueness in the use of words, which presents, indeed, no obstacle to the intercourse of the market, but is absolutely incompatible with the attainment or communication of distinct and precise conceptions."[513]

For Faraday "purified language" was, as for Coleridge, "as much a spiritual as a material necessity."[514] This desire for a "factual rigour of expression" seems resolutely tied in with his religious beliefs (the Sandemanian Church played a key role in Faraday's life and work); he had: "recurrent concern to comprehend the natural world as a divinely created entity."[515] Faraday believed that a word's full and unambivalent meaning must be expressed. It was as if communication about Nature could be shorn of all theoretical baggage, and plain, neutral, and unadorned facts freed to be able 'to speak for themselves'. Cantor writes that:

> "This aspect of Faraday's work may be related not only to the Sandemanian emphasis on plain, true reading of the Bible but also to Genesis 2:20, where Adam contributes to each type of animal its true and accurate name."[516]

This may well correlate with Faraday's fairly consistent use of the 'empirical positive' modality in his published papers, as well as the 'electromagnetism' letters.

In conclusion, Faraday's private and public writings on electromagnetism have been scanned using a variety of textual features and indicators, and the results displayed statistically, and visually as confidence level intervals and histogram profiles. What was found is the ability to discriminate between Faraday's private and public writings on electromagnetism for both 1821 and 1831, and between the 1821 and 1831 texts, as summarized in table 1. It was also found that the letters had similar statistical and visual profiles to the published papers. It turned out that compression, conjunctives, 'empirical positive' modality, and paragraph length were the most consistent in enabling discriminations to be made.

8.7 Normal Ranges, Baseline and Typical Values for Faraday's Writings

Now that the empirical studies are complete, and before moving to my final chapter, I am in a position to produce normal ranges, baseline and typical

[513] Coleridge; 1830: 167, quoted in Schaffer, Simon; 1991 in, Fischer, Menachem and Schaffer, Simon; 1991: 208.
[514] Schaffer, Simon; 1991 quoted in, Fischer, Menachem and Schaffer, Simon; 1991: 208.
[515] Cantor, Geoffrey *et al*; 1991: 7.
[516] Cantor, Geoffery; 1991: 215.

values for Faraday's private and public texts (see chapter four, section 4.3), which was one original aim of my research. The values for Faraday's letters to scientists, non-scientists, and friends are given in table 8.7, and those for Faraday's notebook and published papers in table 8.8. The figures are for the textual features (chapter four, section 4.4) that proved to be the most successful indicators of the discriminations I wanted to make in chapters six to eight. The range consists of the lowest and highest values. A baseline value is a mean value, and a typical value the narrow range within which the majority of scores are found.

Table 8.7 Normal Ranges, Baseline and Typical Values for Faraday's Private and Public Writings.

	Range (ep)	Baseline Value (ep)	Typical Value (ep)	Range (cn)	Baseline Value (cn)	Typical Value (cn)	Range (ellipsis)	Baseline Value (ellipsis)	Typical Value (ellipsis)
Faraday's letters to scientists	2.67 - 15.00	0.95	4 - 7	0.72 - 36.59	0.47	5 - 8	3.23 - 61.09	10.75	10-20
Faraday's letters to non-scientists	1.37 - 29.41	2.63	5 - 9	0.53 - 45.45	1.44	9	2.56 - 57.69	12.22	15-30
Faraday's (early) letters to friends	1.09 - 30.00	2.61	6 - 13	0.34 - 35.71	1.26	2 - 14	3.33 - 61.19	18.83	8 - 22

Table 8.8 Normal Ranges, Baseline and Typical Values for Faraday's Notebooks and Published Papers.

	Range (ep)	Baseline Value (ep)	Typical Value (ep)	Range (cnj)	Baseline Value (cnj)	Typical Value (cnj)	Range (ellipsis)	Baseline Value (ellipsis)	Typical Value (ellipsis)	Range (pl)	Baseline Value (pl)	Typical Value (pl)
1821 notebook	1.32 - 35.71	3.13	3 - 13	2.94 - 26.67	4.61	5 - 12	3.57 - 43.82	17.94	13 - 28	10 -214	42.96	10 - 44
1831 notebook	0.72 - 30.0	3.71	3 - 16	1.72 - 37.50	7.50	2 - 25	3.03 - 51.89	17.91	5 - 30	10 - 219	42.11	10 - 90
Published papers (1821)	1.22 - 36.36	7.25	4 - 10	4.17 - 23.76	9.63	8 - 15	7.62 - 60.20	37.50	10 - 40	10 - 517	158.20	38 - 150
Published paper (1831)	1.31 - 30.0	5.87	2 - 11	2.22 - 33.33	9.41	3 - 17	3.85 - 58.33	28.07	5 - 44	10 - 297	90.72	10 - 90

9 The Final Cut: Conclusions and Discussion

We shall not cease from exploration
And the end of all our exploring
Will be to arrive where we started
And know the place for the first time.

Little Gidding V,
Four Quartets.
-- T.S. Eliot (1943)

Abstract. Having completed a pilot study, two main case studies, and three smaller studies, with texts ranging over approximately thirty years, I have investigated a large volume of private and public writings of Michael Faraday and Charles Darwin, and learnt a lot about their use of language, and its role in the scientific process. In this concluding chapter I will discuss my main findings and say why they are interesting. I shall also defend my methods for producing the results, then discuss limitations and draw implications for science studies, text analysis, and analysis of scientific discourse. I conclude that CLARITY has enabled me to discriminate between private and published writings of Faraday over time. Also compression, conjunctives, and 'empirical positive' modality have proved most effective in allowing discriminations to be made. Additionally I conclude that Faraday's personality, religious beliefs, and need for a controlled, safe environment are inextricably enmeshed with his language use and the science he constructed.

9.1 Introduction

In this final chapter I wish to meet a number of aims, primary amongst which will be an evaluation of my main findings, and the conclusions drawn from them, in relation to, for example, current science studies and text analysis. This chapter will also summarize what is unique about my methods and draw together the central themes of this research. The aims of this chapter will be accomplished in nine main sections. The first of which will provide a summary of each of my previous chapters, to be followed by an evaluation of the research themes, aims, questions, and theory. The next two sections are a summary, and then an interpretation and evaluation of, my main findings. My research had an overriding question to address: 'How is language used as a medium for thinking and communication?' What my results mean in relation to this question is the concern of the next section. The section that follows comprises an evaluation of my methods and theory, a discussion of the limitations of my methods, and an evaluation of Clarity. I then discuss the implications of my findings for science studies, the analysis of scientific discourse, and for text analysis. In the course of my research I have learned a lot about language, Faraday, and the role of text in scientific endeavours. Thus in the final section I extrapolate beyond the quantitative results to give my final thoughts and conclusions. This section has three parts, the first of which concerns meaning, asking questions such as: 'Is it possible to determine the meaning of a text?', and: 'How can quantitative analysis help us with language meaning?' This is followed by a drawing of conclusions about the intermeshing of language, science, and Faraday's religious beliefs. Finally, a discussion of what I have learnt, and concluded, about Faraday's own use of language; how he used language as a tool to construct his world and his science.

Before I move on I have a few remarks about constructivism and science studies. Disciples of science studies have been said to be quite eager to debunk empiricist views of how science works, and to contest whether it really works at all, and to do so solemnly and with no sense of apparent irony or reflexivity. They then cite as evidence that their views are grounded in unimpeachable empirical studies. I believe reflexivity and irony are important. My research has studied Faraday's use of language in the construction of ideas and knowledge. I have used language and rhetoric to do this and frame my findings; the analysis done of Faraday's work could be done on my research. On social constructivism I do not subscribe to its more extreme form, whereby the 'world and it contents' are passive and baggy, until given a 'sense of purpose' and shape through representation by social and discursive purposes; by human agency. This constitutes a dichotomy, a dualism, which is argued against in the writings of Bruno Latour and in Andrew Pickering's *The Mangle of Practice* (1995), for example. Additionally, my perspective on constructivism and history of science does admit that my descriptions are time bound and culture-relative. In terms of language the activist philosophy is central to my arguments, whereby language use is pragmatic, and has an active role in ordering the world and mediating between people and 'reality', rather than existing as structure that represents or reflects a 'true', singular reality.[517]

9.2 Summary of Chapters

The issues and questions my research was concerned with were laid out in chapter one, and which are evaluated in the next section. My theory of language, embracing activist/constructivism and post-structuralism, was then discussed, which emphasised the creative and intimate relationship between language and the world. Languages themselves are not static, free-floating entities but are actively constructed and are embedded within a wider social and historical fabric. Concomitantly, meaning does not reside passively in writing, rather meaning is the outcome of language in use. As I went onto giving evidence for in my case-studies, scientific language is no different. Ideas and facts are not static objects frozen in language, though scientific language is used to give the impression that it is a transparent medium through which incontestable 'truths' can be described and disseminated. Faraday's discourse tries to establish the (false) impression that what is being referred to exists independent of the story being told about it. As Barthes said the:

> "[F]act never has any but a linguistic existence (as the term of discourse), yet everything happens as if this linguistic existence were merely a pure and simple 'copy' of another existence, situated in an extra-structural field, the 'real.'"[518]

For example, the 'fact' of electromagnetism we saw change as Faraday's language use changed in his private and public documents, as well as over time (chapter eight).

The biography of Faraday in chapter one (section 1.3) described a brilliant experimental scientist whose life and scientific work were greatly shaped by the rigid boundaries of the socially isolated Sandemanian sect, whose tenets were based on a literal reading of the Bible. Another key factor affecting Faraday's life and scientific endeavours was his poor health. He suffered chronic physical illness, experienced one mental breakdown, and probably existed on the edge

---

517 Wavell, Bruce; 1986.
518 Barthes, Roland; 1986: 138.

of another.[519] Faraday married into an eminent Sandemanian family, the Bernards, and his marriage to Sarah proved to be childless. Though Faraday is thought to have adored children and written children's science books.

This was followed by a brief discussion of the importance of Faraday's use of language (section 1.4). He was an innovative user of language at a time when scientific language was highly standardized, and was in no little way responsible for the passing of the inductive style. This chapter continued with a discussion of pertinent work undertaken in the fields of analysis of scientific discourse and quantitative discourse analysis, and the relevance of this work to my research. One author I engaged with was Frederick Holmes, who points out the usefulness of studying a scientist's private and unpublished works in terms of understanding thinking and the processes of experimental practice removed from published writings. This is a theme threaded throughout my research. There have been many words written to critique quantitative studies of language. One critique argues that language meaning can only be approached through knowledge of the background context and information (section 1.5). I wholeheartedly agree with this, numerical results are but one part of the interpretative puzzle; a qualitative framework is needed to bring wider meaning to quantitative results. In this chapter I go beyond the quantitative results and apply them to contextual knowledge about Faraday and language use. Section nine is where I bring this to fruition.

Chapters two and three provided the bulk of the historical and contextual material for my research. In chapter two we saw the development of text manipulation from simple words lists to the first use of concordances in the Middle Ages (section 2.2.2). Then we had the first use of statistical text analysis for authorship attribution studies in the late 1800's by Corwin Mendenhall (section 2.2.7), the predominate use of stylistics up to the early 1950's (sections 2.27 and 2.2.8), when we first saw use of computer-assisted analysis. Father Bosa initiated the first computer-assisted humanities project in 1949 (section 2.2.11). Another key part of my context setting, and help with understanding where this research may fit into the wider scheme of things, was to look at how meaning and uses of language and science from a variety of perspectives,

[519] Cantor, Geoffery; 1991.

including, literary studies, feminism, history of science, and scientometrics (section 2.3). The context for my research was further developed in chapter three which contained two case studies: i) the use of discourse analysis as a research tool in science studies and ii) the quantitative analysis of text outside of science studies. The main findings were, first, that my analysis of the journal *Science Studies* from its inception showed a dearth of published research using quantitative methods for discourse/text analysis. Second, quantitative methods are plentiful, and much in use, in the humanities.

Chapter four presented a comprehensive account of the aims, methods, and resources involved in my research project. I also described my main case-studies that were written up in chapters six through eight. The empirical part of my paper began with the write-up of my pilot study in chapter five, which was first testing of textual features as potential indicators of discriminations I wanted to make. Twenty paragraphs from the beginning of seven private and public texts by Faraday and Charles Darwin were analyzed for compression, and using eight modal wordlists and three 'Gilbert and Mulkay' wordlists (from *Opening Pandora's Box*; 1984). The latter lists were composed of terms that Gilbert and Mulkay presented as either 'contingent' or 'empirical' (section 5.1). I found that the text features most sensitive and able to enable discriminations to be made, for example, between private and public texts, were the 'empirical positive' and 'empirical negative' wordlists (appendix four), as well as compression (section 5.4.1).

In chapter six I investigated Faraday's language use in his letters to scientists and non-scientists over time in terms of compression and modality. To recap, the aims of this case study were, first, to test the expectation that Faraday would demonstrate more compression in his more 'informal' texts - those letters dealing with 'non-experimental' issues - and to non-scientists, and that he would use the most 'empirical positive' modality in his more 'formal' letters and in the later texts. Second, I wanted to investigate whether the textual features would act as indicators of three discriminations I wanted to make, namely, Faraday's letters to scientists and non-scientists, early and later letters, and letters dealing primarily with 'experimental' and 'non-experimental' issues (section 6.4.1). What I found was that Faraday demonstrates a consistent use of a 'strong' and confident truth modality in his letters to scientists and non-scientists. Also though when the subject matter is non-experimental the compression scores are generally greater, in Faraday's later letters to the scientists the 'experimental' topic tends to produces the higher compression scores. Another main finding was that three features (compression and 'empirical positive' and 'contingent negative' modalities) acted as sensitive indicators of the discriminations I wanted to make. An evaluation of my main findings is found in section 9.4.

My next chapter contained three studies intended to test the main findings of chapter six, which were: i) Faraday's strong and consistent use of the 'empirical positive', and ii) the sensitivity of compression, and the modalities, 'empirical positive' and 'contingent negative'. I investigated the ability of these textual features to make three further discriminations, namely, Faraday's letters to scientists and to friends, letters from Sandemanians and non-Sandemanians, and Darwin's private (notebooks) and published writings (chapters of *Origin of the Species* and *Voyage of the Beagle*).

In chapter eight I investigated the role of language in the construction of Faraday's discovery of electromagnetic rotation and induction. For this study I also measured conjunctives, paragraph length, and three other wordlists: 'bench_words', 'any_experiment', and 'new_discovery' (section 8.3.2). The texts analyzed were *Diary* writings and the corresponding published papers for 1821 and 1831, as well Faraday's *Sketch* and letters discussing electromagnetism for this decade. I was also testing further the ability of my textual features to act as indicators of discriminations, such as between the Faraday's *Diary* entries and published paper on electromagnetism for 1821.

9.3 Evaluation of Themes, Aims, Questions, and Theory

I began chapter one with a description of two major themes of this research. One concerned the relationship between the production of new knowledge and ideas and language, and the other, an evaluation of the effectiveness of quantitative text analysis tools in studying this relationship. We have seen how Faraday's language changes, when, for example, there was more compression and truth modality in the published papers on electromagnetism for 1821 than the corresponding *Diary* entries (section 8.5.1).

Changes in language use can pinpoint important stages in the articulation of new ideas. Language being used to construct new knowledge rather than being a structure to describe a static body of knowledge and pre-given ideas (section 1.2). So, for example, Faraday's use of language in his papers and letters suggests that new knowledge or ideas become ossified as facts and deemed certain at those times when there is greater use of both the 'empirical positive' modality and of conjunctives, and less use of compression. This construction of facts is not a feature of notebooks where the use 'empirical positive' modality is less prominent and compression is greater. Thus we have an active, constructing relationship between knowledge, ideas, and language use. I have described how electromagnetism was not a concrete entity with specific characteristics for which the published paper is the medium for its unfettered description. Rather we saw how different depictions of electromagnetism were constructed linguistically by Faraday in different texts (sections 8.5 and 8.6).

It is also important to remember that changes in language, as well as the appearance of new knowledge and ideas, occur over time, developing gradually, and are 'buried' in the text (section 1.5.4). Thus the effectiveness of computer-assisted text analysis tools in studying the relationship between knowledge and idea construction and language is in enabling changes and patterns to be identified by facilitating the analysis of a large volume of an author's writings over their life. To do this by 'hand' would be very a tedious and time-consuming task.

This research had four original aims, which were to: i) analyze private and public/published texts, ii) test the sensitivity of a number of textual features to act as indicators to discriminate between different types of text written, primarily by Faraday, but also Charles Darwin. The discriminations I wanted to make in subsequent case-studies were, for Faraday, between: i) early and later letters, ii) letters dealing with 'experimental' and 'non-experimental' issues, iii) letters from scientists and non-scientists, iv) Sandemanian and non-Sandemanian letters, v) letters from friends and scientists, and vi) notebook and published writings on electromagnetism. For Darwin it was between private and published works: notebooks and book chapters. Then to, iii) investigate how amenable linguistic meanings are to quantitative-computational analysis, and iv) evaluate the text analysis programs.

At this point I ask: 'What has happened in trying to meet these original aims?' I have had access to a wide range of private and published writings of Faraday and Charles Darwin, over, in the case of Faraday, approximately thirty years. For example, in chapter eight I analyzed two sets of notebook writings, seven pieces of published material, plus correspondence, all concerned with electromagnetism. The wealth of material I had at my disposal provided a great deal of scope to meet the first aim. The other three aims are more specifically addressed in sections 9.4.3, 9.9.1, and 9.7.6 respectively.

My research asked two principal questions. The first, whether it is possible to investigate the question: 'How do new expressions and ideas arise in language?', using computational-quantitative methods of text analysis. I found that this is possible to an extent for certain reasons. First, we know that language is used to construct new expressions and ideas. The interesting question is whether we can capture features of innovation using quantitative methods. We have seen how Faraday's writings on electromagnetism placed different emphasis on various textual features in his private and published writings. Language use changed as different audiences were written for, and ideas, and knowledge, linguistically presented as facts palatable for public consumption. For example, the notebook for 1821 has approximately twice as much compression as the corresponding published papers (table 9.1). Similarly, the mean paragraph length in Faraday's 1831 electromagnetism paper is a little over double that of the notebook pages. Second, as Holmes has pointed out, ideas and discoveries do not appear at any one moment, rather they are infused throughout a scientist's work.[520] They arise in language as a slow growing, organic progress, as Gruber also believes.[521] Thus this requires the researcher to be able to study large text corpora, which, without computer-assisted analysis, would prove very difficult and time-consuming.

[520] Holmes, Frederick L; 1985.
[521] Gruber, Howard E; 1974

The second principal question had two parts, the first of which asked whether it is possible to expose the differences in laboratory activity and the experimental recounting in published papers.[522] This part of the question was affirmatively answered in chapter eight. We saw how it was possible to discriminate between Faraday's *Diary* entries on electromagnetic rotation and the two corresponding published papers for 1821. The features especially able to show the differences were compression, paragraph length, conjunctive, and truth modality. A similar picture was realized not only for Faraday's 1831 private and public writings on electromagnetic induction, but for the 1821 and 1831 electromagnetic writings. Thus we saw how the experimental work on electromagnetic rotation and induction was written about differently; we were able to use the textual features measured to discriminate between the private and public texts.

With the above in mind we can ask two further questions, first, how this goes beyond what I had done with the pilot study we started with? Secondly, how do the differences, or changes, I found relate to or assist our understanding of language and innovation? The pilot study argued that scientific discourse does not fall only into the dual repertoires devised by Gilbert and Mulkay (1984): the empiricist and the contingent. I constructed lists of modal terms, where those terms expressing the most robust truth-assertion were labelled 'empirical positive', and those indicating the greatest uncertainty, 'contingent negative'. How the findings in chapter eight go beyond what I had in the pilot study is in two ways. First, not only was truth modality able to demonstrate differences between the private and published writings, but other text features were as well, namely, compression, conjunctives, and paragraph length. Second, we saw changes in Faraday's language use in writing about electromagnetism in his notebooks, when describing experimental activity, and when he came to write up his researches for publication. Additionally, my main studies also went further by demonstrating Faraday's changing use of language over time. We saw this when analyzing his letters in chapter six, both in terms of the contents ('experimental' and 'non-experimental') and those sent to scientists and non-scientists.

[522] Gooding, David C; 1989 (a): 63-82.

If we now look at the second of our additional questions. How these changes and differences in Faraday's language use assist our understanding of language and innovation is, first, to show that they have a relationship of interdependence. Innovation, such as Faraday's work on electromagnetism, is not a sudden, single event nor a fully-formed object for which language is used for its description in a peer-reviewed paper. Rather, innovation is a process that develops gradually over time, which has a history and involves preparatory work. Its written representation in a published paper belies this. In chapter eight we saw how in his papers Faraday uses language to describe his work on electromagnetism, and the phenomenon itself, in concrete, certain terms. Whereas in the laboratory notebooks the language is less certain. This is evident when we compare the use of the 'empirical positive' modality in the notebook and published papers for 1821 and 1831, which we find more of in the latter texts (tables 8.2 and 8.3).

In the other part of the second principal question I asked whether it is possible to describe and reconstruct the human activity of scientific work through the study of the micro-history contained in private writings, which:

> "ordinarily follow along close enough behind the activity itself to provide a track still 'fresh' enough to trace the daily actions of those who have left their imprint there."[523]

I do not think this is possible. My results do not indicate that this is realizable. One of the achievements of this research was to use both private and public/published material. This research only investigated certain features of language, but we can never entirely know what comprises language, and language and writing can never completely capture what there is to a skilled performance[524] (though on reflection I do not think it has to be complete to allow something interesting and meaningful to be said). One reason for this is that private writings are always incomplete because, first, there will be gaps in the records of an experiment, and secondly, there will be the omission of tacit knowledge: "the skills, techniques, assumptions of which practitioners were either unaware, or which, by their very nature, could not be recorded in writings or drawings."[525]

---

[523] Holmes, Frederick L.; 1981: 65.
[524] Collins; 1985 and Dreyfus; 1986.
[525] Gooding, David C.; 1989 (a): 64.

An interesting question to arise from this issue is: 'Does the fine personal and contextual details have to be complete before something interesting can be said about a piece of writing or music?' It cannot be expected that the reader will know exactly what the author intended to write without falling into the trap of 'intentional fallacy', which suggests that it is a mistake to read a text in search of the author's intent, and that the text, once written, holds a meaning specific to itself. I do not think that it is necessary to know fine personal details and authorial intent because, first, these can never be know in their entirety, and second, the meaning of a text or piece of music or painting changes with the audience. It is possible to say interesting things about an author and their writing, if, as with this research, a close analysis of how ideas and concepts change as the language changes over time is undertaken.

Lastly, in this section I wish to evaluate the theory underpinning my research. My theory embraced a post-positivist, activist/constructivist philosophy of language, where language is not a static, passive structure used to describe or reflect 'objective' things or events existing 'out there'. Instead language is active and helps to shape, and is shaped by, the world we inhabit and has meaning through its use. The application of this theory to Faraday, science, and his writings is exemplified when we saw Faraday's published depiction of electromagnetism in comparison to that found in his notebooks and letters. There was no-one 'true' and authoritative account of electromagnetism reflected in language, rather Faraday consciously used language to construct electromagnetism as per the intended audience, his feelings towards Christianity, and his psychological need for order and to minimize ambiguity, for example (see sections 9.5.2 and 9.9.2). Language was one of many actants that actively persuaded people of the validity of Faraday's position in the scientific scheme of things, as well as express facts from which came Faraday's science as "an island of security on which he could take refuge from the storms of the outside world." [526]

[526] Cantor, Geoffrey; 1991: 287.

9.4 Summary of Main Findings

Before evaluating my main findings for chapters six to eight and defending my methods I wish to summarize these findings under three headings: i) numerical results, ii) histogram shapes, iii) text discriminations and the success of textual features as indicators of discriminations.

9.4.1 Numerical Results

In the case study in chapter six I investigated the use of compression and modality in Faraday's letters to scientists and non-scientists over a twenty-five year period (1825-1849). It was found that, first, Faraday's use of the 'empirical positive' modality in his letters generally increased over time for both the correspondence to scientists and to non-scientists, with the highest mean percentages in the 1845-49 period (table 6.6). For Faraday's correspondence to scientists these percentages were between 1.69 and 2.64, higher than the baseline value of 0.95 (table 8.7). In the case of letters to non-scientists the percentages ranged from 1.65 to 4.75, values that both undercut and overstep the baseline value of 2.63%. What I also found in chapter six was that Faraday's correspondence, except for the letters to Herschel sent early in Faraday's career, has about five to twenty percent more compression when he is dealing with a 'non-experimental' matter (table 6.8). If we recall the more compressed a text was the less repeated words it contained, the lower the compression score was, and thus the more compressed the text, which is a feature of a piece of informal or personal writing (section 4.4.1).

The compression for Faraday's correspondence to scientists was more prevalent in early letters. The mean compression score was 8.34% and 13.18% for the earliest letters and 13.37% and 16.42% for 1845-49 for Herschel and Whewell respectively (table 6.5). The compression found in Faraday's letters to friends in chapter seven was up to half that found in the letters to scientists (table 7.2). This was contrary to my expectation that Faraday's letters to the scientists would be more formal and assured, and this, *inter alia*, would be reflected in less compressed texts. We find a similar pattern for Faraday's letters to the non-scientists, except for Magrath (table 6.5). These changes in language use are one way to demonstrate the transformation in Faraday's confidence and stature as a scientist over time. These findings also show the importance of analyzing an author's writings over a long period of time.[527]

The first study in chapter seven aimed to check whether the strong use of 'empirical positive' modality in his letters to scientists exists in earlier letters to friends. I had expected that Faraday's correspondence to scientists would contain more 'empirical positive' modality (table 6.3) but it transpired that the mean and maximum scores for this modality were higher in Faraday's early letters to friends (table 7.2). Another unexpected result arose in the next study when it was revealed that the mean 'empirical positive' modality scores for the Darwin's 'M' and 'N' notebooks were 7.56% and 6.32%, compared with 2.71% and 3.78% for chapters one and two of the *Beagle* and *Origin* respectively. Though the amount of (mean) compression was a lot higher in the notebooks (compression scores 14.33% and 14.05%) than the published material (compression scores 27.81% and 38.55%). For Faraday's 1821 notebook and published texts on electromagnetism the former had more than twice the mean content of 'empirical positive' modality than the published paper: 7.98% versus 3.13%. Also the language use was more compressed in the notebook, with a mean compression score of 17.94% compared to 38.35% for the paper (tables 7.3 and 7.4). Correspondence to and from Faraday and non-Sandemanian scientists were investigated in the third, and last, study to comprise chapter seven. My expectation was that more 'empirical positive' modality (mean) would be present in the letters written by Faraday. This turned out to be the case for the correspondence dealing primarily with both experimental and with non-experimental matters. Finally, the letters from Faraday were less compressed than those written by the non-Sandemanian scientists (table 7.5).

Moving onto the chapter eight where my analysis included other textual features, that is, conjunctives, paragraph length, and the wordlists - 'bench_words', 'any_experiment', and 'new_discovery'. In table 9.1 I have summarized the main (mean) results for the texts analyzed in chapter eight (*sketch* is the whole document). I want to bring a few of these results to the fore. First, compression in the Faraday's 1821 notebook (17.94%) was more than

[527] Stubbs, Michael; 1996

twice that of the two published papers (38.35 % and 42.26%). Next, the presence of the other measured features in Faraday's *On some new Electro-Magnetical Motions, and on the Theory of Magnetism* paper is about twice that of the 1821 notebook. If we now look at the results for 1831 compression is greater in the notebook entries, and other features are found more in the published paper. Additionally, the compression in the letters was approximately nine percent less than the *Diary* pages for 1821 and 1831, but similar to the published papers (tables 8.2 to 8.7). Thus the textual features in table 9.1 have not only clearly allowed discrimination between the private and published writings (including the *Sketch*), but the results concur with my expectations for Faraday's private and published writings (matrix 4.1, section 8.32).

Table 9.1 Summary of Main (Mean) Results for Chapter Eight.

	ellipsis	conjunctives	paragraph length	modality ep	'new_discovery'	'any_experiment'
1821 Diary						
Mean [528]	17.94	4.61	42.96	3.13	1.66	2.69
Paper[529]						
Mean	38.35	11.00	156.1	7.98	3.10	5.43
Paper[530]						
Mean	42.26	5.04	211.6	3.45	1.29	2.50
1831 Diary						'
Mean	17.91	7.50	42.11	3.71	1.39	3.60
Paper						
Mean	28.07	9.41	90.72	5.87	2.43	4.67
Sketch						
Mean	35.40	10.43	141.86	8.27	3.01	5.53
Letters (1821-1831)						
Mean	26.47	6.38	104.3	4.09	0.87	2.79

[528] Truncated mean, where the Minitab removes the smallest 5% and the largest 5% of the values, and then averages the remaining values.
[529] *On some new Electro-Magnetical Motions, and on the Theory of Magnetism.*
[530] *Electro-magnetic rotation apparatus.*

9.4.2 Histogram Shapes

In my main studies I also represented my primary results as histogram shapes in order to try and discern whether Faraday's writings could be identified with unique shapes. How successful did this turn out to be?

In chapter six the histogram shapes for the 'empirical positive' and 'contingent negative' modalities, when Faraday is discussing 'experimental' matters in his letters to Herschel, are quite the opposite of those for the 'non-experimental' letters. This is interesting, first, because the histograms give us visual collaboration for the numerical results; the shapes are allowing discrimination to be made between these different texts. Secondly, these two modalities represent the two extreme ends of modality, the most unequivocal use of modality, the use of which frames a statement as very 'cut and dried'; as certain, factual, and authoritative. When we look at the corresponding mean scores for 1845-1849 Faraday does make more use of the 'empirical positive' modality (appendix five; tables 6.9 and 6.11) for his 'experimental' letters, though it is the other way around for the 1825-29 period. One explanation for this is that by 1845-1849 Faraday was one of the most eminent scientists of his day. Continuing with chapter six Faraday's correspondence to Herschel dealing mainly with experimental and non-experimental concerns showed both producing distinctively different shapes for the periods 1825-1829 and 1845-1849 for the compression scores (section 6.5.3; histograms 6.19.1 to 6.19.4).

The histogram shapes for compression scores in chapter seven (section 7.4.1) allowed discrimination between Faraday's correspondence to Herschel and Whewell and the earlier letters to his friends, Benjamin Abbott and Thomas Huxtable (histograms 7.1 to 7.5). The second study investigated private and publics writings of Faraday and Charles Darwin. We produced different histogram profiles for the 'empirical positive' and compression scores for Faraday's 1821 notebook and published writings on electromagnetism (for example, histograms 7.6 and 7.7). For Darwin, histogram shapes allowed discrimination between his 'M' and 'N' notebooks for his use of 'empirical positive' modality and compression (histograms 7.8 to 7.11). Interestingly, chapters from his *Beagle* and *Origins* produced compression score profiles that were very similar (histograms 7.12 and 7.13).

Moving onto chapter eight where I compared histograms profiles for corresponding private and public texts by plotting them on a single graph. This allowed us to visualize the clear differences in Faraday's language use when writing on electromagnetism, for both 1821 and 1831 (histograms 8.1 to 8.3).

To conclude, I think histograms representation of numerical results have proved a useful way of allowing visual discrimination between different types of Faraday's writings. We have seen distinct differences in Faraday's language use when he writes on electromagnetism in his notebooks compared to his letters and published papers, for both 1821 and 1831.

9.4.3 Text Discriminations and the Success of Textual Features as Indicators of Discriminations

A primary aim of my research was to investigate how sensitive the chosen textual features were in acting as indicators of discriminations between different types of writing. How successful were the textual features I measured at enabling these discriminations?

In chapters six, seven, and eight there were three categories of discrimination I wanted to make. Those in chapter six were between: early versus later letters of Faraday, letters to scientists versus to non-scientists, and letters with 'experimental' versus 'non-experimental' topics. In chapter seven the three discriminations attempted were between: Faraday's letters to scientists and to his friends, private and public texts of Faraday and Charles Darwin, and letters from Sandemanian and non-Sandemanian scientists. For chapter eight I tried to discriminate between private and public texts on electromagnetism, early and later writings on electromagnetism, and between letters and published writings on this topic. The results of how successful the textual features (see section 4.4) were as indicators of discriminations are summarized in tables 9.1 to 9.3.

Table 9.1 Ability of Textual Features to Indicate Discriminations for Correspondence Analyzed in Chapter Six.

Feature	ellipsis	ai	ei	ep	en	ci	cp	cn
letters to scientists v non-scientists	yes, generally higher for non-scientists in later letters	no	no	yes, generally higher for non-scientists in later letters	no	no	no	yes, generally higher for non-scientists in later letters
early v later letters	yes, higher in later letters	no	no	yes, higher in later letters	no	no	no	yes, higher in later letters
'experimental' v 'non-experimental' letters	yes, higher in 'non-experimental' letters	no	no	yes, higher in the 'experimental' letters	no	no	no	yes, higher in 'experimental' letters

Table 9.2 Ability of Textual Features to Indicate Discriminations for Chapter Seven.

	ellipsis	ai	ei	ep	en	ci	cp	cn
scientists v friends	yes, higher in letters to friends	no	no	yes, higher in letters to friends	no	no	no	yes, higher in letters to friends
Sandemanian v non-Sandemanian	yes, higher in Faraday's letters	no	no	yes, higher in Faraday's letters	no	no	no	yes, higher in Faraday's letters
Darwin: private v published	yes, higher in the notebooks	no	no	yes, higher in the books	no	no	no	yes, higher in the books

Table 9.3 Ability of Textual Features to Indicate Discriminations Among Faraday's Writings on Electromagnetism (Chapter Eight).

	ellipsis	ep modality	cn modality	paragraph length	conjunctives	'bench-words'	'any-experiment'	'new-discovery'
private v public 1821	yes, highest in the papers	yes, highest in the papers	yes, highest in the papers	yes, highest in the papers	yes, highest in the papers	yes, highest in the papers	yes, highest in the papers	yes, highest in the papers
private v public 1831	yes, highest in the papers	yes, highest in the papers	yes, highest in the papers	yes, highest in the papers	yes, highest in the papers	yes, highest in the papers	yes, highest in the papers	yes, highest in the papers
letters v published	yes, highest in the papers	yes, highest in the papers	yes, highest in the papers	yes, highest in the papers	yes, highest in the papers	yes, highest in the papers	yes, highest in the papers	yes, highest in the papers

Meaning of indicators (modal terms) for Tables 9.1 - 9.3:

ai = 'all inclusive', ei = 'empirical inclusive', ep = 'empirical positive', en = 'empirical negative', ci = 'contingent inclusive', cp = 'contingent positive', and cn = 'contingent negative'. Further information about the modal wordlists can be found in sections 4.4.2 and 6.2, and in appendix four.

In each chapter I was able to make the discriminations between Faraday's private and public texts that I wanted to. Though the textual features had varying success as indicators of the discriminations. A primary finding was that across my case studies compression and the 'empirical positive' modality proved to be consistently able to allow discriminations. This was the case between different genres of writing, documents dealing with a particular theme, whether 'experimental' topics in letters or electromagnetism, and early and later texts, such as correspondence. The 'contingent negative' modality also proved a reasonably sensitive indicator, and measurements of conjunctives and paragraph length were effective as discriminators between text types.

In chapter eight the eight textual features all allowed the discriminations to be made (table 9.3). Though when we look closer the significant findings are, first, that the mean compression score for both of Faraday's 1821 papers on electromagnetism was more than double that of the corresponding notebook entries (table 8.2). Also the mean paragraph length for these papers was over three times the mean length for the notebook, which contained approximately half as much 'empirical positive' modality and conjunctives as Faraday's 1821 *On some new Electro-Magnetical Motions, and on the Theory of Magnetism* paper. All three parts of Faraday *Sketch* were found to have two or three times the compression score, paragraph length, conjunctives, and 'empirical positive' modality than the notebook (section 8.4, tables 8.4 and 8.5). For 1831 the mean compression score, paragraph length, and 'empirical positive' modality in the published paper were up to twice that found in the notebook (section 8.4, table 8.3). Except for paragraph length and the mean compression score for the 1821 *Electro-magnetic rotation apparatus* paper Faraday's letters dealing with electromagnetism contains scores for the textual features that are not very dissimilar to those for the published electromagnetism papers (section 8.4, table 8.6).

Additionally the success with discrimination is seen in appendix six which contains the results for the analysis of variance/confidence levels for chapter eight where we clearly see how Faraday's language use in 1821 changes from his private writing to the public and formal writings. This is seen in terms of a lack of overlap for compression, conjunctives, paragraph length, 'empirical positive' modality', and the 'new_discovery', and 'any_experiment'. As well as for 'bench_words' wordlist between the laboratory notebook (D1) and the *On some new Electro-Magnetical Motions,* and *on the Theory of Magnetism* and *Electro-magnetic rotation apparatus* papers (EMp and EMp2) and the *Sketch* [sk]).[531] We see a similar change in language use for Faraday's private and public writings on electromagnetism for 1831. This is except for the 'bench_words' wordlist, where there is great overlap. Another interesting finding is that not only do the confidence levels for the laboratory notebooks for 1821 and 1831 overlap, but generally neither of them overlap with the 1821 or 1831 confidence levels for the published material. Thus a similar pattern of language use emerges for the private writings on electromagnetism over time, and a change in language use between private and published texts on electromagnetism over time.

9.5 Interpretation and Evaluation of my Main Findings

The main empirical work consisted of a number of studies, namely, those involving Faraday's letters to scientists, non-scientists, friends, and non-Sandemanians in chapters six and seven. The latter chapter also investigated private and public writings of Charles Darwin. These studies used compression and seven types of modality as potential indicators of the discriminations I was interested in. Then analysis of Faraday's writings - letters, notebooks, and published papers - on electromagnetism between 1821 and 1831 comprised the case study in chapter eight, which included textual features such as conjunctives and paragraph length. What follows in this section is an evaluation of my main findings under three headings. The first of which reassesses the private versus public domains dichotomy, then a comparison of the Sandemanian 'language of God' and Faraday's 'language of science' comprises the second, and the third looks at the linguistic construction of certainty.

[531] The exceptions are the overlap between D1 and EMp2 for compression and 'empirical positive' modality.

9.5.1 Reassessment of the Private versus Public Domains

My research findings and interpretations have brought about a rethinking of the perceived ‘private’ and ‘public’ division. What counts as a private versus public spheres will vary with different people. Faraday was very private person and was not socially outgoing, ‘consistently avoiding social gatherings’.[532] He was a private person in another sense, that is, his social space was constructed as a private/public dichotomy. The Royal Institution for Faraday was unequivocally divided up into private and public spaces. The private space, where Faraday constructed a controlled and safe environment, consisted of his living quarters and his basement laboratory, where he worked alone: “I do not think I could work in company, or think aloud, or explain my thoughts at the time.”[533] I would argue that any perceived separation between the private and public domains does not withstand closer inspection. As well as being Faraday’s home the Royal Institution was a public space; members came in to use the Library, to attend lectures, and hold social gatherings on Friday evenings.[534] For Faraday the private and public spheres “were not only adjacent, they overlapped”, there was a “fluid division between the two domains.”[535] Furthermore, Faraday was not working in strict isolation, in some intellectual vacuum. He was connecting and communicating with the world through his laboratory assistant, the management of the Royal Institution, public lectures he gave, correspondence he wrote and received, and papers he published and read. For as Bakhtin writes:

> “Truth is not to be found inside the head of an individual person, it is born between people collectively searching for truth, in the process of their dialogic interaction”.[536]

For Faraday the conventional notion of private versus public writing breaks down. We have seen how he used language in a particular way across private and published texts, as well as with friends and scientists. We have seen how use of the most confident and truth-ascribing modality - the ‘empirical positive’ - is used quite consistently across ‘experimental’ and ‘non-experimental’ topics,

[532] Cantor, Geoffery; 1991: 111.
[533] Faraday to Hansteen, 16 December 1857, in Williams, L. P. *et al* (Eds.); 2; 1971: 673, in Forgan, Sophie; 1985: 60.
[534] Forgan, Sophie; 1985: 61.
[535] Forgan, Sophie; 1985: 61.
[536] Bakhtin, Mikhail; 1984: 110.

with scientists and non-scientists alike, and over time. My original thinking was that the letters were to be classified as private and informal writings. This notion was based on the premise that Faraday would reserve his most stringent, unequivocal language for peer-reviewed publications addressing an audience of other scientists. Now we see, as with the compression results, Faraday's use of truth modality in the letters is as robust as in the prose of his published papers.

9.5.2 Language of God and Language of Science

Measurement of compression and the 'empirical positive' modality continued to act as indicators of the discriminations I wanted to make in all my main case studies. Additionally, conjunctives, as well as paragraph length, proved successful in making discriminations between the private and public electromagnetism texts in chapter eight. The conjunctives were more prevalent in the published papers than the *Diary* for both 1821 and 1831. This was my expectation because the role of conjunctives in a published scientific paper is to create seemingly causal links between actions, event, and objects. The experiment is being presented as a linear unfolding of 'natural' events with the role of the scientist filtered out; a positivistic accounting of experimental practices.

This conclusion, along with Faraday's consistent use of truth-affirming modality, is consistent with Faraday's notion of language which emphasized the use of 'plain style' and 'clear style' in writings, terms used by the Sandemanians in their elucidation of the Bible, to express incorrigible facts.[537] Faraday believed, as a Sandemanian, that God wrote the 'book of nature'.[538] Thus:

> "Just as the Sandemanians prize every word in the Bible as incorrigible truth not to be misinterpreted by man, so Faraday looked at scientific facts as the basic words or signs comprising the book of nature."[539]

These features of Faraday's language, not least the use of 'empirical positive' modality, are a form of rhetoric used not only to unambiguously express facts but also to persuade an audience of "the validity of the author's position",[540] which is intertwined with the truth claims of the discourse. Faraday's

---

[537] Cantor, Geoffery; 1985: 73.
[538] Cantor, Geoffery; 1985: 72.
[539] Cantor, Geoffery; 1985: 73.

psychological need for security and order required his relation to God and to science to be one that "curb[ed] excessive speculation."[541] For this to be successful his audiences would need to be convinced of his experimental accounts and results. This necessitated empirical facts. A highly rhetorical use of language was a prerequisite to construct these facts as authoritative and to persuade the audience of the validity of Faraday's position. I suggest that this is one explanation for the striking contrast in language use between Faraday's private (notebook) and published writings (sections 9.4.1 to 9.4.3), and his generally strong use of the 'empirical positive' modality across his various writings and over time.

Even though this research has not demonstrated a direct link between Sandemanian 'language of God' and Faraday's 'language of science' (with his consistently strong use of 'empirical modality'), I believe it is one possible hypothetical explanation for my findings, based upon a familiarity with the qualitative work of Geoffery Cantor. These two language types are used for the same purpose - to persuade people of the validity of their position, to create certainty and represent unadulterated 'truths'. Language for Faraday must not confuse or obfuscate, lest it lend itself to uncertainties and to misrepresent the word of God, for: "Faraday was committed to the view that all phenomena were produced by unbreakable laws framed by God at the Creation."[542] Concomitantly, Faraday habitually used language to express clear and exact 'facts' and the refinement of language was a hallmark of science so that: "the reference of terms becomes increasingly precise."[543] What is important to observe is the emphasis on a clear writing style in elucidation of the Bible and in Faraday's scientific prose.[544] Through writing about incontestable facts Faraday could avoid immoderate speculation about his hypotheses and his findings, and so keep faith with his believe in God as Creator.[545] This analysis confirms James' belief in the inseparability of Faraday's religion and his science, that "religion did indeed play a crucial role in Faraday's natural philosophy, his

540 Cantor, Geoffery; 1989: 161.
541 Cantor, Geoffery; 1991: 287.
542 Cantor, Geoffery; 1991: 287.
543 Cantor, Geoffery; 1985: 74.
544 Cantor, Geoffery; 1985: 75.
545 Cantor, Geoffery; 1991: 287.

natural science."[546] This is a view shared by Cantor, for whom Faraday "achieved a highly personal synthesis"[547] between science and religion.

Another possible explanation for the lucidity and precision in Faraday's 'language of science' is the accusations of plagiarism in 1820 and 1834. The need to avoid any further possible misinterpretation about the originality of his work and the authority of his findings may well have prompted Faraday to keep detailed laboratory notes and ensure linguistic exactitude in his published writings. Furthermore, I think Faraday's strong commitment to a literal reading of the Bible and the accusations of plagiarism were not separate issues, but a powerful combined influence on Faraday's science and his 'language of science'. Both necessitated minimal speculation and equivocation to be allowed. Faraday's scientific findings were meant to reveal nature as created by God. Any doubts could have put his relationship with the Sandemanian church on rocky ground. The central place and overarching influence of Christianity in Faraday's life made the possibility of expulsion from the Sandemanian sect a grave concern, especially with his brief expulsion in 1844. This event was believed to have damaged Faraday's health.[548] Similarly, I believe the 'scientific community' would not have brooked further reasons to doubt the veracity and uniqueness of Faraday's scientific work.

So we can see how Faraday lived among two communities – the Sandemanian and the scientific, both of which very strongly shaped his life, and tolerated minimal speculation and doubt. Faraday's primary way of communicating his science to the public was through his writings (and his lectures). Furthermore, Faraday could maintain his membership of these two communities through his very precise and focused use of language. This was an authoritative language that dealt in certainty, so reducing the possibility of speculation and doubt.

---

[546] James, Frank; 1991: xxx

[547] Cantor, Geoffery; 1991: 292.

[548] See Faraday's letter to Christian Friedrich Schoenbein, April 12 1844, in James, Frank A. J. L; 1993; letter 1575.

9.5.3 Linguistic Construction of Certainty

Finally in this section I wish to return to another use of scientific language found in published papers, that of being a 'certainty-producing technology'[549] (section 2.3.9), whereupon I posed the question: 'How is certainty being linguistically or textually created as we move along the private-public continuum?' Having now analyzed a number of different types of writings by Faraday and over a considerable period of time, I believe my results show that Faraday is actively constructing certainty both in published writings and over time as he becomes an established, respected, and ground-breaking scientist. Certainty in science writing requires one voice of 'scientific authority' existing throughout. This research has demonstrated that by looking at how an idea or concept is presented in language in different genres, at different stages of accounting and reconstruction, there is no-one 'voice' present throughout. We have revealed a different 'voice' - use of language - when Faraday is writing about electromagnetism in his notebooks as compared with the published material. Another example is the difference in Darwin's language in his notebooks and his published books.

9.6 Interpretation of What the Results Mean in Relation to my Overriding Question ('How is language used as a medium for thinking and communication?')

Language does not communicate to a public audience a private world of ideas and thoughts. Also we must keep in mind that sentences and ideas or thoughts do not neatly map onto each other. Instead there is a dynamic interplay between thought, communication, and action, which cannot fully be laid bare for analysis or judgement. Textual language is used to construct thoughts and ideas, while they in turn can be the impetus for new vocabulary. This conception of language, which is important for post-structuralists, is at the heart of my perception of the research's overarching question:

> "the distinction between language on the one hand and human thought and action on the other must be dissolved if human phenomena are to be understood as what they truly are, that is to say, elements of a communication system."[550]

---

[549] Thompson, James D.; 1967.
[550] White, Hayden; 1978: 230.

This research was occupied by an overriding question: 'How is language used as a medium for thinking and communication?' The belief in the existence of 'objective' knowledge and the possibility of unproblematically communicating it through language is something I got a strong sense of when analyzing Faraday's private and public texts. What we have seen is that Faraday used language as the means of communicating information in different ways. That is, to different groups of people, at different times, and in different types of texts. Language is used to communication thoughts and ideas to the public realm, for only then can consensus and shared 'certainties' be actively sought. For this an especially truth-asserting and authoritative must be consciously employed; we have witnessed Faraday's robust and consistent use of 'empirical-positive' modality. It is from this consensus that knowledge or an idea will gain 'truth' status and their originator social cachet and power:

> "Hardly any original thoughts on mental or social subjects ever make their way among mankind, or assume their proper importance in the mind even of their inventors, until aptly selected words or phrases have as it were nailed them down and held them fast."[551]

The positivistic philosophy of language, reproduced in the rhetoric of peer-reviewed scientific papers, places language as a mirror between us and reality. But the reality described in a scientific monograph is selective and an artefact, as Goethe said: "the words of science hide its substance". This is an argument for the disabling quality of language because it obscures so much of scientific practice, a point made by Gooding when he discusses types of reconstruction of scientific accounting.[552] Faraday in his important 1821 and 1831 papers used words to depict a different electromagnetism than we saw in the diary entries (chapter eight). Thus language is being used as a medium to create a particular version of reality for a specific audience, and to persuade that audience, in the case of a published paper, of the certainty and facticity of what is written. Word choice is not arbitrary, it is shaped by, and is a shaper of, the relationship between the communicators. Concomitantly, through language use a common reality is created, which can become normative, and communication brings about a co-ordination of activities.[553]

551 Mill, John Stuart; A System of Logic; London; 2; 1843: 285.
552 Gooding, David C.; 1990; 1992: 65-112.
553 Flores, F. and Ludlow, J. J.; 1980: 95-118.

Has anything been revealed about Faraday's thinking in terms of language use through the analysis of his writings? Is language a medium through which we can access thoughts? Thought and ideas are constructed and expressed in language, in texts. This relationship is never fully transparent or absolute. Language use can, to varying degrees, liberate or constrain thought. Thoughts: "may not be directly observable, because they are often a matter not of individual words, but of patterns of distribution and frequency."[554] Thus computer-assisted techniques allow us to make visible these patterns since we can now scan large volumes of text. I argue that contrary to distancing the researcher from the text computer-assisted analysis affords the opportunity to get closer to the text.[555] By this I mean that the larger the corpora the more likely patterns or features of language use will emerge and become evident. An analogy is the difference between standing beside a field to try to describe the lie of the land and that field's place in it, and achieving the same from atop a hill. The latter will give you far greater perspective. Two results of mine which demonstrate this are, first, Faraday's generally robust use of the 'empirical positive' modality across both different texts and time. Second, we found unmistakable differences between Faraday's language use in his private and published accounts of electromagnetism, for 1821 and 1831.

The important point here is about being able to 'observe' thoughts through patterns or particular uses of words. I would argue that the quite stark differences in Faraday's language between his private and public writings over time indicate that Faraday's thinking produced clear cut thoughts, wherein confusion, ambiguity, and doubt are minimized. Thinking that could be described as dichotomous. This supports Cantor's contention that Faraday "conceived many situations in terms of stark contrasts between black and white, rather than as slowly varying shades of grey."[556] Additionally a mind striving for order, certainties, exactitudes would fit Faraday's strong and consistent use, through different text types and over time, of the truth-affirming 'empirical positive' modality'.

[554] Stubbs, Michael; 1996: 235.
[555] Stone, Phillips, J; 1997.
[556] Cantor, Geoffrey; 1991: 285.

I found how Faraday used language differently in his private and public writings, and both of these forms of writing are incomplete accounts of experimental practice (section 1.1). Though as the letters and published papers were public documents, in terms of being the primary tool used to persuade and win support of scientific peers, their contents have the most 'distance' from actual laboratory activity. Thoughts do not typically arrive singularly and complete like train carriages, we can be thinking a number of thoughts at different stages of fruition; thoughts go though numerous stages and changes before being realized as words (or not all). Thus I argue that with Faraday's writing in his letters and published papers thoughts are a 'pure' or idealized version of the messier and less 'ripe' thinking at the experimental coal-face. Thinking is used to force the existence of thoughts that are more like unitary, homogenous units. Though language does not render thoughts transparent, for neither is deterministic of the other, and their relationship is dynamic and interactive.

9.7 Evaluation of my Methods

In this section I defend and evaluate my methods. First I explain my choice of methods and tools, and consider whether they were sensitive enough for my aims. Then I discuss how my methods and research were unique and interesting, followed by addressing the limitations of these methods. I conclude this section by evaluating the CLARITY/FAITH software tool.

9.7.1 Choice of Methods and Tools - Reasons and Explanations

In chapter three my study of qualitative and quantitative discourse analysis in the main science studies journal, *Social Studies of Science*, revealed four interesting findings. First, since the 1980's the number of qualitative studies have overtaken those using quantitative methods (section 3.1.3), and secondly, the variety of qualitative discourse analysis methods has expanded. Thirdly, relative to the expansion in the types of qualitative discourse analysis, quantitative methods have developed very little. This takes us onto the fourth finding, the barely visible use of computer-assisted methods in science studies. Certainly within the history of science such methods are not found. So the choice of a computational tool is in itself cutting edge.

Furthermore, the use of computational methods allowed the analysis of a large body of Faraday's private and public writings over a period of about thirty years. This is quite crucial for as Gruber said, the "solidification of new ideas is a relatively slow process".[557] This means that since the development of ideas does not happen overnight but is a gradual growth process throughout a scientist's experimental activities and their body of writings, the more language or text that can be studied the more likely significant and meaningful patterns will be revealed. In summation, to achieve this aim you need to meet two criteria: i) study different types of a writer's texts, and ii) a large volume of documents over a long period of time. This is what is missing from the work by Badalamenti that I reviewed in section 1.5.1.[558] Nor do we see these criteria being addressed in the work of Callon[559] and Latour.[560] Relying on a narrowly focused set of documents can lend analysis to the charges described in Tom Stoppard's *Arcadia* (section 9.8.3). As I have mentioned before, it would have been not practically possible for me to conduct large-scale analysis of the volume of documents I did without the application of a computational tool.

Central to my methods was the use of both 'private' and 'public' documents'. The former were originally perceived as non-published writings, that is, notebooks and letters. The latter were Faraday's papers and chapters of Darwin's books. One outcome of my research is a re-evaluation of the private-public division (section 9.5.1). The use of both these types of documents was for a number of reasons. It is not uncommon that when scientific writings are referred to this means published papers. Also these papers are often seen as the end-product of a linear scientific process, and separate from experimental work. My method highlights the existence of different types of 'scientific writing', which overlap, for example, we have seen Faraday discussing scientific matters in letters, laboratory notebooks, and in published papers.

My method for analyzing a variety of Faraday's private and public texts over approximately thirty years, as well as similar writings of Charles Darwin, did allow me to meet Stubb's methodological stipulation that:

[557] Gruber, Howard E; 1974: 114, quoted in Holmes, Frederick Lawrence; 1981: 66.
[558] Badalamenti, Anthony F. *et al*; 1994: 46-71.
[559] Callon, Michel *et al*; 1983, 1989, 1991.
[560] Teil, Geneviéve and Latour, Bruno; 1995.

> "the major criterion for text analysis is that individual texts or text fragments must be analyzed in ways which allow comparisons to be made to other texts and text corpora."[561]

9.7.2 Assessment of the Sensitivity of my Methods

My method generally was sensitive enough to allow me to make a number of discriminations between various types of text and for different time periods (section 9.4.3). The measure of ellipsis proved to be consistently capable of allowing discriminations to be made. Of the types of modality only the 'empirical positive' and 'contingent negative' turned out to be sensitive enough. We saw how in chapter six it proved possible to discriminate between Faraday's: i) letters to scientists and non-scientists, ii) letters on chiefly experimental and non-experimental matters, and iii) early and later letters over a twenty-five year period. The discriminatory power of compression and the 'empirical positive' modality was further tested in chapter seven, where it proved possible to discriminate between early letters Faraday wrote to friends and later correspondence to scientists. Additionally differentiation was achieved with private and published writings of Darwin, and correspondence to Faraday from non-Sandemanians. Finally, in chapter eight, Faraday's notebook writing's and his published papers on electromagnetism were differentiable from each other, for both 1821 and 1831, when compression, 'empirical positive' modality, and conjunctives, and paragraph length were measured. This was shown through the numerical results, histograms shapes, and analysis of variance graphs. Thus, to sum up, the use of the computational tool and measurement of particular text features and indicators outlined above enabled the exposure of stark differences in Faraday's writings over time.

9.7.3 Uniqueness of My Methods and Research

My research was novel, first, in using a computational tool to help analyze a large body of scientific writings over time. The application of computer-assisted methodology in science studies has been limited to mainly scientometrics and

[561] Stubbs, Michael; 1996: 131.

work done by Callon[562] and Latour.[563] The use of a computational tool allowed me to access, and so investigate, a very large volume of electronic texts, comprising published and unpublished documents written by Faraday and Charles Darwin. My analysis of a variety of writings, both private and published, underlines and substantiates the usefulness of this approach seen in the canonical work on Lavoisier by Frederick Holmes.[564] For instance, Holmes talks about how different documents can give different perspectives on the nature of the investigation, or a concept even in the context of my research. When I analyzed the language Faraday used in his writings on electromagnetism we found that the notebooks writings for both 1821 and 1831 were more compressed, contained less conjunctives and 'empirical positive' modality, and had shorter paragraphs than not only the published papers, but also the *Sketch* (table 9.1). Thus the private writing is a narrative discourse, less certain, didactic, and authoritative about electromagnetism. Whereas the perspective we receive from the published writings is that they constitute an unmediated, direct, pure descriptive account of experimental work and results. Through the language use what is suppressed is that:

> "The published narrative is the outcome of a complex process whereby an extended series of experiments is translated and condensed in prose."[565]

The perspective we find in the published narrative is unreservedly empiricist, wherein Faraday:

> "believed that facts, and only facts, are the basic signs of nature and the foundation on which the whole edifice of science has to be constructed."[566]

My research is also interesting in having advanced beyond work done by Holmes, Gruber, Badalamenti, and Bolton and Roberts (section 1.5). That is to say, the FAITH programs have allowed me to analyze a number of aspects of Faraday's language use over a large and varied corpora of his writings, so allowing the comparison of texts of different types and across a period of twenty five years or so. Additionally, I believe my research gains on computer-assisted work undertaken by Callon[567] and Latour.[568] This is in the sense that I have had

[562] Callon, Michel *et al*; 1983, 1989, 1991.
[563] Teil, Geneviéve and Latour, Bruno; 1995.
[564] Holmes, Frederick; 1984 and 1987.
[565] Cantor, Geoffery; 1989: 160.
[566] Cantor, Geoffrey; 1991: 200.
[567] Callon, Michel *et al*; 1983, 1989, 1991.
[568] Teil, Geneviéve and Latour, Bruno; 1995.

to grapple with a large volume of Faraday's writings quantitatively and qualitatively over time, as well obtain biographical and cultural knowledge. Thus my results have come out of a much richer engagement with my material than I see in the works of Callon and Latour I cite. Additionally, my research is unique among Faraday scholars as this is the first large-scale computer-assisted study of Faraday's writings and language use, about which I have garnered original insights and perspectives.

Different types of discourse have been classified using quantitative methods of text analysis.[569] My methods were able to classify and discriminate between different types of writing for Faraday and Darwin (sections 6.6, 6.7, and 7.5). What other studies have not done is address the ways in which language changes as part of the process of scientific innovation. This was central to my research. My findings demonstrated how Faraday's use of language clearly changed between writing about experimental and non-experimental matters (section 6.6). As well as between letters to scientists and non-scientists (section 6.6), and between later correspondence to scientists and early letters to friends (section 7.4.1). This feature was also demonstrated when I measured and compared the use of textual features for Darwin's 'M' and 'N' notebooks and chapters from his *Origins* and the *Beagle* (section 7.4.2). How language changes as an integral part of the scientific process was strongly seen in the case-study on electromagnetism in chapter eight.

Another relatively unique feature of my research method has been the use of longitudinal analysis; the introduction of a temporal element, which I argue increases the explanatory power of empirical analysis. This is because longitudinal analysis provides a bridge or a means of fusing together quantitative and qualitative methods. What I mean by this is that by studying Faraday's texts over time I was able to analysis a large volume of his writings comprising different types of discourse. Thus able to get a better grasp on Faraday's ideas which grow slowly over time. The computerized tool enabled flexible and quick measurement of various aspects of Faraday's language use, which coupled with biographical/qualitative knowledge of Faraday, has allowed insights and conclusions to be drawn across a wide temporal landscape.

[569] Gilbert and Mulkay; 1984.

9.7.4 Limitations of my Methods and What I Could Not Achieve

A computational tool allows a large volume of texts to be scanned relatively quickly, but this time saving has to be balanced with the vast amount of preparation that is still required before a text is digitized ready for analysis. This involves the initial inputting of text, whether using keyboarding, OCR scanning, or digital imaging, which must then be proof-read. Creating an electronic text can be a time-consuming and expensive process. Another limitation is that texts which are specifically created for one program cannot easily be used with other software. [570] Almost every text analysis system has its own interface and query language, thus it can take time to learn about their capabilities and how to use them.[571]

Originally I proposed a number of textual features to be measured as indicators, that is: i) ellipsis, ii) use of the active and passive voice, iii) paragraph, sentence, and word lengths, iv) structured (nested) keyword list searches (for modal words, 'technical', and 'grammar' lists, such as conjunctives and adverbs), v) the construction of lexicons from words which differentiate texts to be used to generate a frequency distribution of the introduction of new words in any text, and vi) the waiting time (the distance (number of words) between new words or key-words). As my research progressed this list changed. The pilot study demonstrated that the investigation of compression and the 'empirical positive' modality would prove useful in my main case studies. I decided to use conjunctives in chapter eight after reading research done on texts by Fawley.[572] For technical reasons use of the passive and active voice, construction of lexicons, and measurement of the waiting time were not doable, the necessary programs were not developed during this research project. I did not think measurement of word length would prove useful in terms my research aims and questions.

In chapter four (section 4.2) I proposed that this research could test the conjecture that an author's writing style will ossify as she/he progresses through

---

[570] General Handbook; CETH Summer Seminar 1996; Section Two: 61.
[571] General Handbook; CETH Summer Seminar 1996; Section Two: 61.
[572] Fawley, William; 1987.

their career,[573] that is, 'certain words and patterns become increasingly preferred.'[574] It has turned out that my research has not specifically addressed this issue, though my results demonstrate that over the course of his writings Faraday quite consistently and robustly used the 'empirical positive' modality. Elsewhere I argue why this was so (for example, section 9.5.2). Additionally we saw quite stark differences between Faraday's private and published accounts of electromagnetism for both 1821 and 1831 (section 9.4.1).

In terms of my theory, to demonstrate more definitely the 'shared narrative' or dialogic nature of language a more fine-grained approach to empirical work might have achieved this.

9.7.5 Concluding Remarks

The impression I am left with is that computerized-quantitative analysis is a valuable addition to a researcher's chest of text analysis tools. With the software I have used, and others like it, we may be seeing "the emergence of a new view of language, and the technology associated with it".[575] This has to be tempered with realization that computer assistance is only one aspect of any meaningful analysis; it "does not bring pure objectivity to text analysis."[576] The measurements, frequencies, and patterns generated by computerized analysis do not tell us a great deal on their own. They have to be related to wider qualitative, contextual knowledge, such as, in Faraday's case, the powerful predominance of religion or his chronic ill-health. Another aspect that the results of computer assistance need to relate to is the findings of other pertinent research, which I attend to in section 9.8.1.

9.7.6 Evaluation of Clarity

My assessment of the FAITH/CLARITY software will be done through a discussion of five other quantitative computer-assisted text analysis tools

[573] Brainerd, Barron; 1979: 6.
[574] Tallentire, D.R.; 1976: 315.
[575] Sinclair, J. McH; 1991: 1.
[576] Stubbs, Michael; 1996: 154.

currently available. These will be TACT 2.1, TEXTPACK,[577] WordSmith Tools, MonoConc, and the Oxford Concordance Program. This assessment will draw upon my hands-on experience with text analysis software tools whilst attending the Centre For Electronic Texts In The Humanities 1996 Summer Seminar[578] and workshops at the CTI Text Analysis Centre at Oxford University.[579] Before my comparison and evaluation a brief description of each of these tools.

TACT (Text Analysis Computing Tools) is a text-analysis and retrieval system for MS-DOS with fifteen programs, which builds an index of the text from which all subsequent information can be retrieved. The occurrence of the user's selected words are displayed within the KWIC (Key Word In Context) format, frequency distribution, or with collocates:

> "Typically, researchers use TACT to retrieve occurrences of a word, word pattern, or word combination. Output takes the form of a concordance, a list, or a table. Programs also can do simple kinds of analysis, such as sorted frequencies of letters, words or phrases, type-token statistics, or ranking of collocates to a word by their strength of association."[580]

TACT is designed to analyze individual literary texts, "or small to mid-size groups of such texts, such as Chaucer's poetry",[581] which have been tagged or marked up.

TEXTPACK was developed at The Center for Survey Research and Methodology (ZUMA) in Germany, which was established in 1974, and is one of three institutions which constitute GESIS.[582] Computer-assisted text analysis forms one of the basic research areas within ZUMA:

> "The practical goal of the current activities of text analysis in ZUMA is the conception, modeling and implementation of a tool for computer-supported text analysis, which supports both quantitative and qualitative analysis methodologies"[583]

This software has been designed over a number of years to "cope with many aspects of computer aided text analysis and most of all content analysis."[584]

---

[577] Home Page; http://www.zuma-mannheim.de/software/en/textpack/; accessed October 4, 1997.
[578] CETH Home Page: http://www.ceth.rutgers.edu/; accessed March 15,1998.
[579] Home Page: http://info.ox.ac.uk/ctitext/; accessed March 15, 1998.
[580] http://www.epas.utoronto.ca:8080/cch/TACT/tact1.html; accessed March 15, 1998.
[581] http://www.epas.utoronto.ca:8080/cch/TACT/tact1.html; accessed March 15, 1998.
[582] Home Page; http://www.social-science-gesis.de/index-e.htm; June 2, 1997; accessed October 4, 1997.
[583] Computer-assisted text analysis;

The current version 5.0 is designed for MS-DOS, and a version for Windows 95/NT was released this February. A wide variety of procedures are available, and I wish to briefly describe four. The first of which is SENTENCE, which is "the main input procedure to transform ASCII character (text)files into a TEXTPACK system file." Second is FREQ, a procedure that "counts word frequencies for the entire text or its sub-units and calculates the type-token ratio (TTR)." Thirdly, WORDCOMP which "directly compares the vocabularies of two texts", and fourthly, "TEXTPACK contains special-purpose procedures which allow you to categorize/classify/tag any kind of text according to so called "content analytic dictionaries.""[585]

The next package is WordSmith Tools for Windows 3.1/95, a suite of programs for text analysis and manipulation. Version 2.0 appeared in November 1997 and its primary functions are, first, WordList, which generates word lists from one or more texts 'by frequency and by alphabet',[586] and includes: "'cluster' word list[s], in which the entries are made up of sequences of words as they appeared in the texts."[587] Second, Concord can display a concordance for any specified word or part of word. Thirdly, Collocates allow collocates to be created and word clusters to be identified. Additionally, WordSmith Tools can "identify key words in a particular text and create a database of keywords to enable identification of key keywords and associated words."[588]

MonoConc is a commercial concorder for Windows and Mac. It provides a KWIC display and it is possible to initiate concordance searches for words, phrases and parts of words. Lastly, the Oxford Concordance Program is a versatile "general purpose tool for generating concordances, word lists, and indexes from texts in any language or alphabet."[589]

---

http://www.zuma-mannheim.de/research/methods/en/textanalysis/; July 2, 1997; October 4, 1997.

584 http://www.zuma-mannheim.de/software/en/textpack/overview.htm; accessed March 15, 1998.

585 http://www.zuma-mannheim.de/software/en/textpack/overview.htm; accessed March 15, 1998.

586 http://info.ox.ac.uk/ctitext/resguide/resources/w135.html; accessed March 15, 1998.

587 Sardinha, A. P. Berber; http://info.ox.ac.uk/ctitext/publish/comtxt/ct12/sardinha.html; July 1996; accessed March 15, 1998.

588 http://info.ox.ac.uk/ctitext/resguide/resources/w135.html; accessed March 15, 1998.

589 http://info.ox.ac.uk/ctitext/resguide/resources/o125.html; accessed March 16, 1998.

One achievement of this research has been to robustly test the CLARITY text analysis programs using the writings of Faraday. The computer tool I was using was under development, whilst the other computer-assisted text analysis tools I have used, and describe above, are commercial, industrial-strength, in active use by educators and students. Also these programs have pull-down menus or short-cut commands to access functions, while at the time of my research it was necessary to type in the syntax for each query as described in the introduction to my main case-study chapters. Though this was eased by being able to use a 'copy and paste' facility. It is proposed that in a later development each function will be able to be called up in a short-cut manner, for example, using pull-down menus.

The text analysis software I used in this research was marked by its flexibility. The programs can analyze any untagged, plain text, as well as support searches through large corpora, and can be adapted by users. Text can be inputted to the CLARITY directory from other databases, through OCR scanning, keyboarding, or from the Internet. The library of programs is extensive and continues to grow. It is difficult to compare the CLARITY text analysis programs with the packages I have discussed above. CLARITY is the most flexible in terms of being able to add to or modify the library of text analysis functions. The ability to scan texts for their content of user-defined wordlists is not supported by the tools I describe above, a function I found useful in my research.

TEXTPACK is a powerful text analysis application and I think the WORDCOMP function could be very useful. The direct comparison of the vocabulary of two texts would be an effective way, for example, to reveal further differences in Faraday's private and published writings on electromagnetism. Also by comparing the whole vocabulary of texts over time this could show when words are introduced and also disappear from use.

9.8 Implications of my Findings

I now wish to discuss the implications of my findings, first, for science studies, then for the analysis of scientific discourse and, lastly, for text analysis.

9.8.1 Implications for Science Studies: The 'Linguistic Turn' in Science Studies

What are the implications of the research findings for Science Studies? As well as a general discussion, I shall relate my results to the work by Badalamenti, Zahar, Holmes, Bolton and Roberts, and Gooding discussed in chapter one, and also to the discourse analysis of Gilbert and Mulkay.

Science studies, which includes history of science, is one component of modern human studies, for which language - discursive practices - constitute the primary medium through which we shape our environment and our place in it. Science studies has made some progress in executing a 'linguistic turn', which is to:

> "relinquish the view that the language of science....is simply a means of representation, separable from and irrelevant to the 'content' of science. Such a move carries with it the implication that historians should no longer seek to penetrate *through* the linguistic practices of scientists, to isolate ideas and conceptual structures in their minds, and should instead start to investigate those verbal and textual practices themselves.[590]

My research has taken this turn. Language pervades all human behaviour, and, as such, linguistically-orientated studies offer the opportunity to connect science studies with a wide variety of other contemporary disciplines.[591] There have been several textual practices of Faraday's that I have isolated. These rhetorical practices are deployed by Faraday to persuade the audience of the objectivity, or 'factual', status of what he has 'described'. What this research has shown is the need to analyse as large a corpus of a scientist's writing as possible, using both private and public/published texts. What can this achieve? It is much more meaningful to analyse writing over time where any patterns or habits are more likely to be revealed. A concomitant point, which I briefly investigated in chapter seven, is that creativity or the process of 'discovery' is not to be found, fully articulated and transparent, at any one point. Instead it is to be infused throughout a scientist's work. To take this further, a discovery develops slowly over time, it is not a sudden event, rather the discovery has a history and

[590] Golinski, Jan V.; 1990: 111-112.
[591] Golinski, Jan V.; 1990: 110.

numerous tacit factors will be involved in its growth.[592] Put another way, discovery is a human and cultural process.

If we are seeing progress, or evolution, in science, its foundations are linguistic in nature, that is, the application of new linguistic solutions to descriptive problems; the development of descriptions for particular kinds of phenomena.[593] We have followed scientific innovation in the work of Faraday and analyzed how the language changes. This is important because published papers do not tell the full, unpurged story of scientific work.[594] My findings demonstrate that scientific accounts and texts have rhetorical qualities. My research challenges cherished distinctions between scientific fact and textual production.

The study by Badalamenti *et al* was designed to determine whether there are mathematical laws of word usage; language perceived as a closed physical system (section 1.5.1). In a study investigating one work of each author such laws may well be revealed. Though to extrapolate such a finding beyond the confines of the research project, as Badalamenti seems to do, is erroneous. Language meaning is dynamically intertwined with language use, and is constrained by how others use the language; by social and linguistic conventions.[595] As Badalamenti, as well as Callon and Latour believe, this cannot be measured and no properties can be rendered static for interpretation as laws. "The richness of language is in its context", Teil and Latour rightly state.[596] They used a computerized tool (the Hume machine) to create a network of co-occurrences that "enables us to keep open a great number of characteristics of the context."[597] My impression is the context being evoked here is only that of the text, the discursive space of co-occurrences. But that context is shaped by a larger context of social, historical, and pragmatic factors. In contrast, my research studied Faraday's construction of ideas over time in private and public writings, within the wider social and cultural context. I tried to demonstrate in Faraday's language the belief that his membership of the Sandemanian Church had a strong influence on his scientific method and his

592 Gruber, Howard E; 1974; Holmes, Frederick, L.; 1986.
593 Bazerman, Charles; 1988, Bruce, Gregory; 1988.
594 Medawar, Peter B. 1987: 220-221.
595 Stubbs, Michael; 1996: 3.
596 Teil, Geneviéve and Latour, Bruno; 1995: 11
597 Teil, Geneviéve and Latour, Bruno; 1995: 14

construction and use of language.[598] When I compared the degree of compression and truth modality in Faraday's letters with those from non-Sandemanian scientists, these textual features were found to be the most prevalent in Faraday's writing.

The work by Zahar is informed by the constructivist philosophy of language, as is my own research, whereby language is seen as not describing "the way the world is". [599] Rather language is constructed and its use is an integral part of experimental practice, as has been argued by, *inter alia*, Gooding.[600] The way in which Faraday used the textual features measured in my research supports this. In his laboratory notebooks dealing with his experiments on electromagnetism Faraday described his thoughts and new knowledge within a particular literary space. We then saw how this work was re-constructed in his published papers, with different emphasis on the textual features. Compression in his notebook was more than double that found in the corresponding published papers for 1821. Also the mean 'empirical positive' modality and the use of conjunctives in Faraday's 1821 paper - *On some new Electro-Magnetical Motions, and on the Theory of Magnetism*, was more than twice that found in the notebook (table 9.1). Thus the experiment is not complete and then described in a neat, hermetically-closed, and neutral form, which is the impression given when we just consider peer-reviewed scientific papers. Instead the experimental work includes different uses of language which shape and re-shape the physical laboratory experimentation until language is used to bring artificial closure in the form of a peer-reviewed published paper. The greater use of 'empirical positive' modality in published papers is a rhetorical device employed to construct certainty, limit interpretative flexibility, and, in Faraday's case, prevent excessive speculation about God-given 'laws of nature'.

The relationship between writing and scientific thought, and the value of studying personal, private texts, are major concerns of Holmes.[601] One of my findings supports Holmes' belief that different documents can give different

[598] Cantor, Geoffery, 1991.
[599] Rorty, Richard; 1989: 20.
[600] Gooding, David C. 1990.
[601] Holmes, Frederick L.; 1981, 1984, 1987.

perspectives on the nature of the experimentation and the concept being investigated. How this is useful is in terms of the point I made earlier in this section about discovery being a human and cultural process. This is to say that 'discovery', along with experimental work, are not 'all-or-nothing' events which are fully realized at any one moment. By looking at different versions of writings on, say electromagnetism, we see how the knowledge, the idea is being constructed differently through textual language. Another point about 'discovery' is that it also depends on finding a problem significant enough to be labelled an important achievement. Faraday positively selected one of the most important cutting-edge problems of his time: how magnetism could be used to induce electricity.

In chapter one I was not convinced by Bolton and Roberts' hypothesis that scientists write both scientific prose and literature.[602] Having studied work by Faraday and Darwin my views have not changed. Scientific prose and literature are both types of writing; constructed for particular audiences, and instilled with different rhetorical devices. They are both literary and social spaces enabling a variety of agendas through building public consensus and shared certainties.

How does my form of analysis of scientific discourse bear upon the two repertoires employed by Gilbert and Mulkay (1984)? Both of these types of discourse analysis reveal the variability of scientists' discourse; variability in relation to the social context. I argue that when Faraday was writing about glass experiments to Herschel or about electromagnetic induction in his notebooks these literary spaces also constituted different social contexts. These produced differences in Faraday's language which were demonstrated through measurements of the various textual features used. What my research has shown is that there are a number of text genres that encompass Faraday's writings. These include letters to scientific peers, letters to friends, notebooks entries on electromagnetism, and published papers. Evidence was found that Faraday's language use had variations in each of the text genres. It was found possible to use certain textual features as indicators of discriminations between: i) text genres, ii) texts written at different time periods, and iii) texts dealing with

[602] Bolton and Roberts; 1995: 295-302.

different content. I think it is for these reasons that the differences are useful even without them necessarily being statistically significant.

9.8.2 Significance of My Research for The Analysis of Scientific Discourse

I now want to discuss the bearing of my research on the field of text analysis of scientific discourse. The hallmark of my research - computer-assisted study of scientific discourse - is largely unexplored territory in science studies. The majority of discourse and text analyses in science studies have largely been qualitative. More quantitative and computer-assisted analysis of scientific discourse has been done by scientometrics and literary studies (section 3.2.5). This research has also highlighted the importance of paying close attention to earlier drafts of work and private writings on an experiment or concept, in conjunction with published writings.

This research has concentrated upon textual language and science as a literary form. The nature and role of language is not to transmit pre-packaged scientific knowledge, rather it is a system or space for the construction of knowledge and for achieving consensus, and thus truth. It is also pertinent to remember that, as stated by Carnap, “a language is a system of habits of human beings”. [603]

I was able to work with not just one text or even one genre of texts but several. We saw how electromagnetism was constructed differently in laboratory notebooks, letters, and published texts. The ‘final’, published version of an experiment and the results is constructed such that the actions of the scientist are rendered passive, and the site of agency is passed onto non-human entities. This writing form, as with all literature, employs rhetoric, in this case, to remove the writer’s voice and agency linguistically. Woolgar makes the point that:

> “The discovered object is to be apprehended as neither the product nor the artful creation of the scientists; scientists came upon these objects rather than creating them.”[604]

[603] Carnap, Rudolf; 1937: 3.
[604] Woolgar, Steve; 1993 (1988): 75.

With the non-published material Faraday's active role in the making of the science was still there. The author/scientist was returned to the text. This made for quite a different perspective on Faraday's published work.

Also, as discussed in section 9.5.1, I have come to re-evaluate the notion of 'private' and 'public' writings and believe that, first, these are not separate types of writings. All writing is social and all texts are public; writing is always addressed to a particular audience.[605] As Gooding (1990) suggests, supposedly private experience is *inherently social*:

> "Recent historical and sociological studies of science also show that social and procedural aspects of observation in the local, laboratory context are necessary to explain consensus about observations. However they do not address the passage from personal experience to public discourse, so as to be incorporated into our talk and thought about the world. Nor do they explore the role of individual observers in constructing representations of novel aspects of the natural world and attributing common meanings to those aspects."[606]

Writing is also social in the sense that it has social impact which is based on its ability to constitute certain social subjects:

> "The social, political, moral subject is constituted by and exists through certain discourse, or using the terms of Foucault, discursive apparatus produces a certain mode of subjectivity."[607]

I have suggested, in chapter five, how Faraday constituted himself through his writing. One way of mapping Faraday's evolving status as a scientist is through his increasing use of certain textual features, especially truth modality and compression, over time. Such texts exude authority, certainty, and confidence, thus helping to constitute and socially position Faraday as a 'great man of science'.

Another point to consider is that we should not be using our energies to try and reconstruct the investigative pathways *lying behind* published papers. This is because the peer-reviewed paper is only *part* of the scientist's experimental

[605] Bakhtin, Mikhail; 1984: 110.
[606] Gooding; 1986: 206.
[607] Volkov, Vadim; http://lucy.ukc.ac.uk/csacpub/russian/volkov.html; August 27, 1997.

practice and literary effort. There is no separable or quantifiable *behind* or *in front*. To believe otherwise is to run the risk of breathing fresh life into the twin spectres of dualism and logico-positivism, which typically ignore the pre-histories and social processes of experimental practice and discourse. Faraday may have been focusing on electromagnetic rotation - the *foreground*, which sat among the *background* of wider linguistic, scientific, and personal practices and experiences, but you cannot try to give meaning to one without changing the other. Also they can both, at different, times background and foreground each other.

9.8.3 Implications for Text Analysis: In The Beginning Was The Word

> Bernard: Yes, one of my colleagues believed he had found an unattributed short story by D. H. Lawrence, and he analysed it on his home computer, most interesting, perhaps you remember the paper?
> Valentine: Not really. But I often sit with my eyes closed and it doesn't necessarily mean I'm awake.
> Bernard: Well, by comparing sentence structures and so forth, this chap showed that there was a ninety per cent chance that the story was written by the same person as *Women in Love*. To my inexpressible joy, one of your maths mob was able to show that on the same statistical basis there was a ninety per cent chance that Lawrence also wrote the *Just William* books and much of the previous day's *Brighton and Hove Argus*.
> Tom Stoppard, *Arcadia*, 1993: 19.

At the end of my research I find myself grappling with the question: "How amenable are qualitative literary qualities to quantitative/computer-assisted methods?"

As we have seen computer-assisted text analysis began in the 1950s and early 1960s. Before then statistical literary analysis was predominantly used for authorship attribution studies. The inception of machine-based studies and electronic texts swept aside the problem of limitation in corpus size. Though through the 1960's and 1970's computer scientists and software designers were both designing systems constructed on the belief that all text analysis should be implemented to analyse smaller textual units, such as sentences.[608] Contemporary computer-assisted text analysis, from the 1980's, has been much more interested with meaning and the role of language.

The early use of quantitative and computational methods and tools took a logico-positivistic view of the relationship between language, meaning, and the

[608] Renear, Allen; 1992: 221-248 (222).

world. There has been a move away from grand theories to more grounded, pragmatic, and micro-scale research in the social sciences over the past 25 years.[609] This has been the case for text analysis, which in the 1960s attended to the theories of, *inter alia*, Laswell and Kaplan (1950), Parsons and Shills (1951), and Osgood *et al* (1957). My research has been an example of this trend. I have concentrated on specific, focused experiments.

The software has enabled the analysis of large amounts of Faraday's private and public writings. This has allowed me to observe things and uncover patterns which would not have been possible with the eye alone. Thus the text has changed and generated insights.[610] The important point is that use of text analysis software enabled the search for "patterns which could not be observed unaided."[611] That is, "the language looks rather different when you look at a lot of it once."[612] I have analyzed a large volume of data, thus avoiding a criticism made by Stubbs[613] that for a lot of text analysis 'rarely are whole texts analyzed'. Stubbs put this complaint, applicable to both qualitative and quantitative analysis (section 2.2.11), thus:

> "Linguistics has traditionally been restricted to the investigation of the extent of language which can comfortably be accommodated on the average blackboard."[614]

Critics of quantitative studies of language argue that any quantitative analysis of texts is reductionist, in that the material is broken down into manageable chunks, thus resulting in a divorce between meaning and form. I think this criticism is subverted where a 'large amount' of text can be analysed. But the question remains: "What constitutes a 'large amount' of text?' I do not think there is one answer to this issue, for much depends on the research questions and issues. For this research what constitutes a volume of text large enough to justify my particular interpretations of results are, first, that the writing ranges, more or less, over the author's working life (my material ranged from 1811 to 1849), and second, that different types of writing - private and published - are

[609] Stone, Phillip, J; 1997: 47.
[610] Stubbs, Michael; 1996: 232.
[611] Stubbs, Michael; 1996: 44.
[612] Stubbs, Michael; 1996: 100.
[613] Stubbs, Michael; 1996.
[614] Phillips, M.; 1989: 8.

analyzed. For Faraday I investigated letters, notebooks, the *Sketch*, and published papers, as well as notebooks and book chapters of Charles Darwin.
It seems important for any text analysis to look at the social functioning of texts and of language, which is at the core of all texts. This is a conclusion I have reached having analyzed a large corpora of Faraday's private and public writings, and considered how Faraday and his work are culturally and historically situated.

At the end of my study of Faraday and Darwin's writings I find myself agreeing with Roberts, who writes:

> "quantitative text analysis is good for drawing inferences about contextual and text-based variables, where contextual variables may include indicators of source, message, channel, and audience, and where text-based variables [include] the frequency with which themes occur".[615]

9.9 Final Appraisal and Final Thoughts

This research has studied how language use changes in the scientific process through the investigation a number of textual features over time, and tested how able they are to act as indicators of discriminations between, for example, notebook versus published writings on electromagnetism. I have come to understand a lot about what language is and how it can be used; use being the primary criteria in shaping linguistic meaning. Scientific prose, like any other genre of writing, imparts narrative and rhetorical conventions. These conventions are employed in the "meaning-making process"[616] to try and establish interpretative authority for the author, the scientist. We have seen Faraday consistently use certain textual features over time. Two of the most prevalent were compression and strong truth modality. Language is not a neutral space in which unbiased or wholly original intentions can be played out. Nor is language a mirror of reality. As Bakhtin says, there are:

> "no neutral' words or forms - words and forms that can belong to 'no one'; language has been completely taken over, shot through with intentions and accents...Each word tastes of the context and contexts in which it has lived its socially charged; all words and forms are populated by intentions...Language is not a neutral medium that passes

[615] Roberts, Carl W.; 1997: 283.
[616] Bizzel, Patricia; 1986: 176.

> freely and easily into private property of the speaker's intentions; it is populated - overpopulated - with the intentions of others."[617]

I wish to close this final chapter, and my thesis, with three sections as way of final conclusions. The first looks at the issue of meaning, the second, the relationship between Faraday the scientist, religion, and language. While the third addresses my thoughts on how Faraday used language.

9.9.1 What Does It Mean?

A concern of this research has been with meaning. In this section I want to reflect on how amenable linguistic meanings are to quantitative-computational analysis. I consider two questions: 'What I have learned from the numerical results about the content and meaning of written texts?, and: 'How can the analysis of patterns of words in a text contribute to an understanding of the meaning of the text?'

Meaning has only been a primary interest of text analysis since the end of the 1970s. an interpretation of a text is achieved by a peculiar synthesis of reader and author; a text does not have an unchanging, unified meaning. A primary role of the language is to try to achieve this; avoiding the multiplying of the number of legitimate interpretations of a text. We have seen this with Faraday's writings.

It seems to me that Faraday was using language in order to maintain a confidence in absolute standards of meaning. But meaning is not just one thing. What science does is operationalize an oppositional, dualistic, either-or model of meaning; something is scientific or non-scientific. Meaning is relational and provisional; context-dependant. It is more about how something relates to a variety of other things. We have seen how meaning is linguistically constructed in the main case-studies and changes with the content, over time, and with regard to the intended audience. For example, I was clearly able to discriminate between different linguistic constructions of electromagnetism both across text types and in different time periods. These different electromagnetisms meant

[617] Bakhtin, Mikhail; 1981: 293-4.

something different for the different targeted audiences. My conclusion is that the peer-reviewed scientific paper is being used to linguistically centre meaning as unitary, static, and irrefutable; (peer-reviewed) scientific meaning as the default. Whereas what I have demonstrated is that meaning changes with context: the text type, intended audience, and different time periods. Also the larger the corpora a pattern or frequency of words is spread through, and the more often this pattern or frequency arises, the more powerful will be the explanatory power of our analysis.

Another point about meaning concerns the important role played by beliefs. That is, the central role played by beliefs in making meanings possible. This central role of beliefs is highlighted in Faraday's case, whose Sandemanianism pervaded all aspects of his life. In chapter seven the main results of a comparison between letters written to and from Faraday and non-Sandemanians were a constantly higher use of the 'empirical positive' and 'contingent negative' modalities, as well as compression, for Faraday's letters (section 7.4.3).

I argue that the numerical data constitute one 'reading' of Faraday and Darwin's writings. For it is committing an 'intentional fallacy'[618] to believe that an author's intentions and meaning are available or transparent to us for judgement without external corroborating evidence. Though we cannot make meaningful sense of numerical data without investigating the context within which the text originated.[619] The meaning of Faraday and Darwin's writing does not reside in the numerical data. Their meanings would be construed by the audience; meanings are negotiated between the text and the audience. We saw how the textual features measured varied if Faraday's audience was other scientists, friends, or non-scientists. Barthes perceived texts as something without a fixed meaning apart from that produced by readers in the act of reading.[620] This contingency is described by Thompson, for whom meaning:

[618] Wimsatt, W. K. and Beardsley, Monroe C.; 1954.
[619] Stone, Phillip, J; 1997: 37.
[620] Barthes, Roland; 1977.

> "is not a stable or invariant property of a linguistic product, but rather a multi-layered and fluctuating phenomenon which is constituted as much by the conditions of production as by the conditions of reception."[621]

Fish proposes that rather than having a text that contains formal features identifiable in all times and places that it is the reader that projects these features onto the text.[622] Fish claims that it is the 'interpretive community' that creates its own reality. It is the community that invests a text with meaning.

It is the contention of the constructivists, post-structuralists, and feminist critical theories that meaning does not, and cannot, reside in the text:

> "Every utterance is potentially the site of a struggle: every word that is launched into social space implies a dialogue and therefore a contested interpretation....Language cannot be neatly dissociated from social living; it is always contaminated, interleaved, opaquely coloured by layers of semantic deposits resulting from the endless processes of human struggle and interaction."[623]

Contemporary science, from the end of the 1400's, has used language to convince people that the above is not the case. Swales has emphasized the importance of the role of language in science; the "sheer importance of the writing".[624] Scientific writing, that is, the consciously constructed 'final version' of an experiment, has enabled the creation of facts, the membership of the 'scientific' community, and procurement of research funding and promotion. Writing is many things, but one thing all writing has in common is that of being rhetoric; the art of persuasion is a skilled performance. Faraday's scientific writing is no exception as I discussed earlier (section 9.5.2).

Thus we have dialogue taking place between Faraday and his peers. I find myself in agreement with Bakhtin and the Slavic School, for whom meaning does not reside with the individual nor with no-one, but in our dialogic exchange of language. Faraday, in his writings, was in dialogic exchange with numerous audiences, such as, scientific peers, friends and confidants, and those who had accused him of plagiarism. So meaning is both a social process and multivocal.

While it is the case that an understanding of the socio-historical context of the production of the text is required, "the meaning of a text is not wholly constituted

---

621 Thompson, John B.; 1984: 65.
622 Fish, Stanley; 1980.
623 Selden, Raman and Widdowson, Peter; 1993: 127.

by them."[625] For the meaning of a text is also mediated by the internal, structural properties, such as grammar, syntax, and style.[626] We have seen how Faraday, for example, predominantly used a high degree of truth ('empirical positive') modality relative to other scientists and non-Sandemanians. To take this further, meaning is also construed through consensus, thus:

> "if a considerable number of people interpret a text in a certain way then it should be possible to show what it is about the language of the text that causes them to do so."[627]

What is most important in giving texts meaning is intertextuality, that is, the numerous connections made between texts, for example, influences, sources, allusions, traces of previous texts, archetypes. The relevance of this point for my research is that it emphasises the importance of analyzing and comparing a large corpora of texts and writings over time. A larger textual landscape is opened up over which patterns or features, such as the contrast between Faraday's language use in his private and published writings on electromagnetism, are exposed for analysis. In a broader sense intertexuality emphasizes the interconnectedness among the texts of a particular culture, period, or author, for instance. There is no private, autonomous, omniscient authorial authority; no author is an oracle of truth. Nor is it possible to capture or reconstruct all of the connections, though I do not think we should try before claiming to have found something original.

Finally, meaning is more likely to reside in the conventions of meaning, the traditions, the cultural codes which have been handed down, so that insofar as we and other readers (and the author) might be said to agree on the meaning of the text, that agreement would be created by common traditions and conventions of usage, practice and interpretation. In different time periods, with different cultural perspectives (including class, gender, ethnicity, belief and world-view), or with different purposes for reading no matter what the distance in time or cultural situation, competent readers can arrive at different readings of texts.

---

[624] Swales, Jones, M.; 1990: 127.
[625] Thompson, John B.; 1984: 66.
[626] Thompson, John B.; 1984: 66.
[627] Thorne, James, in Peer, Willie Van; 1989: 285.

9.9.2 Michael Faraday - Science, Religion, and Language

Throughout my analysis of Faraday's writings I have seen the presence of a common feature. Faraday wrote in a very precise, assured manner, where uncertainty did not have a large part to play. This has been especially evidenced by a dominant use of truth-asserting modality; a language use employed to reduce antithetical views and persuade people. I think Faraday had great need to be persuasive in and through his public narrative. The accusations of plagiarism in 1820 and 1834 left their mark on Faraday. A public image of honesty, openness, and selflessness is a vital component of being persuasive. For the natural sciences, which have come to be the yardstick for objectivity and truth, the powerful use of persuasiveness is crucial to maintaining this socio-cultural standing, and the concomitant benefits, such as research funding and University Chairs. As Cantor believes:

> "modern scientific prose has become the most potent instrument of persuasion in our culture."[628]

Charges of plagiarism can impact quite dramatically on how your peers and the public judge the validity of your beliefs and statements. Faraday's precise, controlled, and persuasive use of prose was inseparably intertwined with the unyielding discipline of the Sandemanians, for which: "there was almost no role for uncertainty or human fallibility."[629] He used language a mirror of reality: God's reality. Faraday harboured a number of fears, which concerned and required his need: "to maintain control of himself within a safe, ordered environment."[630] These fears included his concern that he might be a hypocrite: "the hypocrite is a split personality who is neither true to himself nor his Father but is consumed by his own lies."[631] Another fear was causing confusion in his writing.

For Faraday language use was of great importance in meeting this need 'to maintain control of himself within a safe, ordered environment.' His description, in precise terms, of experimental work and 'objective facts' assisted Faraday to

---

[628] Cantor, Geoffery; 1996: 161.
[629] Cantor, Geoffery; 1991: 272.
[630] Cantor, Geoffery; 1991: 273.
[631] Cantor, Geoffery; 1991; 273.

convince the community of scientists of the veracity of his august status as an experimentalist. Thus Faraday was able to progress 'deeper' into the controlled and safe world of science and find good reason to return almost daily to his basement laboratory of the Royal Institution. This scientific world of perceived certainties helped to control Faraday's fears and doubts and make his psychological character acceptable to the stringent demands asked by membership of the Sandemanian church. The sect in turn also provided Faraday with a safe haven and controlled environment away from external dangers.

In this chapter I have to tried to demonstrate the active role language use plays in the construction of knowledge and ideas. Also how Faraday's science was certainly shaped by personal and social factors, such as his need for order and avoid ambiguity, and his religious beliefs and membership of the Sandemanian church.

9.9.3 Faraday and Uses of Language

To conclude this chapter and my thesis I wish to summarize Faraday's own perception of his language use and my own conclusions about his use of language use.

Faraday believed, as a Sandemanian, that the 'book of nature' was written by God.[632] Thus:

> "Just as the Sandemanians prize every word in the Bible as incorrigible truth not to be misinterpreted by man, so Faraday looked at scientific facts as the basic words or signs comprising the book of nature."[633]

Language for Faraday must not confuse or obscure, lest it lend itself to uncertainties and to misrepresent the word of God, for: "Faraday was committed to the view that all phenomena were produced by unbreakable laws framed by God at the Creation."[634] Concomitantly, Faraday habitually used language to express lucid and exact 'facts', and the refinement of language was a hallmark

[632] Cantor, Geoffery; 1985: 72.
[633] Cantor, Geoffery; 1985: 73.
[634] Cantor, Geoffery; 1991: 287.

of science so that: "the reference of terms becomes increasingly precise."[635] What is important to observe is the emphasis on a clear writing style in elucidation of the Bible and in Faraday's scientific prose.[636] Through writing about incontestable facts Faraday could avoid immoderate speculation about his hypotheses and his findings, and so keep faith with his belief in God as Creator.[637] By presenting God's laws as unquestionable through his experimental results Faraday avoided any challenge to the Sandemanian literalness in interpreting the Bible.

I believe Faraday's need for a controlled and predictable world was constructed and maintained by a number of factors, of which one of the most important was language use. One way this was achieved was by enabling a simplification of the relationship between the self and the world; a perceived world certainties and facts manipulated through the medium of experimental practice. Through this means, Faraday was able to live out his maxim 'Write Publish Finish'.

---

[635] Cantor, Geoffery; 1985: 74.
[636] Cantor, Geoffery; 1985: 75.
[637] Cantor, Geoffery; 1985: 287.

Appendix One: A Current Computer-Assisted Text Analysis Project

In this appendix I will briefly discuss a recent text analysis project, which involves a disputed 'final' elegy by Shakespeare.

The 1612 578-line Funeral Elegy by W.S.[638] for a murdered friend (thirty-year old Oxford student William Peter) was introduced to Shakespeare studies by Donald Foster, who had suspected in 1983 that it may have been written by Shakespeare. In his *Elegy by W.S.: A study in attribution* (1989) Foster furnished arguments for and against Shakepeare's authorship of the valedictory poem, having studied similarities of diction, vocabulary, versification and syntax.[639] There the matter rested until with prompting from Richard Abrams Foster took up the challenge again, building up new evidence for Shakespeare as the author of the *Elegy*. In April 1995 Foster, presented an augmented case of quantitative evidence to the Shakespeare Association of America, and then on December 30th to a special session of the 111th Convention of the Modern Language Association, backed by Professors Richard Abrams, Stephen Booth, Lars Engle, and Leo Daughty.

For his computer analysis Foster developed a database called SHAXICON, which is:

> "a lexical database that indexes all of the words that appear in the canonical plays 12 times or less, including a line-citation and speaking character for each occurrence of each word...[and it] electronically maps Shakespeare's language so that we can now usually tell which texts influence which other texts, and when.[640]

The lexical, grammatical, and syntactical similarities between the Funeral Elegy and Shakespearean texts were collated using the Retrieve software. The methodology was computer-generated statistical analyses of stylometric variables, such as the use of articles, nouns, and 'em's' (meaning 'them').

---

[638] see appendix seven.
[639] Abrams, Richard; 1996: 25.
[640] Foster, Donald; http://www.clark.net/pub/tross/ws/shaxicon.html
More information is found at the SHAXICON Web site: http://vassun.vassar.edu/~foster/shax/intro.html

Then in February 1996, Foster read a new *roman à clef* about the 1992 Clinton administration, the anonymous political satire, 'Primary Colours'. *New Yorker* magazine asked Foster to uncover the author, who turned out to journalist Joe Klein.[641] In his confession and apology, published by Newsweek, Klein stated that Foster had developed "a pretty good program".[642]

[641] *Primary Culprit*; New York ; 26 Feb. 1996: 50-7.
[642] Klein, Joe; 1996: 76

Appendix Two: Current Computer-assisted Linguistics and Text Analysis Research Projects

The table below lists a selection of computer-assisted linguistics and text analysis projects most relevant to the nature of this research project:

Institution	Where (Including URL)	Current or Primary Project(s), and Software
AI Research Group	School of Computing, Staffordshire University; http://www.soc.staffs.ac.uk/research/groups/ai/	A primary area of research is in Natural Language Understanding, where the aim: "is to develop a method for automatically representing a piece of text as an abstract model...using the cognitive theories of human understanding." The first step is to develop: "a lexicon that is : sense based and rich in world knowledge. The second step is the identification of the concepts within the text. Finally tools for discovering semantic and contextual relationships between these concepts are used to construct a model of the text."

Institution	Where (Including URL)	Current or Primary Project(s), and Software
Bristol Stylometry Research Unit	Faculty of Computer Studies and Mathematics, University of the West of England; http://gate.uwe.ac.uk:8000/csm/stylometry/index.html	The research team are involved in statistical analysis of literary style. The research interests are: 1) "The application of vocabulary richness measures to asphasic speech." 2) "QSUM methods." 3) "The application of neural networks to stylometry." 4) "The provenance of De Doctrina Christiana, attributed to John Milton."
Centre for Computational Linguistics	Katholieke Universtiteit Leuven; http://www.ccl.kleuven.ac.be/about/About.html	During 1992 and 1993 this centre, in conjunction with the Commission of the European Communities (funders), and other EEC linguistics institutes, were involved in a project entitled: 'Formal Semantics for Discourse'. The concern was with, for example, the computational processing of natural language texts, and text linguistics.

Institution	Where (Including URL)	Current or Primary Project(s), and Software
Centre for Computer Analysis of Texts (CCAT)	School of Arts and Sciences, University of Pennsylvania; http://ccat.sas.upenn. edu/ETS-home.html/	This Centre stores electronic texts and has text-management software available for PC, DOS, and Macintosh. An example is the Transcribe program, which is used for global search-and-replace changes in a text file.
Centre for Computing in the Humanities	Faculty of Arts and Sciences, University of Toronto, http://www.cch.epas. Utoronto.ca.8080/cch/ cch.html	MTAS (Micro Text-Analysis System), which: "produces batch word-frequency lists, distribution and density graphs, and type-token statistics." STRAP (Structural Analysis Program): "tool for seeing aspects of the structure of a text not found literally in its surface features, but rather manually encoded in it with "tags", and for counting and graphing these elements."[643] TACT (Text Analysis Computing Tools) a text-analysis and retrieval system for MS-DOS.

[643] http://www.cch.epas.utoronto.ca.8080/cch/tact.html

Institution	Where (Including URL)	Current or Primary Project(s), and Software
Centre for Electronic Texts in the Humanities (CETH)	Universities of Princeton and Rutgers; http://www.ceth.rutgers.edu	One major project is the 'Text Encoding Initiative', sponsored jointly by the Association for Computer and the Humanities (ACH), the Association for Computer Linguistics (ACL), and the Association for Literary and Linguistic Computing (ALLC). The task of which "has been to develop and disseminate guidelines for the encoding and interchange of machine-readable texts among researchers, and to make recommendations for the encoding of new texts."[644]
Centre for Literary and Linguistic Computing	University of Newcastle, Callaghan, Australia; URL: http://www.newcastle.edu.au/departments/lc/index.html	This centre's work involves: "computer assisted comparisons, often statistical, among literary and non-literary texts", which concentrates upon post-medieval texrts in English and French.

644 http://cethmac.princeton.edu/CETH/TEI.html

Institution	Where (Including URL)	Current or Primary Project(s), and Software
Centre for Research in Machine Linguistics and Natural Language Processing	Academy of Sciences, Bucharest, Romania; http://www.cogsci.ed.ac.uk/elsnet.survey/survey.html[645]	GRADE: "A graphical editor for graph-based knowledge representation schemes (taxonomies, semantic nets, and conceptual graphs.)"
Centre for Text and Technology	Academic Computer Centre, Georgetown University; http://www.georgetown.edu/lforte/ACC.html	Offers electronic texts with some textual analysis tools, such as, Word Cruncher, Folio News, and Micro-OCP.
Computational Linguistics/Machine Translation Group	University of Surrey; http://www.surrey.ac.uk/ELI/eli.html	The current project is entitled 'Test Suites for Natural Language'
Computer Fund of Russian Language	Institute of Russian Languages, Russian Academy of Sciences, Moscow; http://www.cogsci.ed.ac.uk/elsnet/survey/survey.html	"Distributional statistical analysis of texts, which is aimed at the discovery of hidden structures lying behind the text"

[645] http://www.ling.rochester.edu/departments.html, http://www.cog.brown.edu/panters/lingdir/universities.html, http://www.ims.uni-suttgart.de/info/SitesEuropes.html

Institution	Where (Including URL)	Current or Primary Project(s), and Software
Computer Linguistics Group	Department of Computer Linguistics, University of Zurich; http://www.ifi.unizh.ch/groups/hess/CLpage.html; http://www. ifi.unizh.ch/groups/hess/CLresearch.html	One current project is the development of: "a robust passage retrieval engine based on linguistic principles" for the linguistic analyse of texts (natural language based documents and passages). As the linguistic analysis of texts is not sufficiently advanced to deal with full texts, "small fragments of text must be chosen that convey the core meanings (i.e. the referential information) of the document yet are amenable to, analysis by computer (i.e. mostly but not exclusively noun phrases)."
Consortium for Lexical Research (CRL)	Computing Research Laboratory, College of Arts and Sciences, New Mexico State University; http://clr.nmsu.edu/Home.html; http://clr.nmsu.edu/clr.bin/clrcat/TOC	Software products include 'Text Categorization': "statistical software for recognising relevant part of texts".

Institution	Where (Including URL)	Current or Primary Project(s), and Software
Corpus Linguistics Group	University of Birmingham, U.K; http://clg1.bham.ac.uk/ tagger.html	This Group offers a 'Part-of -Speech Tagging Service for English texts via E-mail (tagger@clg.bham.ac.uk).
CTI Centre for Textual Analysis	Centre for Humanities Computing, University of Oxford; http://sable.ox.ac.uk/~ ctitext2/resguide/ electron.html	The Oxford University Computing Services is integrated with the 'Centre for Humanities Computing', embracing the 'CTI Centre for Textual Analysis'. CTI, offers textual analysis software in three areas. 1) General textual analysis tools (18), which includes 'Oxford Concordance Program' (OCP) and 'Micro-OCP'. 2) Specific textual analysis tools (9). These include 'Construe', a package that allows on-line parsing primarily with Greek texts. 3) Qualitative analysis tools (8), such as 'Hypersoft', which: "includes features for summarizing, annotating, categorizing, mapping, coding, and quantifying data."[646]

[646] http://sable.ox.ac.uk/~ctitext2/resguide/contents.html

Institution	Where (Including URL)	Current or Primary Project(s), and Software
Dataflight Software	http://concord.lax.primenet.com/	This company offers Concordance for the internet and for Windows.
Department of Computer Science	Cornell University; http:// www.cs.cornell.edu/	One project is called Kenmore: "a general framework for domain-specific knowledge acquisition for conceptual sentence analysis."[647] Kenmore performs on corpora, for example, part of speech tagging, word sense tagging, concept tagging, part of speech disambiguation, and word sense disambiguation.
Department of Informatics	Moscow State University Research Computer Center; http://www.cogsci.ed.ac.uk/elsenet/survey/survey.html	The construction of a: "computer system for derivational analysis" of texts.
English Language Institute (ELI)	University of Surrey; http://www.surrey.ac.uk/ELI/eli.html	The Institute is researching discourse analysis: "in the area of contrastive rhetoric - the relationship between first and second paragraph organization [and] rhetorical strategy."

[647] http://www.cs.cornell/Department/Annual94/Faculty/Cardie.html

Institution	Where (Including URL)	Current or Primary Project(s), and Software
Erasmus Programme On Natural Language Processing	Coordinated by the Institute for Language Technology and Artificial Intelligence (ITK) of Tilburg University. Erasmus; http://itkwww.kub.nl:20280/itk/ITK; http://itkwww.kub.nl:20280/itk/Docs/Projects/Erasmus/brochure.html	This is an Inter-University Cooperation Programme on Natural Language Processing. The research is concerned with linguistic knowledge representation and the design of natural language processing systems.
[Hickey, Raymond]	ftp://www.hd.urb.np/pub/pc/lexa/	Development of software called LEXA: "a suite of programs for tagging, lemmatization, word frequency counts, etc.".
Information Extraction Group	Natural Language Group, Department of Computer Science, University of Sheffield; http://www.dcs.shef.ac.uk/research/groups/nlp/nlp.html; http://www.dcs.shef.ac.uk/research/groups/nlp/extraction.html	Projects include information extraction from natural language texts ("newspapers, journals, patents, electronic mail"), with the construction of GATE, a "General Architecture for Text Extraction", merging the best features of POETIC (University of Sussex) and DIDEROT (New Mexico State University).

Institution	Where (Including URL)	Current or Primary Project(s), and Software
Institute for the Study of Language and Society/Research Sector: Language for Specific Purposes	Faculty of Management, Languages and European Studies, Department of Languages and European Studies, Aston University; http://www.les.aston.ac.uk/isls.html	The support facilities of the Research Sector and 'The Aston Scientific and Technical English Corpus (ASTEC)[648] are investigating five fields: discourse features in scientific and technical text, quantitative pattern analysis, quantitative type-distribution analysis, comparative type-distribution analysis, and genre analysis.[649] Part of ASTEC is the Aston Text Analyzer (ATA).[650]

648 http://www.les.aston.ac.uk/lsu/astec.html

649 http://www.les.aston.uk/lsu/astec.html

650 "ATA, the Aston Text Analyser, is an extension of the UNIX suite of programmes which form part of ASTEC (The Aston Science and Technical English Corpus). It is mounted on a PC and incorporates many of the ASTEC functions in a powerful and user-friendly way, although, being written in C++, it lacks the flexibility of UNIX...It was designed primarily to facilitate the conversion of target corpora into learning materials for marketing on diskettes by M S Technology...ATA has two separate main functions, an indexer and a set of accessing routines". ATA consrtructs 'Frequency Lists.': "These can be selected from the file menu and are all presented in five columns as follows:
"the raw frequency of the word (type);"
"the type itself;"
"the relative frequency of the type (out of 10,000);"
"the relative frequency (out of 10,000) of the type in a reference list (based on a modified version of an earlier COBUILDlisting."
Then 'Frequency List Versions': "are presented in five different versions as follows:
"all types in alphabetical order;"
"all types in numerical order;"
"types with a frequency greater than 1/10,000;"
"types in 5.2, excluding 100 function words;"
"types in 5.3, excluding 100 function words."
ATA also does concordances and a:
Synoptic Profile: "The keyword selected in the frequency list is represented by a asterisk. The four columns to left and right show the types and frequencies (in descending order) occurring in those positions in the corresponding concordance."
As well as statistics: "A separate facility is also available to provide statistical information about the corpus." (http://www.les.aston.uk/lsu/ataguide.html)

Institution	Where (Including URL)	Current or Primary Project(s), and Software
Instituto di Linguistica Comptazional (Institute for Computational Linguistics (ICL-CNR)	University of Pisa; http://www.ilc.pi.cnr.it/; http://www.ilc.pi.cnr.it/ dbt.html/	Developing methods and tools for literary and linguistic text analysis. Currently concentrating on : "developing text and lexical database systems", primarily DBT (textual database system). ILC-CNR is integrated with ELSENET (the Network of Excellence for Speech and Natural Language), and participates in EC projects, such as RENOS.[651] This involves applying statistical methods for information extraction from a 'corpus of legal texts, and the: "auto tagging and lemmatization of the corpus, the creation of a lexicon of the sublanguage, and analysis of compound terms for insertion in this lexicon."
Institut fur Maschinelle Sprachverarbeitung (Institute for Computational Linguistics)	University of Stuttgart; http://www.ims.uni. stuttgart.de/Overview. html	The focus is on theoretical and computational linguistics approaches to natural language.

[651] http://www.ilc.pi.cnr/euro.proj/renos.html

Institution	Where (Including URL)	Current or Primary Project(s), and Software
Laboratory for Language and Speech Technologies	"Joseph Stefan" Institute, Ljublyana, Slovenia; http://www.cogsci.ed.ac.uk/elsnet/survey/survey.html	OKUS: software to allow: "text analysis for concordances, collocations and statistics" (in Pascal).
Language Technology Centre	Human Communication Research Centre, University of Edinburgh; http://www.cogsci.ed.ac.uk/hcrc/wgs/gnp/ltg/ltg.html	Four projects include: 1) Message Understanding· One task is the development of a text processing system which involves text analysis issues, such as word recognition.[652] 2) Text Categorization. The construction of text categorization and routing systems, which involve: the assignment of keywords and categories to texts...and automatically routing texts to interested recipients"[653] 3) Intelligence Text Processing, which automatically assists in text production, checking for stylistic consistency".[654]

[652] http://www.cogsci.ed.ac.uk/hcrc/wgs/gnp/ltg_muc/ltg_muc.html
[653] http://www.cogsci.ed.ac.uk/hcrc/wgs/gnp/ltg_categorization/ltg_categorization.html
[654] http://www.cogsci.ed.ac.uk/hcrc/wgs/gnp/ltg_authoring/ltg_authoring.html

Institution	Where (Including URL)	Current or Primary Project(s), and Software
Library Electronic Texts Research Service	Main Library, Indiana University-Bloomington; http://www.indiana.edu/~letrs/index.html	This service houses an electronic database of primary reference texts for the humanities, which includes on-line access to SGML tagged texts. It also provides a number of text analysis software for DOS and Macintosh. One example is TACT, which, *inter alia*, produces word-length and sentence-length statistics, as well constructing indexes and concordances.
Lingsoft, Inc. and the Research Unit for MultiLingual Language Technology	Helsinki; http://www.lingsoft.fi/ and the Department of General Linguistics, University of Helsinki; http://www.ling.Helsinki.fi/research/rumlat.html	The construction of ENGCG (the 'Constraint Grammar Parser for English'), which performs morphological analysis (tagging) and part of speech disambiguating of running English text. Another Lingsoft product is NPtool, which extracts noun phrases from English texts, as a starting point for term extraction. [655]

[655] http://www.lingsoft.fi/doc/NPtool/

Institution	Where (Including URL)	Current or Primary Project(s), and Software
Microsoft Research Natural Language Group	http://www.research. microsoft.com/research/ nlp, and http://www.research. microsoft.com/research/ nlp/analysis.html	The Group are involved, *inter alia*, in text and natural language analysis which includes: "lexical, morphological, syntactic, and semantic processing."
Natural Language Processing (NLP) Group	IRST - Instituto per la Ricerca Scientifica Tecnologica; http://www.itc/itctext/ sede/testi/IRST.html A research institute of ITC - Instituto Trentino di Culturo; http://www.itc.it/itctext/ sede/testi/indice.html The URL of the NLP Group is: http://.ecate.itc.it.1024/	Current projects include bidirectional parsing analysis for textual analysis.
Rank Xerox Research Centre	Cambridge Laboratory, U.K; http://www.xerox.com/ lexdem/xlt-whitepaper. html	The 'Xerox Lexical Technology' (XLT) offers documents indexing, tokenization, morphological analysis/stemming, morphological generation, and part-of-speech tagging. XLT also provides an on-line French morphological analyzer and part of speech disambiguate.[656]

[656] http://www.xerox.com/lexdemo/xlt-demo.hmtl

Institution	Where (Including URL)	Current or Primary Project(s), and Software
Rank Xerox Research Centre	Grenoble Laboratory, Meylan, Lyon; http://www.xerox.fr/grenoble/mltt/Mos/Tools.html	One current research linguistic analysis project is 'Multi Lingual Theory and Technology' which offers an on-line grammatical tagger and morphological text analyzer.
Research Group of the Microprocessor System Laboratory	St. Petersburgh Institute for Informatics and Automation, Academy of Sciences, St. Petersburgh; http://www.codsci.ed.ac.uk/elsnet/survey/survey.html	An automated parsing system for natural language sentences from various texts: "science, fiction, newspapers, etc."
School of Cognitive and Computing Sciences (COGS)	University of Sussex; http://www.cogs.susx.ac.uk/	Current work includes development of parsing algorithms for tractable grammar formalisms
Sonderforschungsbereich (Special Research Centre)	University of Stuttgart; http://www.ims.uni.stuttgart.de/Overview.html	The Centre is developing 'Discourse Representation Theory'. Where the: "emphasis is on the temporal structure of discourse, the notion of discourse coherence and anaphora.[657]

[657] http://www.ims.uni.stuttgart.de/Overview.html/

Institution	Where (Including URL)	Current or Primary Project(s), and Software
University Centre for Computer Corpus Research on Language (UCREL)	Departments of Linguists and English Modern Language, and of Computing, University of Lancaster; http://www.comp.lancs.uk /computing/ research/ucrel/	The Centre's objective is: "to carry out computer-based research on the analysis and processing of natural data." Research projects include 'Automatic Content Analysis of Market Research Interview Transcripts (ACAMRIT)'
User Interface Technology Group	CRL, State University of New Mexico; http://crl.nmsu.edu/ ETG.html	Work involves constructing object-orientated modules for various natural language applications, e.g.: "standard interfaces for textual analysis including lexical information retrieval (e.g. concordances, lexical databases, and parallel corpora)".

Institution	Where (Including URL)	Current or Primary Project(s), and Software
Natural Language Research and Development Unit	the Institute for Integrated Publication and Information Systems; http://saturn.damstadt.gmd.de/IPS1/ Part of the German National Research Center for Information Technology; http://www.damstadt.gmd.de/	The KONTEXT text analysis system[658] is a natural language processing system based on the KONTEXT text model which: "structures the information of a text into five layers: sentence structure, thematic structure, reference structure (objects), view (facts), background knowledge. The KONTEXT system is a first prototype of a new generation of text analysis systems. This model allows for a novel text oriented processing of texts (instead of isolated sentence oriented processing), where text structure and context are taken into account during processing."
Natural Language Software Registry	German Institute for Artificial Intelligence (DFKI); http://www-dfki.uni-sb-de/cl/ registry/draft.html	This registry has over 100 descriptions of natural language processing software. This includes: 'semantic and pragmatic analysis', 'morphological analysis', and 'syntactic analysis'.

[658] http://www.damstadt.gmd.de/

Appendix Three: The Fifty-Eight Journals Searched, the Number of Articles in Each, and the Number and Year of Articles Selected from Each Journal, for Case Study Two in Chapter Three

Journal Title (n=58)	Number of Articles Perused: 1990 - 1995 (n = 13920)	Number of Articles Selected and Years (n=35)[659]
1. Annals of Science	712	0
2. Applied Linguistics	221	2 (1991; 1993)
3. ariel - a review of international english literature	428	0
4. Behavioral Science	96	1 (1994)
5. Biometrics	660	0
Biometrika	552	0
British Journal of the History of Science	703	0
British Journal for the Philosophy of Science	290	0
Cognitive Linguistics	84	0
Communication Theory	44	0
Computers and the Humanities	337	2 (1992, 1993)
Computational Linguistics	108	0
Discourse & Society	93	0
Discourse Processes	159	2 (1992)
English for Specific Purposes	81	1 (1991)
Forum for Modern Language Studies	177	0

[659] The thirty-fifth paper (Bruce, Donald; *Towards the Implementation of Text and Discourse Theory in Computer-Asissited Textual Analysis*; Computers and the Humanities; 27; 1989: 357-364) is a non-empirical discussion article, included because its contents are very relevant to this current research.

Journal Title (n=58)	**Number of Articles Perused: 1990 - 1995 (n = 13920)**	**Number of Articles Selected and Years (n=35)**
Genre-forms of discourse and culture	44	0
History of Science	118	0
Information Processing & Management	465	2 (1993; 1995)
International Journal of the Sociology of Language	124	0
Journal of Communication	722	0
Journal of Documentation	458	0
Journal of Language and Social Psychology	28	0
Journal of Pragmatics	524	2 (1990; 1992)
Journal of the American Society of Information Science	561	1 (1995)
Journal of the Electronic Writer		0
Journal of Technical Writing and Communication	206	1 (1991)
Journal of the Proceedings of the Annual Conference of Cognitive Science		1 (1993)
Language	421	1 (1991)
Language in Society	181	0
Language Sciences	99	2 (1992)
Linguistics	458	0

Journal Title (n=58)	**Number of Articles Perused: 1990 - 1995 (n = 13920)**	**Number of Articles Selected and Years (n=35)**
Literature and Linguistics Computing	108	2 (1990; 1993)
Knowledge Acquisition	80	1 (1992)
Mind & Language		0
MLN (Modern Language Notes)	539	0
Modern Language Quarterly	310	0
Natural Language and Linguistic Theory	108	0
Notes and Records of the Royal Society of London	202	1 (1995)
Paper on Language and Literature (PLL)	133	0
Papers in Linguistics	120	0
Philosophy and Rhetoric	197	0
Philosophy of Science	357	0
PMLA - publication of the Modern Language Association of America	372	0
Poetics	115	1 (1994)
Quarterly Journal of Speech	114	0
Rhetorica	176	0
Scientometrics	431	7 (1990 x3; 1991; 1994x3)
Science as Culture	32	0
Science In Context	98	1 (1991)
Studies in English Literature 1500-1900	216	0

Journal Title (n=58)	**Number of Articles Perused: 1990 - 1995 (n = 13920)**	**Number of Articles Selected and Years (n=35)**
Studies in the History & Philosophy of Science	154	0
The Technical Writing Teacher	92	1 (1990)
Text	141	0
TEXT Technology		1 (1995)
The Southern Communication Journal	143	0
Verbatim: the language quarterly	428	0
Written Communication	100	2 (1993; 1994)

Appendix Four: Word Lists Used in The Pilot and Main Case-Studies.

1) Modal Terms:

i) all inclusive

“ are" " is " " ought" "always" "always" "always" "aren't" "can't" "cannot" "certain "certainly" "could" "could be" "could not" "couldn't" "does" "does not" "doesn't" "doubtless" "had " "had to" "is not" "isn't" "likely" "may not" "maybe" "maybe not" "must " "must be" "need " "need be" "need to" "never" "ought to" "possibly" "possibly not" "probably" "probably not" "shall " "shall not" "shan't" "shouldn't" "surely" "unlikely" "was not" "were " "were not" "weren't" "will " "will not" "would not" "wouldn't"

ii) empirical positive

“will” “is” “must” “must be” “ought” “ought to” “are” “were” “shall” “need” “need be” “need to” “had” “had to” “certain” “doubtless” “doubtlessly” “certainly does” “would” “always”

iii) empirical negative

“will not” “is not” “was not” “were not” “never” “cannot” “shall not” “does not” “would not”

iv) contingent positive

"maybe" "could" "could be" "likely" "surely" "probably" "possibly"

v) contingent negative

"unlikely" "maybe" not" "may" not" "could" not" "possibly" not" "probably not"

vi) contingent inclusive

"maybe" "could" "could be" "likely" "surely" "probably" "possibly" "unlikely" "maybe not" "may not" "could not" "possibly not" "probably not" "couldn't" "doesn't" "wouldn't" "shouldn't" "can't" "shan't" "aren't" "isn't" "weren't"

vii) empirical inclusive

"always" " are" " is " " ought" "always" "always" "cannot" "certain " "certain " "certainly" "does" "does not" "doubtless" "had " "had to" "is not" "must " "must be" "need " "need be" "need to" "never" "ought to" "shall " "shall not" "was not" "were " "were not" "will " "will not" "would not"

2) conjunctives

"after" "and" "while" "but" "or" "then" "that" "because" "if" "when" "since" "so although" "before" "nor" "than" "till" "until" "unless" "whereas" "yet" "thus" "meanwhile" "subsequently" "at the same time" "nevertheless" "therefore" "by contrast" "however" "also like" "though"

3) bench_words

"eye" "use" "effect" "saw" "see" "sensible" "visible" "touch" "make" "made" "find" "come out" "comes out" "phenomen" "appear" "observe" "work" "sent" "send" "obtain" "join" "place" "push" "put" "making" "tak" "whil" "finger" "hand"

4) any_experiment

"very" "care" "can" "find" "contain" "adulteration" "impur" "pure" "accura" "experiment" "trial" "examin" "sample" "opinion" "view" "substance" "compos" "analys" "weigh" "grain" "ounce" "proportion" "minute" "estimate" "expts" "supported" "adduced" "demonstr" "beautiful" "lovely" "show" "apparatus" "arrang" "instrument" "new" "ascertain" "verify" "establish" "result" "shew" "law" "describe" "repeat" "confirm" "proof" "expect" "anticipate" "made" "make" "sensible" "notice" "obtain" "illustrate" "act" "curious" "attention" "assumption" "suppos" "convinc" "phenomen" "eye" "use" "effect" "saw" "see" "visible" "touch" "come out" "comes out" "appear" "observe" "work" "sent" "send" "join" "place" "push" "put" "making" "tak" "whil" "finger" "hand"

5) chem_analysis

"very" "care" "can" "find" "contain" "adulteration" "impur" "pure" "accura" "experiment" "trial" "examin" "sample" "opinion" "view" "substance" "compos" "analys" "weigh" "contain" "grain" "ounce" "proportion" "minute" "estimate"

6) new_discovery

"experiment" "examin" "expts" "supported" "adduced" "demonstr" "beautiful" "lovely" "show" "apparatus" "arrang" "instrument" "new" "ascertain" "verify" "establish" "result" "shew" "law" "describe" "repeat" "confirm" "proof" "expect" "anticipate" "made" "make" "sensible" "notice" "obtain" "illustrate" "act" "curious" "attention" "assumption" "suppos" "convinc" "phenomen"

7) G & M_empirical

“experimental” “fact” “data” “demonstrate” “test” “results” “model” “rules” “laboratory” “rational” “real” “observation” “see” “truth” “evidence” “correct”

8) G & M_contingent

“bear in mind” “forceful people” “load of nonsense” “damned dogmatic” “personalities” “lack of effort” “dislike” “naivety” “lack of interest” “bad manners” “affecting status” “status” “preposterous” “non-explanatory” “won't do” “unbelievable” “intuition” “doubt” “money” “funding” “manipulative” “non-scientific” “irrational” “bend the data” “personally” “weakness incorrect” “he's” “Oh, my God” “dominant” “pronouncing the gospel” “dishonest” “vitriolic” “unreasonable” “nit-picking” “didn't really understand “ “defect” “think through”

9) G&M_inclusive

“experimental” “fact” “data” “demonstrate” “test” “results” “model” “rules” “laboratory” “rational” “real” “observation” “see” “truth” “evidence” “correct” “bear in mind” “forceful people” “load of nonsense” “damned dogmatic” “personalities” “lack of effort” “dislike” “naivety” “lack of interest” “bad manners” “affecting status” “status” “preposterous” “non-explanatory” “won't do” “unbelievable” “intuition” “doubt” “money” “funding” “manipulative” “non-scientific” “irrational” “bend the data” “personally” “weakness incorrect” “he's” “Oh, my God” “dominant” “pronouncing the gospel” “dishonest” “vitriolic” “unreasonable” “nit-picking” “didn't really understand “ “defect” “think through”

Appendix Five: Numerical Results for Chapter Six

Table 6.5 Modalword Use in Faraday's Correspondence With Whewell Over 5-Year Periods.

Modalwords	Period	no of paragraphs	truncated mean	Min value	Max value
a i	1830-34	91	0.49	1.14	5.80
a i	1835-39	70	0.75	1.02	7.58
a i	1840-44	34	0.58	1.27	6.76
a i	1845-49	139	1.01	1.09	9.52
e i	1830-34	91	0.43	1.08	5.80
e i	1835-39	70	0.74	1.02	7.58
e i	1840-44	34	0.52	1.27	6.76
e i	1845-49	139	0.98	1.09	9.52
c i	1830-34	91	0.00	0.54	3.12
c i	1835-39	70	0.00	0.40	1.38
c i	1840-44	34	0.00	0.00	0.00
c i	1845-49	139	0.00	0.39	2.08
e p	1830-34	91	1.70	1.67	10.00
e p	1835-39	70	2.10	2.94	11.34
e p	1840-44	34	1.55	3.57	11.11
e p	1845-49	139	2.64	1.92	15.00
e n	1830-34	91	0.00	0.95	1.92
e n	1835-39	70	0.05	0.36	2.30
e n	1840-44	34	0.00	0.00	1.35
e n	1845-49	139	0.01	0.37	2.47
c p	1830-34	91	0.00	0.33	3.12
c p	1835-39	70	0.00	0.40	0.53
c p	1840-44	34	0.00	0.00	0.00
c p	1845-49	139	0.00	0.39	2.08
c n	1830-34	91	0.95	2.38	21.74
c n	1835-39	70	1.40	1.03	17.24
c n	1840-44	34	0.61	5.88	36.59
c n	1845-49	139	0.94	1.02	22.73

Appendix Five

Table 6.6 Statistics For Modalword Use in Whewell's Correspondence With Faraday Over Five-Year Periods.

Modalwords	Period	no of paragraphs	truncated mean	Min value	Max value
a i	1830-34	67	0.84	0.91	7.50
a i	1835-39	98	0.82	1.68	14.29
a i	1840-44	21	0.34	2.25	7.37
a i	1845-49	93	0.67	1.24	21.43
e i	1830-34	67	0.91	1.05	8.22
e i	1835-39	98	0.76	1.69	14.29
e i	1840-44	21	0.34	2.25	7.37
e i	1845-49	93	0.65	1.24	21.43
c i	1830-34	67	0.00	0.25	0.68
c i	1835-39	98	0.00	0.45	2.67
c i	1840-44	21	0.00	0.00	0.00
c i	1845-49	93	0.00	0.58	1.26
e p	1830-34	67	4.06	4.21	18.92
e p	1835-39	98	2.74	3.53	14.545
e p	1840-44	21	2.46	4.49	10.04
e p	1845-49	93	2.36	4.76	14.29
e n	1830-34	67	0.05	0.34	2.63
e n	1835-39	98	0.0201	0.21	2.88
e n	1840-44	21	0.00	0.00	2.10
e n	1845-49	93	0.01	0.55	4.00
c p	1830-34	67	0.00	0.25	0.56
c p	1835-39	98	0.0051	0.45	2.67
c p	1840-44	21	0.00	0.00	0.00
c p	1845-49	93	0.00	0.58	1.26
c n	1830-34	67	1.33	2.72	26.32
c n	1835-39	98	0.78	0.95	14.42
c n	1840-44	21	0.70	7.86	15.79
c n	1845-49	93	1.03	1.05	20.00

Appendix Five

Table 6.7 Modalword Use in Faraday's Correspondence With Herschel Over 5-Year Periods.

Modalwords	Period	no of paragraphs	truncated mean	Min value	Max value
a i	1825-29	164	0.43	5.75	8.33
a i	1830-34	13	0.08	0.85	6.60
a i	1835-39	21	0.40	1.10	6.67
a i	1840-44	11	0.37	1.37	3.12
a i	1845-49	82	0.66	1.41	9.09
e i	1825-29	164	0.39	0.99	8.33
e i	1830-34	13	0.08	0.85	5.66
e i	1835-39	21	0.32	1.10	6.67
e i	1840-44	11	0.33	1.37	2.00
e i	1845-49	82	0.60	0.93	9.09
c i	1825-29	164	0.00	0.38	2.38
c i	1830-34	13	0.00	0.00	0.94
c i	1835-39	21	0.00	0.00	1.3889
c i	1840-44	11	0.00	0.00	1.562
c i	1845-49	82	0.00	0.78	1.54
e p	1825-29	164	0.89	2.70	13.40
e p	1830-34	13	0.46	5.60	6.60
e p	1835-39	21	1.50	4.55	8.33
e p	1840-44	11	1.27	2.00	8.22
e p	1845-49	82	1.69	3.75	10.59
e n	1825-29	164	0.00	0.38	2.90
e n	1830-34	13	0.00	0.00	1.89
e n	1835-39	21	0.00	0.00	1.67
e n	1840-44	11	0.00	0.00	2.00
e n	1845-49	82	0.03	0.37	4.55
c p	1825-29	164	0.00	0.35	1.79
c p	1830-34	13	0.00	0.00	0.00
c p	1835-39	21	0.00	0.00	1.39
c p	1840-44	11	0.00	0.00	1.56
c p	1845-49	82	0.00	0.78	1.41
c n	1825-29	164	0.00	2.90	20.00
c n	1830-34	13	0.00	13.56	15.09
c n	1835-39	21	0.00	5.75	10.99
c n	1840-44	11	0.00	0.00	1.56
c n	1845-49	82	0.03	0.93	31.25

Appendix Five

Table 6.8 Modalword Use in Herschel's Correspondence With Faraday Over Five-Year Periods.

Modalwords	Period	no of paragraphs	truncated mean	Min value	Max value
a i	1825-29	164	0.84	1.68	15.39
a i	1830-34	9	0.37	0.00	3.31
a i	1835-39	58	0.29	1.16	2.70
a i	1845-49	82	0.96	1.07	10.81
e i	1825-29	164	0.81	1.68	15.38
e i	1830-34	9	0.368	0.00	3.31
e i	1835-39	58	0.29	1.16	2.70
e i	1845-49	82	0.82	1.07	10.00
c i	1825-29	164	0.00	0.57	8.00
c i	1830-34	9	0.00	0.00	0.00
c i	1835-39	58	0.00	0.00	0.00
c i	1845-49	82	0.04	0.34	4.82
e p	1825-29	164	2.61	2.70	30.77
e p	1830-34	9	1.97	8.65	9.09
e p	1835-39	58	2.10	4.05	6.25
e p	1845-49	82	3.67	1.72	16.22
e n	1825-29	164	0.02	0.36	4.55
e n	1830-34	9	0.03	0.00	0.25
e n	1835-39	58	0.00	0.00	0.00
e n	1845-49	82	0.00	0.73	3.51
c p	1825-29	164	0.00	0.57	8.00
c p	1830-34	9	0.00	0.00	0.00
c p	1835-39	58	0.00	0.00	0.00
c p	1845-49	82	0.03	0.73	3.61
c n	1825-29	164	0.99	1.61	52.63
c n	1830-34	9	0.62	0.00	5.60
c n	1835-39	58	1.06	2.82	11.63
c n	1845-49	82	0.58	0.78	17.54

Appendix Five

Table 6.9 Modalword Use in Faraday's Correspondence With Herschel Over Five-Year Periods (Experimental Topics).

Modalwords	Period	no of paragraphs	truncated mean	Min value	Max value
a i	1825-29	50	1.10	1.04	8.14
a i	1845-49	13	4.49	0.97	11.76
e i	1825-29	50	1.04	1.04	8.14
e i	1845-49	13	4.18	0.97	11.76
c i	1825-29	50	0.00	0.38	2.42
c i	1845-49	13	0.30	0.82	1.57
e p	1825-29	50	2.15	2.70	13.83
e p	1845-49	13	7.00	4.25	11.25
e n	1825-29	50	0.02	0.38	2.27
e n	1845-49	13	0.59	0.38	5.88
c p	1825-29	50	0.00	0.38	1.82
c p	1845-49	50	0.27	0.82	1.51
c n	1825-29	50	1.079	2.90	11.36
c n	1845-49	13	9.17	0.97	29.41

Table 6.10 Modalword Use in Herschel's Correspondence With Faraday Over Five-Year Periods (Experimental Topics).

Modalwords	Period	no of paragraphs	truncated mean	Min value	Max value
a i	1825-29	61	2.19	1.72	15.38
a i	1845-49	14	3.49	1.14	12.50
e i	1825-29	61	1.96	1.72	15.38
e i	1845-49	14	2.87	1.14	9.37
c i	1825-29	61	0.03	1.85	8.00
c i	1845-49	14	0.458	0.35	5.13
e p	1825-29	61	5.566	3.85	30.77
e p	1845-49	14	9.55	1.72	18.75
e n	1825-29	61	0.11	0.36	4.55
e n	1845-49	14	0.11	1.28	3.85
c p	1825-29	61	0.03	1.82	8.00
c p	1845-49	14	0.43	0.82	3.85
c n	1825-29	61	2.61	1.69	52.63
c n	1845-49	14	3.35	0.82	19.23

Appendix Five

Table 6.11 Modalword Use in Faraday's Correspondence With Herschel Over Five-Year Periods (Non-experimental Topics).

Modalwords	Period	no of paragraphs	truncated mean	Min value	Max value
a i	1825-29	11	5.08	1.02	9.09
a i	1845-49	7	3.02	1.51	5.88
e i	1825-29	11	4.76	1.02	9.09
e i	1845-49	7	3.02	1.51	5.88
c i	1825-29	11	0.14	1.27	1.61
c i	1845-49	7	0.00	0.00	0.00
e p	1825-29	11	8.72	5.10	13.79
e p	1845-49	7	5.58	5.05	10.34
e n	1825-29	11	0.45	0.81	3.03
e n	1845-49	7	0.00	0.00	0.00
c p	1825-29	11	0.09	0.81	1.27
c p	1845-49	7	0.00	0.00	0.00
c n	1825-29	11	9.25	3.39	22.73
c n	1845-49	7	7.72	5.88	37.04

Table 6.12 Modalword Use in Herschel's Correspondence With Faraday Over Five-Year Periods (Non-experimental Topics).

Modalwords	Period	no of paragraphs	truncated mean	Min value	Max value
a i	1825-29	12	4.16	2.92	6.59
a i	1845-49	6	4.81	3.05	10.00
e i	1825-29	12	4.03	2.34	6.59
e i	1845-49	6	4.69	2.29	10.00
c i	1825-29	12	0.06	0.58	0.78
c i	1845-49	6	0.13	0.00	0.76
e p	1825-29	12	7.38	2.94	12.90
e p	1845-49	6	9.50	7.32	12.82
e n	1825-29	12	0.26	0.59	3.77
e n	1845-49	6	0.13	0.00	0.76
c p	1825-29	12	0.06	0.58	0.78
c p	1845-49	6	0.13	0.00	0.76
c n	1825-29	12	5.31	3.91	18.87
c n	1845-49	6	3.56	6.10	15.27

Appendix Five

Table 6.13 Statistics For Modalword Use in Faraday's Correspondence With Whewell Over Five-Year Periods (Experimental Topics).

Modalwords	Period	no of paragraphs	truncated mean	Min value	Max value
a i	1835-39	6	3.44	3.00	3.85
a i	1845-49	30	4.02	0.00	9.56
e i	1835-39	6	3.30	3.00	3.65
e i	1845-49	30	3.95	1.45	9.56
c i	1835-39	6	0.14	0.40	0.43
c i	1845-49	30	0.00	0.40	1.75
e p	1835-39	6	7.87	5.47	10.54
e p	1845-49	30	9.41	2.22	20.00
e n	1835-39	6	0.40	0.36	0.86
e n	1845-49	30	0.24	0.22	2.128
c p	1835-39	6	0.04	0.40	0.43
c p	1845-49	30	0.19	0.40	1.75
c n	1835-39	6	6.93	5.00	9.09
c n	1845-49	30	5.79	1.28	25.64

Table 6.14 Statistics For Modalword Use in Whewell's Correspondence With Faraday Over Five-Year Periods (Experimental Topics).

Modalwords	Period	no of paragraphs	truncated mean	Min value	Max value
a i	1835-39	13	4.23	1.73	10.31
a i	1845-49	7	4.16	0.00	8.49
e i	1835-39	13	4.08	1.73	10.31
e i	1845-49	7	3.97	0.00	8.49
c i	1835-39	13	0.04	0.46	1.22
c i	1845-49	7	0.19	0.00	1.30
e p	1835-39	13	8.36	5.56	17.02
e p	1845-49	7	11.52	7.79	15.79
e n	1835-39	13	0.63	0.44	3.09
e n	1845-49	7	0.59	1.30	2.83
c p	1835-39	13	0.04	0.46	1.22
c p	1845-49	7	0.19	0.00	1.30
c n	1835-39	13	5.79	1.38	15.46
c n	1845-49	7	7.40	1.05	18.87

Appendix Five

Table 6.15 Modalword Use in Faraday's Correspondence With Whewell Over Five-Year Periods (Non-experimental Topics).

Modalwords	Period	no of paragraphs	truncated mean	Min value	Max value
a i	1830-34	26	2.80	1.20	6.67
a i	1845-49	10	4.29	1.75	9.09
e i	1830-34	26	2.25	1.11	6.67
e i	1845-49	10	3.96	1.75	9.09
c i	1830-34	26	0.14	0.56	3.57
c i	1845-49	10	0.00	0.00	2.63
e p	1830-34	26	5.67	1.82	13.33
e p	1845-49	10	0.00	4.76	14.63
e n	1830-34	26	0.099	1.00	2.70
e n	1845-49	10	0.00	0.00	1.59
c p	1830-34	26	0.10	0.56	3.57
c p	1845-49	10	0.00	0.00	2.63
c n	1830-34	26	6.73	2.38	29.73
c n	1845-49	10	3.71	1.07	15.79

Table 6.16 Modalword Use in Whewell's Correspondence With Faraday Over Five-Year Periods (Non-experimental Topics).

Modalwords	Period	no of paragraphs	truncated mean	Min value	Max value
a i	1830-34	19	3.12	0.34	8.57
a i	1845-49	12	3.77	1.32	10.00
e i	1830-34	19	3.07	0.34	8.57
e i	1845-49	12	3.60	1.32	10.00
c i	1830-34	19	0.03	0.15	0.48
c i	1845-49	12	0.06	0.60	1.09
e p	1830-34	19	6.33	1.37	12.96
e p	1845-49	12	8.06	5.43	15.79
e n	1830-34	19	0.25	0.23	2.88
e n	1845-49	12	0.384	0.56	5.00
c p	1830-34	19	0.02	0.15	0.38
c p	1845-49	12	0.06	0.60	1.09
c n	1830-34	19	4.56	1.62	28.57
c n	1845-49	12	8.15	6.58	25.00

Appendix Five

Table 6.17 Statistics for Compression Scans of Faraday's Correspondence With Herschel.

Compression scans	co-respondents	no of para-graphs	truncated mean	Min value	Max value
1825-1829	Faraday to Herschel	77	8.34	3.70	43.38
1830-1834	Faraday to Herschel	13	17.14	8.33	27.84
1835-1839	Faraday to Herschel	9	8.95	14.92	24.71
1840-1844	Faraday to Herschel	5	13.75	18.00	27.87
1845-1849	Faraday to Herschel	36	13.37	5.56	61.09
1825-1829	Herschel to Faraday	85	12.11	5.88	38.93
1830-1834	Herschel to Faraday	2	22.9	0.00	45.9
1835-1839	Herschel to Faraday	6	12.82	22.67	27.69
1845-49	Herschel to Faraday	31	14.42	5.71	47.04

Table 6.18 Statistics for Compression Scans of Faraday's Correspondence With Whewell.

Compression scans	co-respondents	no of para-graphs	truncated mean	Min value	Max value
1830-1834	Faraday to Whewell	46	13.18	10.00	47.32
1835-1839	Faraday to Whewell	35	19.58	13.64	50.96
1840-1844	Faraday to Whewell	14	8.34	3.23	37.95
1845-1849	Faraday to Whewell	69	16.42	8.33	54.59
1830-1834	Whewell to Faraday	37	15.40	5.26	57.32
1835-1839	Whewell to Faraday	46	15.30	12.50	51.93
1840-1844	Whewell to Faraday	8	3.33	6.67	47.06
1845-1849	Whewell to Faraday	39	11.56	5.26	44.44

Appendix Five

Table 6.19 Statistics for Compression Scans of Faraday's Letters to Herschel and Whewell ('Experimental' and 'Non-Experimental' Topics).

Compression scans	co-respondents	topic	no of paragraphs	truncated mean	Min value	Max value
1825-1829	Faraday to Herschel	experimental	47	9.33	10.81	43.38
1825-1829	Faraday to Herschel	non-experimental	11	21.76	3.70	39.06
1845-1849	Faraday to Herschel	experimental	13	27.31	5.56	61.09
1845-1849	Faraday to Herschel	non-experimental	7	22.41	10.34	26.32
1835-1839	Faraday to Whewell	experimental	6	43.37	20.88	50.96
1830-1834	Faraday to Whewell	non-experimental	25	23.64	10.00	47.32
1845-1849	Faraday to Whewell	experimental	30	30.28	10.00	54.59
1845-1849	Faraday to Whewell	non-experimental	10	21.40	15.38	38.32

Table 6.20 Statistics for Compression Scans of Herschel and Whewell Letters to Faraday ('Experimental' and 'Non-Experimental' Topics).

Compression scans	co-respondents	primary topic	no of paragraphs	truncated mean	Min value	Max value
1825-1829	Herschel to Faraday	experimental	56	14.06	5.88	38.20
1825-1829	Herschel to Faraday	non-experimental	11	24.56	9.09	38.93
1845-1849	Herschel to Faraday	experimental	14	23.91	5.71	47.04
1845-1849	Herschel to Faraday	non-experimental	6	17.82	6.67	34.35
1835-1839	Whewell to Faraday	experiment	13	32.26	12.50	43.08
1830-1834	Whewell to Faraday	non-experimental	18	33.98	12.50	57.32
1845-1849	Whewell to Faraday	experimental	7	26.88	5.26	42.46
1845-1849	Whewell to Faraday	non-experimental	12	23.28	5.56	44.44

Appendix Five

Table 6.21 Modalword Use in Faraday's Correspondence With Phillips Over Five-Year Periods.

Modalwords	Period	no of paragraphs	truncated mean	Min value	Max value
a i	1835-39	7	1.51	3.03	3.85
a i	1840-44	88	0.14	0.90	8.57
a i	1845-49	29	1.91	1.18	8.49
e i	1835-39	7	1.24	1.92	3.73
e i	1840-44	88	0.11	0.90	8.57
e i	1845-49	29	1.79	1.18	8.49
c i	1835-39	7	0.27	0.00	1.92
c i	1840-44	88	0.00	1.47	2.17
c i	1845-49	29	0.04	0.36	2.31
e p	1835-39	7	4.07	5.05	9.52
e p	1840-44	88	1.48	2.22	12.07
e p	1845-49	29	4.67	4.71	15.09
e n	1835-39	7	0.36	0.62	1.92
e n	1840-44	88	0.00	0.54	0.93
e n	1845-49	29	0.15	0.36	2.38
c p	1835-39	7	0.27	0.00	1.92
c p	1840-44	88	0.00	1.47	1.63
c p	1845-49	29	0.04	0.36	1.39
c n	1835-39	7	3.13	3.85	13.04
c n	1840-44	88	0.15	1.47	11.63
c n	1845-49	29	2.24	0.62	23.81

Table 6.22 Modalword Use in Phillips' Correspondence With Faraday Over Five-Year Periods.

Modalwords	Period	no of paragraphs	truncated mean	Min value	Max value
a i	1835-39	29	0.10	1.03	4.25
a i	1840-44	88	0.29	0.95	4.76
a i	1845-49	7	0.24	0.00	1.70
e i	1835-39	29	0.10	1.03	2.13
e i	1840-44	88	0.23	0.95	4.76
e i	1845-49	7	0.24	0.00	1.69
c i	1835-39	29	0.00	0.00	2.13
c i	1840-44	88	0.00	0.00	2.04
c i	1845-49	7	0.00	0.00	0.00
e p	1835-39	29	2.86	5.40	12.77
e p	1840-44	88	2.23	3.23	11.90
e p	1845-49	7	2.10	6.25	8.47
e n	1835-39	29	0.00	0.00	0.00
e n	1840-44	88	0.00	0.00	0.00
e n	1845-49	7	0.00	0.00	0.00
c p	1835-39	29	0.00	0.00	2.13
c p	1840-44	88	0.00	0.00	2.04
c p	1845-49	7	0.00	0.00	0.00
c n	1835-39	29	0.00	0.00	1.03
c n	1840-44	88	0.14	4.76	16.13
c n	1845-49	7	0.24	0.00	1.69

Appendix Five

Table 6.23 Modalword Use in Faraday's Correspondence With Herbert Over Five-Year Periods.

Modalwords	Period	no of paragraphs	truncated mean	Min value	Max value
a i	1840-44	178	1.35	0.65	10.81
a i	1845-49	237	1.79	0.92	8.65
e i	1840-44	178	1.28	0.65	10.81
e i	1845-49	237	1.66	0.87	8.65
c i	1840-44	178	0.01	0.51	1.72
c i	1845-49	237	0.05	0.26	3.12
e p	1840-44	178	3.33	2.40	18.18
e p	1845-49	237	4.75	1.96	17.44
e n	1840-44	178	0.045	0.36	4.76
e n	1845-49	237	0.07	0.25	2.94
c p	1840-44	178	0.01	0.51	1.72
c p	1845-49	237	0.04	0.26	2.22
c n	1840-44	178	0.80	0.53	33.33
c n	1845-49	237	1.66	1.45	22.06

Table 6.24 Modalword Use in Herbert's Correspondence With Faraday Over Five-Year Periods.

Modalwords	Period	no of paragraphs	truncated mean	Min value	Max value
a i	1840-44	178	0.74	4.71	7.40
a i	1845-49	129	0.41	1.15	9.68
e i	1840-44	178	0.74	4.71	7.41
e i	1845-49	129	0.33	0.71	9.67
c i	1840-44	178	0.00	0.00	0.00
c i	1845-49	129	0.00	0.71	4.35
e p	1840-44	178	3.16	3.70	14.81
e p	1845-49	129	2.68	1.92	13.64
e n	1840-44	178	0.00	0.00	0.00
e n	1845-49	129	0.00	0.93	3.23
c p	1840-44	178	0.00	0.00	0.00
c p	1845-49	129	0.00	0.71	4.35
c n	1840-44	178	0.00	0.00	0.00
c n	1845-49	129	0.40	1.45	16.13

Appendix Five

Table 6.25 Modalword Use in Faraday's Correspondence With Sarah Over Five-Year Periods.

Modalwords	Period	no of para-graphs	truncated mean	Min value	Max value
a i	1819-24	130	0.78	1.51	20.93
a i	1839-44	13	1.02	1.68	5.75
a i	1845-49	10	0.49	1.82	2.44
e i	1819-24	130	0.68	0.82	18.60
e i	1839-44	13	0.90	1.27	4.32
e i	1845-49	10	0.49	1.82	2.44
c i	1819-24	130	0.02	0.32	5.66
c i	1839-44	13	0.12	1.27	1.44
c i	1845-49	10	0.00	0.00	0.00
e p	1819-24	130	2.01	2.98	16.67
e p	1839-44	13	2.77	3.18	13.33
e p	1845-49	10	2.60	6.38	9.76
e n	1819-24	130	0.03	0.32	1.97
e n	1839-44	13	0.00	0.00	2.00
e n	1845-49	10	0.00	0.00	0.00
c p	1819-24	130	0.01	0.32	3.77
c p	1839-44	13	0.06	0.64	0.72
c p	1845-49	10	0.00	0.00	0.00
c n	1819-24	130	0.96	0.83	45.45
c n	1839-44	13	2.64	3.82	11.51
c n	1845-49	10	0.50	4.06	9.09

Appendix Five

Table 6.26 Modalword Use in Faraday's Correspondence With Magrath Over Five-Year Periods.

Modalwords	Period	no of paragraphs	truncated mean	Min value	Max value
a i	1825-29	43	1.09	1.82	13.79
a i	1830-34	52	1.26	2.13	7.14
a i	1835-39	77	0.85	1.85	9.80
a i	1840-44	28	1.19	3.70	7.07
a i	1845-49	4	0.82	0.00	3.30
e i	1825-29	43	1.0	2.13	13.79
e i	1830-34	52	1.07	4.55	7.14
e i	1835-39	77	0.83	1.85	9.80
e i	1840-44	28	1.19	3.70	7.07
e i	1845-49	4	0.82	0.00	3.30
c i	1825-29	43	0.01	0.59	1.82
c i	1830-34	52	0.00	2.22	4.18
c i	1835-39	77	0.00	0.90	1.97
c i	1840-44	28	0.00	0.00	0.00
c i	1845-49	4	0.00	0.00	0.00
e p	1825-29	43	2.26	5.17	29.41
e p	1830-34	52	2.51	2.33	12.50
e p	1835-39	77	1.93	5.26	18.37
e p	1840-44	28	2.49	1.37	15.00
e p	1845-49	4	1.65	0.00	6.59
e n	1825-29	43	0.07	0.54	5.88
e n	1830-34	52	0.00	0.45	1.79
e n	1835-39	77	0.06	0.88	4.89
e n	1840-44	28	0.05	0.46	1.64
e n	1845-49	4	0.00	0.00	0.00
c p	1825-29	43	0.00	0.59	0.90
c p	1830-34	52	0.00	2.22	4.17
c p	1835-39	77	0.00	0.45	1.97
c p	1840-44	28	0.00	0.00	0.00
c p	1845-49		0.00	0.00	0.00
c n	1825-29	43	1.64	3.93	29.41
c n	1830-34	52	1.27	2.22	17.86
c n	1835-39	77	0.82	2.39	24.39
c n	1840-44	28	1.54	4.82	18.52
c n	1845-49	4	5.49	0.00	21.98

Appendix Five

Table 6.27 Modalword Use in Magrath's Correspondence With Faraday Over Five-Year Periods.

Modalwords	Period	no of paragraphs	truncated mean	Min value	Max value
a i	1830-34	15	3.46	4.55	10.00
a i	1835-39	7	1.07	3.57	3.90
e i	1830-34	15	3.46	0.00	10.00
e i	1835-39	7	0.97	3.25	3.571
c i	1830-34	15	0.00	0.00	0.00
c i	1835-39	7	0.09	0.00	0.65
e p	1830-34	15	4.24	3.57	11.76
e p	1835-39	7	4.35	7.69	14.29
e n	1830-34	15	0.00	0.00	0.00
e n	1835-39	7	0.00	0.00	0.00
c p	1830-34	15	0.00	0.00	0.00
c p	1835-39	7	0.65	0.00	0.65
c n	1830-34	15	0.00	0.00	0.00
c n	1835-39	7	3.20	4.55	17.86

Table 6.28 Modalword Use in Faraday's Correspondence With Barlow Over Five-Year Periods.

Modalwords	Period	no of paragraphs	truncated mean	Min value	Max value
a i	1835-39	7	1.38	0.00	9.68
a i		72	0.549	1.28	10.00
a i	1845-49	75	0.89	1.59	7.14
e i	1835-39	7	1.38	0.00	9.68
e i	1840-44	72	0.53	3.23	10.00
e i	1845-49	75	0.80	1.59	6.36
c i	1835-39	7	0.00	0.00	0.00
c i	1840-44	72	0.00	0.00	1.28
c i	1845-49	75	0.00	1.43	2.53
e p	1835-39	7	1.38	0.00	9.68
e p	1840-44	72	1.14	5.00	15.71
e p	1845-49	75	1.91	2.38	20.00
e n	1835-39	7	0.46	0.00	3.23
e n	1840-44	72	0.00	0.94	3.26
e n	1845-49	75	0.04	0.79	2.86
c p	1835-39	7	0.00	0.00	0.00
c p	1840-44	72	0.00	0.00	1.28
c p	1845-49	75	0.00	1.27	2.38
c n	1835-39	7	2.30	0.00	16.13
c n	1840-44	72	0.60	4.72	25.00
c n	1845-49	75	1.71	2.38	27.78

Appendix Five

Table 6.29 Modalword Use in Barlow's Correspondence With Faraday Over Five-Year Periods.

Modalwords	Period	no of para-graphs	truncated mean	Min value	Max value
a i	1845-49	8	1.26	3.10	6.97
e i	1845-49	8	1.14	3.10	6.06
c i	1845-49	8	0.11	0.00	0.91
e p	1845-49	8	3.96	8.63	13.94
e n	1845-49	8	0.26	0.88	1.21
c p	1845-49	8	0.11	0.00	0.91
c n	1845-49	8	1.83	4.65	10.00

Table 6.30 Compression Scans of Faraday's Correspondence With Magrath Over Five-Year Periods.

Co-respondents	Period	no of paragraphs	truncated mean	minimum	maximum
Faraday to Magrath	1825-29	21	11.44	5.88	42.99
Faraday to Magrath	1830-34	39	12.24	6.67	42.01
Faraday to Magrath	1835-39	57	6.32	3.85	37.61
Faraday to Magrath	1840-44	24	9.21	6.25	48.18
Faraday to Magrath	1845-49	4	6.90	0.00	27.59
Magrath to Faraday	1830-34	23	8.75	5.56	23.08
Magrath to Faraday	1835-39	6	12.58	4.55	24.49

Table 6.31 Compression Scans of Faraday's Correspondence With Barlow Over Five-Year Periods.

Co-respondents	Period	no of paragraphs	truncated mean	minimum	maximum
Faraday to Barlow	1835-39	4	5.47	3.70	9.09
Faraday to Barlow	1840-44	27	7.09	5.88	36.69
Faraday to Barlow	1845-49	34	11.55	3.85	41.36
Barlow to Faraday	1845-49	3	28.0	4.55	39.1

Appendix Five

Table 6.32 Compression Scans of Faraday's Correspondence With Herbert Over Five-Year Periods.

Co-respondents	Period	no of paragraphs	truncated mean	minimum	maximum
Faraday to Herbert	1840-44	155	14.76	2.56	50.36
Faraday to Herbert	1845-49	229	20.26	4.65	56.54
Herbert to Faraday	1840-44	15	6.08	4.00	27.50
Herbert to Faraday	1845-49	104	6.98	3.70	57.69

Table 6.33 Compression Scans Of Faraday's Correspondence With Sarah Over Five-Year Periods.

Co-respondents	Period	no of paragraphs	truncated mean	minimum	maximum
Faraday to Sarah	1819-24	58	13.62	3.70	44.89
Faraday to Sarah	1835-39	8	18.53	16.67	28.18
Faraday to Sarah	1845-49	5	18.22	15.91	26.00

Table 6.34 Statistics For Compression Scans of Faraday's Correspondence With Phillips Over Five-Year Periods.

Co-respondents	Period	no of paragraphs	truncated mean	minimum	maximum
Faraday to Phillips	1835-39	6	15.81	18.75	29.05
Faraday to Phillips	1840-44	65	4.66	4.17	38.98
Faraday to Phillips	1845-49	21	24.81	19.75	43.66
Phillips to Faraday	1835-39	7	13.24	20.00	27.96
Phillips to Faraday	1840-44	22	12.59	7.15	28.26
Phillips to Faraday	1845-49	3	15.16	4.55	23.81

Appendix Six: Analysis of Variance/Confidence Levels for the 'Results: Interpretation' Section of Chapter 8

Graph 1. Analysis of Variance/Confidence Levels for the 1821 Private and Published Writings on Electromagnetism.

Compression

```
                                          Individual 95% CIs For Mean Based on
                                          Pooled StDev
Level            N       Mean        StDev -----+---------+---------+--------+-
D1_comp         86      18.15        12.54 (--*--)
EMp_comp        83      37.38        14.31                          (-*--)
EMp2_comp       22      41.30        12.88                           (----*-----)
Sk1_comp        18      35.48        16.68                     (-----*------)
Sk2_comp        53      36.38        11.98                         (--*---)
Sk3_comp        56      33.32        12.49                      (--*---)
Sk_comp        127      34.90        12.92                         (-*-)
                                           -----+---------+---------+--------+-
Pooled StDev =          13.12                   20        30        40       50
```

Conjunctives

```
                                           Individual 95% CIs For Mean Based on
                                           Pooled StDev
Level            N       Mean        StDev  --------+---------+---------+--------
D1_conj        163      5.147        6.411   (--*--)
EMp_conj        88     10.843        4.152                    (---*---)
EMp2_conj       47      5.295        5.918 (-----*----)
Sk1_conj        19     10.992        4.710                (--------*-------)
Sk2_conj        54     11.975        3.792                      (----*----)
Sk3_conj        71      9.111        5.917              (---*----)
Sk_conj        144     10.433        5.203                   (--*--)
                                            --------+---------+---------+-------
Pooled StDev =          5.459                     6.0       9.0      12.0
```

Paragraph Lengths

```
                                        Individual 95% CIs For Mean Based on
                                        Pooled StDev
Level        N       Mean      StDev   ------+---------+---------+--------+
D1_pl       86      47.29      38.17   (--*--)
EMp_pl      83     164.52     116.97                      (--*---)
EMp2_pl     22     212.59     123.69                          (-----*-----)
Sk1_pl      18     141.22      85.39                 (------*-----)
Sk2_pl      53     160.96      83.83                     (---*---)
Sk3_pl      56     129.70      77.23                  (---*--)
Sk_pl      127     144.38      81.85                     (-*--)
                                       ------+---------+---------+-------+-
Pooled StDev =      85.86                    60       120       180     240
```

'Empirical Positive' Modality

```
                                       Individual 95% CIs For Mean Based on
                                       Pooled StDev
Level        N      Mean      StDev   ----+---------+---------+---------+--
D1_ep      163     3.683      5.142    (--*-)
EMp_ep      88     7.963      3.857              (--*--)
EMp2_ep     47     3.704      4.459   (---*---)
Sk1_ep      19    11.969      6.862                    (-----*-----)
Sk2_ep      54     9.167      4.074                (--*---)
Sk3_ep      71     6.912      4.598           (--*--)
Sk_ep      144     8.425      5.044                (-*-)
                                       ----+---------+---------+---------+-
Pooled StDev =     4.799                  3.5       7.0      10.5      14.0
```

'Contingent Negative' Modality

```
                                               Individual 95% CIs For Mean Based on
                                               Pooled StDev
Level          N         Mean        StDev   ---------+---------+---------+-------
D1_cn        163        1.866        5.489         (----*-----)
EMp_cn        88        3.559        6.515                  (-------*------)
EMp2_cn       47        1.742        4.009    (---------*---------)
Sk1_cn        19        3.239        3.846        (---------------*--------------)
Sk2_cn        54        3.993        4.864                   (---------*--------)
Sk3_cn        71        3.496        5.152                 (-------*-------)
Sk_cn        144        3.648        4.869                     (----*-----)
                                              ---------+---------+---------+------
Pooled StDev =          5.270                          1.5       3.0       4.5
```

'New_Discovery' Wordlist

```
                                               Individual 95% CIs For Mean Based on
                                               Pooled StDev
Level          N          Mean       StDev   -----+---------+---------+---------+-
D1_ndis      163        2.072        3.330           (---*---)
EMp_ndis      88        3.206        2.110                   (-----*----)
EMp2_ndis     47        1.505        2.113   (-------*------)
Sk1_ndis      19        2.343        1.876      (-----------*----------)
Sk2_ndis      54        4.022        4.070                         (------*-----)
Sk3_ndis      71        2.949        2.569                 (-----*----)
Sk_ndis      144        3.271        3.192                     (---*---)
                                              -----+---------+---------+---------+-
Pooled StDev =          3.011                     1.2       2.4       3.6       4.8
```

'Any_Experiment' Wordlist

```
                                          Individual 95% CIs For Mean Based on
                                          Pooled StDev
Level         N      Mean     StDev   ---+---------+---------+---------+---
D1_anyx     163     3.276     4.761        (--*--)
EMp_anyx     88     5.467     2.774                  (---*----)
EMp2_anyx    47     2.679     3.246    (----*-----)
Sk1_anyx     19     4.857     2.913          (--------*--------)
Sk2_anyx     54     6.921     4.195                        (-----*----)
Sk3_anyx     71     5.128     3.915                (----*---)
Sk_anyx     144     5.765     3.989                     (--*--)
                                       ---+---------+---------+---------+---
Pooled StDev =      3.999                 2.0       4.0       6.0       8.0
```

'Bench_words' Wordlist

```
                                          Individual 95% CIs For Mean Based on
                                          Pooled StDev
Level         N      Mean     StDev   ------+---------+---------+---------+
D1_bnch     163     1.034     1.869        (-*--)
EMp_bnch     88     2.290     1.602                     (---*---)
EMp2_bnch    47     0.965     1.299   (-----*----)
Sk1_bnch     19     1.601     1.228        (-------*-------)
Sk2_bnch     54     3.056     2.058                            (----*---)
Sk3_bnch     71     2.037     1.888                (---*----)
Sk_bnch     144     2.362     1.953          (--*--)
                                       ------+---------+---------+---------+
Pooled StDev =      1.818                    1.0       2.0       3.0       4.0
```

Graph 2. Analysis of Variance/Confidence Levels for the 1831 Private and Published Writings on Electromagnetism.

Compression

```
                                        Individual 95% CIs For Mean Based on
                                        Pooled StDev
Level          N      Mean     StDev   -+---------+---------+---------+-----
DI_comp      539     18.25     11.43   (--*--)
ERE1_comp    452     27.80     12.33                           (--*---)
                                       -+---------+---------+---------+----
Pooled StDev =       11.85             17.5      21.0      24.5      28.0
```

Conjunctives

```
                                            Individual 95% CIs For Mean Based on
                                            Pooled StDev
Level          N       Mean     StDev   ---+---------+---------+---------+---
D1_conj      719      8.050     7.776     (------*------)
ERE1_conj    532      9.511     5.893                      (--------*-------)
                                        ---+---------+---------+---------+---
Pooled StDev =        7.038                 7.70      8.40      9.10      9.80
```

Paragraph Length

```
                                            Individual 95% CIs For Mean Based on
                                            Pooled StDev
Level          N      Mean     StDev       ----+---------+---------+---------+--
D1_pl        539     45.53     32.80        (-*--)
ERE1_pl      452     93.52     57.38                                     (-*--)
                                            ----+---------+---------+---------+-
Pooled StDev =       45.68                      48        64        80        96
```

'Empirical Positive' Modality

```
                                       Individual 95% CIs For Mean Based on
                                       Pooled StDev
Level          N        Mean     StDev  ---+---------+---------+---------+---
D1_ep        719       4.144     4.681   (---*---)
ERE1_ep      532       6.128     4.590                       (----*----)
                                         ---+---------+---------+---------+--
Pooled StDev =         4.642              4.00      4.80      5.60      6.40
```

'Contingent Negative' Modality

```
                                       Individual 95% CIs For Mean Based on
                                       Pooled StDev
Level          N        Mean      StDev  ---------+---------+---------+-------
D1_cn        719       2.823      7.311      (-------------*------------)
ERE1_cn      532       2.687      4.746  (---------------*--------------)
                                         ---------+---------+---------+------
Pooled StDev =         6.348                    2.45      2.80      3.15
```

'New_Discovery' Wordlist

```
                                       Individual 95% CIs For Mean Based on
                                       Pooled StDev
Level           N       Mean      StDev  --+---------+---------+---------+----
D1_ndis       719      1.749      2.778  (-----*----)
ERE1_ndis     532      2.754      3.053                        (-----*-----)
                                         --+---------+---------+---------+---
Pooled StDev =         2.898             1.60      2.00      2.40      2.80
```

'Any_Experiment' Wordlist

```
                                            Individual 95% CIs For Mean Based on
                                            Pooled StDev
Level          N        Mean       StDev    ------+---------+---------+---------+
D1_anyx      719       4.081       4.861    (------*-----)
ERE1_anyx    532       4.975       4.206                   (------*-------)
                                            ------+---------+---------+---------+
Pooled StDev =        4.594                     4.00      4.50      5.00      5.50
```

'Bench_Word' Wordlist

```
                                            Individual 95% CIs For Mean Based on
                                            Pooled StDev
Level          N        Mean       StDev    -------+---------+---------+---------
D1_bwd       719       2.286       3.222        (-------------*-------------)
ERE1_bwd     532       2.246       2.323   (----------------*---------------)
                                            -------+---------+---------+--------
Pooled StDev =        2.875                      2.10      2.25      2.40
```

Graph 3. Analysis of Variance/Confidence Levels for the 1821 and 1831 Private and Published Writings on Electromagnetism (excluding the *Sketch*).

Compression

```
                                         Individual 95% CIs For Mean Based on
                                         Pooled StDev
Level          N      Mean        StDev  -----+---------+---------+---------+-
D1_comp       86     18.15        12.54   (-*--)
EMp_comp      83     37.38        14.31                  (-*--)
EMp2_comp     22     41.30        12.88                   (----*----)
D1_comp      539      18.25       11.43    (*)
ERE1_comp    452      27.80       12.33            (*)
                                         -----+---------+---------+---------+
Pooled StDev =    12.11                      20        30        40        50
```

Conjunctives

```
                                         Individual 95% CIs For Mean Based on
                                         Pooled StDev
Level        N          Mean      StDev  -------+---------+---------+---------
D1_conj     163         5.147     6.411   (----*---)
EMp_conj     88        10.843     4.152                      (----*-----)
EMp2_conj    47         5.295     5.918 (-------*-------)
D1_conj     719         8.050     7.776               (-*-)
ERE1_conj   532         9.511     5.893                     (-*-)
                                         -------+---------+---------+--------
Pooled StDev =6.811                             5.0       7.5      10.0
```

Paragraph Length

```
                                     Individual 95% CIs For Mean Based on
                                     Pooled StDev
Level        N      Mean      StDev  -----+---------+---------+---------+-
D1_pl       86     47.29      38.17  (-*-)
EMp_pl      83    164.52     116.97                    (-*-)
EMp2_pl     22    212.59     123.69                           (--*---)
D1_pl      539     45.53      32.80   (*
ERE1_pl    452     93.52      57.38          (*
                                     -----+---------+---------+---------+
Pooled StDev =     55.54                 60       120       180       240
```

'Empirical Positive' Modality

```
                                     Individual 95% CIs For Mean Based on
                                     Pooled StDev
Level        N      Mean      StDev  ---------+---------+---------+-------
D1_ep      163     3.683      5.142     (--*---)
EMp_ep      88     7.963      3.857                         (----*----)
EMp2_ep     47     3.704      4.459  (------*-----)
D1_ep      719     4.144      4.681         (-*)
ERE1_ep    532     6.128      4.590                   (-*-)
                                     ---------+---------+---------+-----
Pooled StDev =     4.652                    4.0       6.0       8.0
```

'Contingent Negative' Modality

```
                                     Individual 95% CIs For Mean Based on
                                     Pooled StDev
Level        N       Mean     StDev  -+---------+---------+---------+-----
D1_cn      163      1.866     5.489         (-----*------)
EMp_cn      88      3.559     6.515                 (--------*-------)
EMp2_cn     47      1.742     4.009   (-----------*----------)
D1_cn      719      2.823     7.311                  (--*--)
ERE1_cn    532      2.687     4.746                (---*--)
                                      -+---------+---------+---------+----
Pooled StDev =      6.216            0.0       1.5       3.0       4.5
```

'New_Discovery' Wordlist

```
                                     Individual 95% CIs For Mean Based on
                                     Pooled StDev
Level        N       Mean     StDev  ----+---------+---------+---------+--
D1_ndis    163      2.072     3.330          (----*---)
EMp_ndis    88      3.206     2.110                   (-----*-----)
EMp2_ndis   47      1.505     2.113   (-------*-------)
D1_ndis    719      1.749     2.778         (-*--)
ERE1_ndis  532      2.754     3.053                  (--*-)
                                      ----+---------+---------+---------+-
Pooled StDev =       2.888                1.0       2.0       3.0       4.0
```

'Any_Experiment' Wordlist

```
                                          Individual 95% CIs For Mean Based on
                                          Pooled StDev
Level        N      Mean     StDev  -+---------+---------+---------+-----
D1_anyx     163     3.276    4.761          (----*---)
EMp_anyx     88     5.467    2.774                       (-----*------)
EMp2_anyx    47     2.679    3.246  (--------*-------)
D1_anyx     719     4.081    4.861                 (-*-)
ERE1_anyx   532     4.975    4.206                       (-*--)
                                     -+---------+---------+---------+----
Pooled StDev =      4.494           1.5       3.0       4.5       6.0
```

'Bench_Words' Wordlist

```
                                          Individual 95% CIs For Mean Based on
                                          Pooled StDev
Level        N      Mean     StDev  --------+---------+---------+--------
D1_bwd      163     1.034    1.869       (----*----)
EMp_bwd      88     2.290    1.602                    (------*------)
EMp2_bwd     47     0.965    1.299  (---------*---------)
D1_bwd      719     2.286    3.222                        (--*-)
ERE1_bwd    532     2.246    2.323                       (--*--)
                                     --------+---------+---------+-------
Pooled StDev =      2.692                 0.80      1.60      2.40
```

Graph 4. Analysis of Variance/Confidence Levels for the 1821 and 1831 Published Writings and Letters on Electromagnetism.

Compression

```
                                         Individual 95% CIs For Mean Based on
                                         Pooled StDev
Level          N      Mean     StDev   -------+---------+--------+---------
EMp_comp      83     37.38     14.31                    (---*---)
EMp2_comp     22     41.30     12.88                      (-------*-----)
ERE1_comp    452      27.80     12.33        (-*)
lets_cmp      70     26.52     15.84    (---*---)
                                       -------+---------+--------+---------
Pooled StDev =       13.06                  28.0      35.0     42.0
```

Conjunctives

```
                                          Individual 95% CIs For Mean Based on
                                          Pooled StDev
Level          N       Mean        StDev  ------+---------+---------+---------+
EMp_conj      88      10.843      4.152                         (----*----)
EMp2_conj     47       5.295      5.918  (------*------)
ERE1_conj    532       9.511      5.893                       (-*-)
lets_conj    114       6.764      7.227             (---*---)
                                          ------+---------+---------+---------+
Pooled StDev =         5.944                   5.0       7.5      10.0      12.5
```

Paragraph Length

```
                                         Individual 95% CIs For Mean Based on
                                         Pooled StDev
Level          N      Mean     StDev   ---+---------+---------+---------+---
EMp_pl        83    164.52    116.97                (--*--)
EMp2_pl       22    212.59    123.69                      (------*-----)
ERE1_pl      452     93.52     57.38   (-*)
lets_pl       70    118.09    106.71      (---*--)
                                       ---+---------+---------+---------+---
Pooled StDev =      77.21              100       150       200       250
```

'Empirical Positive' Modality

```
                                          Individual 95% CIs For Mean Based on
                                          Pooled StDev
Level          N       Mean      StDev  --------+---------+---------+--------
EMp_ep        88      7.963      3.857                      (----*----)
EMp2_ep       47      3.704      4.459  (------*-----)
ERE1_ep      532      6.128      4.590                 (-*-)
lets_ep      114      4.365      4.500       (---*---)
                                        --------+---------+---------+-------
Pooled StDev =        4.493                   4.0       6.0       8.0
```

'Contingent Negative' Modality

```
                                         Individual 95% CIs For Mean Based on
                                         Pooled StDev
Level          N       Mean     StDev  --------+---------+---------+--------
EMp_cn        88      3.559     6.515                   (--------*-------)
EMp2_cn       47      1.742     4.009  (------------*-----------)
ERE1_cn      532      2.687     4.746                 (--*---)
lets_cn      114      3.416     5.486                   (------*-------)
                                        --------+---------+---------+-------
Pooled StDev =        5.048                   1.2       2.4       3.6
```

'New_Discovery' Wordlist

```
                                          Individual 95% CIs For Mean Based on
                                          Pooled StDev
Level          N       Mean     StDev   ----+---------+---------+---------+--
EMp_ndis      88      3.206     2.110                    (-----*-----)
EMp2_ndis     47      1.505     2.113   (-------*-------)
ERE1_ndis    532      2.754     3.053                   (--*-)
let_ndis     114      1.117     1.789  (----*----)
                                        ----+---------+---------+---------+--
Pooled StDev =        2.756               1.0       2.0       3.0       4.0
```

'Any_Experiment' Wordlist

```
                                          Individual 95% CIs For Mean Based on
                                          Pooled StDev
Level         N       Mean      StDev   ----------+---------+---------+------
EMp_anyx     88      5.467      2.774                     (----*-----)
EMp2_anyx    47      2.679      3.246   (-------*------)
ERE1_anyx   532      4.975      4.206                     (-*-)
let_anyx    114      3.069      3.485        (---*----)
                                         ----------+---------+---------+-----
Pooled StDev =      3.917                          3.0       4.5       6.0
```

'Bench_word' Wordlist

```
                                          Individual 95% CIs For Mean Based on
                                          Pooled StDev
 Level        N       Mean      StDev   -----+---------+---------+---------+-
EMp_bnch     88      2.290      1.602                     (------*-----)
EMp2_bnch    47      0.965      1.299   (--------*-------)
ERE1_bnch   532      2.246      2.323                         (-*--)
let_bnch    114      1.244      1.733         (-----*----)
                                         -----+---------+---------+---------+-
Pooled StDev =      2.124                 0.70      1.40      2.10      2.80
```

Appendix Seven: A Funeral Elegy for Master William Peter[660]

"W[illiam] S[hakespeare], "A Funeral Elegy." Edited by Donald W. Foster from W.S., A Funerall Elegye in memory of the late vertuous Maister William Peeter (London: G. Eld for T. Thorpe, 1612). [4,600 words.] Common nouns capitalized and italicized in Q are here capitalized but not italicized; italicized quotations in Q are rendered in quotation marks.

Participial endings and ellisions may be normalized for use with a private text archive. DWF (1/15/96)"

TO MASTER JOHN PETER
of Bowhay in Devon, Esquire.

The love I bore to your brother, and will do to his memory, hath crav'd from me this last duty of a friend; I am herein but a second to the privilege of Truth, who can warrant more in his behalf than I undertook to deliver. Exercise in this kind I will little affect, and am less addicted to, but there must be miracle in that labor which, to witness my remembrance to this departed gentleman, I would not willingly undergo. Yet whatsoever is here done, is done to him, and to him only. For whom and whose sake I will not forget to remember any friendly respects to you, or to any of those that have lov'd him for himself, and himself for his deserts.

W. S.

A FUNERAL ELEGY

Since Time, and his predestinated end,
Abridg'd the circuit of his hopeful days,
Whiles both his Youth and Virtue did intend
The good endeavors of deserving praise,
What memorable monument can last
Whereon to build his never-blemish'd name
But his own worth, wherein his life was grac'd-
Sith as [that] ever he maintain'd the same?
Oblivion in the darkest day to come,
When sin shall tread on merit in the dust,
Cannot rase out the lamentable tomb
Of his short-liv'd deserts; but still they must,
Even in the hearts and memories of men,
Claim fit Respect, that they, in every limb
Rememb'ring what he was, with comfort then
May pattern out one truly good, by him.
For he was truly good, if honest care
Of harmless conversation may commend
A life free from such stains as follies are,
Ill recompensed only in his end.
Nor can the tongue of him who lov'd him least
(If there can be minority of love

[660] http://www.etsu-tn.edu/english/elegy.htm; accessed February 5; 1998.

To one superlative above the rest
Of many men in steady faith) reprove
His constant temper, in the equal weight
Of thankfulness and kindness: Truth doth leave
Sufficient proof, he was in every right
As kind to give, as thankful to receive.
The curious eye of a quick-brain'd survey
Could scantly find a mote amidst the sun
Of his too-short'ned days, or make a prey
Of any faulty errors he had done-
Not that he was above the spleenful sense
And spite of malice, but for that he had
Warrant enough in his own innocence
Against the sting of some in nature bad.
Yet who is he so absolutely blest
That lives encompass'd in a mortal frame,
Sometime in reputation not oppress'd
By some in nothing famous but defame?
Such in the By-path and the Ridgeway lurk
That leads to ruin, in a smooth pretense
Of what they do to be a special work
Of singleness, not tending to offense;
Whose very virtues are, not to detract
Whiles hope remains of gain (base fee of slaves),
Despising chiefly men in fortunes wrack'd-
But death to such gives unrememb'red graves.
Now therein liv'd he happy, if to be
Free from detraction happiness it be.
His younger years gave comfortable hope
To hope for comfort in his riper youth,
Which, harvest-like, did yield again the crop
Of Education, better'd in his truth.
Those noble twins of heaven-infused races,
Learning and Wit, refined in their kind
Did jointly both, in their peculiar graces,
Enrich the curious temple of his mind;
Indeed a temple, in whose precious white
Sat Reason by Religion oversway'd,
Teaching his other senses, with delight,
How Piety and Zeal should be obey'd-
Not fruitlessly in prodigal expense
Wasting his best of time, but so content
With Reason's golden Mean to make defense
Against the assault of youth's encouragement;
As not the tide of this surrounding age
(When now his father's death had freed his will)
Could make him subject to the drunken rage
Of such whose only glory is their ill.
He from the happy knowledge of the wise
Draws virtue to reprove secured fools
And shuns the glad sleights of ensnaring vice
To spend his spring of days in sacred schools.

Here gave he diet to the sick desires
That day by day assault the weaker man,
And with fit moderation still retires
From what doth batter virtue now and then.
But that I not intend in full discourse
To progress out his life, I could display
A good man in each part exact and force
The common voice to warrant what I say.
For if his fate and heaven had decreed
That full of days he might have liv'd to see
The grave in peace, the times that should succeed
Had been best-speaking witnesses with me;
Whose conversation so untouch'd did move
Respect most in itself, as who would scan
His honesty and worth, by them might prove
He was a kind, true, perfect gentleman-
Not in the outside of disgraceful folly,
Courting opinion with unfit disguise,
Affecting fashions, nor addicted wholly
To unbeseeming blushless vanities,
But suiting so his habit and desire
As that his Virtue was his best Attire.
Not in the waste of many idle words
Car'd he to be heard talk, nor in the float
Of fond conceit, such as this age affords,
By vain discourse upon himself to dote;
For his becoming silence gave such grace
To his judicious parts, as what he spake
Seem'd rather answers which the wise embrace
Than busy questions such as talkers make.
And though his qualities might well deserve
Just commendation, yet his furnish'd mind
Such harmony of goodness did preserve
As nature never built in better kind;
Knowing the best, and therefore not presuming
In knowing, but for that it was the best,
Ever within himself free choice resuming
Of true perfection, in a perfect breast;
So that his mind and body made an inn,
The one to lodge the other, both like fram'd
For fair conditions, guests that soonest win
Applause; in generality, well fam'd,
If trim behavior, gestures mild, discreet
Endeavors, modest speech, beseeming mirth,
True friendship, active grace, persuasion sweet,
Delightful love innated from his birth,
Acquaintance unfamiliar, carriage just,
Offenseless resolution, wish'd sobriety,
Clean-temper'd moderation, steady trust,
Unburthen'd conscience, unfeign'd piety;
If these, or all of these, knit fast in one
Can merit praise, then justly may we say,

Not any from this frailer stage is gone
Whose name is like to live a longer day-
Though not in eminent courts or places great
For popular concourse, yet in that soil
Where he enjoy'd his birth, life, death, and seat
Which now sits mourning his untimely spoil.
And as much glory is it to be good
For private persons, in their private home,
As those descended from illustrious blood
In public view of greatness, whence they come.
Though I, rewarded with some sadder taste
Of knowing shame, by feeling it have prov'd
My country's thankless misconstruction cast
Upon my name and credit, both unlov'd
By some whose fortunes, sunk into the wane
Of plenty and desert, have strove to win
Justice by wrong, and sifted to embane
My reputation with a witless sin;
Yet time, the father of unblushing truth,
May one day lay ope malice which hath cross'd it,
And right the hopes of my endangered youth,
Purchasing credit in the place I lost it.
Even in which place the subject of the verse
(Unhappy matter of a mourning style
Which now that subject's merits doth rehearse)
Had education and new being; while
By fair demeanor he had won repute
Amongst the all of all that lived there,
For that his actions did so wholly suit
With worthiness, still memorable here.
The many hours till the day of doom
Will not consume his life and hapless end,
For should he lie obscur'd without a tomb,
Time would to time his honesty commend;
Whiles parents to their children will make known,
And they to their posterity impart,
How such a man was sadly overthrown
By a hand guided by a cruel heart,
  Whereof as many as shall hear that sadness
  Will blame the one's hard fate, the other's madness;
Whiles such as do recount that tale of woe,
Told by remembrance of the wisest heads,
Will in the end conclude the matter so,
As they will all go weeping to their beds.
For when the world lies winter'd in the storms
Of fearful consummation, and lays down
Th' unsteady change of his fantastic forms,
Expecting ever to be overthrown;
When the proud height of much affected sin
Shall ripen to a head, and in that pride
End in the miseries it did begin
And fall amidst the glory of his tide;

Then in a book where every work is writ
Shall this man's actions be reveal'd, to show
The gainful fruit of well-employed wit,
Which paid to heaven the debt that it did owe.
Here shall be reckon'd up the constant faith,
Never untrue, where once he love profess'd;
Which is a miracle in men, one saith,
Long sought though rarely found, and he is best
Who can make friendship, in those times of change,
Admired more for being firm than strange.
When those weak houses of our brittle flesh
Shall ruin'd be by death, our grace and strength,
Youth, memory and shape that made us fresh
Cast down, and utterly decay'd at length;
When all shall turn to dust from whence we came
And we low-level'd in a narrow grave,
What can we leave behind us but a name,
Which, by a life well led, may honor have?
Such honor, O thou youth untimely lost,
Thou didst deserve and hast; for though thy soul
Hath took her flight to a diviner coast,
Yet here on earth thy fame lives ever whole,
In every heart seal'd up, in every tongue
Fit matter to discourse, no day prevented
That pities not thy sad and sudden wrong,
Of all alike beloved and lamented.
And I here to thy memorable worth,
In this last act of friendship, sacrifice
My love to thee, which I could not set forth
In any other habit of disguise.
Although I could not learn, whiles yet thou wert,
To speak the language of a servile breath,
My truth stole from my tongue into my heart,
Which shall not thence be sund'red, but in death.
And I confess my love was too remiss
That had not made thee know how much I priz'd thee,
But that mine error was, as yet it is,
To think love best in silence: for I siz'd thee
By what I would have been, not only ready
In telling I was thine, but being so,
By some effect to show it. He is steady
Who seems less than he is in open show.
Since then I still reserv'd to try the worst
Which hardest fate and time thus can lay on me.
T' enlarge my thoughts was hindered at first,
While thou hadst life; I took this task upon me,
To register with mine unhappy pen
Such duties as it owes to thy desert,
And set thee as a president to men,
And limn thee to the world but as thou wert-
Not hir'd, as heaven can witness in my soul,
By vain conceit, to please such ones as know it,

Nor servile to be lik'd, free from control,
Which, pain to many men, I do not owe it.
But here I trust I have discharged now
(Fair lovely branch too soon cut off) to thee,
My constant and irrefragable vow,
As, had it chanc'd, thou mightst have done to me-
But that no merit strong enough of mine
Had yielded store to thy well-abled quill
Whereby t'enroll my name, as this of thine,
How s'ere enriched by thy plenteous skill.
Here, then, I offer up to memory
The value of my talent, precious man,
Whereby if thou live to posterity,
Though't be not as I would, 'tis as I can:
In minds from whence endeavor doth proceed,
A ready will is taken for the deed.
Yet ere I take my longest last farewell
From thee, fair mark of sorrow, let me frame
Some ampler work of thank, wherein to tell
What more thou didst deserve than in thy name,
And free thee from the scandal of such senses
As in the rancor of unhappy spleen
Measure thy course of life, with false pretenses
Comparing by thy death what thou hast been.
So in his mischiefs is the world accurs'd:
It picks out matter to inform the worst.
The willful blindness that hoodwinks the eyes
Of men enwrapped in an earthy veil
Makes them most ignorantly exercise
And yield to humor when it doth assail,
Whereby the candle and the body's light
Darkens the inward eyesight of the mind,
Presuming still it sees, even in the night
Of that same ignorance which makes them blind.
Hence conster they with corrupt commentaries,
Proceeding from a nature as corrupt,
The text of malice, which so often varies
As 'tis by seeming reason underpropp'd.
O, whither tends the lamentable spite
Of this world's teenful apprehension,
Which understands all things amiss, whose light
Shines not amidst the dark of their dissension?
True 'tis, this man, whiles yet he was a man,
Sooth'd not the current of besotted fashion,
Nor could disgest, as some loose mimics can,
An empty sound of overweening passion,
So much to be made servant to the base
And sensual aptness of disunion'd vices,
To purchase commendation by disgrace,
Whereto the world and heat of sin entices.
But in a safer contemplation,
Secure in what he knew, he ever chose

The ready way to commendation,
By shunning all invitements strange, of those
Whose illness is, the necessary praise
Must wait upon their actions; only rare
In being rare in shame (which strives to raise
Their name by doing what they do not care),
As if the free commission of their ill
Were even as boundless as their prompt desires;
Only like lords, like subjects to their will,
Which their fond dotage ever more admires.
He was not so: but in a serious awe,
Ruling the little ordered commonwealth
Of his own self, with honor to the law
That gave peace to his bread, bread to his health;
Which ever he maintain'd in sweet content
And pleasurable rest, wherein he joy'd
A monarchy of comfort's government,
Never until his last to be destroy'd.
For in the Vineyard of heaven-favored learning
Where he was double-honor'd in degree,
His observation and discreet discerning
Had taught him in both fortunes to be free;
Whence now retir'd home, to a home indeed
The home of his condition and estate,
He well provided 'gainst the hand of need,
Whence young men sometime grow unfortunate;
His disposition, by the bonds of unity,
So fast'ned to his reason that it strove
With understanding's grave immunity
To purchase from all hearts a steady love;
Wherein not any one thing comprehends
Proportionable note of what he was,
Than that he was so constant to his friends
As he would no occasion overpass
Which might make known his unaffected care,
In all respects of trial, to unlock
His bosom and his store, which did declare
That Christ was his, and he was Friendship's Rock:
A Rock of Friendship figured in his name,
Fore-shewing what he was, and what should be,
Most true presage; and he discharg'd the same
In every act of perfect amity-
Though in the complemental phrase of words
He never was addicted to the vain
Of boast, such as the common breath affords;
He was in use most fast, in tongue most plain,
Nor amongst all those virtues that forever
Adorn'd his reputation will be found
One greater than his Faith, which did persever,
Where once it was protested, alway sound.
Hence sprung the deadly fuel that reviv'd
The rage which wrought his end, for had he been

Slacker in love, he had been longer liv'd
And not oppress'd by wrath's unhappy sin-
By wrath's unhappy sin, which unadvis'd
Gave death for free good will, and wounds for love.
Pity it was that blood had not been priz'd
At higher rate, and reason set above
Most unjust choler, which untimely Drew
Destruction on itself; and most unjust,
Robb'd virtue of a follower so true
As time can boast of, both for love and trust:
So henceforth all (great glory to his blood)
Shall be but seconds to him, being good.
The wicked end their honor with their sin
In death, which only then the good begin.
Lo, here a lesson by experience taught
For men whose pure simplicity hath drawn
Their trust to be betray'd by being caught
Within the snares of making truth a pawn;
Whiles it, not doubting whereinto it enters,
Without true proof and knowledge of a friend,
Sincere in singleness of heart, adventers
To give fit cause, ere love begin to end:
His unfeign'd friendship where it least was sought,
Him to a fatal timeless ruin brought;
Whereby the life that purity adorn'd
With real merit, by this sudden end
Is in the mouth of some in manner scorn'd,
Made questionable, for they do intend,
According to the tenor of the saw
Mistook, if not observ'd (writ long ago
When men were only led by Reason's law),
That "Such as is the end, the life proves so."
Thus he, who to the universal lapse
Gave sweet redemption, off'ring up his blood
To conquer death by death, and loose the traps
Of hell, even in the triumph that it stood:
He thus, for that his guiltless life was spilt
By death, which was made subject to the curse,
Might in like manner be reprov'd of guilt
In his pure life, for that his end was worse.
But O far be it, our unholy lips
Should so profane the deity above
As thereby to ordain revenging whips
Against the day of Judgment and of Love.
The hand that lends us honor in our days
May shorten when it please, and justly take
Our honor from us many sundry ways,
As best becomes that wisdom did us make.
The second brother, who was next begot
Of all that ever were begotten yet,
Was by a hand in vengeance rude and hot
Sent innocent to be in heaven set-

Whose fame the angels in melodious choirs
Still witness to the world. Then why should he,
Well-profited in excellent desires,
Be more rebuk'd, who had like destiny?
Those saints before the everlasting throne
Who sit with crowns of glory on their heads,
Wash'd white in blood, from earth hence have not gone
All to their joys in quiet on their beds,
But tasted of the sour-bitter scourge
Of torture and affliction ere they gained
Those blessings which their sufferance did urge,
Whereby the grace fore-promis'd they attained.
Let then the false suggestions of the froward,
Building large castles in the empty air,
By suppositions fond and thoughts untoward
(Issues of discontent and sick despair)
Rebound gross arguments upon their heart
That may disprove their malice, and confound
Uncivil loose opinions which insert
Their souls into the roll that doth unsound
Betraying policies, and show their brains,
Unto their shame, ridiculous; whose scope
Is envy, whose endeavors fruitless pains,
In nothing surely prosperous, but hope-
And that same hope, so lame, so unprevailing,
It buries self-conceit in weak opinion;
Which being cross'd, gives matter of bewailing
Their vain designs, on whom want hath dominion.
Such, and of such condition, may devise
Which way to wound with defamation's spirit
(Close-lurking whisper's hidden forgeries)
His taintless goodness, his desertful merit.
But whiles the minds of men can judge sincerely,
Upon assured knowledge, his repute
And estimation shall be rumor'd clearly
In equal worth--time shall to time renew 't.
The Grave-that in his ever-empty womb
Forever closes up the unrespected
Who, when they die, die all-shall not entomb
His pleading best perfections as neglected.
They to his notice in succeeding years
Shall speak for him when he shall lie below;
When nothing but his memory appears
Of what he was, then shall his virtues grow.
His being but a private man in rank
(And yet not rank'd beneath a gentleman)
Shall not abridge the commendable thank
Which wise posterity shall give him then;
For Nature, and his therein happy Fate.
Ordain'd that by his quality of mind
T' ennoble that best part, although his state
Were to a lower blessedness confin'd.

Blood, pomp, state, honor, glory and command,
Without fit ornaments of disposition,
Are in themselves but heathenish and [profaned],
And much more peaceful is a mean condition
Which, underneath the roof of safe content,
Feeds on the bread of rest, and takes delight
To look upon the labors it hath spent
For its own sustenance, both day and night;
Whiles others, plotting which way to be great,
How to augment their portion and ambition,
Do toil their giddy brains, and ever sweat
For popular applause and power's commission.
But one in honors, like a seeled dove
Whose inward eyes are dimm'd with dignity,
Does think most safety doth remain above,
And seeks to be secure by mounting high:
  Whence, when he falls, who did erewhile aspire,
  Falls deeper down, for that he climbed higher.
Now men who in a lower region live
Exempt from danger of authority
Have fittest times in Reason's rules to thrive,
Not vex'd with envy of priority,
  And those are much more noble in the mind
  Than many that have nobleness by kind.
Birth, blood, and ancestors, are none of ours,
Nor can we make a proper challenge to them,
But virtues and perfections in our powers
Proceed most truly from us, if we do them.
Respective titles or a gracious style,
With all what men in eminence possess,
Are, without ornaments to praise them, vile:
The beauty of the mind is nobleness.
And such as have that beauty, well deserve
Eternal characters, that after death
Remembrance of their worth we may preserve,
So that their glory die not with their breath.
Else what avails it in a goodly strife
Upon this face of earth here to contend,
The good t'exceed the wicked in their life,
Should both be like obscured in their end?
Until which end, there is none rightly can
Be termed happy, since the happiness
Depends upon the goodness of the man,
Which afterwards his praises will express.
Look hither then, you that enjoy the youth
Of your best days, and see how unexpected
Death can betray your jollity to ruth
When death you think is least to be respected!
The person of this model here set out
Had all that youth and happy days could give him,
Yet could not all-encompass him about
Against th'assault of death, who to relieve him

Strook home but to the frail and mortal parts
Of his humanity, but could not touch
His flourishing and fair long-liv'd deserts,
Above fate's reach, his singleness was such-
So that he dies but once, but doubly lives,
Once in his proper self, then in his name;
Predestinated Time, who all deprives,
Could never yet deprive him of the same.
And had the Genius which attended on him
Been possibilited to keep him safe
Against the rigor that hath overgone him,
He had been to the public use a staff,
Leading by his example in the path
Which guides to doing well, wherein so few
The proneness of this age to error hath
Informed rightly in the courses true.
As then the loss of one, whose inclination
Strove to win love in general, is sad,
So specially his friends, in soft compassion
Do feel the greatest loss they could have had.
Amongst them all, she who those nine of years
Liv'd fellow to his counsels and his bed
Hath the most share in loss: for I in hers
Feel what distemperature this chance hath bred.
The chaste embracements of conjugal love,
Who in a mutual harmony consent,
Are so impatient of a strange remove
As meager death itself seems to lament,
And weep upon those cheeks which nature fram'd
To be delightful orbs in whom the force
Of lively sweetness plays, so that asham'd
Death often pities his unkind divorce.
Such was the separation here constrain'd
(Well-worthy to be termed a rudeness rather),
For in his life his love was so unfeign'd
As he was both an husband and a father-
The one in firm affection and the other
In careful providence, which ever strove
With joint assistance to grace one another
With every helpful furtherance of love.
But since the sum of all that can be said
Can be but said that "He was good" (which wholly
Includes all excellence can be display'd
In praise of virtue and reproach of folly).
His due deserts, this sentence on him gives,
"He died in life, yet in his death he lives."
Now runs the method of this doleful song
In accents brief to thee, O thou deceas'd!
To whom those pains do only all belong
As witnesses I did not love thee least.
For could my worthless brain find out but how
To raise thee from the sepulcher of dust,

Undoubtedly thou shouldst have partage now
Of life with me, and heaven be counted just
If to a supplicating soul it would
Give life anew, by giving life again
Where life is miss'd; whereby discomfort should
Right his old griefs, and former joys retain
Which now with thee are leap'd into thy tomb
And buried in that hollow vault of woe,
Expecting yet a more severer doom
Than time's strict flinty hand will let 'em know.
And now if I have level'd mine account
And reckon'd up in a true measured score
Those perfect graces which were ever wont
To wait on thee alive, I ask no more
(But shall hereafter in a poor content
Immure those imputations I sustain,
Learning my days of youth so to prevent
As not to be cast down by them again)-
Only those hopes which fate denies to grant
In full possession to a captive heart
Who, if it were in plenty, still would want
Before it may enjoy his better part;
From which detain'd, and banish'd in th' exile
Of dim misfortune, has none other prop
Whereon to lean and rest itself the while
But the weak comfort of the hapless, Hope.
And Hope must in despite of fearful change
Play in the strongest closet of my breast,
Although perhaps I ignorantly range
And court opinion in my deep'st unrest.
But whether doth the stream of my mischance
Drive me beyond myself, fast friend, soon lost,
Long may thy worthiness thy name advance
Amongst the virtuous and deserving most,
Who herein hast forever happy prov'd:
In life thou liv'dst, in death thou died'st belov'd.

FINIS

Bibliography

Aarseth; Espen; *The Field of Humanistic Informatics and its Relation to the Humanities*; http://www.hf-fak.uib.no/hi/espen/HI.html; September 24, 1996; accessed September 21, 1997.

Abrams, Richard; *In Defence of W.S*; The Times Literary Supplement; February 9; 1996: 25-26.

Addis, T. R. and Addis, J. J. T.; Clarity; http://www.sis.port.ac.uk/research/clarity/index.html; February 26, 1997; accessed March 17, 1998.

Agassi, Joseph; Faraday as a natural philosopher; The University of Chicago Press: Chicago and London; 1971.

Ager, D. E. *et al* (Eds.); Advances in Computer-Aided Literary And Linguistic Research; AMLC, for the Department of Modern Languages, University of Aston in Birmingham; 1979.

Aitken, A. J. *et al*; The Computer and Literary Studies; Edinburgh University Press: Edinburgh; 1973: 85-99.

Anderson, James M. (Ed.); The Linguistics Encyclopedia; Routledge: London and New York; 1991.

Aviva, Freedman and Medway, Peter (eds.); Genre And The New Rhetoric; Taylor and Francis: London and Bristol, USA; 1994.

Badalamenti, Anthony F., Langs, Robert J., and Robinson, James; *Lawful Systems Dynamics In How Poets Choose Their Words;* Behavioral Science, Vol. 39; 1994: 46-71.

Bailey, Nathanial; Dictionarium britannicum, or A compleat etymological English dictionary; S.N: London; 1721.

Bailey, Richard, W; *Statistics and Style: A Historical Survey*, in Dolezel, Lubomír and Bailey, Richard, W. (Eds.); Statistics and Style; American Elsevier Publishing Company, Inc.: New York; 1969: 218, 219, 220, 223.

Bakhtin, Mikhail; The Dialogic Imagination: Four Essays; Holquist, M. (Ed.); Holquist, M. and Emerson, C. (trans.); University of Texas: Austin; 1981.

Barker; Christopher; The Bible; translated according to the Ebrew and Greek; Printer to the Queenes most excellent maiestie: London; 1588/1594 (with a concordance of eighty-two leaves at the end).

Barker, Christopher; Concordance to the Bible; Published in France (?); 1300.

Barker, Christopher; A Midsummer's Night Dream: A concordance of the text of the first quarto of 1600; Clarendon Press: Oxford; 1600.

Barthes, Roland; Writing Degree Zero; Lavers, Annette and Smith, Colin (trans.); Jonathan Cape Ltd.: London; 1967.

Barthes, Roland.; Image-Music-Text; Miller, R. (trans.); Hill and Wang: New York; 1977.

Barthes, Roland; *From Work to Text,* in Harari, J. (Ed.); Textual Strategies; Cornell University Press: Ithaca; 1979: 79.

Barthes, Roland; *Theory Of The Text* in, Young, Robert; Untying The Text: A Post-Structuralist Reader; Routledge & Kegan Paul: London; 1981: 39. Barthes, Roland; *The Discourse of History* in, The Rustle of Language, trans. Howard, Richard; Hill and Wang: New York; 1986.

Bauer, E.; *Electricity and Magnetism* in, Taton, René (Ed.); A General History Of The Sciences: Science in the Nineteenth Century; Pomerans, A. J. (trans.); Thames and Hudson: London; 1965: 178-234.

Bazerman, Charles; *Scientific Writing as a Social Act: E Pluribus Unum?*; 4 S Review; 3:2; 1985: 13-20.

Bazerman, Charles; Shaping Written Knowledge: the Genre and Activity of the Experimental Article; University of Wisconsin Press: Madison, Wisconsin and London; 1988.

Bence, Jones Henry; Life and Letters of Faraday; Ist edition; London; 1870.

Bennett, Paul E.; *The Statistical Measurement of A Stylistic Trait in Julius Caesar and As You Like It* in, Dolezel, Lubomír and Bailey, Richard W. (Eds.); Statistics and Style; American Elsevier Publishing Company, Inc.: New York; 1969: 30 and 32.

Biber, Douglas; *On the complexity of Discourse Complexity: A Multidimensional Analysis*; Discourse processes; 15:2; April 1992: 133-164.

Bizzel, Patricia; *On the Possibility of a Unified Theory of Composition and Literature*; Rhetorical Review; 4:2; 1986: 176.

Bloom, Ronald L. *et al* (Eds.); Discourse Analysis And Applications: Studies in Adult Clinical Education; Lawrence Erlbaum Associates, Publishers: Hillsdale, New Jersey and Hove, UK; 1994.

Bolter, David J.; Writing Space: The Computer in the History of Literacy; Lawrence Erlbaum: Hillsdale, N. J.; 1990.

Bolton, H.C. and Roberts, Alan; *On The Comparison Of Literary And Scientific Styles: The Letters And Articles Of Max Born, F. R. S.*; Notes and Records of the Royal Society of London; 49:2; 1995: 295-302.

Borden, George, A. and James, Watts J.; *A Computerized Language Analysis System*; Computers and the Humanities; 5:3; January 1971; 129-141.

Boscovich, Roger; A Theory of Natural Philosophy; First Edition, 1758; English Translation, 1763. Reprint of the English Edition; MIT Press: Cambridge, Massachusetts; 1966.

Brainerd, Barron; *Prounouns and Genre in Shakespeare's Drama*; Computers and the Humanities; 13:3; 1979; 3-16.

Brannigan, Augustine T.; The Social Basis of Scientific Knowledge; Cambridge University Press: Cambridge; 1981.

Bright, William (Editor in Chief); International Encyclopedia Of Linguistics; Volume 2; Oxford University Press: New York and Oxford; 1992.

Bright, Wiliam (Editor in Chief); International Encyclopedia of Linguistics, Volume 3; Oxford University Press: New York and Oxford; 1992.

Brinegar, Claude S.; *Mark Twain and the Quintus Curtius Snodgrass Letters: A Statistical Test of Authorship*; Journal of the American Statistical Association; 58; 1963: 85-96.

Bruce, Donald; *Towards The Implementation of Text and Discourse Analysis in Computer-Assisted Textual Analysis*; Computers and the Humanities; 27: 1993: 357-364.

Bruce, Gregory; Inventing reality: physics as language; Wiley: New York: 1988.

Burke, Kenneth; A Rhetoric of Motives; University of California Press: Berkeley, CA.; 1950.

Busa, Roberta S. J.; Index Thomisticus; vol. 1; Fromann-Holzboog: Stuttgart-Bad Connstatt; 1974.

Byatt, A. S; Babel Tower; Vintage Books: London; 1996.

Callon, Michel and Courtial, J. P.; Co-Word Analysis: A Tool for the Evaluation of Public Research Policy; Ecole Nationale Supérieure des Mines: Paris; 1989.

Callon, Michel and Law, John; *On Interests and their Transformation; Enrolement and Counter-Enrolement*; Social Studies of Science; 12; 1982: 615-625.

Callon, Michel *et al* (Eds.); Mapping The Dynamics of Science And Technology; The Macmillan Press Ltd: London; 1986.

Callon, Michel *et al*; *Co-Word Analysis As A Tool For Describing The Network Of Interactions Between Basic And Technological Research: The Case Of Polymer Chemistry*; Scientometrics; 22:1; 1991: 155-205.

Callon, Michel *et al*; *From Translation to Network: An Introduction to Co-Word Analysis*; Social Science Information; 1983: 191-235.

Cameron, Deborah (Ed.); The feminist critique of language. A reader; Routledge, London and New York; 1990.

Cantor, Geoffery *et al*; Faraday: Macmillan: London; 1991.

Cantor, Geoffery; Michael Faraday: Sandemanian and Scientist; Macmillan: London; 1991.

Cantor, Geoffery; *Reading The book Of Nature: The Relation between Faraday's Religion and his Science* in, Gooding, David and James, Frank A. J. L. James (Eds.); Faraday Rediscovered; Stockton Press: New York; 1985: 69-81 (70).

Cantor, Geoffery; *The rhetoric of experiment* in, Gooding, David, Pinch, Trevor, and Schaffer, Simon (Eds.); The Uses Of Experiment; Cambridge University Press: Cambridge, New York, New Rochelle, Melbourne, and Sydney; 1989: 159-180.

Carley, Kathleen; *Extracting Culture From Textual Analysis*; Poetics; 22:4; 1994: 291-312.

Carnap, Rudolf; Foundations of Logic and Mathematics; University of Chicago Press: Chicago; 1937.

Cassirer, Ernest; Language and Myth; Langer, S.K (trans.); Dover Publications: New York; 1946.

Cherry, L. L.; *PARTS - A system for assigning word classes to English text*; Computing Science Technical Report No. 81; Bell Laboratories; Murray Hill: New Jersey; 1980.

Cherry, L. L.; *Writing Tools*; IEEE Transactions on Communications; 30:1; 1982; 100-105.

Coleridge; 1830: 167 in, Schaffer, Simon; 1991 in, Fischer, Menachem and Schaffer, Simon; 1991: 208.

Collins, Harry M.; *The TEA set: Tacit knowledge and scientific net works*; Science Studies; 4; 1974: 165-185.

Collins, Harry M.; Changing order: replication and induction in scientific practice; University of Chicago Press: Chicago; 1985.

Coulthard, Malcolm; An Introduction to Discourse Analysis; Longman Group Limited: London and New York; 1985.

Coulthard, Michael (Ed.); Advances In Written Text Analysis; Routledge: New York and London; 1994.

Coupland, Nikolas (Ed.); Styles of Discourse; Croom Helm: London, New York, and Sydney; 1988.

Courtial, J. P.; *Qualitative models, quantitiative tools and network analysis*; Scientometrics; 15; 1989: 527-534.

Courtial, J. P. *et al*; *A Model For Social Interaction Between Cognition And Action Through A Key-Word Simulation Of Knowledge Growth*; Scientometrics; 31:2; 1994: 173-192.

Crane, Gregory; *Composing Culture: The Authority of an Electronic Text*; Current Anthropology; 32: 3; June 1991: 294.

Day, P; Proceedings of the Royal Institution of Great Britain; Oxford University Press: Oxford, New York, London; 68; 1997.

de Morgan, Augustus; A Budget of Paradoxes; Longmans, Green, and Co.: London; 1859.

de Morgan, Sophie; Memoir of Augustus de Morgan by his Wife Sophie Elizabeth de Morgan with Selections from his Letters; Longmans, Green, and Co.: London; 1882.

Deleuze, Gilles and Guattari, Felix; www.scsn.net/~efolley/; accessed September 19, 1997.

DeRocher, James E. *et al*; The Counting of Words: A Review of the History, Techniques, and Theory of Word Counts with Annotated Bibliography, Special Report; Syracuse University Research Corporation: New York; Ist July 1972 to 15th May 1973.

Dewey, John; Experience and nature; Norton: New York; 1925.

Dijk, Van Teun A. (Ed.); Discourse and Literature; John Benjamins Publishing Company: Amsterdam and Philadelphia; 1985.

Dijk, Van Teun A.; *Introduction: The Common Roots Of The Studies Of Literature And Discourse* in, Dijk, Van Teun A. (Ed.); Discourse and Literature; John Benjamins Publishing Company: Amsterdam and Philadelphia; 1985: 2.

Dolezel, Lubomír and Bailey, Richard, W. (Eds.); Statistics and Style; American Elsevier Publishing Company, Inc.: New York; 1969.

Dombrowski, Paul M.; *Some perspectives on rhetoric, science, and history* in, Haywood's Humanistic aspects of technical communication (technical communications series); 1994.

Doorman, S. J.; The Image of Science; Gower Publishing Company Limited: Aldershot, England; 1980.

Dreyfus, Hubert L.; Mind over machine: the power of human intution and expertise in the era of the computer; Free Press: New York; 1986.

Ducan, John; *Approaches to the Study of Discourse in the Social Sciences*, in Bloom, Ronald L. *et al* (Eds.); Discourse Analysis And Applications: Studies in Adult Clinical Education; Lawrence Erlbaum Associates, Publishers: Hillsdale, New Jersey and Hove, UK; 1994: 4.

Dyer, Geoff; Private Eye; January 14, 1994: 9.

Estoup, J. B.; Gammes stenographies; Insitut Stenographique de France: Paris; 1916.

Fabb, N *et al* (Eds.); The linguistics of writing, arguments between language and literature; Methuen: New York; 1987.

Faraday, Michael; *Historical Sketch of Electro-Magnetism*, Annals of Philosophy; 18: 195-200, 274-290; 19: 107-121; October 1821 and November 1822. Reprinted in Experimental Researches in Electricity vol. 2; 1839-55: 127-147.

Faraday, Michael; *Electro-magnetic Rotation Apparatus*; Quarterly Journal of Science; 12; 1822; 147-158, reprinted in Experimental Researches in Electricity vol. 2; 1839-55.

Faraday, Michael; *Experimental Researches in Electricity* (3 vols., 1839 - 1855); vol.2; London; 1844: 127-147.

Faraday, Michael; *On some new Electro-Magnetical Motions, and on the Theory of Magnetism*; Quarterly Journal of Science; 12; 1821: 74-96.

Faraday, Michael; *On the induction of electric currents. On the evolution of electricity from magnetism. On a new electrical condition of matter. On Arago's magnetic phenomena*; Philosophical Transactions; 122; 1832: 125-162. Republished in Experimental Researches in Electricity vol.1; 1839.

Faraday; *On some new Electro-magnetical motions, and on the Theory of Magnetism*; ERE 2: 127-147, and *Electro-magnetic Rotation Apparatus*; *ibid*: 147-158.

Fawley, William; Text and Epistemology; Ablex Publishing Company: Norwood, New Jersey; 1987.

Feigl, Herbert and Scriven, Michael (Eds.); Minnesota Studies in The Philosophy of Science, Vol. 1: The Foundations of Science and the Concepts of Psychology and Psychoanalysis; University of Minnesota Press Minneapolis: Minnesota; 1956: 253-329.

Fick, G. and Spraque, H. (Eds.); Decision Support Systems: Issues and Challenges; Pergamon: New York; 1980.

Finnegan, R; Literacy and Orality; Blackwell: Oxford; 1988.

Fischer, Menachem and Schaffer, Simon (Eds.); William Whewell: A Composite Portrait; Clarendon Press: Oxford; 1991.

Fish, S. E. (Ed.); Seventeenth Century Prose: Modern Essays in Criticism; Oxford University Press: New York; 1971.

Fish, Stanley; Is There a Text in This Class?; Harvard University Press: Cambridge, Massachusetts, and London; 1980.

Fleck, Ludwick; Genesis and Development of a Scientific Fact; The University of Chicago Press: Chicago and London; 1979 (1935).

Flores, F. and Ludlow, J. J.; *Doing and Speaking in the Office*, in Fick, G. and Spraque, H. (Eds.); Decision Support Systems: Issues and Challenges; Pergamon: New York; 1980: 95-118.

Fodor, Jerry A. and Katz, Jerrold J.; The Structure of Language: Readings in the Philosophy of Langage; Prentice-Hall, Inc.: Englewoods Cliffs, New Jersey; 1964.

Forgan, Sophie; *Faraday - From Servant to Savant: The Institutional Context*, in Gooding, David and James, Frank A. J. L. James (Eds.); 1985: 51-67.

Foster, Donald; Elegy by WS: A Study in Attribution; University of Delaware Press: Delaware; 1989.

Foster, Donald, SHAXICON '95; http://www.clark.net/pub/tross/ws/shaxicon.html; accessed March 2; 1998.

Foucault, Michael; The Order of Things: An Archeology Of The Human Sciences; Tavistock Publications: London; 1970.

Fowler, Roger; Language In The News: Discourse and Ideology in the Press; Routledge: London and New York; 1991.

Frase, Lawrence T. *et al*; *Computer Aids for text Assessment and Writing Instruction*; NSPI Journal; November 1981: 21-24.

Fuhrman, Ellesworth R. and Oehler, Kay; *Discourse Analysis and Reflexivity*; Social Studies of Science; 16; 1986: 293-307.

Garrett, Leroy; http://www.dm.net/~ginb/scotland.htm; accessed September 10, 1997.

Germain, Ellen; *Introducing Natural Language Processing*; AI Expert; 7:8; 1992: 30-35.

Giere, Ronald N.; Cognitive Models in the Philosophy of Science: Minnesota Studies Of Science; University of Minnesota Press: Minneapolis; 1992.

Gilbert, G Nigel and Mulkay, Michael; Opening Pandora's Box: Sociological Studies of Scientific Discourse; 1984.

Gilbert, G Nigel and Mulkay, Michael; *Experiments Are The Key: Participants' Histories and Historians' Histories of Science*; ISIS; 75;1984: 105-125

Ginzberg, Ruth; *Uncovering Gynocentric Science* in, Tuana, Nancy; Feminism & Science; 1989: 69-84.

Godsen, Hugh; *Discourse Functions of Subject in Scientific Research*; Applied Linguistics; 14:1; 1993: 56-75.

Goetz, W. Philip (Editor-in-Chief); The New Encyclopaedia Brittanica; Volume 18; Macropaedia; 1986.

Geoffrey the Grammarian (Galfridus Grammaticus), *Promptorium parvulorum sive clericorium* ("Storehouse for Children or Clerics"); circa 1440.

Golinski, J. V.; *Language, Discourse And Science* in, Olby *et al*; 1990: 110-123.

Golinski, Jan; Science as public culture: chemistry and enlightenment in Britain, 1760-1820; Cambridge University Press: New York : 1992.

Gooding; David C.; *How do scientists reach agreements about novels observations?*; Studies in History and Philosophy of Science; 17; 1986: 205-30.

Gooding, David C.; *History in the Laboratory: Can We Tell What Really Went On?* in, James, Frank. A. J. L. (Ed); 1989 (a): 63-82.

Gooding, David C.; *How to be a Good Empiricist*; British Journal for the History of Science; 22; 1989 (b): 419-427.

Gooding, David C.; Experiment And The Making Of Meaning: Human Agency in Scientific Observation and Experiment; Kluwer Academic Publishers: Dordrecht, Boston, and London; 1990.

Gooding, David. C.; *The Procedural Turn: or, Why Do Thought Experiments Work?* in, Giere, Ronald N.; 1992: 65-112.

Gooding, David C.; Technical Report - Clarity-Faith Text Analysis Functions; January 1995.

Gooding, David C.; 'Clarity' Text Analysis Programs Manual; 9 February 1996.

Gooding, David C; *Text Analysis Tools for Modelling, Ellipsis* in, Gooding, David C. and Addis, T. R.; 5.1; April 1996 (c).

Gooding, David C. and Addis, Tom R.; *A Text Classification System for Source-constrained Models* (Working Draft); February 1995.

Gooding, David C. and Addis, Tom R.; Final Report to the ESRC: Graphic Programming Methods for Modeling Cognitive and Social Processes of Science; April 1996 (a).

Gooding, David C. and Addis, Tom R.; Draft Manual: 'Using FAITH Text Search & Analysis Programs' in CLARITY; April 1996 (b).

Gooding, David C. and Addis, Tom R.; *Graphic Programming Methods for Modelling Cognitive and Social Processes of Science, Supporting Documents: Papers*; 5.1; April 1996 (c).

Gooding, David C. and James, Frank A. J. L. James (Eds.); Faraday Rediscovered; Stockton Press: New York; 1985.

Gooding, David C.; Pinch, Trevor, and Schaffer, Simon (Eds.); The Uses Of Experiment; Cambridge University Press: Cambridge, New York, New Rochelle, Melbourne, and Sydney; 1989.

Goodman, Nelson; Ways of Worldmaking; Harvester: Brighton, Sussex; 1976.

Gorlicjk, M. *et al*; Texture User's Manual; Department of Computer Science, University of British Colombia, Vancouver; 1975.

Gorman, Michael E.; Discovery As Invention: Michael Faraday; http://jefferson.village.virginia.edu/~meg3c/id/id_disc/id_disc_13.html; accessed September 11, 1997.

Gower, Barry; *Speculation in Physics: The History and Practice of Naturphilosophie*; Studies in the History and Philosophy of Science; 3; 1973: 305-356.

Green, Jonathon; Chasing the Sun: Dictionary Makers and the Dictionaries They Made; Cape: London; 1996.

Greenstein, Daniel I; A Historian's Guide to Computing; Oxford University Press Inc.: New York; 1994.

Gregory, Michael; Notes on Communication Linguistics: A Framework for the Description of English Discourse; Glendon College, York University: Toronto; 1982.

Gruber, Howard E.; Darwin on Man: A psychological study of scientific creatvity, 2nd ed.; University of Chicago Press: Chicago; 1981.

Gruber, Howard E; 1974: 114, quoted in Holmes, Frederick Lawrence; 1981: 66.

Habermas, Jurgen; Moral consciousness and communicative action; Polity Press: Cambridge; 1990.

Habgood; John; *Michael Faraday: Sermon by the Archbishop of York, Dr John Habgood, Westminster Abbey, Friday, 20th September, 1991*; Royal Institution Proceedings; 64; 1992: 1-3.

Hacking, Ian; The Taming of Chance; Cambridge University Press: Cambridge, New York, Port Chester, Melbourne, and Sydney; 1990.

Halliday, M. A. K.; *Language and the order of nature* in, Fabb, N *et al* (Eds.); 1987: 135-154.

Halliday, M. A. K; *Some Grammatical Problems In Scientific English*; ARAL Series S; 6; 1989: 97.

Halliday, Michael A. K.; *The Construction of knowledge and value in the grammar of scientific discourse, with reference to Charles Darwin's The Origin of Species*, in Coulthard, Michael (Ed.); Advances In Written Text Analysis; Routledge: New York and London; 1994: 137.

Halloran, Michael S.; *Technical Writing And The Rhetoric Of Science*; Journal of Technical And Communication; 8:2; 1978: 79.

Harari, J. (Ed.); Textual Strategies; Cornell University Press: Ithaca; 1979.

Harris, Zellig S.; *Discourse Analysis*; Language; 8; 1952: 1-30.

Hatch, J. A and Wisniewski, R.; Life History and Narrative: Falmer: London: 1995.

Helmholtz, Hermann von; On the modern development of Faraday's conception of electricity; Harrison and Sons: London; 1881.

Herbert, Simon; *Literary Criticism: a Cognitive Approach*; part 4; SEHR, 4:1; http://shr.stanford.edu:80/shreview/4-1/text/simon4.html; updated 8 April, 1995; accessed September 18, 1997.

Herdan, Gustav; *A New Derivation and Interpretation of Yule's 'Characteristic' K*; Journal of Applied Mathematics and Physics (ZAMP); 6; 1955: 333.

Herdan, Gustav; *The mathematical relation between Greenberg's index of linguistic diversity and Yule's characteristic*; Biometrika; 45; 1958: 269.

Hesse, Mary B.; Revolutions And Reconstructions In The Philosophy Of Science, The Harvester Press Limited: Britain; 1980.

Hirsch, Eric; in Crane, Gregory; 1991: 305; Finnegan, R; Literacy and Orality; Blackwell: Oxford; 1988.

Hockey, Susan *et al*; General Handbook; CETH Summer Seminar; 1996.

Hodge, Robert and Kress, Gunther; Social Semiotics; Polity Press in association with Blackwell: Oxford; 1988.

Holmes, Frederic Lawrence; *The Fine Structure Of Scientific Creativity*; History of Science; 18; 1981: 60-70.

Holmes, Frederic Lawrence; *Lavoisier and Krebs: The Individual Scientist in the Near and Deeper Past*; ISIS; 75:276; 1984: 131-142.

Holmes, Frederic Lawrence; Lavoisier and the Chemistry of Life: An Exploration of Scientific Creativity; The University of Wisconsin Press: Madison, Wisconsin; 1985.

Holmes, Frederic Lawrence.; *Patterns of Scientific Creativity*; Bull. Hist. Med.; 60: 1986: 19-35.

Holmes, Frederic Lawrence; *Scientific Writing and Scientific Discovery*; ISIS; 78:2; 1987: 220-235.

Hubbard, Ruth; *Science, Facts, and Feminism* in, Tuana, Nancy; Feminism & Science; Indiana University Press: Bloomington and Indianapolis; 1989: 119-131.

Humboldt; Karl Wilhelm von, cited by Cassirer, Ernest; 1946: 9.

Hutchinson, I. M; http://venus.pfc.mit.edu/Faraday.html (*Background*); accessed September 10, 1997.

Hutchinson, I. M; Michael Faraday: Scientist and Nonconformist; http://venus.pfc.mit.edu/Faraday.html; accessed September 10, 1997.

Ide, Nancy and Walker, Donald; *Introduction: Common Methodologies in Humanities Computing and Computational Linguistics*; Computers and the Humanities; 26; 1993: 327, 328, 330.

Ide, Nancy; *The Relevance of Computational Linguistics to Textual Studies*; Computers and Texts; 2; May 1991: 7-9.

Iker, H. P. and Harway, N. I; *A computer approach towards the analysis of content*; Behavioral Science; 10; 1965: 173-183.

Iker, H. P. and Harway, N. I; *A computer systems approach to the recognition and analysis of content*; Computer Studies in the Humanities and Verbal Behavior; 1; 1968: 134-154.

Iker, H. P.; *WORDS: A computer system for the analysis of content*; Behavior Research Methods & Instrumentation; 6:4; 1974: 430-438.

James, Frank. A. J. L. (Ed); The Development Of The Laboratory, Essays on the Place of Experiment in Industrial Civilization; 1989.

James, Frank A. J. L. (Ed.); The Correspondence of Michael Faraday, volumes 1-3; Institute of Electrical Engineers; 1991, 1993, 1995.

James, Frank A. J. L; *Faraday in the pits, Faraday at sea: the role of the Royal Institution in changing the practice of science and technology in nineteenth-century Britain*, in Day, P; Proceedings of the Royal Institution of Great Britain; Oxford University Press: Oxford, New York, London; 68; 1997: 277-301.

Jameson, Frederic; The Political Unconscious; Cornell University Press: Ithaca; 1981.

Jandt, F. E.; *Sources for computer utilization in interpersonal communication instruction and research*; Today's Speech; 20; 1972: 25-31.

Jane Barnard to Bence Jones, 3 September 1867, in Henry, Bence Jones; 1870: 482, in Forgan, Sophie; 1985: 51.

Jauss, Hans Robert.; Toward An Aesthetic of Reception; Bahti, Timothy (trans.); Harvester Press: Brighton; 1982.

Johnson, E.; *Counting Words and Computing Word Frequency: Project Report: WORDS*; TEXT Technology; 5:1; Spring 1995: 8-17.

Jones, Allan and Churchhouse, R.F.; The Computer in Literary and Linguistic Studies; The University Of Wales Press: Cardiff; 1976.

Jones, Bence Henry; The Life and Letters of Faraday; 2 volumes; London; 1870.

Jones, R. F.; *Science and English Prose Style, 1660-75* in Fish, S. E. (Ed.); Seventeenth Century Prose: Modern Essays in Criticism; Oxford University Press: New York; 1971: 63.

Joose, Martin; *The Five Clocks*; International Journal of Linguistics; 28: 2; 1962.

Jordanova, Ludmilla; Sexual visions: images of gender in science and medicine between the eighteenth and twentieth centuries; University of Wisconsin Press: Madison, Wisconsin; 1993.

Kaeding, Wilhelm Friedrich; Haufigkeitsworterbuch der deutschen Sprache, Festgestellet durch einen Arbeitsausschuss der deutschen; Selbsverlag des Herausgebers, E. S. Mittler: Steglitz bei Berlin; 1897.

Kenny, Anthony; The Computation of Style: An Introduction To Statistics For Students Of Literature And Humanities; Peragamon Press: Oxford, New York, and Toronto; 1982.

Kiefer, Kathleen E and Smith, Charles R.; *Textual analysis with computers: Tests of Bell Laboratories' computer software*; Research in the Teaching of English; 17; 1983: 201-214.

Klein, Harald; INTEXT 4.0 - Text Analysis Software; http://ifsws.soziologie.uni-jena.de:80/home/klein/intexte.html; September 12, 1997; accessed September 18, 1997.

Klein, Joe; *A Brush with Anonymity*; Newsweek; July 29, 1996: 76-77.

Knight, D. M.; *The Physical Sciences and the Romantic Movement*; History of Science; 9; 1970: 54-75.

Knorr-Cetina, Karin D. and Mulkay, Michael (Eds); Science Observed; 1982.

Knuth, Donald E.; TEX: A System for Technical Text; Technical Report AIM-217; Stanford University; November 1978.

Kopple, William J. Vande; *Some Characteristics and Functions of Grammatical Subjects in Scientific Discourse*; Written Communication; 11:4; October 1994: 534-564.

Krebs, Hans A. and Shelley, Julian A. (Eds.); The creative process in science and medicine; American Elsevier: New York, 1975.

Kristeva, Julia; Language: The Unknown, An Initiation Into Linguistics Menke, Anne M (trans.); Harvester Wheatsheaf: London, Sydney, and Tokyo; 1989: 225-226.

Kroeber, Karl; *Perils of Quantitification: The Exemplary Case of Jane Austin's Emma* in, Dolezel, Lubomír and Bailey, Richard W. (Eds.); Statistics and Style; American Elsevier Publishing Company, Inc.: New York; 1969: 206-207.

Lakoff, George; Women, Fire, and Dangerous Things: What Categories Reveal about the Mind; University of Chicago Press: Chicago; 1987.

Lambropoulos, Vassilis and Miller, David Neal; *Michel Foucault, "What Is an Author?" Twentieth-Century Literary Theory*; State University Press of New York: Albany; 1987: 124-42; http://mh.cla.umn.edu/ebibjd1.html; accessed September 18, 1997.

Lampson, Butler; Bravo Manual; Xerox Corporation; Palto Alto, California; 1978.

Lance J. W.; The mechanism and management of headache; 2nd edition; London; 1973.

Latour, Bruno and Woolgar, Steve; Laboratory Life: the Social Construction of Facts; 1979.

Latour, Bruno; *Drawing things together* in, Lynch, Michael and Woolgar, Steve (Eds.); Representation In Scientific Practice; The MIT Press: Cambridge, Massachusetts and London, England; 1990: 19-68.

Latour, Bruno; *A Reply to Carlson and Gorman*; Social Studies of Science; 22:1; February 1992: 91-95.

Law, John and Courtial, Jean-Pierre; *A Co-Word Analyis of Artificial Intelligence*; Social Studies of Science; 19; 1989: 301-311.

Lehnert, Wendy G. *et al*; *BORIS - An Experiment in In-Depth Understanding of Narratives*; Artificial Intelligence; 20; 1983: 15-62.

Lentricchia, Frank and McLaughlin, Thomas (Eds.); Critical Terms for Literary Study (Second Edition); The University of Chicago Press: Chicago and London; 1995.

Leydesdorff, Loet; *The Relations Between Qualitative Theory And Scientometric Methods In Science And Technology Studies*; Scientometrics; 15:5-6; 1989: 33-347.

Leydesdorff, Loet.; *Relations Among Science Indicators Or More Generally Among Anything One Might Wish To Count About Texts, 1. The Static Model*; Scientometrics; 18:3-4; 1990: 281-307.

Leydesdorff, Loet.; *Relations Among Science Indicators Or More Generally Among Anything One Might Wish To Count About Texts, 11. The Dynamics Of Science*; Scientometrics; 19:3-4; 1990: 271-296.

Leydesdorff, Loet.; *The Prediction Of Science Indicators Using Information Theory*; Scientometrics; 19:3-4; 1990: 297-324.

Leydesdorff, Loet.; *In Search of Epistemic Networks*; Social Studies of Science; 21:1; August 1991: 75-110.

Leydesdorff, Loet.; *The Prediction of Science Indicators Using Information Theory*; *Knowledge Bayesian inferences and empirical science studies*, Social Science Information; 31:2; 1992: 213-238.

Leydesdorff, Loet.; The Challenge of Scientific Scientometrics: the development, measurement, and self-organization of scientific communications; DSWO Press: Leiden; 1995.

Lord, R. D.; *Studies in the History of Probability and Statistics VIII: De Morgan and the Statistical Study Of Literary Style*; Biometriks; 45; 1958: 282.

Lovekin, David; Technique, discourse, and consciousness: an introduction to the philosophy of Jacques Ellul; Lehigh University Press: Bethlehem (PA); 1991.

Lovie, A. D.; Context and Commitement: A Psychology of Science; Harvester Wheatsheaf: New York, London, Toronto; 1992.

Luhn, H. P.; *The Automatic creation of literature abstracts*; IBM Journal of Research and Development; 2; 1958: 159-165.

Lynch, Michael; Art and artifact in laboratory science: a study of shop work and shop talk in a research laboratory; Routledge & Kegan Paul: London and Boston; 1985.

Lynch, Michael and Woolgar, Steve (Eds.); Representation In Scientific Practice; The MIT Press: Cambridge, Massachusetts and London, England; 1990.

M.; *Curves of Literary Style*; Science; 22: 322; 1889: 269.

M, A. B.; *Curves of Literary Style*; Science; 22:320; March 1889: 226.

MacDonald, Nina H. *et al*; *"The Writer's Workbench"™: Computer aids for textual analysis*; IEEE Transactions on Communication; 30; 1982: 105-110.

MacRoberts, Michael H. and MacRoberts, Barbara R.; *A Revaluation of Lotka's Law of Scientific Productivity*; Social Studies of Science; 12:3; November 1982: 535-558.

Mailloux, Steven; *Interpretation* in, Lentricchia, Frank and McLaughlin, Thomas (Eds.); Critical Terms for Literary Study (Second Edition); The University of Chicago Press: Chicago and London; 1995: 127.

Mair, Miller; Between Psychology and Psychotherapy. A poetics of experience; Routledge: London and New York; 1989.

Martindale, Charles and Stoddart, Simon; in Crane, Charles; 1994: 305.

Martindale, Martin and Stoddart, Simon; in Crane, Charles; *Composing Culture: The Authority of an Electronic Text*; Current Anthropology; 32:1; June 1991: 305-306.

Matsuba, Stephen Naoyuki.; *Finding the Range: Linguistic Analysis and Its Role in Computer-Assisted Literary Study*; Computers and the Humanities; 27:5-6; 1993: 331-340.

McCarty, Willard, Literary and mechanical thinking: a position-paper on the boundaries of computer-assisted literary analysis; University of Toronto; November 1995; http://www.kcl.ac.uk/kis/schools/hum/ruhc/wlm/essays/litmech.html; accessed January 14th, 1997.

Medawar, Peter B. in, Holmes, Frederick, L.; 1987: 220-221.

Mendenhall, Thomas C.; *The Characteristic Curves Of Composition*; Science; 9: 214; Friday, March 11, 1887: 213.

Mendenhall, Thomas C.; *A Mechanical Solution Of A Literary Problem*; The Popular Science Monthly; December 1901: 97-105.

Miall, David S.; *Estimating Changes in Collocations of Key Words Across a Large Text: A Case Study of Coleridge's Notebooks*; Computers and the Humanties; 26; 1992: 1-12.

Mill, John Stuart; A System of Logic; London; 2; 1843.

Mitchell, J. L. (Ed.); Computers in the Humanities; University of South Carolina Press: Columbia, S. C; 1974.

Mitchell, T. F.; *The language of buying and selling in Cyrenaica*; Hesperis; 44; 1957: 31-71.

MLLT Home Page; http://www.rxrc.xerox.com/research/mltt/home.html; 12 September 1997; accessed September 23, 1997.

Morgan, Baynard Quincy; German frequency word book based on Kaeding's Haufigkeitsworterbuch der deutschen Sprache; The Macmillan Company: New York; 1928.

Morton, Andrew Queen; Literary Detection: How to prove authorship and fraud in literature and documents; Bowker Publishing Company: United Kingdom; 1978.

Mulkay, M., Potter, J. and Yearley, S.; *Why an Analysis of Scientific Discourse is needed*, in Knorr-Cetina, Karin D. and Mulkay, Michael (Eds); Science Observed; 1982: 171-204.

Mulkay, Michael and Gilbert, G Nigel; *What is the Ultimate Question? Some Remarks in Defence of the Analysis of Scientific Discourse*; Social Studies of Science; 12; 1982: 309-319.

Mulkay, Michael; *Action and Belief or Scientific Discourse? A Possible Way of Ending Intellectual Vassalage in Social Studies of Science*; Philosophy of the Social Sciences; 11; 1981: 163-172.

Mulkay, Michael; *The Scientist Talks Back: A One-Act Play with a Mora, About Replication in Science and Refelexivity in Sociology*; Social Studies of Science; 14; 1984: 265-283.

Mulkay, Michael; *The Ultimate Compliment: A Sociological Analysis of Ceremonial Discourse*; Sociology; 18; 1984: 531-549.

Mulkay, Michael; The Word and the World: Explorations in the Form of Sociological Analysis; Allen and Unwin: London; 1985.

Myers, Greg; *"In this paper we report...": Speech Acts and Scientific Facts*; Journal of Pragmatics; 17; 1992: 295-313.

Neurath, Otto; Philosophical papers, 1913-1946; D. Riedel Publishing Co.: Dordrecht, Holland; Boston: Hingham, MA; 1983.

Nickles, Thomas; *Discovery* in, Olby, R. C. *et al* (Eds.); Companion To The History Of Modern Science; Routledge: London and New York; 1990; 148-165.

Nickles, Thomas; Scientific Discovery: Case Studies; D. Reidel Publsihing Company: Dordrecht: Holland, Boston: U.S.A, and London, England; 1980.

Olby, R. C. *et al* (Eds.); Companion To The History Of Modern Science; Routledge: London and New York; 1990.

Olsen, Mark; *Signs, Symbols and Discourses: A New Direction for Computer-aided Literature Studies*; Computers and the Humanities; 27:5-6; 1993; http://tuna.uchicago.edu/homes/mark/Signs.html; accessed September 18, 1997.

Ossanna, J. F.; TROFF User's Manual; Tech. Rept. 54; Bell Laboratories; 1977.

Parker, H. A.; *Curves of Literary Style*; Science; 23:321; March 29, 1889: 246.

Paul; Christaine; *Reading/Writing Hyperfictions: The Psychodrama of Iinteractivity*; Leonardo; 28:4; August 1995: 265-272.

Pease, E. Donald; *Author*, in Lentricchia, Frank and McLaughlin (Eds.); Critical Terms for Literary Study (Second Edition); The University of Chicago Press: Chicago and London; 1995.

Peer, Willie van; *Quantitative Studies of Literature. A Critque and an Outlook*; Computers and the Humanities; 23; 1989: 301-307.

Peer, Willie van; The Taming of the text: Explorations In Language, Literature And Culture; Routledge: London and New York; 1991.

Pickering, Andrew; Constructing Quarks. A sociological History of particle Physics; Chicago University Press: Chicago; 1984.

Pickering, Andrew; The Mangle of Practice. Time, Agency, & Science; Chicago University Press: Chicago; 1995.

Pickering, Andy; *Cyborg History And The WWII Regime*; Perspectives on Science; 3; 1995; 1-48; http://gaia.lis.uiuc.edu/leep3/stim/pickcyb.html; accessed September 19, 1997.

Polanyi, Michael; Personal Knowledge; Doubleday; New York; 1964.

Porter, Roy, and Teich, Mikuláš (Eds.); Romanticism in National Context; Cambridge University Press: Cambridge and New York; 1988.

Potter, Jonathan *et al*; Social Texts And Contexts: Literature and Social Psychology; Routledge & Kegan Paul: Boston, London, Melbourne, and Henley; 1984.

Potter, Jonathan and Margaret, Wetherall; Discourse And Social Psychology: Beyond Attitudes And Behaviour; Sage Publications: London, Newbury Park, Beverly Hils, and New Delhi; 1987.

Primary Culprit; New York ; 26 Feb. 1996: 50-7.

Privett, Ronna; *Stripping off the Layers: Evaluating Allegory in Thomas Pynchon's The Crying of Lot 49*; http://pegasus.acs.ttu.edu/~ykfrp/pynchon.html; December 1st, 1996; accessed September 19, 1997.

Putnam, Hilary; Realism with a Human Face (Edited and Introduced by Conant, James); Harvard University Press; Cambridge, Massachusetts, and London, England; 1990.

Quine, Willard van Orman; Word and Object; MIT Press: Cambridge, Massachusetts; 1960.

Quine, Willard van Orman and Julian, J. S.; The Web Of Belief; Random House: New York; 1970.

Quine, Willard van Orman; Theories and Things; The Belknap Press of Harvard University Press: Cambridge, Massachusetts, and London, England; 1981.

Raben, Joseph; *Humanities Computing 25 Years Later*, 25; Computers and the Humanities Computing; 1991: 343.

Rabon, Don; Investigative Discourse Analysis; Carolina Academic Press: Durham, North Carolina; 1994.

Rank Xerox's 'Language Tools' Web site: http://www.rxrc.xerox.com/research/mltt/toolhome.html; accessed January 14th, 1997.

Raskin, N. H. and Appenzeller, O.; Headache; Philadelphia; 1980: 172-84.

Reed, Alan and Schonfelder, J. Laurie; *CLOC: a general-purpose concordance and collocations generator* in Ager, D. E. *et al* (Eds.); Advances in Computer-Aided Literary And Linguistic Research; AMLC, for the Department of Modern Languages, University of Aston in Birmingham; 1979: 59.

Reid, Brian K.; Scribe Users' Manual; CMU Computer Science Department; 1978.

Renear, Allen; *Representing Text On The Computer: Lessons For And From Philosophy*; Bulletin of the John Rylands University Library; 74:3; Fall 1992: 221-248.

Rice, Mayer Joseph; The rational spelling book; American Book Co.: New York; 1898.

Roberts, Carl W. (Ed.); Text Analysis For The Social Sciences; Lawrence Earlbaum Associates: Mahwah, New Jersey; 1997.

Robins, R. H.; A Short History of Linguistics (Third Edition); Longman: London and New York; 1990.

Rorty, Richard; Contingency, irony, and solidarity; Cambridge University Press: Cambridge, New York, New Rochelle, Melbourne, and Sydney; 1989.

Rorty, Richard; *Is Natural Science a Natural Kind?* in idem, Objectivity, Relativism, and Truth: Philosophical Papers vol. 1; Cambridge University Press: Cambridge; 1991: 46-62.

Rose, Hilary; Love, power and knowledge : towards a feminist transformation of the sciences; Polity Press:Cambridge; 1994.

Rose, Hilary; *My Enemy's Enemy Is - Only Perhaps My Friend*; Social Text; 14:1 & 2; Spring/Summer; 1996.

Ross, A. S. C.; *Philological Probability Problems*; Journal of the Royal Statistical Society, Series B; 12; 1950: 19-59.

Ross, D. Jr and Rasche, R. H.; *EYEBALL: a computer program for desription of style*; Computers and the Humanities; 6; 1972: 213-221.

Ross, D.; *Beyond the concordance: algorithms for description of English clauses and and phrases*, Aitken, A. J. *et al*; The Computer and Literary Studies; Edinburgh University Press: Edinburgh; 1973: 85-99.

Ross, D.; *An EYEBALL view of Blake's Song of Innocence and Experience,* in Mitchell, J. L. (Ed.); Computers in the Humanities; University of South Carolina Press: Columbia, S. C; 1974: 94-108.

Ross, Sydney; *The Search For Electromagnetic Induction 1820-1831*; Notes & Records of the Royal Society; 20:2; 1965: 184-219.

Ross, Sydney; Nineteenth-Century Attitudes: Men of Science; Kluwer Academic Publishers: Dordrecht, Boston, London; 1991.

Rudwick, Martin J. S.; *Charles Darwin in London: The Integration of Public and Private Science*; ISIS; 73: 267; 1982: 186-206.

Ruhleder, Karen; *Reconstructing artifacts, reconstructing work: From textual edition to on-line databank*; Science, Technology & Human Values; 20; Winter 1995: 7.

Sardinha, A. P. Berber;
http://info.ox.ac.uk/ctitext/publish/comtxt/ct12/sardinha.html; July 1996;
accessed March 15, 1998.

Saussure, Ferdinand de; Course in General Linguistics (Trans. Baskin, Wade); McGraw-Hill: New York; 1966.

Schaffer, Simon; *The History and Geography of the Intellectual World: Whewell's Politics of Language* in, Fischer, Menachem and Schaffer, Simon; 1991: 201-231.

Schils, Erik and de Haan, Pieter; *Characteristics of Sentence Length in Running Text*; Literary and Linguistic Computing; 8:1; 1993: 20-26.

Scholes, Robert; Textual Power: Literary theory and the Teaching of English; Yale University Press: New Haven; 1986.

Searle, J. R.; Speech acts: an essay in the philosophy of language; Cambridge University Press: London; 1969.

Seigel, Daniel M.; Innovation in Maxwell's electromagnetic theory; Cambridge University Press: Cambridge, New York, Port Chester; 1991.

Selden, Raman and Widdowson, Peter; A Reader's Guide to Contemporary Literary Theory, Third Edition; Harvester WheatSheaf: New York, London, and Toronto; 1993.

Sellars, Wilfred; *Empiricism and The Philosophy of Mind* in, Feigl, Herbert and Scriven, Michael (Eds.); Minnesota Studies in the Philosophy of Science, Volume I: The Foundations of Science and the Concepts of Psychology and Psychoanalysis; University of Minnesota Press; 1956: 253-329.

Shapin, Steven; *Pump and Circumstance: Robert Boyle's Literary* Technology; Social Studies of Science; 14:4; November 1984: 481-520.

Shapin; Steven; *Talking History: Reflections on Discourse Analysis*; ISIS; 75; 1984: 125-130.

Shapin, Steve and Schaffer, Simon; Leviathan and the Air-Pump: Hobbes, Boyle and the experimental life; Princeton University Press: New York; 1985.

Sherman, L. A.; *On Certain Facts and Principles in the Development of Form in Literature*; University of Nebraska Studies; 1: 4; 1892: 337-366.

Sherman, L. A.; *Some Observations upon the Sentence-Length in English Prose*; University of Nebraska Studies; 1: 2; 1888: 119-130.

Sibbald, Kern E.; DPS User's Guide, Technical Report CN-16.0; University of Maryland; 1976.

Simon, Herbert; http://shr.stanford.edu:80/shreview/4-1/text/simon4.html; April 1995; accessed September 17, 1997.

Simons, H. W.; The Rhetorical Turn; University of Chicago Press: Chicago; 1990.

Sims, Brenda R. and DiMaaggio, Donna; *Using Computerized Textual Analysis with Technical Writing Students: Is It Effective?*; The Technical Writing Teacher; XVII:1; Winter 1990: 61-68.

Sinclair, John McH.; *Trust the Text*, in Coulthard, Malcolm (Ed.); Advances In Written Text Analysis; Routledge: London and New York; 1994: 19.

Singer, Charles; A Short History Of Scientific Ideas To 1900; Oxford University Press: Oxford; 1959.

Snelgrove, Teresa; *A Method for the Analysis of the Structure of Narrative Texts*; Literary and Linguistic Computing; 5:3; 1990: 221-225.

Steiner, Peter (Ed.); The Prague School: Selected Writings; University of Texas Press: Austin; 1982.

Steiner, Peter; Russian Formalism: A Metapoetics; Cornell University Press: Ithaca; 1984.

Sterkei, Karen S; *Textual analysis with computers to improve the writing skills of business communication students*; The Journal of Business; 23; 1986: 43-61.

MacDonald, Nina H. *et al*; *"The Writer's Workbench"™: Computer aids for textual analysis*; IEEE Transactions on Communication; 30; 1982: 105-110.

Stern, Robert; Introduction to F. W. J.Schelling, Ideas for a Philosophy of Nature; transl. Harris; E.E. and Heath; P. transl.; Cambridge University Press: Cambridge; 1988.

Stone, Phillip. J.; *Thematic Text Analysis: New Agendas for Analyzing Text Content* in Roberts, Carl W.; 1997: 35-54.

Stoppard, Tom; Arcadia; Faber and Faber: Boston and London; 1993.

Street, B; Literacy in theory and practice; Cambridge University Press: Cambridge; 1984.

Stubbs, Michael; Discourse Analysis: The Sociolinguistic Analysis of Natural Language; Basil Blackwell Publishing Limited: Oxford; 1983.

Stubbs, Michael; Text And Corpus Analysis: Computer-assisted Studies of Language and Culture; Blackwell Publishers Ltd.: Oxford; 1996.

Sturrock, John (Ed.); Structuralism and Since: From Lévi Strauss to Derrida; Oxford University Press: Oxford and New York; 1979.

Surette; Leon; Deconstruction: Whence It Came And Wither It Tendeth; Papers in Political Economy; 49; May 1994: 1-19.

Sutton, Brett; 1994 quoted in, Hockey, Susan *et al*; 1996: 58.

Swales, John M.; Genre Analysis; Cambridge University Press: Cambridge; 1990.

Swales, John M.; Academic writing for graduate students: commentary: a course for nonnative speakers of English; Ann Arbor: University of Michigan Press, 1994.

Tallentire, D. R.; *Confirming Intututions About Style Using Concordances* in, Jones, Allan and Churchhouse, R.F.; The Computer in Literary and Linguistic Studies; The University Of Wales Press: Cardiff; 1976.

Taton, René (Ed.); A General History Of The Sciences: Science in the Nineteenth Century; Pomerans, A. J. (transl.); Thames and Hudson: London; 1965.

Teil, Geneviéve and Latour, Bruno; *the hume machine: can association networks do more than formal rules?*; SEHR; 4:2; (updated) 4 June 1995; shr.stanford.edu/shreview/4-2/text/teil-latour.html; January 21, 1998.

Tesler, Larry; PUB: The Document Compiler; Technical Report ON-70; Stanford University Artificial Inteligence Project; September 1972.

Thompson, James D.; Organization in Action; McGraw-Hill: New York; 1967.

Thompson, John B.; Studies in the Theory of Ideaology; Polity Press: Cambridge; 1984.

Thorne, James; *What is a poem?* in, Peer, Willie van; 1989: 285-291.

Tibbetts, Paul and Johnson, Patricia; *The Discourse and Praxis Models in Recent Reconstructions of ScientificKnowledge*; Social Studies of Science; 15; 1985: 739-749.

Tijssen, R. J. W. and van Raan, A. F. J.; *Mapping co-word structures: a comparison of multidimensional scaling and LEXIMAPPE*; Scientometrics; 15; 1989: 285-297.

Tobey, Ronald C.; *Beginner's Guide to the History of Science. Field Theory;* http://www.kaiwan.com/~lucknow/horus/guide/ph106.html; accessed September 20, 1997.

Tweney, Ryan D. and Gooding, David C; Michael Faraday's 'Chemical notes, hints, suggestions, and objects of pursuit' of 1822; P. Peregrinus in association with the Science Museum: London; 1991.

Tyndall, John, *On the Scientific Use of the Imagination.* A discourse delivered before the British Association at Liverpool on 16th September, 1870; London; 1870.

Volkov, Vadim; *Towards A Genealogy Of Revolutionary Practices*; http://lucy.ukc.ac.uk/csacpub/russian/volkov.html; August 27, 1997.

Vygotsky, Lev, S.; Thought and Language; MIT Press: Cambridge, Mass.; 1962.

Wajcman; Judy; Feminism Confronts Technology; Polity Press: Cambridge; 1991.

Wales, Kathleen; *Back to the future: Bakhtin, stylistics and discourse* in van Peer, Willie; The Taming of the text: Explorations In Language, Literature And Culture; Routledge: London and New York; 1991: 176-192.

Walters, Frank D.; *Scientific Method And Prose Style In The Early Royal Society*; Journal of Writing And Communication; 23:3; 1993: 239-258.

Wavell, Bruce; Language and Reason; Mouton de Guyter: Berlin; 1986.

Webster, Andrew; Science, Technology And Society; MacMillan: London; 1991.

White, Harry; *The problem of pulp fact*, in SPA: Science and Public Affairs; Winter 1995: 7-8.

White, Hayden; *Foucault Decoded: Notes from Underground*, in Topics of Discourse: Essays in Cultural Criticism; John Hopkins University Press: London; 1978.

Widdowson, Peter and Selden, Raman; A Reader's Guide to Contemporary Literary Theory; 3rd edition; Harvester Wheatsheaf; 1993.

Williams, C. B.; *A Note On The Statistical Analysis Of Sentence-Length As A Criterion Of Literary Style*; Biometrika; 31; 1940: 360-361.

Williams, L. P. *et al* (Eds.); The selected correspondence of Michael Faraday (2 volumes); Camridge University Press: Cambridge; 1971.

Williams, L. P. *et al* (Eds.); 2; 1971: 673, in Forgan, Sophie; 1985: 60.

Williams, Pearce L; Michael Faraday; Basic Books: New York; 1965.

Wills, J. C.; Age and area, The University Press: Cambridge; 1922.

Wimsatt, W. K. and Beardsley, Monroe C.; The Verbal Icon: Studies in the Meaning of Poetry; University of Kentucky: Lexington; 1954.

Wise, Norton M.; *Electromagnetic Theory In The Nineteenth Century* in, Olby, R. C. *et al* (Eds.); Companion To The History Of Modern Science; 1990: 347.

Whittaker, John; *Creativity and Conformity in Science; Titles, Keywords and Co-Word Analysis*; Social Studies of Science; 19; 1989: 473-496.

Wittgenstein, Ludwig; Philosophical Investigations; Anscombe, G.E.M. transl.; Basil Blackwell: Oxford; 1967 (third edition).

Woolf, Daniel; *Disciplinary History and Historical Culture. A Critique of the History of History: the Case of Early Modern England*; Cromohs: Cyber Review of Modern Historiography; 2: 1-25; http://home.cc.umanitoba.ca/~sprague/woolf.htm#*; note 38; August 29, 1997; accessed October 4, 1997.

Woolgar, Steve; *On the Alleged Distinction Between Discourse and Praxis*; Social Studies of Science; 16; 1986: 309-317.

Woolgar, Steve; Science: The Very Idea; Routledge: London and New York; 1993 (1988).

Wright, Will: Wild Langauge: Science, Language, and Social Life in a Fragile Environment; University of Minnesota Press: Minneapolis; 1992.

Yearley, Steven; *Textual Persuasion*; Philosophy of the Social Sciences; VII; 1981: 409-435.

Youmaus, Gilbert; *A new tool for Discourse Analysis - The vocab-management profile*; Language; 67:4; December 1991: 763-789.

Young, Robert; Untying The Text: A Post-Structuralist Reader; Routledge & Kegan Paul: London; 1981.

Yule, Udny G.; The Statistical Study Of Literary Vocabulary; The University Press: Cambridge; 1944: 44-53.

Zahar, Alexander; *The Constructivist Conception Of Language And Its Functions In Science*; PhD Thesis; University College London; 1995.

Zipf, George Kingsley; The psycho-biology of language; Haughton Miffin: Boston; 1935.

Zipf, George Kingsley; Human behaviour and the principles of least effort; Addison-Wesley: Reading, Massachusetts; 1949.