

# The influence of the Royal Society on 17th-century scientific writing

*Lilo Moessner*  
*RWTH Aachen*

## **Abstract**

*An outline of the language policy of the Royal Society leads to the hypotheses that it contributed to the development of a new linguistic profile and a homogeneous linguistic structure of scientific texts in the second half of the 17th century. These hypotheses are empirically tested on a corpus of ca. 76,000 words, which consists in equal parts of texts from the first and the second half of the 17th century. The research method follows the principles of multi-dimensional analysis. The hypothesis of the new linguistic profile is supported on all five dimensions, whereas the homogeneity hypothesis is only partly supported. The results of the study allow a tentative sketch of the long-term development of scientific writing.*

## **1 The Royal Society and its language policy**

The intellectual movement which led to the foundation of the Royal Society (RS) started in the first half of the 17th century. On the authority of John Wallis, Thomas Birch (1756/1968, vol. I: 1) traces its beginnings to the year 1645 when “several worthy persons residing in London, who were inquisitive into natural, and the new and experimental philosophy, agreed to meet weekly on a certain day, to discourse upon such subjects”. The meeting on November 28th, 1660, when it was decided to follow the example of other countries and found “a college for the promoting of physico-mathematical experimental learning” (ibid.: 3), is usually taken as the foundation date of the RS.

From the very beginning the RS took a vivid interest in the accessibility and the dissemination of research results: “It was resolved, that every member, who hath published or shall publish any work, give the society one copy.” (ibid.: 25). It was planned to set up a library, in which all publications handed in to the society should be kept.

The RS also considered it its duty to contribute to the shaping of an adequate writing style for scientific publications. Texts submitted for publication had to have the approval of the president, and prospective authors could be asked to change whole passages before their works were accepted for publication.

The first historiographer of the RS, bishop Thomas Sprat (1667/1959: 113), mentions two fields in which the RS contributed most to the advancement of learning, namely the experiments themselves and the language in which they were reported. The language policy of the RS is summarized in these words:

a constant Resolution, to reject all the amplifications, digressions, and swellings of style: to return back to the primitive purity, and shortness, when men deliver'd so many *things*, almost in an equal number of *words*. They have exacted from all their members, a close, naked, natural way of speaking; positive expressions; clear senses; a native easiness: bringing all things as near the Mathematical plainness, as they can: and preferring the language of Artizans, Countrymen, and Merchants, before that, of Wits, or Scholars.

In 1664, the RS set up a language committee, which was to meet once or twice a month and be chaired by Dr. Wilkins.<sup>1</sup> Unfortunately, there is no record about the precise rules this committee set up “for improving the English language”. Their general tenor can be gleaned from Chapter V of the 1663 Statutes of the RS, which is entitled: “Of Experiments, and the Reports thereof” and reads as follows: “In all reports of Experiments to be brought into the Society, the matter of fact shall be barely stated, without any prefaces, apologies, or rhetorical flourishes.”

Robert Boyle, one of the most eminent founding members of the RS, set out the principles of an appropriate writing style in his *Proëmial essay*. He advocates a “plain and unadorn'd way of Writing” (Hunter and Davis 1999–2000, Vol. 2: 33) with “expressions [which] should be rather clear and significant, than curiously adorn'd” (ibid.: 16), and he admits the use of foreign words only “when Custom has not only Denizon'd them, but brought them into request” (ibid.: 17).

Against this background it can be expected that under the influence of the RS a characteristic discourse type developed with a largely homogeneous linguistic form. Since the approach to the study of natural phenomena promoted by the RS was expressly different from that of previous periods,<sup>2</sup> it is also to be assumed that the linguistic structure of the texts produced by its members differed from that of earlier scientific writing.

These two hypotheses about the homogeneity and the novelty of the linguistic structure of scientific texts produced under the auspices of the RS will be empirically tested in this study.

## **2 Previous research on scientific writing**

The language of Robert Boyle, the best-known representative of the RS, has been described in some detail in Gotti (1996). In the context of the genre ‘experimental essay’, Gotti (2001: 227–237, 2003: 227–241, 2005: 227–241) summarizes the characteristic features of Boyle’s language under the headings brevity, lack of assertiveness, perspicuity, simplicity of form, and objectivity. Gotti’s descriptions, which are based on Boyle’s *New pneumatical experiments about respiration*, contain plenty of illustrative examples, but no quantitative data.

Members of the *Scientific Thought Styles* project in Helsinki studied various aspects of scientific writing in the Late Middle English and the Early Modern English (EModE) period. Taavitsainen and Pahta (1995: 525ff.) noted a “period style in the latter half of the seventeenth century” characterized by features expressing personal experience (first and second person pronouns, imperatives, modal expressions and past tense active verbs). Taavitsainen (1993) applied factor analysis to the texts of the Middle English sub-periods ME3 and ME4 of the *Helsinki Corpus of English Texts* (HC). Its input was a set of putatively emotional and interactive linguistic features (first and second person pronouns, interjections, direct *wh*-questions, private verbs). The focus of the study was on religious texts, and they proved a rather heterogeneous text category, whereas the two medical texts did not differ too much on the emotional scale, being placed at its lower end. The frequency figures of personal pronouns presented in Taavitsainen (1994: 332) allow the conclusion that in the EModE period medical as well as non-medical scientific texts developed a more personal interactive writing style, but formed a linguistically less homogeneous text category.

Biber’s model of multi-dimensional analysis (MD analysis), although originally set up for the analysis of Present-Day English (PDE) texts, has been applied in studies of the diachronic development of scientific writing by Atkinson (1996, 1999), Biber and Finegan (1997), and González-Álvarez and Pérez-Guerra (1998).

Biber and Finegan (1997) based their research on the science texts contained in *A Representative Corpus of Historical English Registers* (ARCHER). They come from the *Philosophical Transactions of the Royal Society* (PTRS). Starting with the period 1650–1699, the texts are grouped in 50-year periods, and each period is represented by ten texts of about 2,000 words each. In their 1997 study,

where they investigated dimensions 1, 2, 3, and 5 (i.e. involved vs informational production, narrative vs non-narrative concerns, situation-dependent vs elaborated reference, non-impersonal vs impersonal style), they found that scientific writing became more informational, less narrative, more elaborated, and more impersonal after the 17th century. Whereas the first two developments proceeded unidirectionally, the other two were reversed in the 20th century.

Atkinson's corpus which he subjected to MD analysis also contains texts from the PTRS arranged in 50-year intervals, but his sampling technique was different, and the intervals do not coincide with those of ARCHER (for a comparative description of the sampling methods for ARCHER and Atkinson's corpus cf. Atkinson 1999: 68–71). The chronological range of the corpus covers the years 1675 to 1975. The articles of the relevant years of the PTRS were assigned to the research fields 'physical sciences and mathematics', 'biological/life science', and 'others'.<sup>3</sup> Each category is represented in Atkinson's corpus with the same proportion of texts as in the corresponding PTRS volumes. As a consequence of this stratification, as Atkinson calls this part of his sampling strategy, the research field 'physical sciences and mathematics' represents 50 per cent of his corpus, the field 'biological/life science' constitutes 37 per cent, and the rest comes from 'others'. His whole corpus comprises 243,304 words, so that each of his seven intervals contains nearly 35,000 words.<sup>4</sup> On the whole, the results of Atkinson's MD analysis correspond to those of Biber and Finegan (1997). The development of scientific writing after the 17th century proceeds unidirectionally on dimensions 1 and 2, in a zigzaggy way on dimensions 3, 4 (overt expression of persuasion), and 5. But the details of the more complex developments on dimensions 3 and 5 are not in line with Biber and Finegan's description. In Atkinson's corpus the reversal of the development towards more elaborated reference sets in already after 1825, and a second change of direction begins after 1925. His texts become less abstract between 1675 and 1725, before the development towards greater abstractness sets in, which is reversed again after 1925.

The development of scientific writing between the 15th and the 17th century is the object of an article by González-Álvarez and Pérez-Guerra (1998). They investigated the development of the genres 'science', 'fiction', 'drama', and 'letters'. Three of their four science texts are taken from the HC; the fourth, Lanfrank's *Cirurgie*, was provided by the *Scientific Thought Styles* project. It is now accessible as part of the *Middle English Medical Texts* corpus (Taavitsainen *et al.* 2005). The scientific sub-corpora contain 12,726 (15th century) and 11,149 (16th century) words. The MD analysis of this corpus yielded interesting results for the development of scientific writing on dimensions 1, 3, and 5 from the

beginning of the 15th to the middle of the 16th century.<sup>5</sup> On dimensions 1 and 3, the science texts of this corpus showed a clear development towards more involvement and less elaborated reference. On dimension 5, a comparison of the relevant figures justifies the claim of a moderate development towards a more impersonal style.<sup>6</sup> In an additional step, the authors compared their genre dimension scores to Atkinson's for the year 1675, and this comparison led them to claim that on all three dimensions there was a unidirectional development from the 15th to the 17th century.

Taken together, these studies support the hypothesis that the linguistic structure of scientific writing in the second half of the 17th century was different from that of earlier and later texts of the same genre and that on dimensions 1 and 3 this period even constituted a turning-point in the development of the genre scientific writing.

### **3 Corpus and research method**

The corpus on which this study is based consists of six science texts, which form two sub-corpora of three texts each. The texts of the earlier sub-corpus date from the first half of the 17th century (non-RS texts), those of the later sub-corpus were written in the second half of the 17th century by eminent members of the RS (RS texts).

These are the texts in chronological order:

William Barlow: *Magnetical aduertisements* (1616)

Gabriel Plattes: *A discovery of subterraneall treasure* (1639)

John Pecquet: *New anatomical experiments* (1653)

Robert Hooke: *An attempt for the explication of the phænomena, observable in an experiment published by the Honourable Robert Boyle* (1661)

Henry Power: *Experimental philosophy* (1664)

Robert Boyle: *Continuation of new experiments physico-mechanical, touching the spring and weight of the air, and their effects* (1669)

Barlow's text consists of 12 chapters, the first eight of which are included here. This amounts to 15,615 words.

The text by Plattes is about several methods of discovering and treating different kinds of minerals. It comprises 10,831 words and was analysed as a whole.

Pecquet’s text is the English translation of his Latin original *Experimenta nova anatomica* (published in Paris in 1651). Several treatises are combined in this volume, and the passage included in the corpus is part of *Experiments physico-mathematical of vacuity*. It comprises 8,825 words.

As the title specifies, Hooke’s text is a response to a report published by Boyle. The whole text was analysed; it contains 10,392 words.

Power’s text is divided into three books “Containing new experiments microscopical, mercurial, magnetical”. The text chosen for analysis comprises all mercurial experiments and consists of 15,606 words.

Boyle’s text contains 50 experiments and comprises about 63,000 words. The passage analysed here comes from the beginning of the text and comprises 15,238 words.

Electronic files were produced from all texts. The textual bases were the facsimile editions of the texts by Barlow and Power,<sup>7</sup> microfilm versions of the original texts in all other cases. All texts had to be keyed in manually, because the letters of the early prints are not recognized by the standard software of scanners. Table 1 maps the structure and size of the corpus.

Table 1: Structure and size of the corpus

sub-corpus	non-RS texts			RS texts		
author	Barlow	Plattes	Pecquet	Hooke	Power	Boyle
number of words	15,615	10,831	8,825	10,392	15,606	15,238
	35,271			41,236		
	76,507					

The present corpus differs from the corpora of earlier studies in two ways. With its 76,507 words it is bigger than those, and the number of texts is bigger than in the relevant sub-corpora investigated by the members of the *Scientific Thought Style* project and by González-Álvarez and Pérez-Guerra, but smaller than the ‘science’ register in ARCHER and Atkinson’s modified version of ARCHER. Therefore it cannot be excluded that idiosyncratic traits of individual authors are reflected more than is desirable in the results of the analysis.<sup>8</sup>

The research method adopted here follows the principles of MD analysis. The choice of linguistic features was partly governed by pragmatic considerations (the corpus is not tagged, and at least some of the features should be automatically recoverable), partly by an effort to include those features which were found relevant in earlier MD analyses and which marked different dimensions.

Care was also taken that the features should be among those with high factor loadings in Biber's (1988: 89f.) list.

The following 17 features were considered: present tense verbs, 2nd person pronouns, 1st person pronouns, *be* as main verb, possibility modals, past tense verbs, perfective verbs, 3rd person pronouns, relative clauses with relative markers in subject position, relative clauses with relative markers in object position, pied piping constructions, prediction modals, conditional subordination, necessity modals, passive constructions, past participle constructions with WHIZ deletion, other (= non-conditional) subordination.

The production date of the texts made some adaptations necessary. They concern the features 3rd person pronouns and relative clauses.

Biber's feature third person personal pronouns (1988: 225) comprises the following forms: *she, he, they, her, him, them, his, their, himself, herself, themselves*. The forms *his* and *their* indicate that possessive determiners are to be included in the list of search items. All singular elements of the list relate to substantives with personal referents. In EModE, the PDE possessive determiner *its* was still often replaced by the more archaic form *his*, so that the occurrences of *his* had to be checked whether they referred to a personal or to a non-personal substantive. Occurrences of 'neutral' *his* were excluded. Additionally, *his* could occur as a marker of genitive (type: *Linus his principles*). Although this is usually considered a minor alternative, the corpus contains quite a few occurrences of *his* in this function; they, too, had to be excluded.

In the 17th century the distribution of the relative marker *that* on the one and *which* and *who* on the other hand did not yet follow PDE rules with respect to the animateness of the antecedent and the degree of restrictiveness of the relative clause. It seemed therefore inappropriate to count *wh*-relative clauses and *that*-relative clauses separately, although in Biber's model the former are markers of dimension 3, the latter of dimension 4.<sup>9</sup>

The tokens of all features were counted in each text, and tables of raw frequencies were set up. These figures were normalised on the basis of 1,000 words, and the mean values were calculated. For the following computation of standardized frequencies, the steps described by Biber (1988: 93–97) were followed. Summing up the standardized frequencies of the features yielded dimension scores of each text for each dimension (text dimension scores). Comparisons of the two sub-corpora also required the calculation of genre dimension scores, i.e. calculations of the mean values of the text dimension scores.

## **4 Analysis and discussion**

### **4.1 Methodological issues**

The most comprehensive description of the model of MD analysis is given in Biber (1988). The model is based on data from the LOB Corpus and the *London-Lund Corpus of Spoken English* plus a collection of professional and personal letters. The set of linguistic features which serve as input for the factor analysis comprises 67 elements. In the last chapter, where Biber outlines possible fields of application of the model (1988: 203), he explicitly mentions the study of the “historical evolution of written texts in English”. This possibility was exploited in the publications mentioned before and several others (Biber and Finegan 1989, 1992; Atkinson 1992). There are, however, some problems involved in the application of the model to diachronic studies especially when they deal with texts of earlier periods.

Such an approach tacitly implies that the features included in the 1988 model were part of English in former periods as well, that they co-occurred in the same way, and that their co-occurrence reflected the same communicative purposes, i.e. that automatically produced factors allow the same interpretation as the 1988 factors. This methodological complication was recognized by Atkinson (1999: 64). The only publication I am aware of which steps away from the well-trodden path of MD analysis of the 1988 type is Biber (2001).<sup>10</sup> Here the author uses data from the 18th century; he works with the same linguistic features as before, but he identifies different factors, and he interprets them partly differently.<sup>11</sup> He identifies three dimensions, and only the first of them is characterized by the same communicative purpose as the corresponding dimension in the 20th century model. Dimension 2 in the 18th century model has no counterpart in the 20th century model.<sup>12</sup> It sets off dramatic texts of the 18th century against all other genres of the corpus. The third 18th century dimension corresponds to dimension 2 in the 20th century model; it reflects narrative concerns.

The present study follows the example of Atkinson (1996, 1999), Biber and Finegan (1997), and González-Álvarez and Pérez-Guerra (1998) in taking over the first five dimensions and their linguistic characterization as established in Biber (1988). A separate factor analysis is planned for the future.

A less serious, but yet annoying problem concerns the comparability of results achieved in former studies based on the model of (20th century) MD analysis. Comparisons are often difficult because the lists of features investigated need not be identical; sometimes they are not even specified. Originally Biber (1988) worked with 67 features, and they are neatly listed and illustrated



in his book (Appendix II). In Biber and Finegan (1997) 57 features are listed, but it remains unclear which of them served as input to the factor analysis. Since González-Álvarez and Pérez-Guerra (1998) established dimension scores only for dimensions 1, 3, and 5, it is not surprising that their list of linguistic features is shorter, but it does not contain all features of these dimensions listed in Biber and Finegan (1997), and others are added instead. The feature list in Atkinson (1999: 66) is not a proper subset of that in Biber and Finegan (1997: 258f.) either. For the features used in the present study cf. section 3.

A last point to be considered is the status of ‘science’ or ‘scientific writing’ as a register. Since registers/genres are prototypical text categories, they contain core and peripheral members. Therefore reliable statements about changes of the linguistic structure of this register can only be made after an analysis of texts which occupy similar positions in the register. The texts analysed in this study share the register properties ‘written by an expert about a topic of the natural sciences and published for the benefit of experts in the same field’.

#### 4.2 Dimension 1: Involved vs. informational production

The following features are interpreted as indicators of the degree of involvement: present tense verbs, 2nd person pronouns, 1st person pronouns, *be* as main verb, and possibility modals. In Biber’s 1988 list they have factor loadings between .86 and .50. High frequencies of these features indicate a high degree of involvement.

The values obtained for the present corpus are entered in Tables 2a (mean normalised frequency of each feature and its standard deviation in the two sub-corpora) and 2b (standardized frequencies of each feature, text and genre dimension scores).<sup>13</sup>

Table 2a: Mean normalised frequency and standard deviations (dimension 1)

	mean norm. frequency	stand. dev. non-RS	stand. dev. RS
present tense verb	35.92	11.41	8.11
2nd person pronoun	6.61	1.93	4.54
1st person pronoun	13.68	4.18	4.82
main verb <i>be</i>	14.43	4.73	0.98
possibility modal	7.37	3.88	2.71

The standard deviation provides a measure for the distribution of the frequency values of the individual linguistic features. When a linguistic feature has a low

standard deviation, its frequencies in the individual texts are grouped in a narrow band around the mean. Consequently, a comparison between the standard deviations of the two sub-corpora allows statements about the degree of their linguistic homogeneity.

The standard deviation of three out of the five linguistic features is lower in the RS sub-corpus than in the texts from the first half of the 17th century. The most evenly spread feature of the RS texts is *be* as a main verb. Here the standard deviation is below 1.0. The development of the distribution of the individual features leads to an overall increase of the homogeneity of the linguistic structure of RS texts compared to the earlier texts.

Table 2b: Standardized frequencies, text and genre dimension scores

	non-RS sub-corpus			RS sub-corpus		
	Barlow	Plattes	Pecquet	Hooke	Power	Boyle
present tense verb	0.45	-0.14	1.80	-0.58	-0.25	-2.13
2nd person pronoun	1.61	1.11	-0.31	-0.54	0.74	-1.22
1st person pronoun	-2.45	-0.55	-0.96	1.05	0.20	2.19
main verb <i>be</i>	0.95	0.55	-0.94	0.18	-1.77	-1.12
possibility modal	-0.38	0.76	-1.22	1.46	-0.52	0.27
text dimension score	0.18	1.73	-1.63	1.57	-1.60	-2.01
genre dim. score	0.09			-0.68		

Since high frequencies of the linguistic features on dimension 1 are interpreted as an involved interactive style, the genre dimension scores indicate a development towards a less involved style. This development is mirrored by the text dimension scores of the RS texts; Boyle’s text has the lowest dimension score. The genre dimension score of the earlier texts situates them slightly above the dividing-line on the involved side, whereas the RS texts are located below the dividing-line on the informative side. This result clashes with González-Álvarez and Pérez-Guerra’s claim (1998: 330f.) that science texts became more involved from the middle of the 15th to the second half of the 17th century. This discrepancy can be due to one or several of the following reasons.

González-Álvarez and Pérez-Guerra’s 15th century texts belong to the domain of medicine and those of the 16th century to the fields of anatomy and geometry, whereas the texts analysed here come from the natural sciences (physics/geology).

González-Álvarez and Pérez-Guerra's analysis of dimension 1 is based on 23 positive and five negative features,<sup>14</sup> whereas in the present analysis only five features are used. Starting from the standardized frequencies listed in González-Álvarez and Pérez-Guerra's appendix, I calculated the corresponding normalised frequencies of their science texts and integrated them into my own corpus. The resulting genre dimension scores supported González-Álvarez and Pérez-Guerra's claim of a development towards a more involved style between the 15th and the 16th century, but not between the 16th and the 17th century.

González-Álvarez and Pérez-Guerra's genre dimension score for science texts of the 17th century was taken over from Atkinson (cf. González-Álvarez and Pérez-Guerra 1998: footnote 9). But his score is based on different mean values of the linguistic features (e.g. first person pronouns: Atkinson: 24.7; González-Álvarez and Pérez-Guerra: 55.0), and since they are used in the calculation of standardized frequencies and of text and genre dimension scores, a direct comparison of González-Álvarez and Pérez-Guerra's 16th century and Atkinson's 17th century genre dimension scores may lead to skewed conclusions.

According to Biber and Finegan (1997) and Atkinson (1999), science texts became less involved after the 17th century. The results of the present analysis allow the conclusion that this development was already under way in the first half of the 17th century, and the RS texts followed this development.

### **4.3 Dimension 2: Narrative vs non-narrative concerns**

Biber's dimension 2 contains six positive features. Three of them, which have the highest factor loadings (between .48 and .90) were counted. The relevant values are given in Tables 3a and 3b:

*Table 3a:* Mean normalised frequency and standard deviations (dimension 2)

	mean norm. frequency	stand. dev. non-RS	stand. dev. RS
past tense verb	14.61	2.36	11.16
perfective verb	3.44	0.74	2.67
3rd person pronoun	9.81	2.05	1.86

Contrary to dimension 1, here the distribution of the relevant linguistic features became less concentrated around the mean between the first and the second half of the 17th century. This is primarily due to the wide frequency range of past tense verbs in the RS texts. Here the standard deviation contrasts sharply with

that of the earlier texts. Only third person pronouns became more evenly distributed, and they reached a standard deviation below 2.0.

*Table 3b:* Standardized frequencies, text and genre dimension scores

	non-RS sub-corpus			RS sub-corpus		
	Barlow	Plattes	Pecquet	Hooke	Power	Boyle
past tense verb	-3.28	-1.34	-2.73	-0.24	0.14	1.65
perfective verb	-1.80	-0.29	0.08	-0.60	-0.14	1.30
3rd person pronoun	1.45	0.02	1.95	-0.10	-1.78	-1.87
text dim. score	-3.63	-1.61	-0.70	-0.94	-1.78	1.08
genre dim. score	-1.98			-0.54		

The genre dimension scores of both the non-RS and the RS texts have values below zero, i.e. narrative concerns do not play a role in these texts. Yet the values show that a change took place in the direction towards a higher degree of narrativity. This development during the 17th century contrasts with that mentioned in Biber and Finegan (1997: 271) and Atkinson (1999: 133) after the 17th century; here the degree of narrativity decreases. The results of the present analysis suggest that the RS texts marked a turning-point in the development on this dimension.

#### ***4.4 Dimension 3: Situation-dependent vs elaborated reference***

Dimension 3 in Biber’s 1988 model contains positive and negative features. Negative features are time adverbials, place adverbials, and adverbs derived from adjectives with the suffix *-ly*; they are indicative of situation-dependent reference. Establishing lists of corresponding items in EModE would have been a very time-consuming process. Therefore these features were not considered. The positive features with the highest factor loadings are relative clauses. Three types of relative clauses were counted: relative clauses with a relative marker functioning as subject, relative clauses with a relative marker functioning as object, and pied piping constructions. The corresponding values are listed in Tables 4a and 4b:

Table 4a: Mean normalised frequency and standard deviations (dimension 3)

	mean norm. frequency	stand. dev. non-RS	stand. dev. RS
rel. marker subject	7.13	0.48	0.40
rel. marker object	1.82	0.74	0.58
pied piping	3.00	0.31	0.93

All three linguistic features are fairly evenly spread across the texts of both sub-corpora with standard deviations below 1.0. But the lower values of the standard deviations of two of the three linguistics features in the RS sub-corpus support the hypothesis of a growing linguistic homogeneity of science texts under the influence of the RS.

Table 4b: Standardized frequencies, text and genre dimension scores

	non-RS sub-corpus			RS sub-corpus		
	Barlow	Plattes	Pecquet	Hooke	Power	Boyle
rel. marker subject	-1.73	-2.14	-0.24	2.82	1.20	1.00
rel. marker object	-0.05	-0.35	1.50	-1.14	-0.92	0.68
pied piping	-2.72	-0.79	-1.30	1.51	-0.48	0.57
text dim. score	-4.50	-3.28	-0.04	3.19	-0.20	2.25
genre dim. score	-2.60			1.74		

The genre dimension scores demonstrate a change from situation-dependent to elaborate reference. This seems to contrast with González-Álvarez and Pérez-Guerra's claim (1998: 333) that science texts underwent a unidirectional development from more elaborate to more situation-dependent reference between the 15th and 17th century. The same explanations which I offered for the discrepancy between their results and mine on dimension 1 may be valid here. Yet there is the additional complication that the integration of their frequency values for relative constructions into my corpus does not support their claim of a development towards less elaborate reference between the 15th and the 16th century, and only a comparison of the recalculated genre dimension scores of the 16th and those of the second half of the 17th century seems to suggest a development towards a lower degree of elaborate reference. If, however, the new value for the early 17th century texts is also taken into consideration, the development turns out to be wavelike (15th c. > 16th c.: more elaborate; 16th c. > early 17th c.: less

elaborate; early 17th c. > late 17th c.: more elaborate; late 17th c. > early 18th c.: more elaborate). Further studies are needed to decide whether the change from situation-dependent to elaborate reference was initiated by the writers of the RS, as my results suggest, or whether this was a later development.

**4.5 Dimension 4: Overt expression of persuasion**

Dimension 4 is marked by positive features only. Three of them with factor loadings between .46 and .54 were considered: prediction modals (*will, shall, would*), conditional subordination, and necessity modals (*should, must*). The relevant figures are given in Tables 5a and 5b:

Table 5a: Mean normalised frequency and standard deviations (dimension 4)

	mean norm. frequency	stand. dev. non-RS	stand. dev. RS
prediction modal	9.86	4.02	3.52
cond. subordination	4.48	2.84	0.31
necessity modal	2.50	0.53	0.51

The values on this dimension show an increase in the homogeneity of the linguistic structure. The standard deviation of two of the three features is even below 1.0 in the RS texts.

Table 5b: Standardized frequencies, text and genre dimension scores

	non-RS sub-corpus			RS sub-corpus		
	Barlow	Plattes	Pecquet	Hooke	Power	Boyle
prediction modal	0.84	-0.04	-1.15	0.96	0.41	-0.97
cond. subordination	1.12	0.56	-0.81	-2.63	-1.61	-3.61
necessity modal	-0.02	1.71	-0.02	-1.69	-0.25	0.22
text dim. score	1.94	2.23	-1.98	-3.36	-1.45	-4.36
genre dim. score	0.73			-3.05		

The genre dimension score of the non-RS texts situates them a little above the baseline on the scale of persuasiveness. The degree of open persuasion decreases markedly afterwards, so that the RS texts have a genre dimension score which is well below the baseline. At the moment it is difficult to decide if this development was initiated by the RS writers, because only Atkinson studied

changes on this dimension, and the earliest texts of his corpus date from 1675. He noted a general development towards a non-persuasive style, starting in the first half of the 18th century.

#### **4.6 Dimension 5: Impersonal vs non-impersonal style**

On dimension 5 the following positive features were counted: passive constructions, past participle constructions with WHIZ deletion, and other (= non-conditional) subordination. Since agentless and agented passives mark the same dimension in Biber's 1988 model, they were not treated as separate features. Tables 6a and 6b contain the relevant values:

*Table 6a:* Mean normalised frequency and standard deviations (dimension 5)

	mean norm. frequency	stand. dev. non-RS	stand. dev. RS
passive	13.54	3.72	4.38
past part. WHIZ del.	4.39	1.64	1.17
other subordination	5.96	0.80	3.18

As on dimension 2, the homogeneity of the linguistic structure decreases on this dimension. The values of the feature 'other subordination' contribute most to this development.

*Table 6b:* Standardized frequencies, text and genre dimension scores

	non-RS sub-corpus			RS sub-corpus		
	Barlow	Plattes	Pecquet	Hooke	Power	Boyle
passive	-1.16	0.82	0.04	0.63	-1.07	0.68
past part. WHIZ del.	-1.15	0.41	0.70	-0.54	-0.57	1.17
other subordination	-2.62	-1.32	-0.65	-0.63	0.43	1.36
text dim. score	-4.93	-0.09	0.09	-0.54	-1.21	3.21
genre dim. score	-1.64			0.48		

The general development which can be inferred from the genre dimension scores indicates a change from a less to a more abstract style. A comparison of the text dimension scores of the earliest and the latest text (Barlow vs Boyle) would suggest a really dramatic change, but the values of the intermediate texts diminish the size of the change considerably.

According to Biber and Finegan (1997: 268) and Atkinson (1999: 129–133), science texts of the 20th century are more abstract than those of the 17th century. Between these two poles there is no unidirectional development. Atkinson, whose timescale is more fine-grained, noted a change from more to less abstract between the second half of the 17th and the first half of the 18th century.

According to González-Álvarez and Pérez-Guerra (1998: 335), science texts became less abstract between the 15th and the 16th century.<sup>14</sup> Comparing their 16th century values to Atkinson's 17th century values, they came to the conclusion that the degree of abstractness rose in this period.

When the genre dimension score of the texts of the first half of the 17th century is also taken into consideration, the claim of increasing abstractness can be formulated more precisely. This change took place between the first and the second half of the 17th century, and the RS writers were responsible for it. Since Atkinson's figures (1999: 111) show a reversal of this development between the second half of the 17th and the first half of the 18th century, the obvious conclusion is that the writing style of the RS authors marks a turning-point in the development of abstractness in science texts.

## **5 Summary and evaluation of results**

In this study a corpus of six scientific texts of the 17th century comprising about 76,000 words was analysed with the method of MD analysis. The texts form two sub-corpora, which differ with respect to their publication dates and authors. One sub-corpus contains texts written in the second half of the 17th century and authored by eminent members of the RS, the other sub-corpus texts from the first half of the 17th century, which were produced by less famous natural scientists. The validity of two hypotheses was tested, namely the homogeneity and the novelty of the linguistic structure of the texts produced under the auspices of the RS.

The first hypothesis was supported with respect to dimensions 1, 3 and 4, but rejected with respect to dimensions 2 and 5. This confirmatory interpretation was based on the number of linguistic features whose standard deviation values were smaller in the RS-texts than in the non-RS texts. When the size of the difference between the two standard deviation values is also taken into account, a different interpretation seems equally adequate. Figure 1 visualizes the degrees of linguistic homogeneity as the sums of the standard deviations of all features on the different dimensions of both sub-corpora. Higher columns represent a lower, lower columns a higher degree of homogeneity.



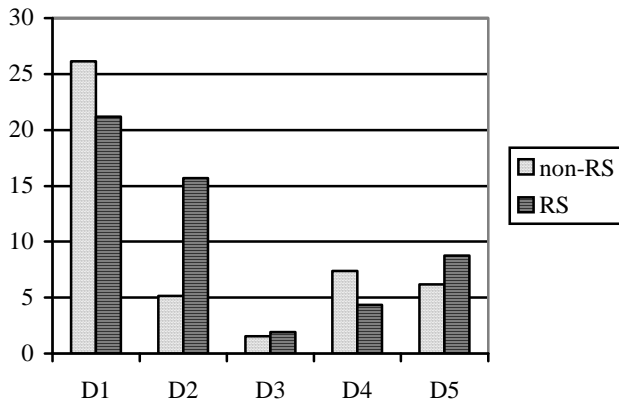


Figure 1: Degrees of linguistic homogeneity in the sub-corpora

The difference concerns dimension 3, where the small differences between the standard deviation values of the features relative marker with subject function and relative marker with object function are outweighed by the big difference of the corresponding values of pied piping constructions. The cumulative effect is a slightly lower degree of homogeneity of the RS texts.

The hypothesis of the novelty of the linguistic structure of RS texts could be supported with respect to all five dimensions. A comparison of the sub-corpora revealed that the texts of the RS sub-corpus was less involved, more narrative, more elaborate in their reference, less openly persuasive, and more abstract than the texts of the earlier sub-corpus.

The differences between the corresponding genre dimension scores are more or less pronounced. The biggest changes could be noticed on dimensions 3 and 4, i.e. the developments towards more elaborate reference and less persuasiveness are most noteworthy; they are followed by the changes on dimensions 2 and 5 towards more narrativity and more abstractness; the change on dimension 1 towards less involvement is least conspicuous. Yet it also contributes to the overall picture which shows that the language policy of the RS bore first fruits already a few years after its foundation.

The attempt at relating the results of the present analysis to those of previous research proved difficult. Direct comparisons with genre dimension scores of earlier studies are excluded, because they were established with different mean

values for the linguistic features and consequently different standardized frequencies, which form the basis for the calculation of text and genre dimension scores.

Another difficulty results from the different sets of linguistic features used in former studies and the present one.

Lastly, the corpora of all studies differ in size and composition. González-Álvarez and Pérez-Guerra's corpus of nearly 24,000 words contains two texts from the 15th and another two from the 16th century. Biber and Finegan's corpus, ARCHER, is composed of ten short extracts of about 2,000 words each for every 50-year period, which amounts to about 70,000 words of science texts for the period 1675–1975. All of them are taken from the PTRS. The corpus used by Atkinson is comparable to ARCHER. The corpus of the present study is bigger than all the others; one century is represented by more words than four centuries in ARCHER type corpora. Yet they have the advantage that they are not restricted to a small set of authors like the present corpus. Whether they are therefore more representative of scientific writing is a problem that needs to be explored in more detail.

For the reasons just specified, only tentative lines of the long-term development of scientific writing can be sketched. On dimensions 1, 3 and 4, the structure of the RS texts fits into a development which started in the first half of the 17th century (in the case of dimension 1 already in the middle of the 16th century) and which continued in the same direction after the 17th century, whereas on dimensions 2 and 5 the RS texts mark a turning-point in the respective developments. Figures 2a and 2b display the developments in the 17th century and indicate the directions of subsequent developments. The values under the label "later" are estimated, not real values; they are derived from scores calculated by Atkinson.

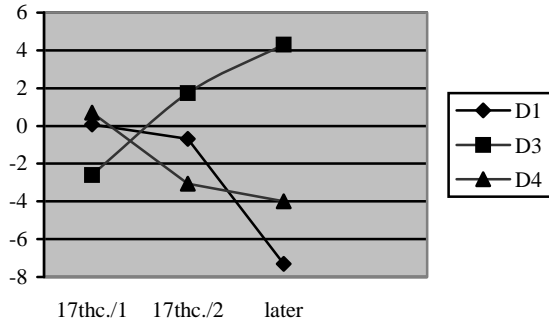


Figure 2a: Genre dimension scores of dimensions with a unidirectional development

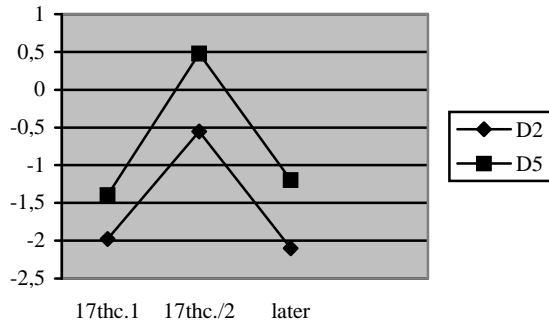


Figure 2b: Genre dimension scores of dimensions on which the direction of development changed after the 17th century

Vague as the estimates about developments of scientific writing after the 17th century may be, even less do we know at present about the developments which led up to the structure of scientific writing at the beginning of the 17th century. The data which will help to solve this problem are now more easily accessible in the *Middle English Medical Texts* corpus.

### **Primary sources**

- Barlow, William. 1616/1968 [facsimile edition]. *Magneticall aduertisements*. Amsterdam and New York: Da Capo Press.
- Hall, Marie Boas (ed.). 1966. *Experimental philosophy, in three books: Contain- ing new experiments microscopical, mercurial, magnetical, by Henry Power*. New York and London: Johnson Reprint Corporation.
- Hooke, Robert. 1661. *An attempt for the explication of the phaenomena, observ- able in an experiment published by the Honourable Robert Boyle*. London: Sam. Thomson.
- Hunter, Michael and Edward B. Davis (eds.). 1999–2000. *The works of Robert Boyle*, 14 vols. London: Pickering and Chatto.
- Pecquet, John. 1653. *New anatomical experiments*. London: Octavian Pulleyn.
- Plattes, Gabriel. 1639. *A discovery of subterraneall treasure*. London: Iasper Emery.

### **Notes**

1. “It being suggested, that there were several persons of the society, whose genius was very proper and inclined to improve the English tongue, and particularly for philosophical purposes; it was voted, that there be a committee for improving the English language; and that they meet at Sir Peter Wyche’s lodgings in Gray’s-Inn, once or twice a month, and give an account of their proceedings to the society, when called upon.” (Birch 1756/1968, vol. I.: 499)  
“It was ordered that Dr. Wilkins meet the first time (at least) with the committee for improving the English tongue; and that particularly he intimate to them the way of proceeding in that committee, according to the sense of the council, viz. chiefly to improve the philosophy of language.” (Birch 1756/1968, vol. II: 7)
2. “The truth is, the Science of Nature has been already too long made only a work of the *Brain* and the *Fancy*: It is now high time that it should return to the plainneß and soundneß of *Observations on material and obvius things*.” (Robert Hooke: *Micrographia*, Preface)
3. The first category comprises the following sub-categories: astronomy, physics, mechanical philosophy, chemistry, geology, geological natural history, mathematics, meteorology, scientific instrumentation. The second category contains texts from the following fields: physiology, anatomy, botany, morphology, embryology, zoology, biological natural history, taxonomy,

- medicine. The category ‘others’ contains texts from the following fields: nonbiological and nongeological natural history, technology/inventions, applied research. (Atkinson 1999: 70)
4. If each interval is represented by exactly the same number of words, the figure is 34,757 words.
  5. The first text (Lanfrank) dates from 1420, the last (Record) from 1551.
  6. The dimensions ‘situation-dependent vs elaborate reference’ and ‘non-impersonal vs impersonal style’ are called dimensions 2 and 3 in their article. I will stick to the usual numbering of dimensions.
  7. The Hunter/Davis edition of Boyle’s works was not accessible to me at the time of compiling the corpus.
  8. The ARCHER science section of 1675 contains texts of 15 identifiable authors, who contribute between 126 and 1,829 words; about 40 per cent of the texts are anonymous. I am grateful to Christian Mair (Freiburg), who made these texts accessible to me.
  9. It must be admitted that the strategy adopted here is not without problems, since it tacitly assumes that all relative clauses have the same function. A more refined analysis would have to distinguish between restrictive and non-restrictive relative clauses irrespective of the relative marker.
  10. Taavitsainen (1993) carried out a factor analysis of the texts of the sub-periods M3 and M4 of the HC, but she used only seven features, which loaded on the same factor.
  11. In the 20th century model, factor 1 is marked by 23 positive and five negative features with a sufficiently high factor loading to be considered for interpretation. In the 18th century model, factor 1 is marked by only 13 positive and eight negative features. The high-loading features contraction and analytic negation for example are absent from the 18th century factor 1. On the other hand, necessity modals and prediction modals, which mark factor 4 of the 20th century model, figure in factor 1 in the 18th century model.
  12. Its characteristic positive features are discourse particles, *wh*-questions, and contractions; it has no negative features.
  13. Tables 3a and 3b – 6a and 6b are organised accordingly.
  14. In their terminology, the features indicating involvement are called negative and those indicating information are called positive. I will stick to the convention followed in the publications by Biber and Atkinson and use the opposite terminology.
  15. Although this statement cannot be derived from the genre dimensions scores in González-Álvarez and Pérez-Guerra (1998: 328f.), it is justified when the same recalculation procedure is carried out as for dimensions 1 and 3.

## References

- Atkinson, Dwight. 1992. The evolution of medical research writing from 1735 to 1985: The case of the *Edinburgh Medical Journal*. *Applied Linguistics* 13: 337–374.
- Atkinson, Dwight. 1996. The *Philosophical Transactions of the Royal Society of London, 1675–1975*: A sociohistorical discourse analysis. *Language in Society* 25: 333–371.
- Atkinson, Dwight. 1999. *Scientific discourse in sociohistorical context. The Philosophical Transactions of the Royal Society of London, 1675–1975*. London and Mahwah, NJ: Laurence Erlbaum.
- Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber, Douglas. 2001. Dimensions of variation among 18th-century speech-based and written registers. In H.-J. Diller and M. Görlach (eds.), *Towards a history of English as a history of genres*, 89–109. Heidelberg: Winter.
- Biber, Douglas and Edward Finegan. 1989. Drift and the evolution of English style: A history of three genres. *Language* 65: 487–517.
- Biber, Douglas and Edward Finegan. 1992. The linguistic evolution of five written and speech-based English genres from the 17th to the 20th centuries. In M. Rissanen, O. Ihalainen, T. Nevalainen, and I. Taavitsainen (eds.), *History of Englishes: New methods and interpretations in historical linguistics*, 688–704. Berlin: Mouton de Gruyter.
- Biber, Douglas and Edward Finegan. 1997. Diachronic relations among speech-based and written registers in English. In T. Nevalainen and L. Kahlas-Tarkka (eds.), *To explain the present: Studies in the changing English language in honour of Matti Rissanen*, 253–276. Helsinki: Société Néophilologique.
- Birch, Thomas. 1576 [1968]. *The history of the Royal Society of London for improving of natural knowledge*, 4 vols. London: Millar [facsimile reprint Hildesheim: Olms].
- González-Álvarez, Dolores and Javier Pérez-Guerra. 1998. Texting the written evidence: On register analysis in late Middle English and early Modern English. *Text* 18 (3): 321–348.
- Gotti, Maurizio. 1996. *Robert Boyle and the language of science*. Milano: Guerini.
- Gotti, Maurizio. 2001. The experimental essay in Early Modern English. *European Journal of English Studies* 5 (2): 221–239.

- Gotti, Maurizio. 2003. *Specialized discourse. Linguistic features and changing conventions*. Bern: Peter Lang.
- Gotti, Maurizio. 2005. *Investigating specialized discourse*. Bern: Peter Lang.
- Hooke, Robert. 1667 [1975]. *Micrographia: Or some physiological descriptions of minute bodies made by magnifying glasses with observations and inquiries thereupon*. London: John Martyn [facsimile reprint Stuttgart: Edition Medicina Rara].
- Sprat, Thomas. 1667. *The history of the Royal-Society of London, for the improving of natural knowledge*. London: J. Martyn. [Jackson I. Cope and Harold Whitmore Jones (eds.) 1959; repr. 1966. London: Routledge & Kegan Paul].
- Taavitsainen, Irma. 1993. Genre/subgenre styles in Late Middle English?. In M. Rissanen, M. Kytö, and M. Palander-Collin (eds.). *Early English in the computer age: Explorations through the Helsinki Corpus, 171–199*. Berlin and New York: Mouton de Gruyter.
- Taavitsainen, Irma. 1994. On the evolution of scientific writings from 1375 to 1675: Repertoire of emotive features. In F. Fernandez, M. Fuster, and J. J. Calvo (eds.). *English historical linguistics 1992*, 329–342. Amsterdam: John Benjamins.
- Taavitsainen, Irma and Päivi Pahta. 1995. Scientific ‘thought-styles’ in discourse structure: Changing patterns in a historical perspective. In B. Wårvik, S-K. Tanskanen, and R. Hiltunen (eds.). *Organization in discourse. Proceedings from the Turku conference*, 519–529. Turku: University of Turku.
- Taavitsainen, Irma and Päivi Pahta (eds.). 2004. *Medical and scientific writing in late medieval English*. Cambridge: Cambridge University Press.
- Taavitsainen, Irma, Päivi Pahta and Martti Mäkinen. 2005. *CD: Middle English medical texts*. Amsterdam and Philadelphia: John Benjamins.

