

Multi-Dimensional Analyses of Authors' Styles: Some Case Studies from the Eighteenth Century

DOUGLAS BIBER and EDWARD FINEGAN

1. Introduction

Considerable attention has been paid to the linguistic analysis of genres (or registers) in English, with special attention to the functional associations of distinctive linguistic features. Many of these studies also consider diachronic analyses of particular registers. Few studies, though, have taken a diachronic approach to the linguistic analysis of register variation, that is, a comparative analysis of the linguistic development of functionally different registers. In Biber and Finegan (1989, 1990) we present the results of the first stage of a research project that we are undertaking for this purpose. In those articles, we trace the developments of three written genres—essays, fiction, and personal letters—from the late seventeenth century to the present. We show that, although these genres have been evolving at different rates along three independent linguistic dimensions, the developments have not been random; rather, they reflect a single underlying pattern of drift towards more oral linguistic characterizations. We adopted the term 'drift' from Sapir to suggest a cumulative series of gradual linguistic developments in a consistent direction; we extended the notion, however, by proposing functional and attitudinal motivations—conscious as well as unconscious—for the observed developments among these genres.

In the present paper, we show how this multidimensional framework can be profitably used to analyse the linguistic characteristics of authors' styles. Although there have been numerous quantitative analyses of style (e.g. Burrows 1986, Oakman 1975; cf. Hockey 1980, ch. 6), there are methodological problems associated with such analyses. For example, Leech and Short (1981, 42–73) raise two fundamental issues: (1) As there are no absolute analyses of style, particular styles must be analysed in relative terms by comparison with other texts and styles; this raises the issue of what standard of comparison to use for particular analyses. (2) As there is no exhaustive list of linguistic features that might be relevant in defining the style of a text, and little agreement on which features

might be stylistically relevant, the appropriate linguistic characteristics chosen for analysis are a further issue. The multidimensional approach addresses both these issues. We do not claim, of course, that it provides the only correct approach to quantitative analyses of style, but rather that it offers a well-developed analytical framework that enables description of particular texts and authors relative to a wide range of comparable texts, genres, historical periods, and linguistic features.

2. Three Dimensions of Variation

Our analysis is based on the model of register variation developed in Biber (1988). The major analytical constructs in that model are dimensions of variation. Each dimension is a continuous parameter of variation comprising a group of co-occurring linguistic features; co-occurrence patterns are identified empirically, rather than being proposed on a priori functional bases. Our previous studies have shown that no single dimension is adequate in itself to account for the range of linguistic variation in English; rather, a multidimensional analysis is required.

Dimensions have both linguistic and functional content. The linguistic content is defined by a group of features (such as nouns, attributive adjectives, prepositional phrases) that co-occur with a markedly high frequency in texts. These groups of co-occurring linguistic features are identified using a factor analysis. On the assumption that co-occurrence reflects shared functions, the co-occurrence patterns are interpreted to assess the situational, social, and cognitive functions most widely shared by the features.

We focus here on three dimensions that interact directly with the contrasts among oral and literate genres in English. The linguistic features comprising these dimensions are given in Table 1, which also includes the functional labels associated with each.¹

- Dimension A — Informational versus Involved Production
- Dimension B — Elaborated versus Situation-Dependent Reference
- Dimension C — Abstract Style

Biber (1988) provides further methodological and theoretical discussion: methodological details concerning the texts and linguistic features analysed, as well as the computational and statistical techniques used; and theoretical discussion of the linguistic characteristics of present-day spoken and written English genres with respect to these dimensions of variation. Dimensions can be used to analyse the linguistic characteristics of English texts and genres by computing dimension scores for texts. These scores are computed by summing the frequencies of the features having

Contemporary Texts representing 23 Spoken and Written Genres

DIMENSION A Informational v. Involved Production		DIMENSION B Elaborated v. Situation-Dependent Reference	
nouns	0.80	WH-relative clauses on object positions	0.63
word length	0.58	pied-piping constructions	0.61
prepositions	0.54	WH-relative clauses on subject positions	0.45
type/token ratio	0.54		
attributive adjs.	0.47		
private verbs	-0.96	time adverbials	-0.60
that-deletion	-0.91	place adverbials	-0.49
contractions	-0.90	other adverbs	-0.46
present-tense verbs	-0.86		
2nd-person pronouns	-0.86		
DO as pro-verb	-0.82		
analytic negation	-0.78		
demonstrative pronouns	-0.76		
general emphatics	-0.74		
1st-person pronouns	-0.74		
pronoun IT	-0.71		
BE as main verb	-0.71	DIMENSION C	
causative subordination	-0.66	Abstract Style	
discourse particles	-0.66		
indefinite pronouns	-0.62	conjuncts	0.48
general hedges	-0.58	agentless passives	0.43
amplifiers	-0.56	past-participial adverbial clauses	0.42
sentence relatives	-0.55	BY-passives	0.41
WH-questions	-0.52	past-participial WHIZ deletions	0.40
possibility modals	-0.50		
non-phrasal co-ordination	-0.48		
WH-clauses	-0.47		
final prepositions	-0.43	[no negative features]	

the dimension score for dimension B is computed by adding together the frequencies of WH-relative clauses on object positions, pied-piping constructions, WH-relative clauses on subject positions, phrasal coordinators, and nominalizations—the features with positive loadings on dimension B (Table 1)—and then subtracting the frequencies of time adverbials, place adverbials, and other adverbs: the features with negative loadings. In this way, a score for each dimension is computed for each text; then the means of the dimension scores for any genre can be

computed. Plots of these mean dimension scores enable linguistic characterizations of each genre and comparison of the relations between any two genres. For example, Fig. 1 plots the mean dimension A scores (Informational versus Involved Production) of five written and speech-based genres in each of four historical periods (based on Biber and Finegan 1989, 1990). The genres with large positive values (such as nineteenth-century fiction and essays) have markedly high frequencies of nouns, prepositional phrases, and long words (the positive features on dimension A), together with markedly low frequencies of present-tense verbs, private verbs, first-

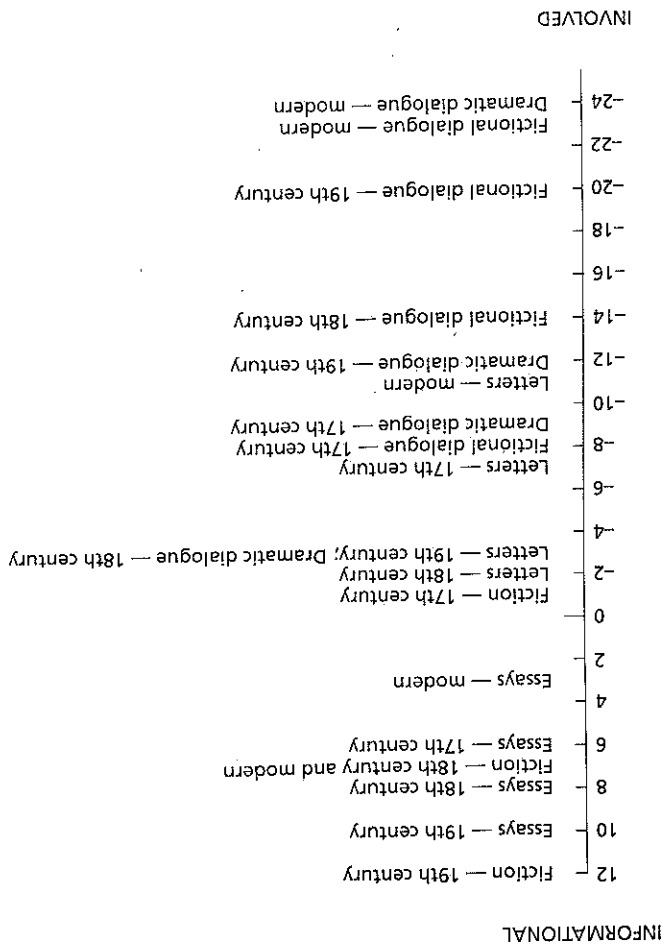


Fig. 1. Mean Scores of Dimension A for Fiction, Essays, Letters, Fictional Dialogue, and Drama in Four Centuries ('Informational versus Involved Production')

on dimension A). Genres with large negative values (such as modern fictional and dramatic dialogue) have very low frequencies of nouns and prepositional phrases, plus very high frequencies of private verbs and contractions.

The overall distribution of genres shown in Fig. 1 reflects the function of dimension A as distinguishing among texts according to the demands and possibilities of informational production. Although literary dialogue is carefully produced, it represents face-to-face conversation and thus is characterized by linguistic features of interaction and involvement. In

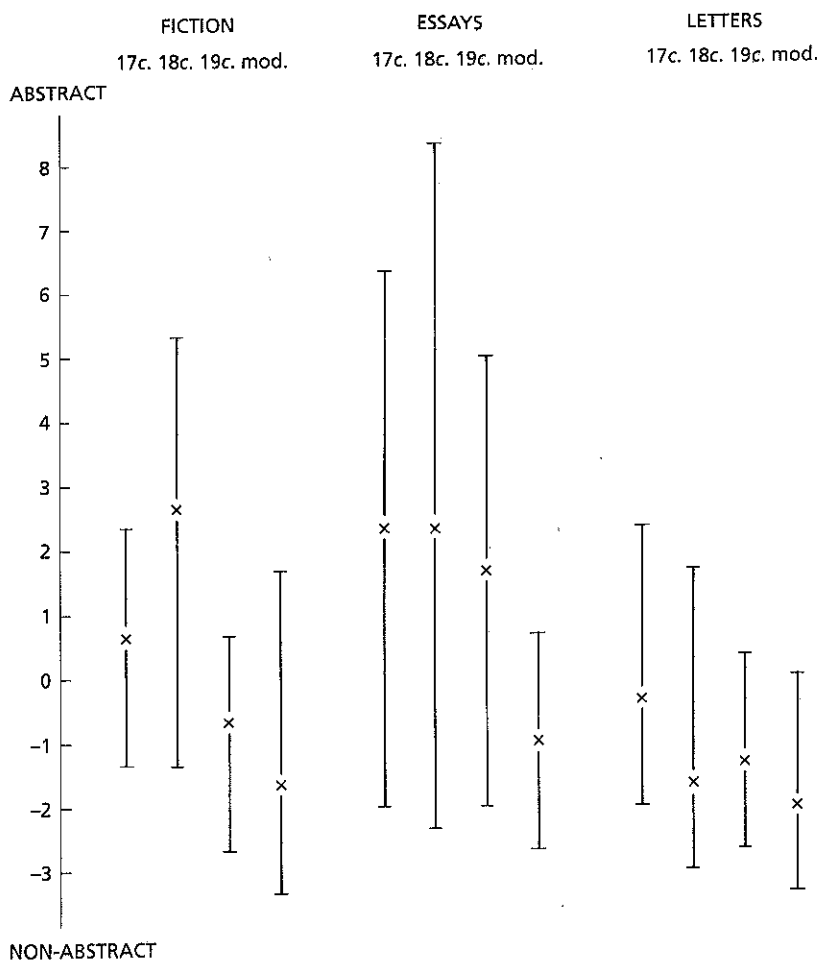


Fig. 2. Drift of norms along Dimension C ('Abstract versus Non-Abstract Style') for Fiction, Essays, and Letters (x marks the mean score of each genre)

essays and fictional narrative, on the other hand, texts are carefully produced to integrate high amounts of information, and there is no direct or immediate interaction among participants. In a similar way, dimension scores for dimension B and dimension C can be computed for different genres. By considering genres relative to this three-dimensional space of variation, it is possible to analyse their synchronic relations and the course of their diachronic developments.

It is also possible to consider the actual spread of scores within each genre for each period. Fig. 2 plots the dimension C scores, reflecting Abstract versus Non-Abstract Style, for the 33 fiction texts, 51 essays, and 31 letters analysed in Biber and Finegan (1989). The mean score of each genre in each period is marked by an asterisk, and the range of scores is shown by the vertical bar. High scores on this dimension, such as for some eighteenth-century essays, show a very frequent use of conjuncts, agentless passives, past participial adverbial clauses, and BY-passives (the positive features on dimension C), while low scores, such as for modern letters, show markedly infrequent use of these same features. The figure shows an overall movement towards less abstract scores in more recent periods, and an extremely large range of variation characterizing the eighteenth century (particularly in fiction and essays). Similar patterns are found with respect to the Elaborated v. Situated Reference dimension and the Informational v. Involved Production dimension, plotted in Figs. 3 and 4 respectively. On the basis of these patterns, we concluded in our earlier study that there has been a steady drift to more oral styles in these three written genres, as measured by changes towards more involved styles (on dimension A), more situated styles (on dimension B), and less abstract styles (on dimension C). The study further found, on all three dimensions, an eighteenth-century reaction to this general pattern of drift; in that period, there was an extreme range of variation among the texts within a genre (i.e. some texts were markedly oral, while others were extremely literate), and the overall genre characterization actually became somewhat more 'literate'. We interpret these patterns by reference to conscious and unconscious attitudes towards style in the various periods, which interact with larger developments of standardization, popular literacy, and mass education.

This framework can be used for stylistic analyses, providing principled solutions to the two issues raised by Leech and Short. First, a principled basis of comparison is provided by the multidimensional analysis of texts and genres, which characterizes both the typical (i.e. average) texts and the extremes (shown by the range of variation) within and across genres and historical periods. Second, a principled selection of linguistic features is provided by the dimensions of variation, which represent a wide range of the linguistic features found in English. Further, several previous studies (e.g. Biber 1988, 1990; Biber and Finegan 1989) have shown that

ELABORATED

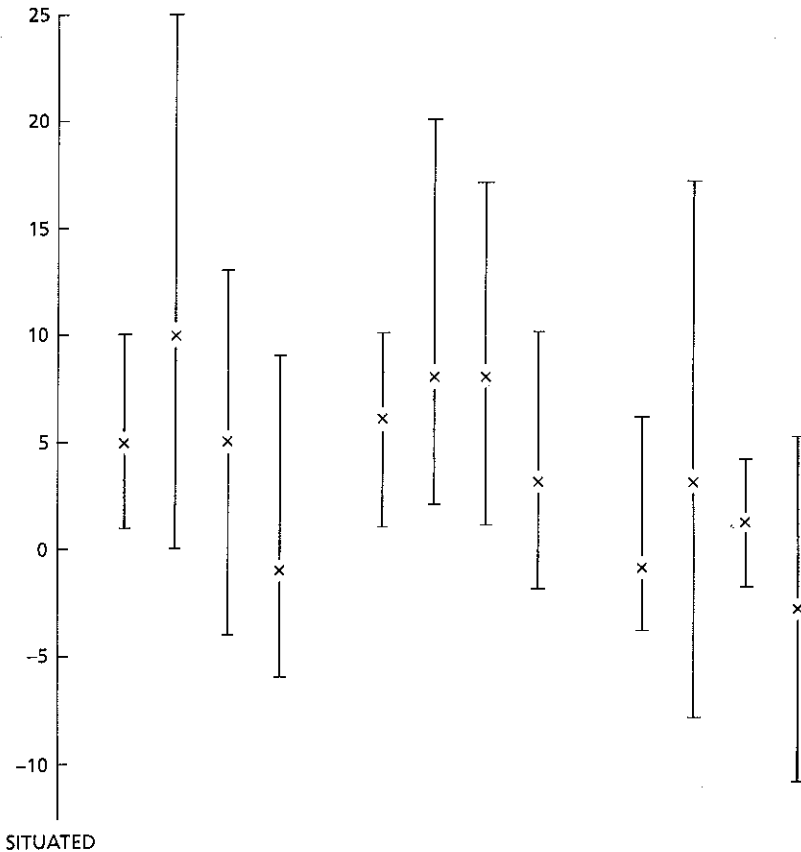


Fig. 3. Drift of Norms along Dimension B ('Elaborated versus Situated Reference') for Fiction, Essays, and Letters (x marks the mean score of each genre)

the dimensions are functionally coherent and important predictors of genre differences. We thus undertook the present study with the expectation that the dimensions would provide useful perspectives on the salient linguistic characteristics of an author's style.

3. A Pilot Study on the Styles of Four Authors

Based on our earlier diachronic analyses of genre drift, we decided to focus here on eighteenth-century authors, to investigate further the

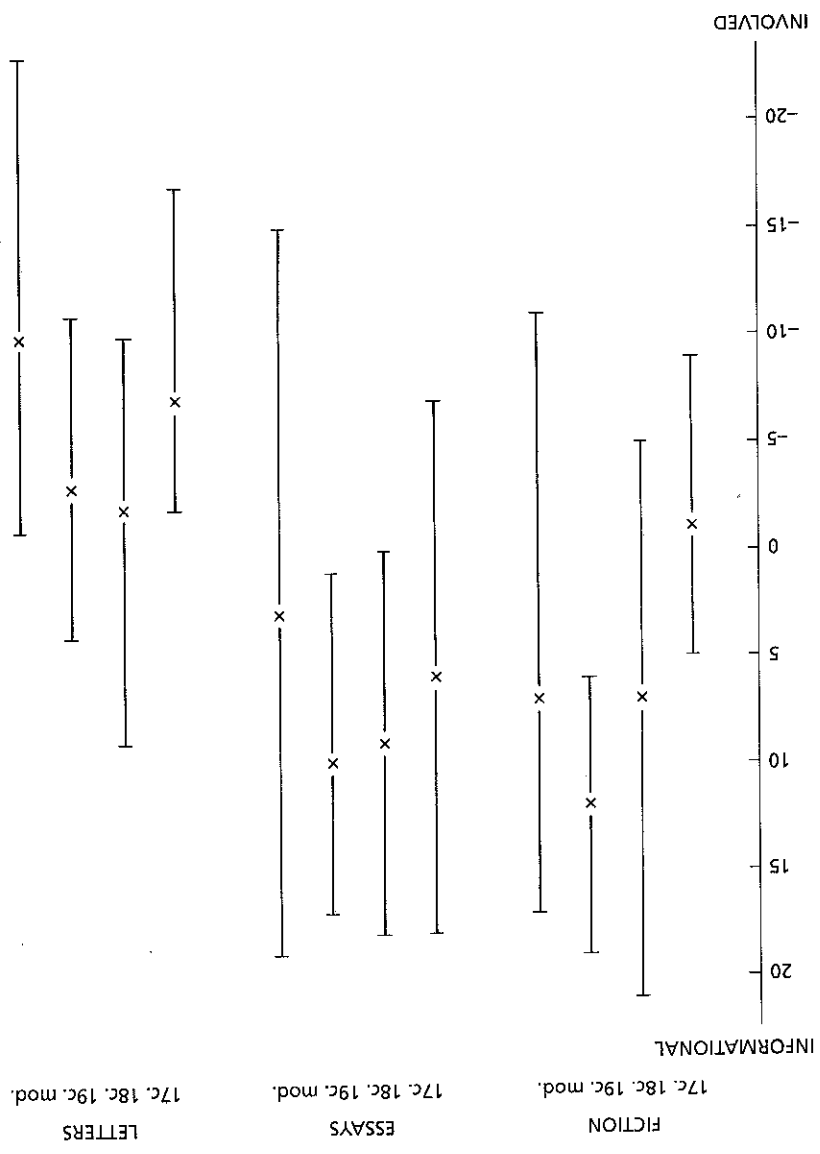


Fig. 4. Drift of Norms along Dimension A ('Informational versus Involved') for Fiction, Essays, and Letters (x marks the mean score of each genre)

extremely wide range of variation found in that period. We consider the writing styles of four influential authors: Addison, Defoe, Swift, and Johnson. The first three were contemporaries in the early part of the eighteenth century, while Johnson wrote during the second half of the

Addison

Essays

- 'The Royal Exchange' 1711—AD1
 'Wit: True, False, Mixed' 1711—AD2
 'On the Scale of Being' 1712—AD3

Defoe

Essays

- 'Some Thoughts on . . . Commerce' 1713—DE1
 'Essays on Projects: Of Academies' 1698—DE2

Fiction

- Life of Robinson Crusoe* (ch. 1) 1719—DE1
Moll Flanders (ch. 1) 1722—DE2

Swift

Essays

- 'A Modest Proposal' 1729—SW1
 'Abolishing Christianity in England' 1708—SW2
 'A Tale of a Tub' 1704—SW3

Fiction

- Gulliver's Travels* 1726
 Part II, Ch. 1—SW1
 Part II, Ch. 2—SW2

Johnson

Essays

- 'On Fiction' (*Rambler* no. 4) 1750—JO1
 'Preface to Dictionary of the English Language' 1755—JO2
 'Lives of the Poets: Milton' 1779—JO3

Fiction

- The History of Rasselas: Prince of Abyssinia* 1759
 Narrative passage—JO1
 Descriptive passage—JO2

century. For each author, we analyse several prose samples from fiction and essays. The particular texts analysed are summarized in Table 2.

Figs. 5-7 plot the dimension scores for each of these text samples relative to the linguistic characteristics of comparable texts. For example, Fig. 5 plots the scores of each text with respect to the Abstract Style dimension (C). The distribution of texts along this dimension generally corresponds to the mid-eighteenth-century stylistic shift described by Gordon (1966, 133-52), with early eighteenth-century prose being speech-based and even conversational, while late eighteenth-century prose is classical in style, representing a deliberate reaction against speech-based

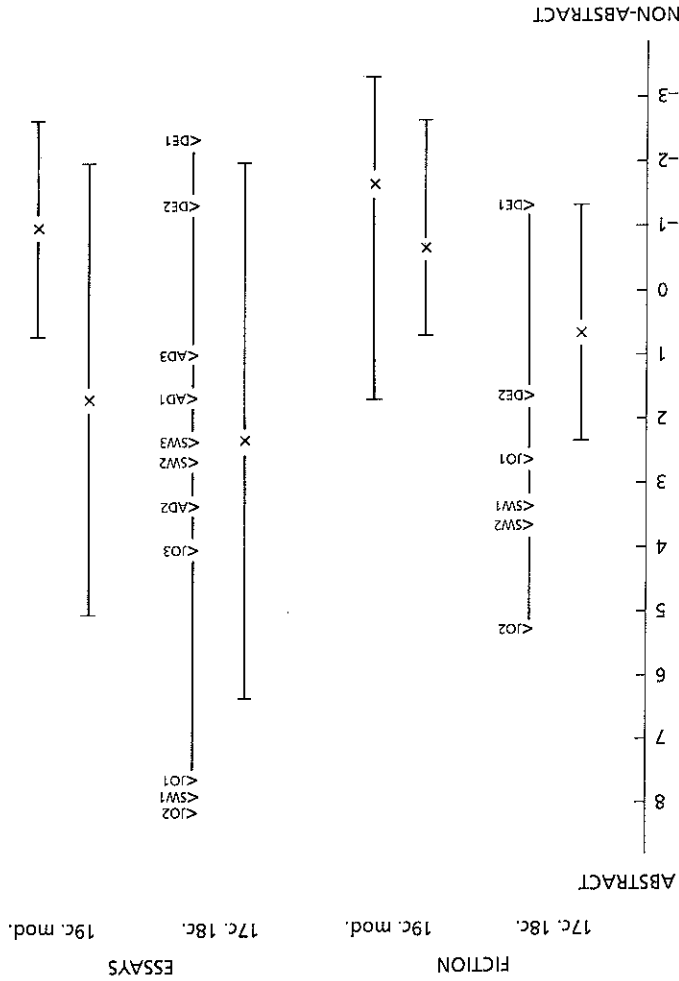


Fig. 5. Characterization of Particular Texts along Dimension C ('Abstract versus Non-abstract Style')

Thus, Johnson, representing the classical style, has the highest scores with respect to both fiction (the descriptive sample from *History of Rasselas*) and essays (from his Preface to the *Dictionary*); these scores represent the most abstract values in this study, marked by very frequent use of passives, past-participial clauses, etc. In contrast, the other three authors tend to have considerably lower scores on this dimension. However, a simple dichotomy between speech-based and classical styles cannot account for the differences among the three early authors shown on Fig. 5—Swift tends to high scores, Defoe has

markedly low scores, and Addison has moderate scores. These linguistic differences generally agree with previous literary characterizations of the three authors (e.g. Fowler 1987, Gordon 1966, Lonsdale 1986). Defoe was from the middle class and explicitly wrote for a newly literate middle-class audience, and so he consciously strove for simplicity and clarity (reflected on this dimension in the marked absence of passive constructions). Addison and Swift were both described as practising a 'middle style'. Swift advocated a middle style that avoids the 'barbarity' of vulgar abbreviations while at the same time avoiding the 'abundance of polysyllables' (*Tatler*, 1710; reported in Gordon 1966, 136). Johnson characterizes Swift's style as having an 'easy and safe conveyance of meaning' (*Life of Swift*; reported in Gordon 1966, 145). Johnson wrote about Addison that his prose is 'the model of the middle style; on grave subjects not formal, on light occasions not grovelling; pure without scrupulosity, and exact without apparent elaboration; always equable, and always easy, without glowing words or pointed sentences' (*Life of Addison*, reported in Fowler 1987, 172). The tendencies shown on the Abstract Style dimension (Fig. 5) correspond fairly well to these literary characterizations: Defoe consistently has the least abstract, and therefore most speech-like, style; Swift and Addison both have intermediate (middle) styles; and Johnson tends towards markedly abstract styles.

A comparison of these text samples on the other two dimensions, however, does not follow previous characterizations as neatly. Fig. 6 plots the scores for the Elaborated versus Situated Reference dimension (B), which represents the use of various relative clause constructions versus place and time adverbials. With respect to fiction, Johnson is clearly the most elaborated (and, therefore, literate), with the descriptive fiction sample being extremely marked—by far the most elaborated text sample in our study. The fiction passages from both Defoe and Swift are moderately non-elaborated, although Swift's chapter 1 is the most situated of any text in our study.

With respect to essays, there is considerable overlap among authors along this dimension (B), and considerable diversion from the patterns seen with respect to the Abstract Style dimension (C; Fig. 5). One of Defoe's essays is the least elaborated (and most situated) of all these texts, as would be expected from previous characterizations, but the other Defoe sample is relatively elaborated. Swift's three essays are consistently middle, and the three Johnson samples have moderately elaborated scores on this dimension. Addison's essays, though, show wide variation here, ranging from a markedly elaborated score to quite situated. It is interesting in this regard that Johnson in his description of Addison's style makes reference to its adaptability. As several commentators have noted, Addison wrote on an extremely wide range of topics for a variety of purposes, and his prose is praised for being well-suited to this range of text types.

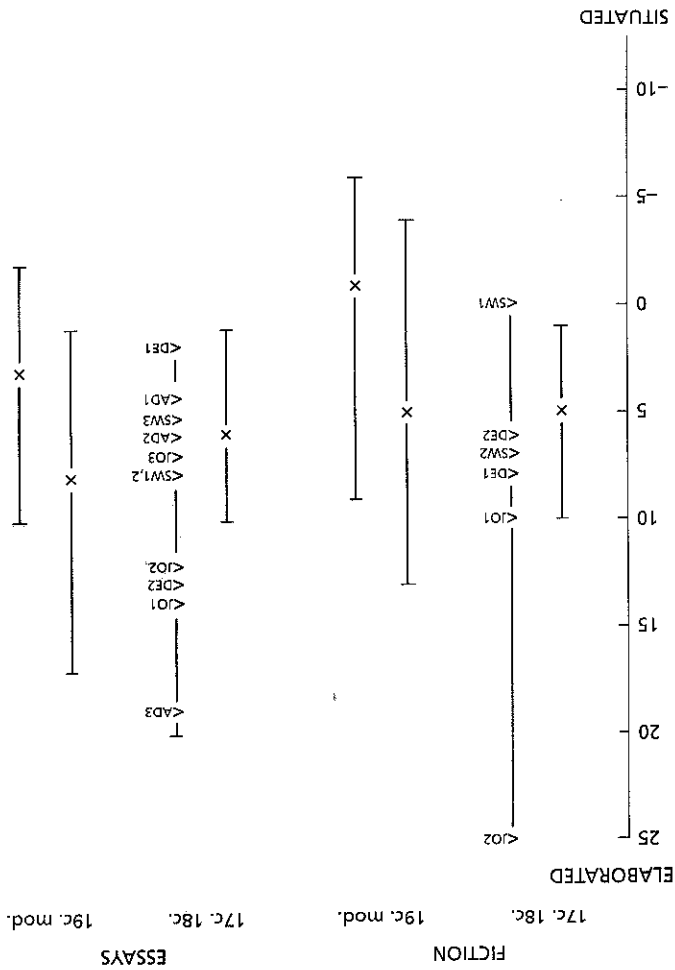


Fig. 6. Characterization of Particular Texts along Dimension B ('Elaborated versus Situation-Dependent Reference')

The wide range of linguistic variation shown for Addison's essays is perhaps a reflection of his stylistic adaptation to various topics and purposes.

On Fig. 7, which plots the scores for the Informational versus Involved Production dimension (A), there are other departures from the stylistic stereotypes for these authors. As expected, with respect to fiction, Johnson has the most informational text sample (again for the descriptive passage from *Rasselas*), while Defoe has the most involved text sample. The other

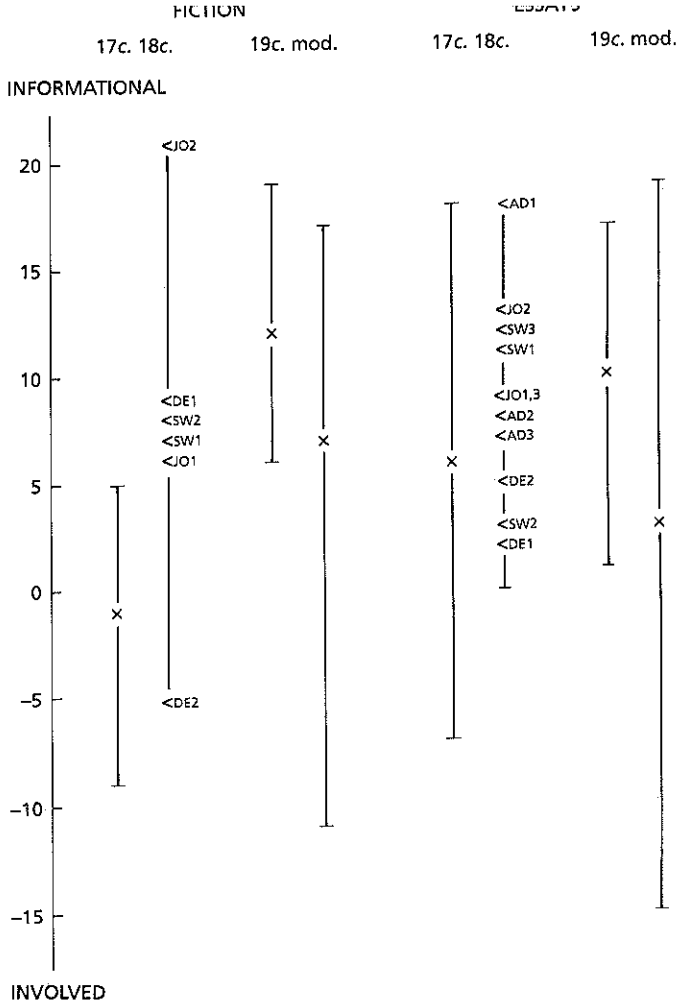


Fig. 7. Characterization of Particular Texts along Dimension A ('Informational versus Involved Production')

with Defoe's *Robinson Crusoe* sample actually being somewhat more informational than Johnson's narrative passage from *Rasselas*.

With respect to essays, Defoe has the most 'involved' text sample, but Addison rather than Johnson has the most informational text (similar to the pattern on Fig. 6 for the 'Elaboration' dimension). All three Johnson essays are moderately informational here, but both Swift's and Addison's essays extend over relatively wide ranges of variation, again perhaps reflecting the adaptability of the middle style.

Biber, D. (1988), *Variation across Speech and Writing* (Cambridge).

— (1990), 'Methodological Issues regarding Corpus-Based Analyses of Linguistic Variation', *Literary and Linguistic Computing* 5: 257–69.

— and Finegan, E. (1989), 'Drift and the Evolution of English Style: A History of Three Genres', *Language* 65: 487–517.

— (1992), 'The Linguistic Evolution of Five Written and Speech-Based English Genres from the 17th to the 20th Centuries', in M. Rissanen, O. Thälsten, T. Nevalainen, and I. Taavitsainen (eds.), *History of Englishes: Interpretations in Historical Linguistics* (Berlin), pp. 688–704.

References

1. Each dimension is a 'factor', identified by a factor analysis of the distribution of 67 linguistic features across 481 contemporary texts from 23 spoken and written genres. The 'loadings' (on Table 1) indicate the strength of the association between particular features and factors. Only loadings over 0.40 are included in the table, as these were the primary features used in the interpretation of each factor. There are positive and negative loadings on Dimensions A and B; these represent two groupings of co-occurring features that have a complementary distribution to one another, and thus they operate together in defining a dimension. The polarity of Dimension 1 has been reversed here from that presented in Biber (1988), to highlight the textual parallels across dimensions. 2. All frequencies are standardized to a mean of 0.0 and a standard deviation of 1.0 before the dimension scores are computed (see Biber 1988, 93–5).

Notes

Our primary goal in this paper has been to illustrate a promising methodology for linguistic analyses of authors' styles. For this purpose, we carried out a pilot investigation of seventeen text samples from four eighteenth-century authors across two genres. A complete investigation of these issues would require: (1) a fuller overall analysis of fictional and essayist genres in the eighteenth century, and (2) longer samples from individual texts and a more complete representation of the works of each author. For example, novels could be represented by several 1,000-word passages, to determine the extent of internal linguistic variation. Authors could be represented by numerous texts, to determine their range of variation and their typical style. Diachronic change over the course of an author's life could be investigated, as well as the similarities and differences in an author's style across genres. And with respect to analyses of eighteenth-century prose, several additional authors from the latter part of the century should be analysed, to determine the extent to which Johnson's style is representative of his period.

4. Conclusion

- Burrows, J. P. (1988), 'Modal Verbs and Modal Auxiliaries in Austen's Style', *Literary and Linguistic Computing* 1: 9-23.
- Fowler, A. (1987), *A History of English Literature* (Oxford).
- Gordon, I. A. (1966), *The Movement of English Prose* (London).
- Hockey, S. (1980), *A Guide to Computer Applications in the Humanities* (Baltimore).
- Leech, G. N. and Short, M. H. (1981), *Style in Fiction: A Linguistic Introduction to English Fictional Prose* (London).
- Lonsdale, R. (ed.) (1986), *Dryden to Johnson* (Worcester).
- Oakman, R. L. (1975), 'Carlyle and the Machine: A Quantitative Analysis of Syntax in Prose Style', *ALLC Bulletin* 3: 100-14.