

# Modal verbs in written Indian English: A quantitative and comparative analysis of the Kolhapur corpus using correspondence analysis

*Andrew Wilson*  
*Lancaster University*

## ***1 Introduction***

The central modal verbs (*can, could, may, might, must, ought, shall, should, will* and *would*) are some of the most common yet most semantically complex words in the English language. Consequently, these verbs have frequently attracted the attention of corpus linguists: major corpus-based studies include those of Coates (1983) on British English, Ehrman (1966) on American English, Katikar (1984) on Indian English, and, from a pedagogically-oriented perspective, Mindt (1995). These detailed empirical studies have added enormously to our understanding of how the modal verbs behave. However, many corpus studies – and not only of modal verbs – also possess a certain deficit in that they tend quite often to concentrate on using the corpus as a unitary body of text, in order to arrive at an ‘averaged out’ picture of the entire language variety from which the corpus is constructed. Thus, *intra*-varietal variation across text-types is often backgrounded. A similar criticism can be made of the study of variation in modal usage across national and sub-national varieties of English: although Johansson and Hofland (1982) discovered noteworthy differences between British and American English in the frequencies of a number of the modal verbs, it is only relatively recently that such quantitative differences between corpora have begun more systematically to be explored.

In a paper which can be considered ground-breaking for corpus-based studies of the modal verbs, Nakamura (1993) carried out a quantitative comparison of modal frequencies in the American Brown and the British Lancaster/Oslo-Bergen (LOB) corpora. Although the basic frequency counts for these corpora had been made already by Johansson and Hofland (1982), Nakamura took these a step further by applying a technique called Hayashi’s Quantification Method Type III, which is very closely related to the more familiar (to Europeans) technique of correspondence analysis.

Correspondence analysis, and Hayashi's Quantification Method Type III, are essentially multivariate mapping techniques, closely related to factor analysis, which aim to represent the data contained in complex frequency tables of many rows and columns in terms of a smaller number of hypothesized axes. The techniques assume that the frequencies for the many samples and variables included in basic frequency tabulations may, in fact, be masking a smaller number of 'hidden' variables which can help to explain better why the observed frequency differences occur. The analyses thus aim mathematically to 'summarize' the variables and samples in terms of this smaller number of hypothesized reference axes.<sup>1</sup> Each variable receives a loading on each axis, signifying its closeness to that axis. Different variables will have larger or smaller loadings on each of the hypothesized axes, so that it is possible to see which variables are most characteristic of a given axis. For example, in comparing individual word frequencies between a number of texts, it may be found that words in the religious domain load highly on the positive end of one axis, whereas words in the government domain load highly on the negative end of another. These two statistical methods also plot the samples on the same set of axes as the variables, so that it is possible to see how far certain variables are linked to particular samples. Although correspondence analysis can extract more than two hypothesized axes, in practice, the first two axes normally suffice to give a visual picture of the interrelationships within the data.

Using Hayashi's Quantification Method Type III, Nakamura was able to observe a number of important patternings of modal verbs across genres and varieties. In the cases of both Brown and LOB, he found that the main distinction underlying the frequency distributions was between imaginative and informative writing. Axis 1 of his study reliably separated the imaginative genres from the informative genres in both the single-corpus plots and in the combined corpus plot. In the case of Brown, the modals *would*, *used*, *ought* and *might* were primarily characteristic of imaginative writing, whilst *dare*, *need*, *may*, *shall* and *could* were characteristic of informative writing; *can*, *should* and *must* showed a weaker tendency towards informative writing, whilst *will* was neutral. Within the LOB corpus, *used*, *might* and *ought* remained characteristic of imaginative writing, but were joined by *dare*, *could* and *would*. *May*, *should* and *shall* remained characteristic of informative writing, but were also joined by *will*, which had been rather neutral between the two writing types in Brown. In LOB, *can* and *need* were relatively neutral, though still with a slight tendency towards informative writing. When the data for the two corpora were combined in a single plot, Nakamura found that the plot showed two primary distinctions: first, as

in the single corpus plots, between imaginative and informative writing along Axis 1, and second, between the samples from Brown and LOB along Axis 2. In effect, the plot formed a quadrant with the LOB imaginative genres dominating one quarter, the Brown imaginative genres the second, the LOB informative genres the third, and the Brown informative genres the fourth. The modals which primarily distinguished the Brown from the LOB genres were *could*, *would* and *will*. As just noted, *could* and *would* 'switch places' in the Brown and LOB samples with regard to their association with imaginative or informative writing, whilst *will* moves from being neutral in Brown to being characteristic of informative writing in LOB.

In this paper, I propose to extend Nakamura's style of multivariate quantitative analysis to the Kolhapur Indian English corpus, to discover how the modals behave both within the corpus itself (i.e. across text-types) and in comparison with the parallel British LOB and American Brown corpora (i.e. across national varieties). Katikar's study of modal usage in Indian English (1984, cited in Shastri 1988) noted few qualitative differences from British and American English, but, as Leitner (1991) has pointed out, differences between Indian English and other Englishes have tended most often to be quantitative rather than qualitative. It thus seems likely that any differences in modal use between Indian English, British English and American English will also prove to be quantitative.

## 2 Method

The corpora employed for the study were, as stated, the Kolhapur Corpus of written Indian English, the Brown Corpus of written American English, and the LOB Corpus of written British English. Each corpus contains approx. 1 million words of text, distributed across 15 genre categories. The corpora are sampled broadly according to the same criteria. The most substantial difference in sampling criteria is that the texts in the Brown and LOB corpora date from 1961, whereas the texts in the Kolhapur corpus date from 1978. As regards text typology, the differences between the Kolhapur corpus and the Brown and LOB corpora are, for the most part, comparatively minor, although the number of individual text samples in some of the fictional categories vary considerably. More notably, however, the category of Mystery and Detective Fiction differs in content, in that it contains samples of stories which are about mysterious or miraculous happenings, not just detective mystery stories; similarly, Adventure and Western Fiction contains no samples of westerns (Shastri, Patilkultarni and Shastri 1986).

Concordances of all the variants of the central modal lemmata *can, could, may, might, must, ought, shall, should, will* and *would* in the Kolhapur corpus of written Indian English were extracted using the WordCruncher software. The variants extracted included all the enclitic forms of the modals; for example, *I'll* was counted as an instance of *will* and *shan't* was counted as an instance of *shall*. Since, at the time of the data collection, only the British LOB corpus was available to me in corrected part-of-speech tagged form, the concordances were read manually to exclude non-modal uses of the ambiguous graphic word forms (e.g. *will* as a noun or as a verb of desire [*I will you to do it*]). The corrected concordances were then used to extract frequency counts of the modal lemmata in each of the 15 genre categories in the corpus. The cross-tabulation of modal frequencies for the 15 genres was finally loaded into the CORRESP program of Alastair McKinnon's FORMATK software (McKinnon 1989, 1993) and subjected to correspondence analysis.

The corresponding modal frequencies for written British English (based on the LOB corpus) and written American English (based on the Brown corpus) were taken from the tables in Nakamura (1993), which are based on the same lemmatization procedures that are used here for the Indian data. Based on these three sets of counts, a further cross-tabulation was produced showing the frequency counts of the modal lemmata for each of the 15 genre categories in all three corpora under examination. Nakamura's data were re-run because he included the semi-modals *need, dare* and *used* in his analyses: since these were excluded in the present investigation, my analysis of the Indian corpus and his analysis of Brown and LOB would not otherwise have been strictly comparable. In any case, although correspondence analysis and Hayashi's Quantification Method Type III produce very similar graphic plots, the techniques are not identical and it is therefore preferable to run all three data sets using the same statistical method.

### **3 Results and discussion**

The frequencies of the modal verbs in the Indian corpus are shown in Table 1.<sup>2</sup> The modal frequencies for the Brown and LOB corpora can be found in Nakamura (1993: 32).

Table 1: Modal frequencies in the Kolhapur Corpus

	can	could	may	might	must	ought	shall	should	will	would
KA	210	120	100	24	43	3	12	117	227	199
KB	148	49	86	29	34	4	4	131	218	104
KC	59	36	15	13	12	0	0	33	37	36
KD	89	23	46	4	28	1	10	27	43	46
KE	211	84	93	10	42	4	9	150	164	135
KF	178	117	112	41	59	5	7	114	154	176
KG	264	186	130	29	126	12	28	144	156	306
KH	106	39	204	10	58	2	162	203	315	123
KJ	424	164	249	52	114	5	42	183	206	186
KK	202	340	48	52	110	4	38	95	283	512
KL	55	134	21	22	59	2	17	44	153	205
KM	9	6	3	0	4	0	17	0	21	9
KN	47	81	6	7	20	0	13	27	56	90
KP	67	108	8	9	33	1	8	33	94	164
KR	55	54	14	9	19	0	3	13	52	65

### 3.1 Intra-varietal variation: genres and modals in the Indian corpus

The first two axes calculated by the correspondence analysis of the modals in the Indian corpus account for 84.1 per cent of the total variation in the data.

#### 3.1.1 Genres

A plot showing the coordinates of the 15 genre categories on these two axes is reproduced as Figure 1:

### Kolhapur Corpus: Genres

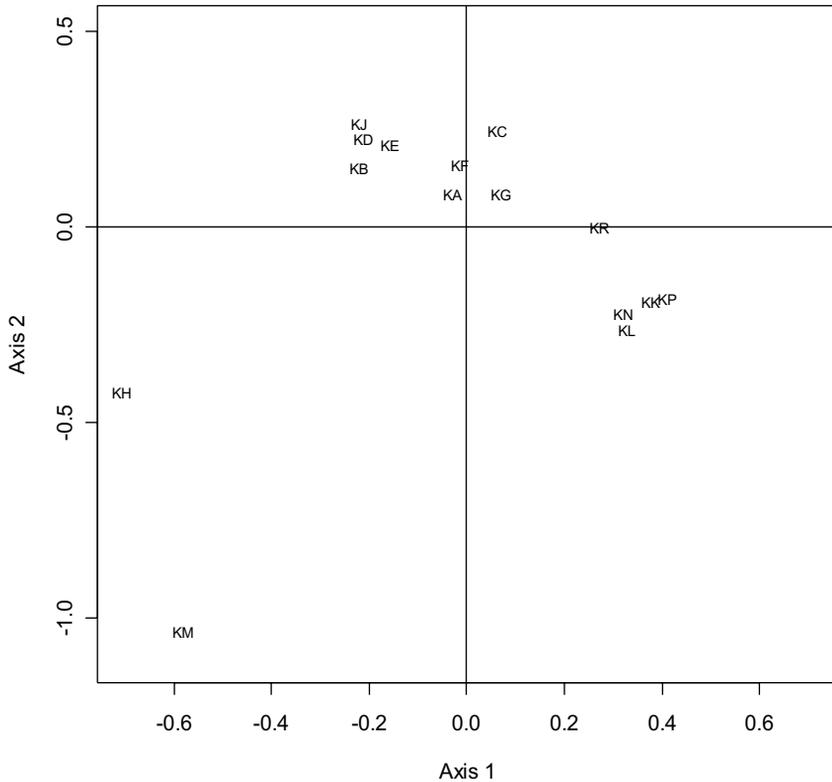


Figure 1: Correspondence analysis plot of genres in the Kolhapur Corpus

Looking first at the distribution of the genres along Axis 1, it will be seen that genres KK, KL, KN, KP and KR are positioned in the right-hand (positive) region of the axis.<sup>3</sup> KC and KG also appear to the right of the origin, although closer to it. The remaining genres KA, KB, KD, KE, KF, KH, KJ and KM fall in the left-hand (negative) region of the axis.

Genres KK to KR are ‘imaginative’ genres, whilst genres KA to KJ are ‘informative’ genres. It will thus be seen that five out of the six imaginative genres show a strong tendency to the positive region of Axis 1, whereas all but two of the informative genres fall in the negative region of Axis 1. The two

informative genres that do tend towards the positive region of Axis 1 (viz. KC [Reviews] and KG [Belles Lettres]) do so comparatively weakly and are notably of a more literary-imaginative nature than most other informative genres, dealing as they do with topics such as the theatre, music, books, etc. Axis 1 of the correspondence analysis of the Indian corpus thus seems quite clearly to discriminate the informative and imaginative genre categories. This replicates the findings of Nakamura on Brown and LOB. The only notable exception to this neat pattern in the Kolhapur Corpus is the Science Fiction genre, KM, which falls well within the negative region of the axis. We shall return to the matter of this genre category when considering the distribution of the modals.

Looking next at the distribution of genres along Axis 2, it will be seen that KA, KB, KC, KD, KE, KG and KJ fall in the upper (positive) region. KR falls on the origin. KH, KK, KL, KM, KN, and KP fall in the lower (negative) region.

Genres KL to KP are genres of fiction, and all show a strong affinity to the negative region of Axis 2. Of the genres in this region of the axis, only genre KH is a non-fictional genre. The remaining imaginative genre, KR (Humour), falls on the origin. All informative genres (apart from KH [Miscellaneous]) fall in the positive region of the axis. Here again, therefore, the imaginative/informative distinction appears to be the primary discriminator, although the Humour and Miscellaneous genres stand out as exceptions.

Before moving on to interpret these results in terms of the coordinates of the modal lemmata themselves, let us finally just take a look at the two-dimensional clustering of the genres. In the lower right quadrant of the plot, four of the five fictional genres cluster closely together. In the lower left quadrant, genres KH and KM appear, but show a strong independence both from each other and from the other genres in the corpus.

In the upper left quadrant, a cluster is formed from KB, KD, KE and KJ. Genres KA, KC, KF, and KG form a looser cluster around Axis 2, whereas KR stands out as more independent. Genre E (Skills and Hobbies) is not an easy genre category to interpret: for example, in relation to the Brown and LOB Corpora, Wikberg (1992) has questioned its internal consistency. However, genres KB (Press: Editorials), KD (Religion) and KJ (Learned and Scientific) do have something in common. Making use of Werlich's (1983) taxonomy of text-types, we can view these genres as being made up largely of argumentative or expository texts, in contrast to texts which are mostly narrative or descriptive. That these genres form a close cluster in relation to their use of modal verbs is thus, perhaps, not without import and likely to be related to their dominant communicative functions.

### 3.1.2 Modals

Let us now bring the modals themselves into the picture and see why the genres have clustered in the way that they have. A plot showing the coordinates of the modal verbs on the same two axes is reproduced as Figure 2. A combined plot of modal verbs and genres is shown in Figure 3.<sup>4</sup> Note that *shall* is missing from the plot: it falls outside the visible plot area, to the extreme lower left of the lower left quadrant.

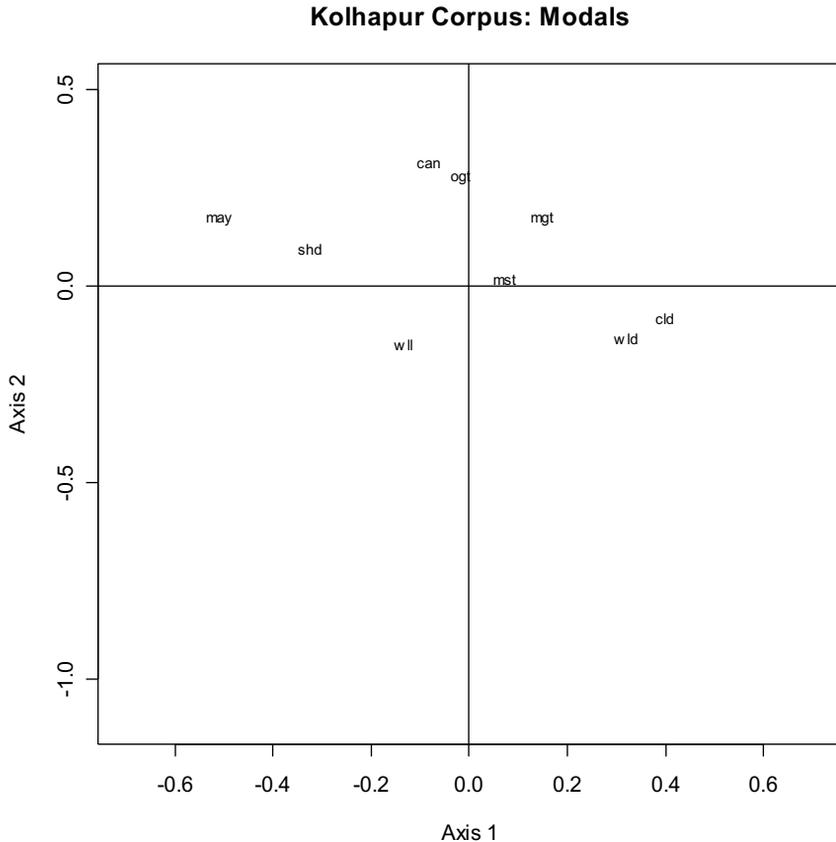


Figure 2: Correspondence analysis plot of modals in the Kolhapur Corpus

**Kolhapur Corpus: Genres and Modals**

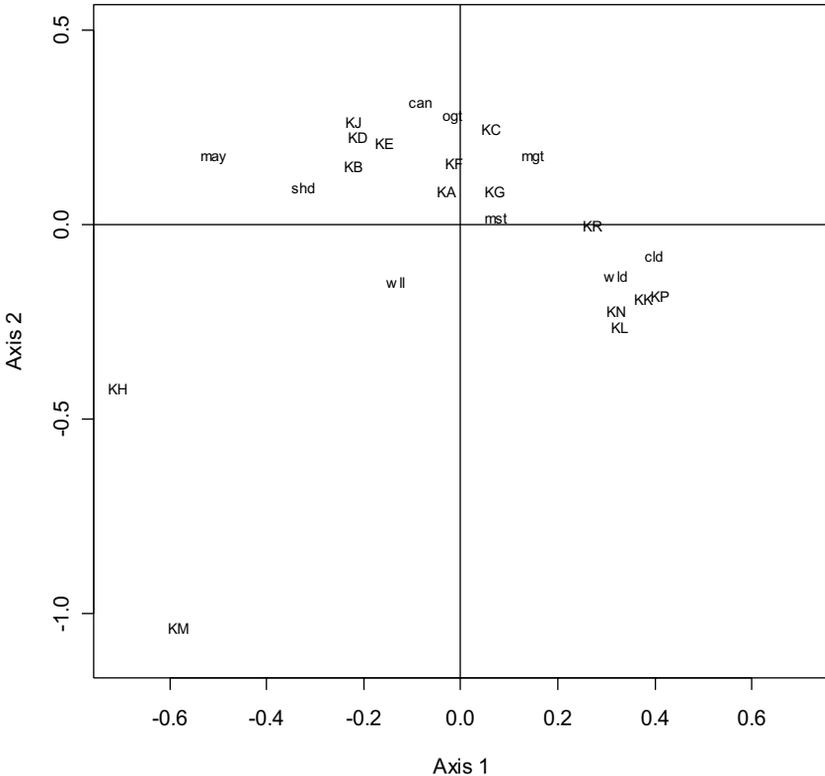


Figure 3: Correspondence analysis plot of modals and genres in the Kolhapur Corpus

Axis 1 discriminates between the modals *could*, *would*, *might* and *must* in the positive region, and *can*, *will*, *should*, *may* and *shall* in the negative region. *Ought* falls almost on the origin, just within the negative region. Superimposing the genre distribution on Axis 1, we can see that the first group of modals is characteristic of imaginative prose, whilst the second group is characteristic of informative prose. This corresponds closely to Nakamura’s findings on LOB, slightly less so to his American findings, where *could* was not so closely linked to imaginative writing.

Axis 2 distinguishes *can*, *ought*, *may*, *might*, *should* and *must* from *could*, *would*, *will* and *shall*. Here again, the main genre discriminator was found in

section 3.1.1 to be the imaginative/informative distinction, but with Miscellaneous texts and Humour forming exceptions. This is interpretable in that *will* and especially *shall* – both contributors to the negative region of Axis 2 – are prototypical markers of both futurity and obligation. Fictional prose, which is the overarching category of all imaginative corpus texts apart from Humour, typically contains substantial amounts of direct speech and thought, and these are likely to contain many more future-oriented statements than informative texts. The presence of *would* in the negative region may find a similar explanation, in that it is, *inter alia*, the back-shifted form of *will*, which is likely to be encountered in indirect speech and thought presentation. As for the presence of the Miscellaneous texts category, this is made up partly of rules and regulations, which dictate what people *shall* or *will* do. This finding again mirrors Nakamura's findings on Brown and LOB.

Looking at the two-dimensional clustering, the two 'past-tense' modals *could* and *would* seem to be the verbs which are most characteristic of fictional prose, probably arising from a mixture of indirect speech and thought presentation and past-tense narrative (see above). *Shall* and *will* appear to characterize the Miscellaneous category, as just mentioned above. *Shall* seems also, for some unclear reason, very characteristic of Science Fiction. This finding is unique to the Indian corpus and may be due to some oddity of a single authorial style: the Indian Science Fiction category is not a large one, consisting of only two texts as compared to six in Brown and LOB. *May* and *should*, along with *can* and *ought* again, dominate the upper left quadrant and thus appear to be especially characteristic of Learned and Scientific writing, Religion, Skills and Hobbies, and Press Editorials. As noted earlier, these are mostly argumentative/expository text-types. Myers (1989) has discussed the use of the modals as hedging devices in scientific writing, where strong claims are being made. He cites *would* and *could* as prototypical examples of this practice, but the Indian corpus, like the Brown and LOB corpora, suggests that *may* and *can* may be more prototypical of scientists' hedging.

### **3.2 *Inter-varietal variation: Indian, British and American Englishes compared***

The first two axes calculated by the correspondence analysis of the combined modals data account for 61.5 per cent of the total variation in the data.

#### **3.2.1 *Genres***

A plot showing the coordinates of the 15 genre categories on these two axes is reproduced as Figure 4:

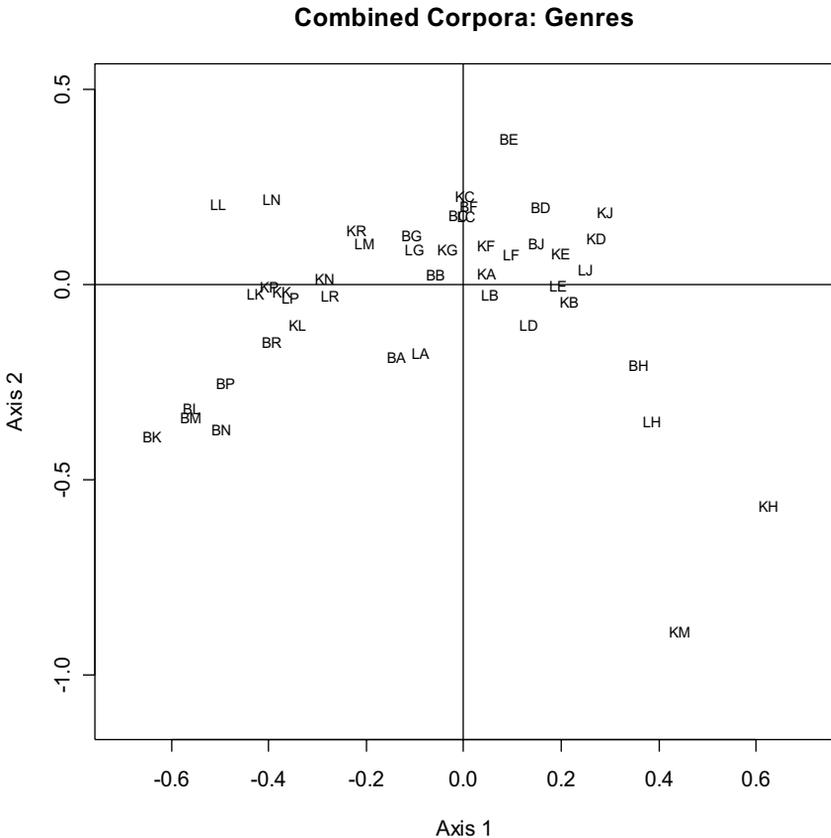


Figure 4: Correspondence analysis plot of genres in the combined corpora

Owing to the larger number of data points, the three-corpus plot is less easy to interpret at first sight. Nevertheless, close examination reveals clear patterns.

The negative region of Axis 1 contains the plots for BA, BB, BG, BK, BL, BM, BN, BP, BR, LA, LG, LK, LL, LM, LN, LP, LR, KG, KK, KL, KN, KP, KR. In other words, all the imaginative genres K-R (with the exception of KM) fall in the negative region of Axis 1. Category G in all corpora also falls in the negative region: as we observed with the results of the Indian corpus in isolation, G (Belles Lettres) has a more literary/imaginative content than most other informative genres. The other informative genres in this region are BA and LA (two out of the three Reportage categories), and BB. All the other informative

genres fall in the positive region of Axis 1, with the very individualistic KM being the only imaginative genre in that region. Again, therefore, we see Axis 1 broadly distinguishing informative and imaginative genres, with rather little large-scale difference evident between the three national varieties.

Axis 2 on the combined corpus plot is more complex. Primarily, it distinguishes the imaginative genres of the Brown corpus (BK-BR), plus the Miscellaneous genre of all three corpora (K), from the other genres. However, BA, LA, KB, LB, and LD also appear here, as do some (but not all) imaginative genres from Kolhapur and LOB. The main reasons for this more complex picture will become clear shortly, when we examine the distribution of the modal verbs.

In two-dimensional terms, the lower left and upper left quadrants are dominated by the imaginative genres of all three corpora. The lower left is dominated by Brown, whilst the upper left contains more genres from LOB. The Indian genres hover around the horizontal origin, some in the lower quadrant, some in the upper. The upper quadrant also contains a cluster of the Belles Lettres category (G) of all three corpora. The Reportage samples from Brown and LOB (BA and LA) appear in the lower left quadrant, and Brown Editorials (BB) appear just in the upper left quadrant near the origin. The lower right quadrant is dominated by the Miscellaneous genre (H) from all three corpora. The aberrant Indian Science Fiction genre (KM) also falls in this quadrant, with three further genres (KB, LB and LD) also occurring, close to the origin. The remaining genres fall in the upper right quadrant. The Reviews samples (C) from all three corpora form a tight cluster around Axis 2 in the upper half of the plot.

Within the two-dimensional plot, a number of genres show close similarities across the three national varieties. Categories G (Belles Lettres) and C (Reviews) cluster especially closely, whilst categories F (Popular Lore), J (Learned and Scientific) and H (Miscellaneous) also form identifiable groupings, albeit less close. Other genres show greater variation in distribution. For example, the newspaper reportage samples from British and American English (LA and BA) are located quite closely together but are some distance away from the corresponding Indian sample (KA). For genre E (Skills and Hobbies) the picture is a little different: Indian and British English are quite close together, but American English is quite distant. For genre D (Religion) the picture is different again: this time Indian and American English are both situated in the upper right quadrant, but British English is separated in the lower right. Underlying reasons for some of these differences can be hypothesized, but would require further investigation. For example, in the case of Skills and Hobbies, it may be that the long symbiosis of British and Indian culture has left a stronger imprint of Britishness on the style of Indian English publications in this area.

This accounts for the distribution of the informative genres: the imaginative genres are a little more complex. As already observed, all the imaginative genres fall in the negative region of Axis 1. However, Axis 2 differentiates them. The American fictional genres (BK-BP) cluster fairly tightly in the lower left quadrant with the Humour genre (BR) located just to the right above them and forming an extended cluster. The Indian and British imaginative genres cluster less tightly together and are also interspersed with one another, beginning just above the American genres and moving well into the upper half of Axis 2. The two genres of General Fiction (K) and Romantic Fiction (P) in British and Indian English cluster especially tightly together: again, the long symbiosis of British and Indian English may underlie this mixing of the two national varieties, when set against the parallel samples from American English.

### 3.2.2 Modals

A plot showing the coordinates of the modal verbs on the same two axes is reproduced as Figure 5. A combined plot of modal verbs and genres is shown in Figure 6.

As with the Indian corpus, *could*, *would* and *might* fall in the negative (imaginative) region of Axis 1. This time they are also joined by *ought*. *Must* falls on the origin. In the positive (informative) region, the characteristic modals are *can*, *will*, *may*, *should* and *shall*.

In the negative region of Axis 2, *would*, *will*, *should* and *shall* appear, with *could*, *can*, *must*, *ought*, *might* and *may* in the positive region. The main difference here from the Indian corpus alone is the switching of positions between *should* and *could*, with *could* moving to the positive region of Axis 2 and *should* to the negative.

Looking again at the two-dimensional clustering, *shall* and *will*, along with *should*, dominate the lower right quadrant, where, apart from the aberrant KM, the Miscellaneous genre is the dominant text-type. *Shall*, characteristic of the register of regulations, exerts the strongest pull in this quadrant, which tallies with observations made by Nakamura about Brown and LOB and those made above on the Indian corpus. The lower left quadrant, where the imaginative genres of Brown are situated, is dominated by *would*, whereas *could* is most central in the upper left quadrant.

Nakamura found that *could* and *would* were good discriminators between British and American English fiction, and this finding is replicated here. It is interesting to note, therefore, the ambiguous position of the corresponding Indian genres: they are positioned more closely to British than American

English, but they do not – especially in genres L (Mystery and Detective Fiction) and N (Adventure Fiction) – show quite the same degree of attraction towards the position of *could*. Aside from the very idiosyncratic Science Fiction genre, it is notable that it is the Indian L and N genre samples which differ the most from their British counterparts (with K – General Fiction – and P – Romantic Fiction – being located very close to the corresponding British samples). These are the genres which were noted above (section 2) as being most divergent in content from LOB and Brown, with L being based on a broader interpretation of the word ‘mystery’ and N containing no samples of westerns.

**Combined Corpora: Modals**

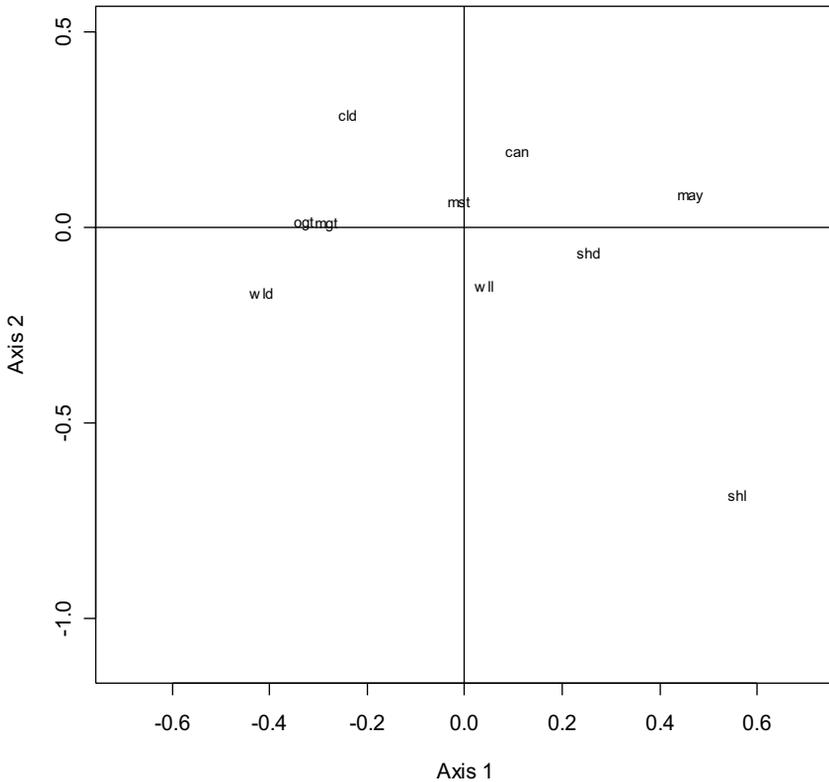


Figure 5: Correspondence analysis plot of modals in the combined corpora

**Combined Corpora: Genres and Modals**

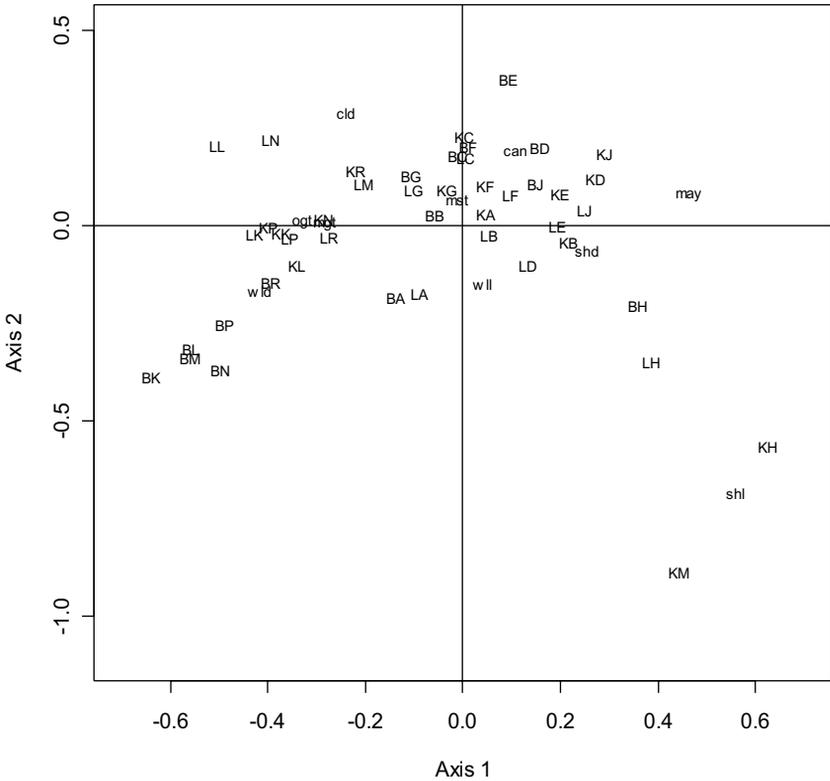


Figure 6: Correspondence analysis plot of modals and genres in the combined corpora

*May* and *can*, in the upper right quadrant, are the modals most characteristic of religious and of learned and scientific writing, with the exception of British English Religion (LD). The lower position of LD on Axis 2 is most probably due to the pull of *shall* and may possibly reflect a larger number of Biblical quotations or the presence of more archaic religious language.

**4 Conclusions**

Perhaps not surprisingly, the gross overall findings on the Indian corpus show a close similarity to those obtained by Nakamura for the Brown and LOB corpora,

especially with regard to the importance of the informative/imaginative distinction. This appears to corroborate the observation of Katikar (1984) that few systematic differences exist between the three types of English. It also seems to corroborate Kirk's (1994) claim that text-type is a more important factor in variation than regionality. It is interesting to note in this context that more formal text-types (learned writing, rules and regulations, etc.) seem to show a greater degree of homogeneity across national varieties than other text-types, such as fiction.

However, a number of interesting quantitative differences do emerge from the combined correspondence analysis, and these may deserve further investigation. For instance, in terms of text-types, it should be ascertained through wider-ranging stylistic studies, and also content analyses, how indigenous cultures and/or British and American influences may have affected the various Indian genres. This would be particularly interesting in relation to fiction, which showed an intermediate position between British and American English. With the advent of a greater number of corpora of World Englishes arising from the International Corpus of English project, their multivariate interrelationships in terms of modal verb use, and especially the positioning of their imaginative writing components with regard to the *could/would* dimension, will likewise be worthy of study.

One final factor which could be affecting the results in the present study is that of date. As observed earlier, the texts in both the Brown and the LOB corpora are sampled from the year 1961, whereas those in the Indian corpus are sampled from the year 1978. Even within two or three decades, significant changes can occur in modal verb usage, as Leech (2003) has demonstrated through comparisons of LOB and Brown with the corresponding Freiburg Frown and FLOB corpora, both sampled from the year 1991. A replication of this study using the Frown and FLOB corpora, which were not available to me at the time the data for this piece of research were extracted, may help to establish whether date is a significant factor in these results. However, it should be noted that a good decade (1978–1991) also separates the Kolhapur Corpus from Frown and FLOB, so that any such 'triangulation' findings would be suggestive rather than conclusive.

### Notes

1. See Alt (1990) for a relatively non-technical explanation of how these kinds of technique work.

2. The frequencies in this table differ from those selectively quoted by Leitner (1991). Leitner appears to have omitted at least the negated contractions (*can't*, *cannot* etc.). These should, however, be included, if only for consistency with Nakamura's Brown and LOB counts, which are based on a morphological analysis procedure that separates off enclitic forms and treats them as instances of the lemma.
3. The codes used in this paper to represent corpus genres are explained in the appendix.
4. Abbreviations used on the plots are: *cld* (*could*), *mgt* (*might*), *mst* (*must*), *ogt* (*ought*), *shd* (*should*), and *wld* (*would*).

### ***Appendix – Corpus genres and their codes***

Corpus genres are represented in this paper, especially on graphic plots, by two-letter codes, for example BC, KA, LP: the first letter in a code indicates the corpus (L = LOB, B = Brown, K = Kolhapur); the second letter represents one of the following genres:

<b>Code</b>	<b>Genre</b>
A	Press: Reportage
B	Press: Editorials
C	Press: Reviews
D	Religion
E	Skills and Hobbies
F	Popular Lore
G	Belles Lettres
H	Miscellaneous
J	Learned and Scientific
K	General Fiction
L	Detective Fiction
M	Science Fiction
N	Adventure or Western Fiction
P	Romantic Fiction
R	Humour

## References

- Alt, Mick. 1990. *Exploring hyperspace: A non-mathematical explanation of multivariate analysis*. London: McGraw-Hill.
- Coates, Jennifer. 1982. *The semantics of the modal auxiliaries*. London: Croom Helm.
- Ehrman, Madeline E. 1966. *The meaning of the modals in present-day American English*. The Hague: Mouton.
- Hofland, Knut and Stig Johansson. 1982. *Word frequencies in British and American English*. Bergen: The Norwegian Computing Centre for the Humanities.
- Katkar, P.B. 1984. The meanings of the modals in Indian English. Unpublished Ph.D. thesis. Shivaji University, Kolhapur (India).
- Kirk, John M. 1994. Using VARBRUL for studying modal auxiliary verbs? *Hermes Journal of Linguistics* 13: 71–85.
- Leech, Geoffrey. 2003. Modality on the move: The English modal auxiliaries 1961–1992. In R. Facchinetti, M. Krug and F. Palmer (eds.). *Modality in contemporary English*, 223–240. Berlin: Mouton de Gruyter.
- Leitner, Gerhard. 1991. The Kolhapur Corpus of Indian English – intra-varietal description and/or intervarectal comparison. In S. Johansson and A-B. Stenström (eds.). *English corpus linguistics: A selection of papers and research guide*, 215–232. Berlin: Mouton de Gruyter.
- McKinnon, Alastair. 1989. Mapping the dimensions of a literary corpus. *Literary and Linguistic Computing* 4(2): 73–84.
- McKinnon, Alastair. 1993. The multi-dimensional concordance: A new tool for literary research. *Computers and the Humanities* 27: 165–83.
- Mindt, Dieter. 1995. *An empirical grammar of the English verb: Modal verbs*. Berlin: Cornelsen Verlag.
- Myers, Greg. 1989. The pragmatics of politeness in scientific articles. *Applied Linguistics* 10: 1–35.
- Nakamura, Junsaku. 1993. Quantitative comparison of modals in the Brown and the LOB corpora. *ICAME Journal* 17: 29–48.
- Shastri, Srikant V. 1988. The Kolhapur Corpus of Indian English and work done on its basis so far. *ICAME Journal* 12: 15–26.

- Shastri, Srikant V., C.T. Patilkultarni and Geeta S. Shastri. 1986. *Manual of information to accompany the Kolhapur Corpus of Indian English, for use with digital computers*. Department of English, Shivaji University, Kolhapur.
- Werlich, Egon. 1983. *A text grammar of English*. 2nd edition. Heidelberg: Quelle & Meyer.
- Wikberg, Kay. 1992. Discourse category and text type classification: Procedural discourse in the Brown and LOB corpora. In G. Leitner (ed.). *New dimensions in English language corpora*, 247–61. Berlin: Mouton de Gruyter.

