

Future in Present-day English: Corpus-based evidence on the rivalry of expressions

Ylva Berglund
Uppsala University

1 Expressions of future and the scope of the study

The aim of the present paper is to provide a general, quantitative picture of how some expressions of future are used in three regional varieties of Present-day English, British, American and Indian English. The expressions studied include *will/won't+inf.*, *'ll+inf.*, *shall+inf.*, *BE(pres) going to+inf.* and *gonna+inf.* (henceforth 'expressions of future' or FUT).

will+inf.

History *will* not remember it.

(Kolhapur Corpus, K01:17)

'll+inf.

She *'ll* live, he said.

(LOB Corpus, K03:74)

shall+inf.

We *shall* still get plenty of horses.

(LOB Corpus, A36:10)

BE(pres) going to+inf.

Rhode Island *is going to* examine its Sunday sales law...

(Brown Corpus, A05:24)

gonna+inf.

'I don't think it's *gonna* last very long'

(LLC, 4.1)

Since the issue of this study is neither to argue about the origin of the contracted form 'll nor to claim, or disclaim, similarities between *will* and *shall*, all five expressions are treated as separate forms on their own. To limit the scope of this study, other ways of expressing future reference, such as the simple present or progressive, (I *go/am going* to London tomorrow) are not taken into consideration. The extralinguistic factors that have been considered are regional variation, text type/genre variation and variation brought by the medium. Apart from some observations on the meaning of *shall* and some references to the first and third person subjects, the role of the linguistic factors will be discussed in a follow-up study.

2 The data

The primary data for this study were drawn from three corpora of written English; *The Lancaster-Oslo/Bergen Corpus of Present-Day British English* (LOB), *The Standard Corpus of Present-Day American English* (Brown) and *The Kolhapur Corpus of Indian English* (Kolhapur). Additional data were collected from *The London-Lund Corpus of Spoken English* (LLC).

All instances of *will+inf.*, *'ll+inf.*, *shall+inf.*, *BE(pres)going to+inf.* (*BGT*) and *gonna+inf.* were collected by searches in the database. They were then screened for the relevant examples and these were counted manually. Comparisons were made between the corpora as wholes, between different text types within and between the corpora, and also between some of the text types combined into larger units, here referred to as *hypercategories* (A-J = Informative Prose, K-R = Imaginative Prose). The text type division follows that adopted for the Brown Corpus (see Francis and Kucèra (1982)), reflected in the LOB and Kolhapur corpora.

3 The rivalry of expressions of future (FUT)

In the following subsections I will discuss the distributions of my expressions of future in the three written corpora. Figure 1 illustrates how the number of expressions (per 2,000 words) varies across the different categories (A–R) in the corpora. Variation is considerable and seemingly irregular. The number of expressions varies across and within both categories and corpora. There is, however, no pattern indicating

that all categories in one corpus or one category in all corpora should contain the same number of FUT expressions.

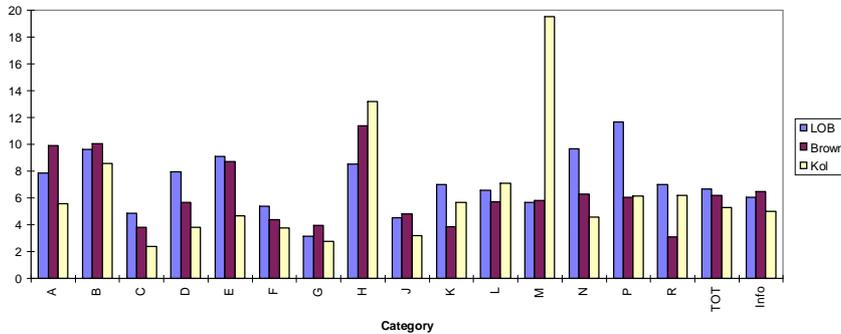


Figure 1: FUT/2,000 words per category and corpus

3.1 The LOB Corpus – British English

The overall number of FUT expressions in the LOB Corpus is 3,348, which is the highest overall number attested in this study. Variation is, however, considerable, and it is interesting to see that only in seven of the 15 categories does the British English corpus actually have the highest number of FUT constructions (see Figure 1). Differences in the *number* of FUT/text between the categories seem to indicate that neither Informative Prose (categories A-J) nor Imaginative Prose (categories K-R) are homogeneous hypercategories in this respect.

Figure 2 illustrates how the proportion of the different expressions of future varies across the categories in the LOB Corpus.

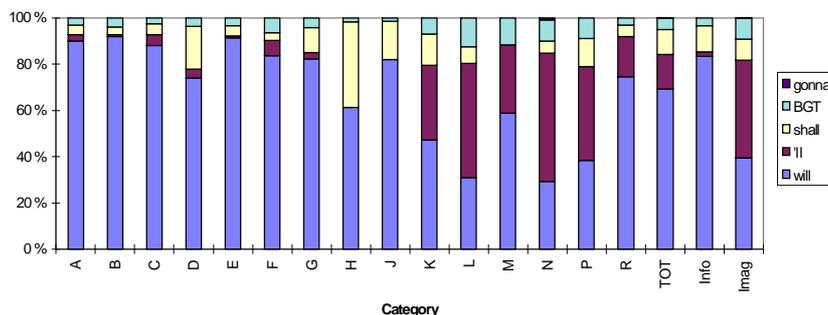


Figure 2: Proportion of different expressions of future per category in the LOB Corpus

The verb *will* is the most frequent FUT in the LOB Corpus. It is about twice as frequent in the Informative categories as in the Imaginative.¹ A very small proportion of the FUT are with *BE(pres) going to (BGT)*. There are generally somewhat more *BGT* constructions in the Imaginative categories than in the Informative (about 9% and 3% respectively). The distribution of *'ll* follows a similar pattern with more *'ll* in the Imaginative categories. Here the skewed distribution is even more obvious – only some 2 per cent of the FUT in the Informative hypercategory occur with *'ll*, to be compared to over 42 per cent *'ll* in the Imaginative Prose texts.

Since the proportion of *shall* in the Informative categories is larger than in the Imaginative, it might be assumed that *shall* is primarily used in Informative Prose. However, variation is great between the single categories and there is no clear pattern indicating a difference in frequency between the Informative and Imaginative categories. It might therefore be safest to say that, though more use is made of *shall* in Informative Prose than in Imaginative, this difference is not a hypercategory feature but derived from variation between the single categories.

3.2 The Brown Corpus – American English

The number of FUT constructions is similar in the LOB Corpus and the Brown Corpus of American English. The difference in number is

not statistically significant, which means that there seems to be no difference in frequency of these constructions between American and British English.

The number of FUT constructions per text does not vary consistently between the American and British text categories. There is no indication that the categories A-J are similar within or between the corpora. Neither can Imaginative Prose be regarded as a uniform hypercategory in either of the corpora if only the number of FUT it contains is considered.

Figure 3 illustrates the proportions of the expressions of future in the different categories of the Brown Corpus. It appears that the similarities between the LOB and Brown corpora are considerable (compare Figure 2).

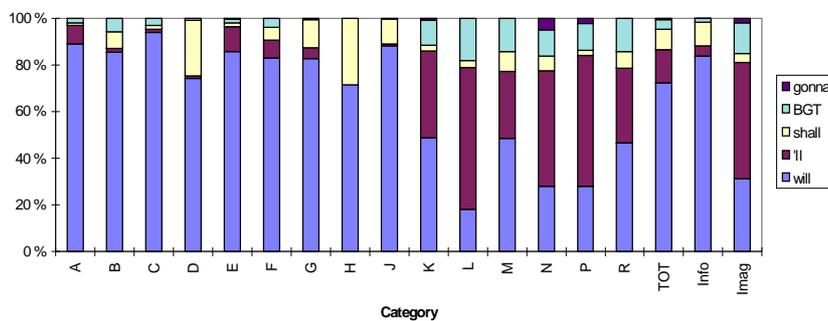


Figure 3: Proportion of different expressions of future per category in the Brown Corpus

will is the most frequent expression of future in both corpora and is used in the Informative categories to a greater extent than in the Imaginative (compare Figure 3). The proportions of *BGT* and the contracted form *'ll* are lower in the Informative than in the Imaginative categories in both corpora though the difference is more marked in the American English corpus.

Shall is used less in the Brown Corpus than in LOB. The expression is most used in the H (Miscellaneous) and D (Religion) categories in both corpora and not used much at all in the other categories.²

3.3 The Kolhapur Corpus – Indian English

The Kolhapur Corpus of Indian English as a whole contains fewer instances of expressions of future than the LOB and the Brown corpora. This seems to indicate that there is a difference between Indian English on the one hand and American and British English on the other, at least as far as the frequency of these expressions of future is concerned. Among the factors that might explain this difference are the use of alternative expressions, selection of texts or cultural factors. Perhaps it is as Shastri (1988:18) suggests: ‘Maybe the Indian mind is not given to thinking much in terms of the future ...’ Unfortunately, conclusive statements in this respect are beyond the scope the present study. What needs to be pointed out, though, is that a lower number of FUT is not a consistent feature that can be applied to all categories of Indian English texts. In four of the categories (H, K, L, M), the number of FUT per text is actually higher in the Kolhapur Corpus than in the other corpora (compare Figure 1).³

As in the British and American corpora, there is nothing in the Indian English corpus to suggest that the number of FUT should vary with text hypercategory; there is no clear difference in the number of FUT/text between the Informative Prose texts on one hand and the Imaginative Prose texts on the other.

Figure 4 illustrates how the proportion of the FUT expressions varies between the different categories in the Kolhapur Corpus.

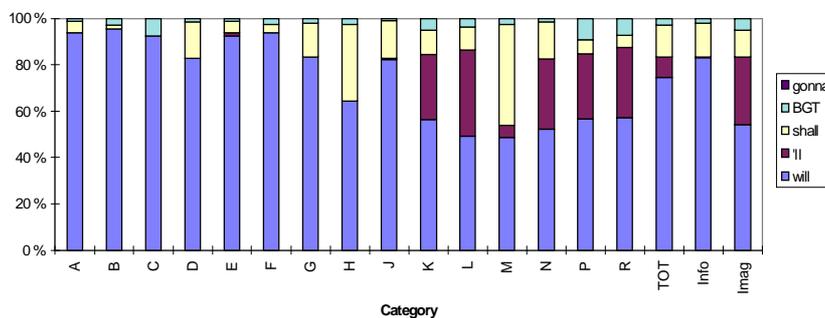


Figure 4: Proportion of different expressions of future per category in the Kolhapur Corpus

The proportion of *will* is higher in the Kolhapur Corpus than in the LOB and the Brown corpora. The expression is more frequent in the Informative categories, just as in the other two corpora. The distribution of *BGT* in the Indian English corpus shows a pattern fairly similar to that found in the British and American corpora: there is less *BGT* in the Informative categories and more in the Imaginative ones. The difference is not as great as in the other corpora, though. This can to some extent be due to the fact that the overall number of *BGT* in the Kolhapur Corpus is very low – there are only 75 occurrences altogether, 0.15 per text (LOB 0.34, Brown 0.26).

The contracted form *'ll* is also less common in the Indian English texts than in the other two corpora. In the Kolhapur Corpus, *'ll* is almost exclusively used in Imaginative Prose texts; there are only 3 examples of the contracted form in the 374 Informative texts.

Shall is used more frequently in the Indian English texts than in the British and, in particular, the American corpora seen as wholes. Category H had the highest proportion of *shall* in the LOB and the Brown corpora and the expression is frequent in that category in the Kolhapur Corpus too; about one third of the FUT in H have *shall*. The frequency of *shall* does not seem to vary between the hypercategories in the Kolhapur Corpus or in the other two corpora. In all three corpora, *shall* is mainly used with a first person subject, except in the categories H and D, where third person subjects are frequent. Quirk et al. (1985: 4.58n) claim that 'a further restricted use of *shall* with a third person subject occurs in legal or quasi-legal discourse, in stipulating regulations or legal requirements. Here *shall* is close in meaning to *must*.' It is interesting to note that category H contains a diverse selection of government documents and other formal and legal documents. The higher frequency of *shall* in category H can thus be explained by the fact that the expression does not primarily indicate future reference in these texts but is used in a modally coloured sense, for example, indicating obligation or duty:

The Commission *shall* comply with the provisions of the Administrative Procedure Act of 1946...
(Brown, H12:15)

A discussion about the extent to which *shall* with first or third person subjects expresses future and whether it should be included at all in a

study of expressions of future would undoubtedly prove fruitful, but is unfortunately outside the scope of the present study.⁴

3.4 Conclusions concerning the written corpora

The three written corpora in this study display fairly similar patterns as far as the distribution of the studied expressions is concerned. There are, however, some dissimilarities between the corpora as wholes as regards the total number of these expressions. The Kolhapur Corpus strikes us as the most different one with the lowest overall number of FUT. However, variation within that corpus suggests that expressions of future are not always infrequent in Indian English.

As far as the proportions of the expressions studied are concerned, it has been shown in this survey that the similarities between the three corpora are bigger than the differences. *will* is the most frequent expression while *BGT* is altogether very infrequent. The most significant differences found when comparing the proportions of the FUT expressions, are differences between hypercategories rather than between regional varieties. The expressions *'ll* and *BGT* are more frequent in the Imaginative hypercategories in all three corpora. The difference is more marked in the American English corpus than in the British English corpus. In the Indian corpus, the difference between the hypercategories is very large where the distribution of *'ll* is concerned and comparatively small as regards the *BGT* expression.

The expression *shall+inf.* is more common in the Indian English corpus than in the other two corpora. The distribution of *shall* does not vary with text hypercategory and thus turns out to be somewhat different from the other expressions in this study.

4 Comparisons with spoken language (LLC)

The number of FUT per text (2,000 words) is almost the same in the three written corpora in this study. Turning to the spoken London-Lund Corpus, however, we find about twice as many instances of FUT per 2,000 words there. It is difficult to find an explanation for that difference without considering the fundamental differences between the spoken and written media in greater detail. It can, however, be speculated whether expressing future is, perhaps, primarily a feature of spoken language, or whether the expressions in this study are more frequent in spoken discourse while other means of referring to the future are used in written texts.⁵

Table 1: Number of FUT per corpus* /2,000 words

	LOB	Brown	Kolhapur	LLC
will+inf. total number / per 2,000 words (% of FUT in the corpus)	2316 / 4.63 (69%)	2237 / 4.47 (73%)	1974 / 3.95 (75%)	1066 / 4.26 (35%)
'll+inf. total number / per 2,000 words. (% of FUT in the corpus)	505 / 1.01 (15%)	441 / 0.88 (14%)	230 / 0.46 (8%)	1101 / 4.40 (36%)
shall+inf. total number / per 2,000 words (% of FUT in the corpus)	363 / 0.73 (11%)	267 / 0.53 (8%)	363 / 0.73 (14%)	234 / 0.94 (8%)
BE(pres) going to +inf. total number / per 2,000 words (% of FUT in the corpus)	170 / 0.34 (5%)	130 / 0.26 (4%)	75 / 0.15 (3%)	62 / 2.52 (21%)
gonna +inf. total number / per 2,000 words (% of FUT in the corpus)	2 / 0 (0%)	14 / 0 (0%)	0 / 0 (0%)	14 / 0.06 (0%)
TOTAL total number / per 2,000 words (% of FUT in the corpus)	3348 / 6.70 (100%)	3089 / 6.18 (100%)	2642 / 5.29 (100%)	3044 / 12.18 (100%)

* LOB, Brown, Kolhapur = 1,000,000 words each
LLC = 500,000 words

If we compare the two British English corpora, the (written) LOB Corpus and the (spoken) London-Lund Corpus, we find that not only the frequency but also the proportions of the FUT attested are different. The proportion of *will* is twice as high in the written corpus as in the spoken (see Table 1). The proportion of the contracted form *'ll*, on the other hand, is considerably lower in the LOB Corpus than in the spoken corpus as a whole. The *BGT* expression is more common in the LLC than in any category in the written corpora, while *shall* is less frequent in the London-Lund Corpus than in the LOB Corpus.

A comparison between the two hypercategories of written text and the spoken data produces some interesting results. The distribution of FUT in the spoken corpus is more similar to the distribution in the written Imaginative hypercategory; there is more *'ll* and *BGT* and less *will* in the spoken data and in the Imaginative hypercategory. A closer look at the texts in the LOB Corpus points to an explanation for this. A very high proportion of the instances of *'ll* is found in contexts similar to spoken language, such as in dialogues, quotes, reported and imagined speech etc. This is the case in both Informative and Imaginative texts. Since Imaginative Prose contains more imagined speech and dialogues than Informative Prose, the *'ll* form is more frequent in the former texts. The difference in distribution of the contracted form between the Informative and Imaginative categories can thus be explained by the *medium*: *'ll* occurs to a greater extent in speech-like texts.

The proportion of *BGT* is also much higher in the spoken corpus than in the written ones. Is the use of *BGT* then mainly a trait of spoken language? If we look at the LOB Corpus, we find that most of the occurrences of *BGT* (about 58%) are actually found in reported speech (within quotation marks). However, in the Informational Prose texts, only some 18 per cent of the *BGT* are found in 'spoken' contexts. In the Imaginative hypercategory that figure is higher, 87 per cent. It seems, then, that though *BGT* indeed occurs more frequently in spoken/speech-like language overall, it is also found in purely written texts and cannot, in this study, be said to be a characteristic of spoken language only.

5 Discussion and conclusions

In this survey I have shown that the expressions of future included in the study are unevenly distributed across the different corpora and categories (see Figures 1 and 2 and Table 1). It can be concluded that as far as the *frequency* (number of occurrences/2,000 words) of FUT is

concerned, no consistent patterns of variation with the variables region, category or hypercategory can be discerned. There is, however, a clear difference between the written and spoken corpora studied. As regards the frequency of the expressions of future, then, the major variation seems to be variation according to medium: difference between written and spoken language.

As far as the *proportions* of the expressions of future are concerned, this study indicates that there is a difference between Informative Prose texts on one hand and Imaginative Prose texts on the other. Though the hypercategories are far from homogenous, it nevertheless seems to be the case that *'ll* and *BGT* are more frequent in the Imaginative categories while *will* is more frequent in Informative texts. Part of the explanation of this can be found in a comparison with spoken language. As regards the proportions of these expressions of future then, the main variation seems to be variation according to medium, written vs. spoken language, and also variation between the text hypercategories, the latter to some extent a consequence of the former.

Linguistic factors (such as subject, main verb, clause type) that might influence the choice of a certain expression of future, have largely been disregarded in this study. These factors must, however, be taken into consideration for a fuller understanding of the choice between the different expressions of future, and they will be given due attention in my forthcoming research. What this study shows is that there is indeed a difference between the occurrence of the expressions of future examined. In my opinion this difference depends primarily on medium and, as a consequence of that, is reflected as a difference between text hypercategories. I have found no evidence pointing to major, consistent variation due to regional factors.

Notes

- 1 The categories M (science fiction) and R (humour) diverge somewhat from this pattern and display a performance somewhere between the Imaginative and Informational hypercategories. It must be pointed out that those categories are the smallest ones with only 9/6 texts each, which might influence the result.
- 2 These results are also obtained by Nakamura (1993) as referred to in McEnery and Wilson (1996:82).

- 3 Note that category M in the Kolhapur Corpus is represented by two texts only. Any conclusions drawn from that data are therefore preliminary.
- 4 For discussion see, for example, Ljunggren (1893), Boyd and Thorne (1968), Wekker (1976), Close (1977), Coates (1983), Nehls (1988).
- 5 For example, the simple present (*I go to London tomorrow*), the progressive (*I am going to London tomorrow*) or other expressions, such as *I am about to go...*, *I intend to go...*, *I am to go...* etc.

References

- Boyd, Julian and J.P. Thorne. 1969. The semantics of modal verbs. *Journal of Linguistics* 5:57–74.
- Close, R.A. 1977. Some observations on the meaning and function of verb phrases having future reference. In W. Bald and R. Ilson (eds) *Studies in English Usage: The Resources of a Present-Day English Corpus for Linguistic Analysis*. Forum Linguisticum. Frankfurt/M: Peter Lang, Bern: Herbert Lang 125–156.
- Coates, Jennifer. 1983. *The semantics of the modal auxiliaries*. London & Canberra: Croom Helm.
- Francis, W. Nelson, and Henry Kučera. 1982. *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Griffin Company.
- Ljunggren, W.P.F. 1893. *The auxiliaries shall and will in the English language, especially with regard to Modern English*. Carlskrona: Länsboktryckeriet.
- McEnery Tony and Andrew Wilson. 1996. *Corpus linguistics*. Edinburgh: Edinburgh University Press.
- Nakamura, Junsaku. 1993. Quantitative comparison of modals in the Brown and the LOB corpora. *ICAME Journal* 17:29–48.
- Nehls, Dietrich. 1988. Modality and the expression of future time in English. *International Review of Applied Linguistics in Language Teaching* 26/4: 295–307.
- Quirk, Randolph, Greenbaum, Sidney, Leech, Geoffrey and Jan Svartvik. 1985. *A comprehensive grammar of the English language*. New York: Longman Inc.

- Shastri, S.V. 1988. The Kolhapur Corpus of Indian English and work done on its basis so far. *ICAME Journal* 12:15–26.
- Wekker, Herman. 1976. *The expression of future time in contemporary British English*. Amsterdam: B.V. Noord-Hollandsche Uitgeversmaatschappij.

