

22 SEPT 1981

NAVF's EDB - senter for  
humanistisk forskning

# ICAME NEWS

Newsletter of the International Computer  
Archive of Modern English (ICAME)

Published by: The Norwegian Computing Centre for the Humanities, Bergen  
The Norwegian Research Council for Science and the Humanities



Machine-readable  
texts in  
English language  
research

**No. 5**  
Nov. 1981



## CONTENTS

Word Frequencies in Different Types of English Texts	
<i>Stig Johansson</i>	1
The s-Genitive with Non-Personal Nouns in British and American English	
<i>Mette-Cathrine Jahr</i>	14
Shall, Will, Should, and Would in British and American English	
<i>Inger Krogvig and Stig Johansson</i>	32
ICAME Projects	57
Material Available from Bergen	62

Editor: Dr. Stig Johansson, Department of English,  
University of Oslo, Norway.



## WORD FREQUENCIES IN DIFFERENT TYPES OF ENGLISH TEXTS

*Stig Johansson*  
University of Oslo, Norway

### INTRODUCTION

One of the main advantages of the LOB Corpus, as of its American counterpart, is that its composition makes possible a comparison of the characteristics of different types of texts. Kučera and Francis (1967:275-93) give the distribution of the 100 most frequent words in the fifteen text categories of the Brown Corpus. Frequencies are shown to vary considerably even with these very common words. Detailed observations on word frequencies in the different text categories of the LOB Corpus are reported in Hofland and Johansson (forthcoming). This paper presents some information from the book.

### MAJOR CATEGORY GROUPS

A study of word frequencies can be used to reveal the relationship between different types of texts. Rank correlations were computed for the distribution of the 39 most frequent words in the text categories of the LOB Corpus (Table 1). The results are presented in Table 2 and Figures 1-2. Though the correlations are high in general, we find distinct groupings of categories. There is a major division between informative (categories A-J) and imaginative prose (categories K-R), with some categories of 'essayistic' prose bridging the gap (G, M, R).

The study of rank correlations, combined with more subjective criteria, made us divide the fifteen categories of the Corpus into four groups:

A-C (88 texts): newspaper text  
D-H (206 texts): miscellaneous informative prose  
J (80 texts): learned and scientific English  
K-R (126 texts): fiction

Word frequencies in the four category groups were examined in detail. Table 3 gives some examples of frequency differences for

groups of words defined by grammatical or semantic criteria.

The relationship between the four category groups revealed a fairly consistent pattern, with K-R and J appearing as the extreme 'poles'. This was brought very clearly in a study of the frequencies of 35 prepositions. The category groups were ranked 1 to 4 for each word, and an average rank difference was calculated. The results are given in Figure 3.

#### SOME DIFFERENCES BETWEEN CATEGORIES J. AND K-R

As categories J and K-R seem to form the extreme 'poles', they may be especially interesting to compare. Tables 4 and 5 give the nouns, lexical verbs, adjectives, and adverbs with the highest 'distinctiveness coefficient' in the two category groups (for the calculation of the coefficient, see the forthcoming book by Hofland and Johanson). The differences are so clear that they need no comment. A study of the types of words with the highest distinctiveness value is also revealing. The degree of distinctiveness varies with grammatical class, as shown by a comparison of the 100 words with the highest distinctiveness coefficient in the two category groups:

	J	K-R
nouns (except proper names and abbreviations)	58	23
lexical verbs	1	31
adjectives	12	2
adverbs	0	4
others <sup>1</sup>	<u>29</u>	<u>40</u>
	100	100

The most distinctive forms in category J are thus nouns and adjectives, while lexical verbs predominate in K-R, which even contains a sprinkling of adverbs. These figures can be interpreted as a reflection of contrasting stylistic features, in particular nominal vs. verbal style.

#### SOME DIFFERENCES BETWEEN INFORMATIVE AND IMAGINATIVE PROSE

The final example to be taken up in this brief presentation is a comparison of the relative frequency of the hundred most frequent words (in the Corpus as a whole) in text categories A-J vs. K, L, N, and P (see Table 6).<sup>2</sup> The number of words where there is a

consistent difference between the two category groups is surprisingly high. The differences reflect important grammatical and stylistic characteristics of the texts, such as differences in tense choice (*is* vs. *was*, *has* vs. *had*, etc.), use of the passive (*by*), relative clauses (*which*), and postmodification of nouns (*of*). The personal pronoun frequencies partially reflect the degree of 'personality' of style (or, simply, the proportion of dialogue), partially differences in subject matter, while differences in article frequency are an indication of 'nouniness' (*the*) or the proportion of Latinate vocabulary (*an*).<sup>3</sup>

#### CONCLUDING REMARKS

As shown above, there are important differences in word frequency between different types of English texts. Yet this is an area which has been very poorly studied. The forthcoming book by Hofland and Johansson includes some general discussion of these matters as well as further detailed information on word frequencies in the text categories of the LOB Corpus and a comparison with other corpora, in particular the Brown Corpus.<sup>4</sup>

Table 1 The text categories of the LOB Corpus

	Number of texts in each category
A Press: reportage	44
B Press: editorial	27
C Press: reviews	17
D Religion	17
E Skills, trades, and hobbies	38
F Popular lore	44
G Belles lettres, biography, essays	77
H Government documents etc.	30
J Learned and scientific writings	80
K General fiction	29
L Mystery and detective fiction	24
M Science fiction	6
N Adventure stories	29
P Romance and love story	29
Humour	9
Total	500



Table 2 Rank correlations between text categories based on the rank numbers of the 89 most frequent words in the LOB Corpus.

	A	B	C	D	E	F	G	H	J	K	L	M	N	P	R	Total
A	1.00															
B	.83	1.00														
C	.84	.84	1.00													
D	.76	.90	.79	1.00												
E	.75	.88	.77	.89	1.00											
F	.81	.82	.86	.86	.86	1.00										
G	.79	.82	.83	.87	.79	.90	1.00									
H	.74	.88	.71	.85	.88	.78	.75	1.00								
J	.74	.87	.76	.89	.90	.85	.82	.90	1.00							
K	.64	.50	.56	.60	.51	.72	.76	.38	.45	1.00						
L	.62	.44	.52	.53	.47	.67	.70	.33	.42	.94	1.00					
M	.75	.68	.69	.77	.70	.83	.83	.60	.65	.88	.82	1.00				
N	.59	.41	.50	.50	.43	.64	.66	.28	.36	.95	.96	.81	1.00			
P	.61	.45	.52	.53	.48	.67	.68	.32	.39	.97	.96	.84	.96	1.00		
R	.77	.70	.72	.78	.74	.85	.88	.64	.65	.87	.80	.89	.80	.82	1.00	
Total	.86	.80	.83	.85	.82	.94	.92	.73	.78	.85	.80	.90	.78	.81	.92	1.00

Table 3 The relative frequencies (in words per million) of some groups of words in categories A-C, D-H, J, and K-R. The highest value for each item is given in *italics*.

	A-C	D-H	J	K-R
the	67,368	71,510	<i>81,239</i>	53,649
a	<i>23,567</i>	22,631	19,940	23,518
an	3,814	3,581	<i>3,831</i>	2,709
and	24,865	<i>29,351</i>	25,513	27,594
but	4,711	4,759	3,701	<i>6,113</i>
or	2,234	<i>5,062</i>	4,507	2,195
although	366	402	<i>607</i>	229
though	434	592	446	<i>900</i>
by	6,691	6,276	<i>7,986</i>	2,827
of	33,678	39,617	<i>48,878</i>	21,390
can	1,579	1,861	<i>2,008</i>	1,176
could	1,365	1,192	892	<i>2,894</i>
may	1,100	1,758	<i>2,443</i>	505
might	536	732	713	<i>1,054</i>
I	3,650	5,113	1,202	<i>14,739</i>
you	1,111	1,955	303	<i>10,000</i>
he	8,090	5,450	3,199	<i>18,017</i>
she	1,687	1,657	260	<i>11,386</i>
it	8,993	9,309	7,794	<i>13,009</i>
we	2,042	<i>3,449</i>	2,802	2,685
they	3,757	3,873	1,866	<i>3,965</i>
this	5,055	5,742	<i>7,037</i>	3,443
that	10,748	11,074	11,142	<i>11,406</i>
these	1,303	1,828	<i>2,263</i>	576
those	840	<i>1,182</i>	1,091	556
also	1,128	1,112	<i>1,593</i>	296
too	857	809	527	<i>1,382</i>
maybe	28	48	6	211
perhaps	304	356	310	<i>608</i>
possibly	73	86	99	110
big	<i>355</i>	103	49	252
great	778	<i>888</i>	558	414
large	253	505	<i>682</i>	276
fairly	56	<i>113</i>	99	43
quite	327	484	347	<i>687</i>
rather	287	387	<i>508</i>	430
so	1,828	2,355	2,089	<i>3,056</i>
such	908	1,295	<i>1,612</i>	655
thus	146	267	<i>675</i>	43
believe	203	183	55	<i>284</i>
think	355	416	161	<i>1,204</i>
appear	107	158	<i>198</i>	55
seem	220	255	<i>297</i>	209

Table 4 Plus-words in categories J vs. K-R: nouns and lexical verbs. The words are listed in the order of their distinctiveness coefficient.

Nouns		Verbs	
J	K-R	J	K-R
constants	mister	measured	kissed
axis	sofa	assuming	heaved
equations	wallet	calculated	leaned
oxides	cheek	occurs	glanced
equation	living-room	assigned	smiled
theorem	café	emphasized	hesitated
coefficient	wrist	obtained	exclaimed
ions	darling	executed	murmured
correlation	sigh	tested	gaped
electrons	gun	corresponding	hurried
impurities	gaze	vary	flushed
oxidation	clip	bending	cried
parameters	fist	varying	eyed
nickel	trail	loading	staring
electron	lounge	measuring	paused
impurity	cheeks	determine	whispered
diagram	lips	isolated	waved
ion	cigarette	dissolved	nodded
parameter	stairs	resulting	frowned
coefficients	footsteps	defined	shivered
oxygen	dad	occur	muttered
sodium	lawn	stressed	stared
equilibrium	receiver	illustrates	flung
oxide	madam	recognized	grinned
variable	jacket	identified	laughed
evaporation	fool	testing	shrugged
contamination	pistol	follows	jerked
approximation	envelope	observed	tapping
alloy	shoulders	tend	laughing
hydrogen	door	demonstrated	swung
ratios	forehead	exposed	pretended
data	phone	containing	leaning
component	knees	deposited	wondered
symmetry	tears	using	shook
curve	bedroom	forming	kiss
displacement	fingers	indicates	straightened
computer	patch	examine	rang
cells	skirt	associated	sounded
curves	eyes	indicate	gripped
particle	pocket	obtain	smiling

Table 5 Plus-words in categories J vs. K-R: adjectives and adverbs.  
The words are listed in the order of their distinctiveness coefficient.

Adjectives		Adverbs	
J	K-R	J	K-R
thermal	damned	theoretically	impatiently
linear	asleep	significantly	softly
radioactive	sorry	approximately	hastily
structural	gay	hence	nervously
finite	miserable	relatively	upstairs
transient	dear	respectively	faintly
physiological	silly	commonly	quietly
numerical	empty	separately	abruptly
magnetic	stiff	consequently	eagerly
conceptual	dreadful	similarly	upright
residual	afraid	rapidly	tomorrow
differential	deadly	thus	downstairs
stationary	sweet	furthermore	gently
statistical	ashamed	sufficiently	anyway
negative	lovely	therefore	maybe
relative	faint	secondly	swiftly
experimental	calm	ultimately	presently
theoretical	silent	readily	suddenly
integral	nice	effectively	somewhere
mechanical	funny	generally	back
chemical	worried	widely	slowly
internal	tired	strictly	desperately
initial	stupid	mainly	sharply
reliable	polite	directly	away
significant	savage	partly	barely
continuous	quiet	previously	backwards
relevant	tall	specifically	somehow
prior	lonely	chiefly	utterly
intermediate	glad	presumably	aboard
liquid	damp	closely	down
equal	dark	accordingly	lightly
rapid	mad	frequently	quickly
constant	pretty	however	inside
imperial	quick	moreover	carefully
consistent	pink	nevertheless	again
positive	clean	unfortunately	off
upper	sudden	briefly	then
aesthetic	desperate	considerably	never
statutory	loud	purely	sooner
external	ugly	originally	scarcely

Table 6 A comparison of the relative frequency of the 100 most frequent words in text categories A-J vs. K, L, N, and P. For words given within parentheses all the text categories in a group show higher values than the mean relative frequency of the other category group taken as a whole; for the others all the individual values in one of the groups are higher than those of the other group.

Type of words	Categories A-J consistently higher frequency	Categories K, L, N, P consistently higher frequency
Articles	the (an)	
Pronouns and quantifiers	this these many most (those) (its) (our) (their) (more) (some) (such)	I me my you he him his she her it (all) (no)
Prepositions	by from in of (for)	about before into (after) (at) (over)
Conjunctions	than	(but) (if)
WH-words	which	what (when) (where)
Adverbs	also	now out then up (even) (there) (well)
Auxiliaries	are is has may (being) (will)	could do had was (would)
Others	new years (two)	like said (man) (time)

Figure 1 Rank correlations between text categories (cf. Table 2)

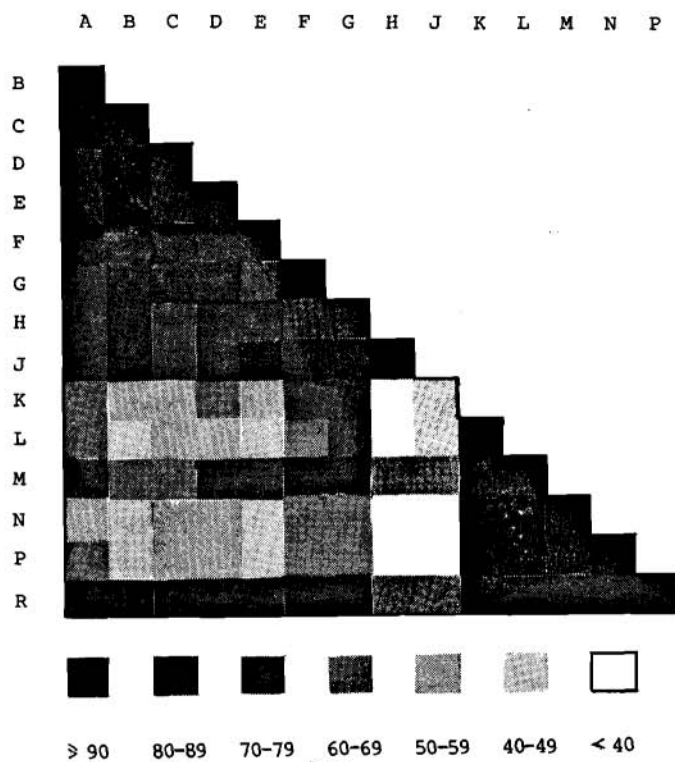


Figure 2 Rank correlations between text categories (cf. Table 2)

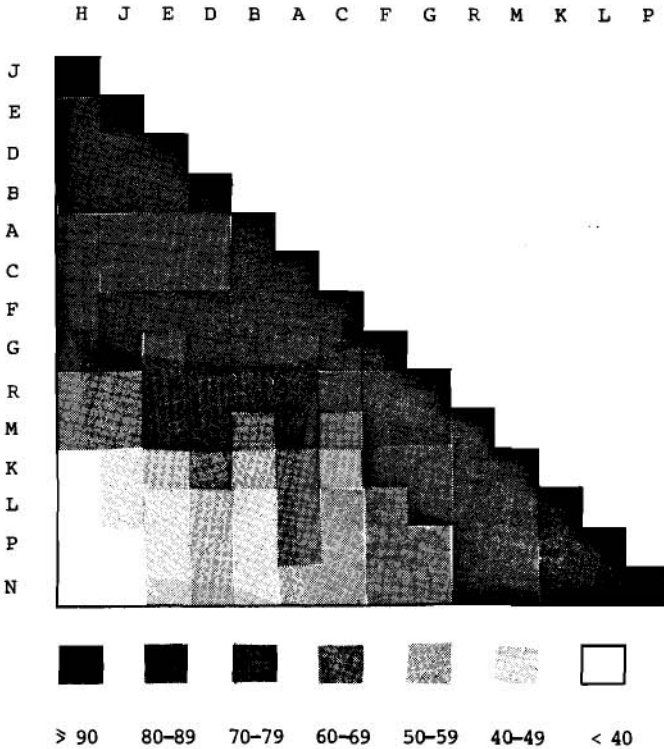
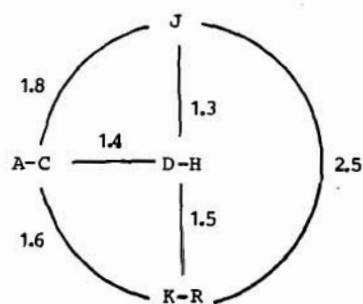


Figure 3 The relationship between category groups





#### NOTES

- 1 The forms categorized as 'others' in category J were mostly abbreviations, scientific symbols, numerical expressions, and letters, while those in K-R were proper names, contractions, interjections, and non-standard forms.
- 2 In this table we have excluded categories M and R from the fiction group in order to sharpen the contrast between informative and imaginative prose. A special reason for excluding these categories is that they are the smallest in the Corpus (6 and 9 text samples, respectively). They are therefore especially sensitive to sampling error.
- 3 These findings agree very well with those reported for the Brown Corpus in Johansson (1978:34f.).
- 4 Note, in conclusion, that all the figures presented in this paper and in the forthcoming book by Hofland and Johansson are based on *graphic words*. Lemmatized lists are in preparation. Where forms are classified according to grammatical function (as in Tables 4-6), we refer to the main function of the form. In cases of doubt we have inspected the concordance of the LOB Corpus.

#### REFERENCES

- Hofland, Knut and Stig Johansson. (forthcoming). *Word Frequencies in Present-Day British and American English*. Bergen: Norwegian Computing Centre for the Humanities.
- Johansson, Stig. 1978. *Some Aspects of the Vocabulary of Learned and Scientific English*. Gothenburg Studies in English, 42. Gothenburg: Acta Universitatis Gothoburgensis.
- Kučera, Henry and W. Nelson Francis. 1967. *Computational Analysis of Present-Day American English*. Providence, Rhode Island: Brown University Press.

# THE S-GENITIVE WITH NON-PERSONAL NOUNS IN PRESENT-DAY BRITISH AND AMERICAN ENGLISH

*Mette-Cathrine Jahr*

University of Oslo, Norway

## INTRODUCTION

For centuries there has been a rivalry between the inflected genitive (the *s*-genitive) and its prepositional equivalent (the *of*-construction). The *s*-genitive has gradually had to yield to the *of*-construction until, by the beginning of the twentieth century, the use of the inflected genitive had been so restricted that it was mainly found with personal or personified nouns, apart from certain idioms and adverbial expressions of measure, time, and space. As early as 1920, however, Zachrisson (1920:39f.) claimed that there is a 'growing tendency towards a more extensive use of the genitive in *s*' in present-day English. Several later writers have supported Zachrisson, among them Potter (1969:105f.): 'Until recently ... we inclined to limit inflected genitives to animate objects. ... Today, however, this distinction between animate and inanimate nouns is slowly disappearing.'

## AIM

The present paper reports some results from my thesis on the *s*-genitive in present-day English (Sørheim 1980). In the first part of the thesis I looked at the frequency of *s*-genitives with non-personal nouns to see if Zachrisson's and Potter's observation holds good, viz. that the use of the inflected genitive has increased and expanded. My aim has been to examine what kinds of nouns, other than those denoting human beings, occur with the *s*-genitive in present-day English. At the same time I wanted to see if there is a difference between British and American English in the use of the non-personal inflected genitive, and between different genres of written English.

## MATERIAL

The material for the investigation was drawn from the Lancaster-

Oslo/Bergen Corpus of British English (the LOB Corpus), since this offers a representative material gathered from 15 different genres or text categories, and since it is available in machine-readable form. All forms containing 's or s' were extracted from the corpus with the aid of the computer. Having excluded contracted forms (e.g. *that man's* in the sense *that man is*), other non-genitives (e.g. *O'Sullivan*), and genitives of non-nouns (e.g. *somebody else's*), I was left with 4,857 genitives of nouns. 1,191 (or 24.5%) of these were s-genitives of non-personal nouns.

The comparison between British and American English was based on my study of the LOB Corpus and a previous investigation of the Brown Corpus by Ingrid Aronsson (1975).

#### CLASSIFICATION

To be able to compare my material with Aronsson's (1975), it was necessary to follow her system of classification (first used by Liisa Dahl in 1971), in which the material is divided into 15 classes according to the meaning of the inflected noun.<sup>1</sup> I shall concentrate on possible extensions in the use of non-personal s-genitives (compared with the established use mentioned in most grammar books) giving one or two examples from each class/subclass, and - where relevant - adding brief comments on the relationship between British and American English usage.

#### PRESENTATION

Under each class/subclass I shall give some information on the total number of instances in the relevant class/subclass and the number of different non-personal nouns found with the s-genitive in the corpus (the nouns are listed in full in the Appendix), as well as state how many text categories are represented. Corresponding figures for the Brown Corpus will also be given (quoted from Aronsson 1975), e.g.

Class V (names of animals):

Total: 52 instances / BROWN: 5  
40 different words / BROWN: 5  
14 categories represented / BROWN: 5

All examples will be given an identification code, e.g.  
the *Government's* decision (B15:42)

which means that the example was found in text category B (Press: editorial), text sample number 15, line 42. The inflected nouns will be italicized.

Summing up the development, I shall also comment briefly on how the use of non-personal *s*-genitives differs from one category or category group to another.<sup>2</sup>

#### I NOUNS DENOTING COLLECTIVE COMMUNITIES

##### a) *Authoritative and other organized bodies*<sup>3</sup>

These nouns have strong human associations. The aspect of individuals making up a group is emphasized.

1) *Authoritative bodies*, i.e. 'bodies making decisions or having some power of control over people' (Dahl 1971:143).

Total: 109 instances / BROWN: 104  
15 different words / BROWN: 27  
10 categories represented / -

Examples: the *Committee's* hats and coats (K01:71)  
the *Government's* decision (B15:42)

##### 2) *Nouns denoting other organized bodies*

Total: 154 instances / BROWN: 221  
47 different words / BROWN: 64  
15 categories represented / -

Examples: *Nato's* military planning committee (B06:60)  
the *Women's International Art Club's* exhibition (C15:175)

##### b) *The complete or shortened name of companies or comparable formations*

Total: 43 instances / BROWN: 43  
35 different words / BROWN: 24  
8 categories represented / -

Examples: at *Boots'* branches (E05:191)  
*BBC's* 'the Little Key' (C04:231)

Subgroup 1b holds a unique position within Class I in that the nouns are proper names of some sort, and as such come very close to proper names of persons. But it is the economic aspect of the firm which is emphasized, not the personal.

##### c) *Nouns which do not primarily denote human beings*

Total: 66 instances / BROWN: 81  
34 different words / BROWN: 31  
12 categories represented / -

Examples: the *Administration's* position (J43:146)  
the *Bank's* money (A06:210)

d) *Group-genitives*

Total: 2 instances / BROWN: 7  
2 different words / BROWN: 7  
2 categories represented / -

Examples: *the Council of Local Authorities' Services* (F43:9)  
*the U.K. Ministry of Aviation's decision* (A15:13)

The only examples included in Id are those where the head noun of the group belongs to Class I.

The use of the *s*-genitive is said to be very common with nouns belonging to Class I when the idea of a group of individual persons is emphasized (Ia). In the present material the *s*-genitive is found quite frequently also when the notion of individual persons in the group is faint or not felt at all (Ib and Ic).

II + III GEOGRAPHICAL PROPER NAMES AND COMMON NOUNS <sup>4</sup>

a) *Political or sociological meaning emphasized*

Total Class IIa:	Total Class IIIa:
191 instances / BROWN: 179	59 instances / BROWN: 77
61 different words / BROWN: 77	10 different words / BROWN: 10
11 categories represented / -	10 categories represented / -

Examples: *Britain's team* (E17:58)  
*the town's reactions* (C16:12)

b) *Purely geographical meaning emphasized* <sup>5</sup>

Total Class IIb:	Total Class IIb:
48 instances / BROWN: 69	15 instances / BROWN: 24
37 different words / BROWN: 49	6 different words / BROWN: 11
11 categories represented / -	7 categories represented / -

Examples: *India's soil* (D15:161)  
*their country's coastline* (F22:193)

c) *Names or nouns without a distinction between political/sociological and geographical meaning*

Total Class IIc: <sup>6</sup>	Total Class IIc:
3 instances / -	10 instances / BROWN: 7
2 different words / -	6 different words / BROWN: 5
2 categories represented / -	6 categories represented / -

Examples: *the Adriatic's most benign month* (K22:97)  
*the desert's flat surface* (N20:210)

d) *Geographical names used to denote football clubs etc.*

Total Class IId:  
27 instances / -  
20 different words / -  
1 category represented / -

Example: to *Forfar's* credit (A41:231)

It is commonly agreed that when geographical nouns denote political or sociological units, i.e. when they function as *collective nouns*, the *s*-genitive is established and frequent (Classes IIa and IIIa). In my material, however, the inflected genitive form was found relatively often also when the purely geographical aspect of these nouns was emphasized (IIb and IIIb) and sparingly even with nouns which do not distinguish between political/sociological and geographical meaning (IIc and IIIc). The *s*-genitive with Class IIb and IIIb nouns seems to be somewhat more freely used in the Brown Corpus than in the LOB Corpus.

#### V NAMES OF ANIMALS <sup>7</sup>

Total: 52 instances / BROWN: 5  
40 different words / BROWN: 5  
14 categories represented / BROWN: 5

Examples: 'The *Lion's* Mantle' (A32:149)  
the *snake's* rear (R19:179)  
the *bug's* tendency to turn deep purple (F06:39)

The use of the *s*-genitive with higher animals (e.g. *lion*) is of long standing. The LOB Corpus contains a fairly large number of *s*-genitives used with names of animals ranging from *elephant* to *bug*.

#### VI NOUNS DENOTING MEANS OF LOCOMOTION AND MACHINES

Total: 40 instances / BROWN: 31  
26 different words / BROWN: 18  
10 categories represented / BROWN: 10

Examples: the *boat's* prow (K12:105)  
the *plane's* doors (A28:182)  
the *pump's* capacity (E27:108)

The *s*-genitive with *ship*, *boat* and *vessel* is mentioned by most grammarians, and the traditional use is also predominant in the LOB material. More than one third of the examples in Class VI are expressions with *ship* and *boat*. Including synonyms and proper names with comparable reference, the figure is 67.5% of all the occurrences in Class VI in the LOB Corpus. In the Brown Corpus, on the other hand, the traditional use with *ship/boat* occurs only 5 times out of 31, and even if we include synonyms for these nouns and proper names of ships or boats, the result is only 9 instances out of 31, or 29.0%. Nouns denoting different kinds of machines, however, account for about one third of the total number of instances in Class VI in

the Brown Corpus, whereas in the LOB Corpus such nouns occur three times only (7.5%).

#### VII THE SUN, THE PLANETS, THE STARS, AND OTHER HEAVENLY BODIES

Total: 10 instances / BROWN: 24  
5 different words / BROWN: 5  
4 categories represented / BROWN: 9

Examples: the *sun's* rays (N07:207)  
the *comet's* tail (J02:149)

With nouns in this class the use of the *s*-genitive is established and has frequently been noted by grammarians.

#### VIII NOUNS DENOTING BUILDINGS AND LOCALITIES

Total: 14 instances / BROWN: 28  
13 different words / BROWN: 20  
7 categories represented / BROWN: 8

Examples: the *cinema's* advertisement (C02:114)  
the *stable's* calm (G26:5)

Nouns like *church*, *university*, *school*, etc. denote organized bodies (Class I) as well as the buildings in which these bodies meet and work (Class VIII). In the former case the *s*-genitive is traditionally used when the human aspect is emphasized (Ia), in the latter case the inflected genitive is slowly beginning to gain ground and spread to nouns which unambiguously denote buildings or localities, e.g. *saloon*, *rectangle* (used in the sense of a *nave* in a church). American English usage may have had some influence on British English with nouns in this class.

#### IX NEWSPAPERS AND PERIODICALS

Total: 9 instances / BROWN: 9  
6 different words / BROWN: 9  
5 categories represented / BROWN: 5

Examples: the *magazine's* editor (F36:69)  
the *London Observer's* science fiction contest (G36:154)

As in Classes II, III, and VIII, the use of the *s*-genitive may be ascribed to an extension of the established use of the inflected genitive with nouns denoting organized bodies (Class I). From e.g. *newspaper* used as a collective noun, meaning 'editing staff' etc., the use of the *s*-genitive seems to have been extended to the same noun used for the actual publication.

#### X ABSTRACT NOUNS

Total: 40 instances / BROWN: 76  
28 different words / BROWN: 44  
10 categories represented / BROWN: 14

Examples: *Death's* kingdom (J62:170)  
the *dream's* warning (F12:170)

Zachrisson (1920:40f.) states that the use of the inflected genitive with abstract nouns must be regarded as exceptional.

There is a considerable difference between the number of instances in Class X in the two corpora. Apart from the established use of the *s*-genitive with personifications of abstract nouns illustrated in the first example above, it was found also when there was no idea of personification attached to the noun. This extended use seems to be gaining ground faster in American English than in British English, and American English influence on British English journalistic style cannot be discounted.

#### XI CURRENCIES

There are no instances of the *s*-genitive occurring with names of currencies in the two corpora.

#### XII MATERIAL NOUNS AND CONCRETE THINGS

Total: 20 instances / BROWN: 50  
17 different words / BROWN: 28  
6 categories represented / BROWN: 9

Examples: the *bed's* occupant (F06:61)  
the *bullet's* exit point (L03:96)

Both Zachrisson (1920:42-45) and Jespersen (1949:327) claim that the use of the *s*-genitive with Class XII nouns is growing, and that this construction is gaining ground in the works of younger writers and in journalism. Old English idioms or the use of *the book* referring to its author have been given as the source of the use of the *s*-genitive with concrete nouns. The inflected genitive is used more freely in the Brown Corpus than in LOB.

#### XIII IDIOMATIC EXPRESSIONS <sup>8</sup>

Total: 35 instances / BROWN: 16  
23 different words / BROWN: 11  
10 categories represented / BROWN: 10



Examples: for *goodness's* sake (N02:89)  
           for *photography's* sake (N15:129)  
           *razor's* edge (A21:116)  
           his *wit's* end (K13:21)

Most of these idiomatic expressions are of very long standing, but here too, we find expansion, probably by analogy. In the second example above, for instance, the comparatively modern *photography* has been put into the old idiomatic frame *for - sake*.

#### XIV EXPRESSIONS OF TIME AND MEASURE <sup>9</sup>

##### a) *Expressions of time*

Examples: *today's* distance (A32:169)  
           the *morning's* paper (K19:158)

##### b) *Expressions of measure*

Examples: a *day's* work (F18:5)  
           a modest *half-crown's* worth (E38:184)

Total: 244 instances / BROWN: 197  
 (a+b) 39 different words / BROWN: 37  
       15 categories represented / BROWN: 14

There is a high degree of conformity as to the types of nouns occurring in Class XIV: 26 nouns (plural forms included) are the same in both corpora. This fact, together with the distribution in all 15 text categories (BROWN: 14), indicates an old and well-established use of the *s*-genitive.

#### DISCUSSION

As this brief survey of my material shows, the use of the *s*-genitive with non-personal nouns is quite extensive and seems to have been extended beyond the established uses noted in most grammar books. Zachrisson (1920:45f.) suggests a development by analogy from nouns which traditionally take the *s*-genitive when used to denote a group of individual persons (i.e. used collectively, e.g. in Classes Ia, IIa, and IIIa), through the same nouns used in a purely non-personal sense (e.g. Classes IIb, IIIb, and VIII), to related nouns which do not have any human associations (e.g. Classes Ic, IIc, IIIC, and VIII). In my view this is a very plausible explanation, illustrated by e.g. *church* (= congregation, Class Ia2) → *church* (= the building in which the congregation meets, Class VIII) → *chapel* (= the building only, Class VIII). A comparison of the LOB Corpus and the Brown Corpus reveals a somewhat freer use of

inflected non-personal genitives in American English than in British English, and the possibility of American English influence on British English (especially in newspaper language) cannot be ruled out. These results agree with the observation by Kirchner (1970:114), who says about the increase in the use of the *s*-genitive with inanimate nouns in British English that 'Vielleicht ist diese rapide Zunahme auf den Einfluss des zeitgenössischen AE. zurückzuführen'.

Although there are some differences between the British and American corpora, the most notable thing is the high degree of agreement. Table 1 below shows that the frequency of the different classes (I-XV) varies considerably, and in a similar way in the two corpora. (cf. also Fig. 1 below, which gives information on all the individual categories of the Corpus). There are marked differences between the various category groups *within* both LOB and Brown, ranging from the very high figures in the newspaper categories (A-C) to the low values for Religion (category D) and Fiction (categories K-R) where non-personal *s*-genitives are rarely used and chiefly occur in adverbial expressions of time and measure (Class XIV). The only notable differences *between* the two corpora are found in categories C and H. In C the discrepancy may be due to stylistic differences, i.e. a more journalistic style in American English newspaper reviews compared with British English, while in H (mainly Government documents) the reason is simply that identical expressions recur repeatedly within the same text samples in the Brown Corpus and cause a high *s*-genitive frequency.

Note, in conclusion, that a frequency study of the *s*-genitive alone gives a biased account. This is a limitation which applies to all previous studies of *s*-genitive frequencies. A comparative study of the *s*-genitive and the construction it competes with, viz. the *of*-construction, is absolutely necessary. In the latter part of my thesis, I examined all instances of 48 nouns representing personal nouns and the different non-personal classes taken up above, all occurring with the *s*-genitive as well as the *of*-construction in the LOB Corpus. A comparison of the two constructions showed that e.g. in Class I, where the *s*-genitive is said to be common and seems to be very frequent (cf. Table 1), the inflected genitive was preferred in only 24.3% of the examples (Fig.2), while in Classes IIb

and IIIB, where the *of*-construction has been regarded as the only 'possible' choice, *s*-genitives account for 23.1% and 34.4% respectively, i.e. roughly the same percentage and even higher than for Class I. The results must, however, be treated with some caution, as they are based on a very limited material.

Another fact which has been ignored by investigators doing frequency studies of the *s*-genitive alone is that the type of modifying noun is only one of the factors influencing the choice of genitive construction. When the border-line between personal and non-personal nouns is no longer closely observed, other factors increase in importance, e.g. style and thematic considerations. These were dealt with in my thesis but cannot be taken up in this brief presentation.<sup>10</sup>

Table 1 The absolute and relative frequency of non-personal *s*-genitives in all classes and all category groups in LOB (L) and Brown (B) and the average number of *s*-genitives per text in all category groups in both corpora.

Categories			ABC	D	EFG	H	J	K-R	TOTAL	Frequency in each class in % of all instances in the Corpus
Number of texts		B L	88	17	159	30	80	126	500	
NUMBER OF INSTANCES IN EACH CLASS	I	B	194	7	83	114	47	11	456	36.4
		L	165	8	89	50	40	22	374	31.4
	II	B	139	4	49	12	22	22	248	19.8
		L	156	4	79	7	13	10	269	22.6
	III	B	44	2	34	12	6	10	108	8.6
		L	30	-	35	-	3	16	84	7.0
	V	B	1	-	1	-	-	3	5	0.4
		L	10	1	24	-	2	15	52	4.4
	VI	B	6	-	11	-	-	14	31	2.5
		L	4	-	22	1	3	10	40	3.4
	VII	B	5	-	9	-	7	3	24	1.9
		L	1	-	-	-	6	3	10	0.8
	VIII	B	8	-	8	-	-	12	28	2.2
		L	2	2	6	-	-	4	14	1.2
	IX	B	1	1	5	-	2	-	9	0.7
		L	6	-	3	-	-	-	9	0.7
	X	B	30	4	28	1	6	7	76	6.1
		L	12	2	16	-	7	3	40	3.4
	XII	B	5	-	24	-	10	11	50	4.0
		L	2	-	10	-	2	6	20	1.7
	XIII	B	3	-	5	1	1	6	16	1.3
		L	3	1	8	-	2	21	35	2.9
	XIV	B	83	2	40	12	18	42	197	15.7
		L	91	1	64	22	19	47	244	20.5
	XV	B	-	-	5	-	-	-	-	0.4
		L	-	-	-	-	-	-	-	0.0
Number of instances in each category group		B L	519 482	20 19	302 356	152 80	119 97	141 157	1253 1191	100.0
Frequency in each category group in % of all instances in the Corpus		B L	41.4 40.5	1.6 1.6	24.1 29.9	12.1 6.8	9.5 8.1	11.3 13.2		100.0
Average number of s-gen per text		B L	5.9 5.5	1.2 1.1	1.9 2.2	5.1 2.7	1.5 1.2	1.1 1.2	2.5 2.4	

Fig. 1 Average number of non-personal genitives per text in each text category.

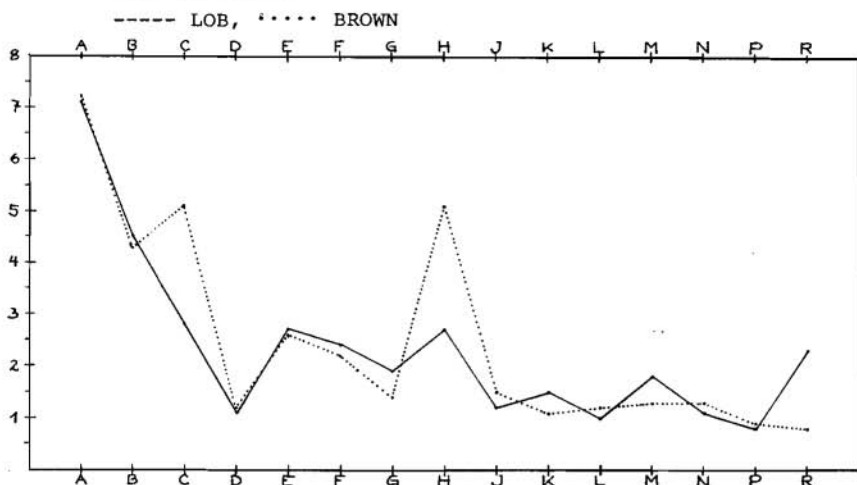
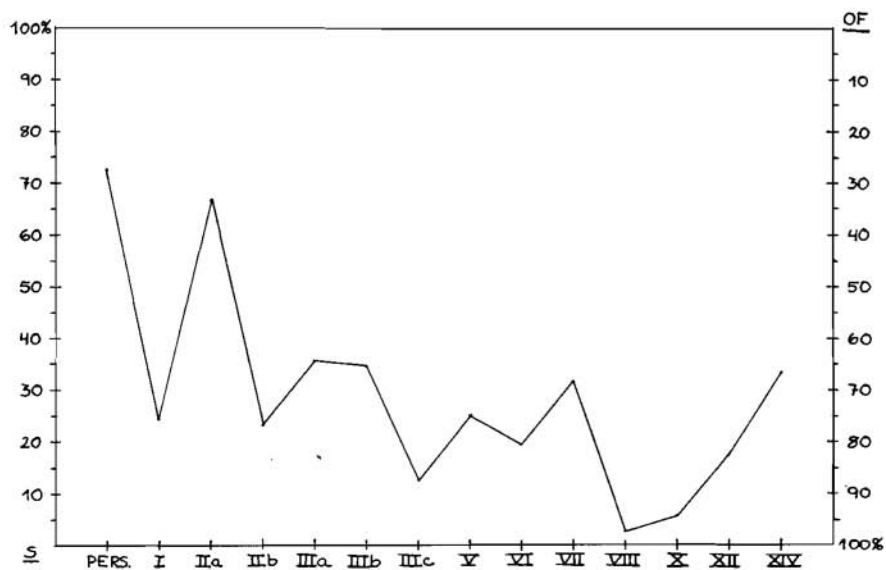


Fig. 2 Relative frequency of *s*-genitives and *of*-constructions with singular noun forms of different classes.



## NOTES

- 1 Some subgroups have been added. The resulting classification is as follows:

### I NOUNS DENOTING COLLECTIVE COMMUNITIES

- a) Authoritative and other organized bodies
- b) The complete or shortened name of companies or comparable formations
- c) Nouns which do not primarily denote human beings
- d) Group-genitives

### II NAMES OF CONTINENTS, COUNTRIES, TOWNS AND OTHER AREAS

- a) Political or sociological meaning emphasized
- b) Purely geographical meaning emphasized
- c) Names without a distinction between political/sociological and geographical meaning
- d) Geographical names used to denote football clubs etc.

### III COMMON NOUNS DENOTING GEOGRAPHICAL CONCEPTS

- a) Political or sociological meaning emphasized
- b) Purely geographical meaning emphasized
- c) Nouns without a distinction between political/sociological and geographical meaning

### IV S-GENITIVES BEFORE SUPERLATIVES

Not included in the present paper.

### V NAMES OF ANIMALS

### VI NOUNS DENOTING MEANS OF LOCOMOTION

### VII THE SUN, THE PLANETS, THE STARS, AND OTHER HEAVENLY BODIES

### VIII NOUNS DENOTING BUILDINGS AND LOCALITIES

### IX NEWSPAPERS AND PERIODICALS

### X ABSTRACT NOUNS

### XI CURRENCIES

### XII MATERIAL NOUNS AND CONCRETE THINGS

### XIII IDIOMATIC EXPRESSIONS

### XIV EXPRESSIONS OF TIME AND MEASURE

- a) Expressions of time
- b) Expressions of measure

### XV MISCELLANEOUS

Not included in the present paper.

- 2 For a list of the text categories, see p. 4 of this issue of *ICAME NEWS*.

In this brief survey I have - for practical reasons - grouped together newspaper texts (A-C), general expository prose (E-G), and fiction (K-R). A more detailed survey of the differences between single categories is given in Sørheim (1980:91-94). See also Fig. 1. For further information on the two corpora, see Johansson et al. (1978) and Francis (1979).

- 3 Aronsson (1975) gives no information on the distribution in categories for the subclasses (here marked by a dash).

- 4 I have chosen to deal with Classes II and III together, since they are closely connected and have usually been treated as one class (e.g. Jespersen 1949:315, Poutsma 1914:50, Zachrisson [1920:38]).
- 5 It is very difficult to make a clear distinction between cases where geographical proper names and common nouns are regarded as organized bodies and those where they are looked upon as geographical areas only, since 'something of the first is apt to creep into the second' (Svartengren 1949:141). In the present investigation the examples have been classified according to their paraphrasability with *in*. Examples which cannot be paraphrased with *in* without changing the meaning of the genitive construction have been classified under IIa and IIIa, those which can have been classified under IIb and IIIb. For the treatment of some problematic cases, see Sørheim (1980:36f., 39, 47).
- 6 Aronsson's (1975) system of classification does not include sub-classes IIc and IId.
- 7 By checking the Brown concordance, I found that Aronsson (1975) had failed to classify the majority of instances with names of animals in the Brown Corpus. The figures for Class V in the two corpora are therefore not comparable. (Cf. Sørheim 1980:55, 152f.).
- 8 Aronsson (1975) has omitted all examples with - *edge* and - *end*. Besides, there are at least 17 instances of other set expressions in the Brown Corpus which have not been classified at all. Consequently, a comparison between LOB and Brown in Class XIII is impossible.
- 9 Aronsson (1975) does not distinguish between genitive expressions which denote *measure of time* and those which - in a very wide sense - denote *point of time*. I have listed the two types separately, but for comparative purposes they have been treated as one class.
- 10 Cf. Sørheim (1981).

## APPENDIX

### LIST OF ALL NON-PERSONAL NOUNS OCCURRING WITH THE S-GENITIVE IN THE LOB CORPUS

Singular and plural forms of a noun (e.g. *authority*, *authorities*) are listed separately, as the two forms seem to behave differently with regard to the choice of genitive construction (cf. Sørheim 1980:110-146). The nouns in each class/subclass are given in the order of highest frequency, with the number of instances for words occurring more than once specified within parentheses after the word, e.g. *Britain* (51). Words occurring only once are listed alphabetically at the end of the list for each class/subclass. In Class XIII (idiomatic expressions) the expressions are listed under different headings such as *for - sake*, *- edge*, etc.

#### I NOUNS DENOTING COLLECTIVE COMMUNITIES

##### a) *Authoritative and other organized bodies*

1) *Authoritative bodies*: government (38), council (21), commission (11), committee (9), church (7), board (6), authority (5), authorities (2), Court (2), Ministry (2), Parliament (2), boards, C.E.G.B. (= Central Electricity Generating Board), Chamber (= the Chamber of Commerce), Gestapo.

2) *Nouns denoting other organized bodies*: company (26), party (16), group (12), Labour (9), nation (8), people (= nation) (7), association (6), club (4), mankind (4), union (4), band (3), family (3), firm (3), KANU (= Kenya African National Union) (3), League (3), society (3), staffs (3), command (2), Commonwealth (2), federation (2), guard (= body of persons) (2), orchestra (2), unions (2), congregation, corporation, corporations, crew, folk (= family), Garden (= the Covent Garden Company), Hilfsverein, I.L.P. (the Independent Labour Party), Legion (= the British Legion), Loyals (= the Loyal Regiment), the Mudlarks (a pop group), Nato (= the North Atlantic Treaty Organization), neutrals (= the neutral powers/countries), N.K.P. (= the New Kenya Party), organisation, Regiment, the Shadows (a pop group), state, subsidiaries, UN (= the United Nations), unit, W.E.A. (= Workers' Educational Association).

##### b) *The complete or shortened name of companies or comparable formations*

*Complete names*: Boots (3), Bent (2), Bents (2), Farley (2), John Smith (2), Amalgamated Limestone Corporation, Atlantic Aviation Corporation, Central Bank, Clacton, Cooper, Cortaulds, Edge Tool, Fry, Gilson & Freeman, Glaxo Laboratories, Grattan Warehouses, Hall & Co., Harris, Lars Halvorsen and Sons, Longdon, Olympic Airways, Pearce, Robinson, Scott, Stewart and Lloyd, Sturrock, Sunderland Shipbuilding Group, Threlfall, Yates.

*Abbreviated names*: BBC (2), B.E.A. (2), CWS, Glaxo, ITA, ITV.

c) *Nouns which do not primarily denote human beings*: school (11), industry (6), administration (5), home (5), branch (3), theatre (3), department (2), library (2), Revenue (= the Inland Revenue Department) (2), section (2), TV (2), airlines, banks, college, conference,



division, foundation, hospital, libraries, mission, movement, opéra, Radio Peking, profession, prosecution, Moscow radio, railways, stable, radio station, stores, Treasury, TUC (= Trades Union Congress), universities, university.

d) *Group-genitives*: the Council of Local Authorities, the U.K. Ministry of Aviation.

## II NAMES OF CONTINENTS, COUNTRIES, TOWNS AND OTHER AREAS

a) *Political or sociological meaning emphasized*: Britain (51), America (12), (West) Germany (12), Avon (10), Russia (10), France (6), India (5), Manchester (4), Moscow (4), Spain (4), West (4), Huddersfield (3), London (3), (Northern) Rhodesia (3), United States (3), Canada (2), Denmark (2), England (2), Ghana (2), Katanga (2), Kenya (2), New Zealand (2), Nord (2), South Africa (2), Tysoe (2), (West) Berlin, Blackpool, Bonn, Britannia, Cheshire, China, Cuba, Dublin, Erin, Guinea, Hungary, Israel, Japan, Lincoln, Madrid, Malacca, Marton, Morocco, New Zealand (team), Nottingham, Nyasaland, Pakistan, Pasai, Rome, Sheffield, Somalia, Soviet Union, Sudan, Sunderland, Surinam, Tanganyika, Tring, Warwick, Watford, Yugoslavia, New York.

b) *Purely geographical meaning emphasized*: London (7), Manchester (3), Ascot (2), India (2), Rome (2), Aberdeen, Accra, Australia, Ayr, (West) Berlin, Birmingham, Brighton, Britain, Budapest, Egypt, Epsom, Hollywood, Huddersfield, Leamington, Lebanon, Lincoln, Malaya, Nottingham, Princes Risborough, Ramsgate, Russia, Soho, Sweden, Sydney, Tonto, Warwick, Windsor, Wirral, New York.

c) *Names without a distinction between political/sociological and geographical meaning*: Gramp (2), Adriatic.

d) *Geographical names used to denote football clubs etc.*: Arsenal (2), Brentford (2), Coventry (2), Forfar (2), West Ham (2), Southampton (2), Wimbledon (2), Blackpool, Chelsea, Fulham, Grimsby, Newcastle, Oxford, Plymouth, Reading, Southend, Swansea, Swindon, Tottenham, Villa (= Aston Villa).

## III COMMON NOUNS DENOTING GEOGRAPHICAL CONCEPTS

a) *Political or sociological meaning emphasized*: world (28), country (15), city (5), town (5), area, countries, county, district, protectorate, side.

b) *Purely geographical meaning emphasized*: city (5), world (5), country (2), area, colony, land.

c) *Nouns without a distinction between political/sociological and geographical meaning*: desert (3), river (3), pond, racecourse, shaft, uplands.

## V NAMES OF ANIMALS

bird (4), animal (3), hyena (3), animals (2), dog (2), lion (2), mare (2), pidgeon (2), Alcoa (= the name of a horse), blackbird, bug, bull, calf, cow, creature, drake, eagle, elephant, fox, Beldon Hall (= the name of a horse), hens, horse, horses, hounds, jackal, lamb, lions, mantis, monster, nightingales, pig, Avon's Pride (= the name of a horse), roan, snake, spider, squirrels, traverser, wolf, worm.

#### VI NOUNS DENOTING MEANS OF LOCOMOTION AND MACHINES

ship (10), boat (4), Magda (2), plane (2), airliner, barge, Brescia Bugatti, Callender, car, Citroen, destroyer, drill, Easterner, life-boat, lorry, Pericles, pump, R34, R38, R101 (= names of airships), Sandpiper, Sceptre, ships, ex-trawler, Warden, Whitehall.

#### VII THE SUN, THE PLANETS, THE STARS, AND OTHER HEAVENLY BODIES

earth/Earth (5), comet (2), globe, Moon, sun.

#### VIII NOUNS DENOTING BUILDINGS AND LOCALITIES

saloon (2), church, cinema, Everglade (= the name of a club), hospital, hotel, house, rectangle (= nave), engine-room, shop, stable, station, White (= the name of a club).

#### IX NEWSPAPERS AND PERIODICALS

paper (= newspaper) (3), Pic (= Sunday Pictorial) (2), magazine, newspaper, London Observer, Punch.

#### X ABSTRACT NOUNS

life (5), nature (4), law (3), medium (= science fiction) (2), music (2), work (2), Death, dream, exports, farce, Fascism, Fascismo, festival, film, horse-racing, love, Common Market, Gallup Poll, pools, self, body-self, show, sin, soccer, subsidies, war, weather, youth.

#### XI CURRENCIES

No instances in either LOB or Brown.

#### XII MATERIAL NOUNS AND CONCRETE THINGS

heart (3), dolls (2), bed, blood, body, book, booklet, bullet, doll, figure, harpsichord, lamp, report, tree, wall, water, weapon.

#### XIII IDIOMATIC EXPRESSIONS

*for - sake*: for argument's sake, for decency's sake, for economy's sake, for goodness' sake, for his health's sake, for Ireland's sake, for photography's sake, for sanity's sake, for old times' sake, for work's sake.

*- edge*: the pond's edge, razor's edge, the river's edge, the water's edge (5).

*- end*: his wit's end.

*Other idioms*: at arms' length, at arm's length/at arm's-length (6), the bull's eyes, a hair's-breadth, your heart's desire, a lion's share (2), your/my/her/the mind's eye (4).

#### XIV EXPRESSIONS OF TIME AND MEASURE

a) *Expressions of time*: year (30), week (16), today/to-day (14), night (10), yesterday (7), month (6), day (5), tomorrow/to-morrow (5), morning (4), Saturday (4), tonight/to-night (4), afternoon (3), evening (3), season (3), autumn (2), Monday (2), Sunday (2), winter (2), century, epoch, period, summer, term, Wednesday.

b) *Expressions of measure*: years (18), year (12), day (11), days (11), minutes (11), months (9), hour (8), week (8), moment (6), fortnight (5), weeks (5), hours (2), month (2), half-crown, instant, lifetime, miles, night, seasons, term, winters.

# REFERENCES

- Aronsson, I. 1975. 'The s-genitive of non-human and collective nouns in present-day American English'. Unpublished term paper, Department of English, Lund University.
- Dahl, L. 1971. 'The s-genitive with non-personal nouns in Modern English journalistic Style', *Neuphilologische Mitteilungen* LXXII, Helsinki. 140-172.
- Francis, W.N. 1979. *Manual of Information to Accompany a Standard Sample of Present-Day Edited American English, for Use with Digital Computers*. Rev.ed. Dept. of Linguistics, Brown University.
- Jespersen, O. 1949. *A Modern English Grammar on Historical Principles*, Vol. VII, Copenhagen.
- Johansson S., Leech, G.N., and H. Goodluck. 1978. *Manual of Information to Accompany the Lancaster-Oslo/Bergen Corpus of British English, for Use with Digital Computers*. Department of English, University of Oslo.
- Kirchner, G. 1970. *Die syntaktischen Eigentümlichkeiten des amerikanischen English*, Vol. 1, Leipzig: VEB Verlag Enzyklopädie.
- The Lancaster-Oslo/Bergen Corpus of British English, for Use with Digital Computers*. 1978. Department of English, University of Oslo.
- Potter, S. 1969. *Changing English*, London.
- Poutsma, H. 1914. *A Grammar of Late Modern English*, Part II 1a, Groningen.
- The Standard Corpus of Present-Day Edited American English*. 1964. Providence: Brown University.
- Svartengren, H. 1949. 'The -s Genitive of Non-Personal Nouns in Present-Day English', *Studier i Modern Språkvetenskap* utgivna av Nyfilologiska Sällskapet i Stockholm, Stockholm Studies in Modern Philology. XVII, Uppsala. 139-180.
- Sørheim, M.-C.J. 1980. *The s-Genitive in Present-Day English*. Unpublished 'hovedfag' thesis, Department of English, University of Oslo.
- Sørheim, M.-C.J. 1981. 'The Genitive in a Functional Sentence Perspective'. In S. Johansson and B. Tysdahl, eds., *Papers from the First Nordic Conference for English Studies*, Oslo, 17-19 September, 1980, Institute for English Studies, University of Oslo. 405-423.
- Zachrisson, R.E. 1920. 'Grammatical Changes in Present-Day English', *Studier i Modern Språkvetenskap* utgivna av Nyfilologiska Sällskapet i Stockholm, Stockholm Studies in Modern Philology VII, Uppsala. 37-49.

## SHALL, WILL, SHOULD, AND WOULD IN BRITISH AND AMERICAN ENGLISH

Inger Krogvig

Stig Johansson

University of Oslo, Norway

### BACKGROUND

No single issue has received more attention in discussions of British-American differences than the use of *shall* and *will*. It has been taken up in general descriptions of American English, such as Krapp (1925), Mencken (1936), Fries (1940), Zandvoort (1968), Forgue and McDavid (1972), and Švejcer (1978). The topic has been dealt with in usage books (e.g. Fowler 1965), grammars (e.g. Quirk et al. 1979), and articles and monographs dealing with the English verb (e.g. Joos 1964, Leech 1971). There are also special studies of the use of *shall* and *will* in American vs. British English, notably Fries (1925) and Taubitz (1978).

In spite of all the attention given to the topic, uncertainty remains --for a variety of reasons. In the first place, the semantic complexity of the modals makes them notoriously difficult to describe. A particular problem with *shall* and *will* is the long-standing conflict between attitude and use, between prescriptive rules and speaker performance. It is further uncertain whether and to what extent observations on *shall* and *will* are applicable to *should* and *would*. Finally, it has become increasingly clear that language varies according to a range of dimensions, such as medium, regional and social dialect, register, and style. To be adequate, statements on British-American differences must specify what *type* of American English differs from British English, and in what respect. This necessitates a satisfactory basis of comparison, preferably with a broad representation of comparable text types for each of the two national forms of English.

The availability of the Brown Corpus of American English texts and its British English counterpart, the LOB Corpus, has made it possible to investigate the problem of *shall* and *will* on the basis of compar-

able material representing a variety of text types (all from printed sources). Coates and Leech (1980) deal briefly with the modals, including the forms we focus on here, using the two corpora. Krogvig (1981) is a more detailed investigation of *shall*, *will*, *should*, and *would* made on the basis of the same material. The present paper summarizes some of the main results of Krogvig's study.

#### TOTAL FREQUENCIES

The frequencies of *shall*, *will*, *should*, and *would* as well as of the contracted forms 'll and 'd are given in Table 1. Irrelevant homographs have been eliminated from the count, such as *will* when used as a noun or 'd when representing *had*. The capitalized forms represent the total occurrences of each main type. Relative frequencies are expressed in per cent of all the words in each corpus (about a million words). The difference coefficient, which expresses the degree of difference between the corpora, is calculated in the following way (cf. Yule 1944):

$$\frac{\text{frequency LOB} - \text{frequency Brown}}{\text{frequency LOB} + \text{frequency Brown}}$$

The coefficient varies from +1 to -1. A positive figure indicates a higher frequency in the British material, a negative figure a higher frequency in the American material. On the basis of Table 1, we can make the following observations:<sup>2</sup>

- a) SHALL and SHOULD are more frequent in the LOB than in the Brown Corpus. This is what could be expected. A more unexpected finding is perhaps that the difference turned out to be larger with SHOULD than with SHALL. SHALL is far less frequently used in both corpora than SHOULD:
- b) WILL and WOULD are much more frequent in both corpora than SHALL and SHOULD. There is no appreciable difference between the two corpora in the use of WILL and WOULD.
- c) Contractions are on the whole rarer than the full forms, as is to be expected in written, fairly formal prose. The contracted form 'll occurs much more often than its preterite counterpart 'd in both corpora. Both contractions are a bit more common in the British material.
- d) Negative contractions make up but a minor part of the total number

of occurrences. They have a similar distribution in the two corpora. A study of concordances for the corpora shows that negation is more frequently expressed by full forms, with the exception of *will not* and *won't*, which have almost the same frequency.<sup>3</sup> *Shan't* is extremely infrequent in both the British and the American material, though there are a few more examples in the LOB Corpus.

It is of particular interest to note that the over-representation of SHALL and SHOULD in the LOB Corpus is not compensated for by an equivalent decrease of WILL and WOULD (which are about equally common in both corpora) or of the contracted forms (most of which are, in fact, more frequent in the British material). The total number of these auxiliary forms is therefore higher in the LOB Corpus than in the Brown Corpus.

In the rest of this paper we shall look more closely at the use of SHALL and SHOULD, in an attempt to specify more precisely where the differences are found.

#### FREQUENCY IN RELATION TO TYPE OF TEXT

As pointed out in the beginning of the paper, it is important to take the type of text into account in discussions of British-American differences. This is confirmed by Tables 2 and 3, which show the distribution of SHALL and SHOULD across the text categories of the two corpora.<sup>4</sup> We can make the following observations:

a) SHALL is mainly a feature of informative prose (i.e. categories A-J), especially of legal, scientific, and religious language, represented by categories H, J, and D. In this respect there is little difference between the two corpora. The main difference is found in imaginative prose (i.e. categories K-R), where the LOB Corpus, at least in two categories (K: General fiction, P: Romance and love story), has a strikingly higher frequency of SHALL. The distribution of SHALL across text categories is visualized in a diagram showing the relative deviation of the absolute frequency from the expected frequency (Figure 1). Similarities as well as differences between the two corpora are clearly seen in the diagram.<sup>5</sup>

b) SHOULD is more frequent in the LOB Corpus than in the Brown Corpus, and the over-representation is found in all the text categories apart

from D (Religion), where the figure for the American material is slightly higher. In both corpora the largest proportion of the occurrences is found in informative prose. The most conspicuous differences between the relative frequencies of SHOULD in the two corpora are found in categories A (Press:reportage), B (Press: editorial), E (Skills, trades and hobbies) in the informative prose section, while K (General fiction) and N (Adventure and western fiction) show the most notable differences in the imaginative prose section. In the same way as with SHALL, a diagram has been set up to illustrate differences and similarities of distribution in the two corpora, seen in relation to the expected frequency (Figure 2). The relationship between the two corpora is remarkably close, which testifies to basic agreement in the relationship between text categories, in spite of the difference in frequency.<sup>6</sup>

We can now refine the statements based on our observations of total frequency. There is overall similarity between the corpora in the distribution of SHALL and SHOULD across text categories. The over-representation of SHALL in the LOB Corpus is found in imaginative prose, while there is a general over-representation in the text categories of the LOB Corpus for SHOULD.

#### FREQUENCY IN RELATION TO PERSON

Grammatical person (of the subject) is usually said to influence the choice of auxiliary (*shall* vs. *will*, *should* vs. *would*), and it is often pointed out that usage in British and American English differs in this respect. The distribution in relation to person is specified in Tables 4 and 5. On the basis of the tables, we can make the following observations on the use of SHALL and SHOULD:

a) The over-representation of SHALL in the LOB Corpus is due to its occurrence with a first person subject. This is in agreement with what we could expect from previous observations. There is a surprisingly close correspondence between the two corpora in the figures for the second and third persons. SHALL with a second person subject is extremely infrequent in both corpora. While the majority of the examples of SHALL are found with a first person subject in the British material, most of the examples in the American material are in the third person.

b) The figures for SHOULD confirm previous statements that this modal is more common with the first person in British than in American English. The figures for the first person are very close to those for SHALL. In the second person SHOULD is more frequent than SHALL in both corpora, with an almost equal number of occurrences. In the third person, on the other hand, there is a striking difference between the American and British material, with an over-representation of more than 300 examples in the LOB Corpus.

We can then conclude that differences in the use of SHALL are due to a higher occurrence with the first person in the LOB Corpus, while the over-representation of SHOULD in the British material is to be found both in the first and the third person. These observations will be further refined in the next section.

#### FREQUENCY IN RELATION TO CLAUSE TYPE

The type of clause is another factor which may affect the choice of auxiliary. This has long been recognized and is usually dealt with in grammatical descriptions. We shall follow the division set up by Fries (1925), and use the terminology of Quirk et al. (1979:386, 721): independent declarative clauses, questions, and subordinate clauses. The identification of the three clause types is based on the criteria set forth in Quirk et al. (1979). Borderline conjunctions such as *for* and *so* have thus been regarded as subordinators. Declarative questions, identical in form to statements, but with final rising question intonation (indicated in writing by a question mark), have been registered as questions.

The distribution of SHALL and SHOULD with each grammatical person in the three types of clauses is presented in Table 6. Tables 7-10 give a combined survey of the distribution a) with grammatical person, b) in the three types of clauses, and c) across text categories. We shall now comment on the use of SHALL and SHOULD in the three clause types, using the tables as our point of departure.

##### *Independent declarative clauses*

While the figures in Table 6 seem to indicate that the use of SHALL with second and third person subjects in independent declarative clauses is fairly similar in British and American English, confirmed



by the detailed survey in Tables 7 and 8, there is a clear difference with regard to the first person. The British material has more than twice as many examples of *I/we shall*, 132 as against 60 in the American material. The over-representation is found especially in imaginative prose (cf. Tables 7 and 8). It is the difference here, rather than the difference in subordinate clauses, that is responsible for the discrepancy between the two corpora in the use of SHALL with the first person. This is contrary to Fries's observations, based on drama material (Fries 1925:1016). According to Fries, the main difference was found in clauses of reported speech. In our material there are very few instances of SHALL with a first person subject in reported speech.

As in the case of SHALL, the difference between the two corpora in the figures for SHOULD in the second and third person is negligible in independent declarative clauses. With a first person subject SHOULD is clearly more frequent in the British material (65 examples in Brown and 110 in LOB). The over-representation in the British material, which is found in both imaginative and informative prose, is probably due to a more frequent use of SHOULD as a stylistic variant of WOULD in the main clause of a conditional sentence, or in a sentence with implicit conditional context.

#### Questions

There is fairly close agreement between the two corpora in the use of SHALL with a first person subject in questions, though the British material has some more examples.<sup>7</sup> The figures confirm previous statements that *shall* is a normal auxiliary with the first person in questions in American English. *Will* can also be used in questions with a first person subject. The difference between the two auxiliaries is that *shall* can have two meanings--it may ask for instructions or it may have only future reference--while *will* can only refer to a non-volitional future. The use of *will* is said to be typical of American English, but is lately also becoming more frequent in British English (Jacobsson 1962a and b, Quirk et al. 1979:99-100). In the LOB Corpus as well as in the Brown Corpus there are a few examples of *will I/we*, five (of which two were tag questions) and six, respectively. *Will* with a first person subject in questions is apparently used to much the same extent in printed British and American English, but

is less frequent than SHALL.

In our material there are very few questions with SHALL in combination with a third person subject, in the American as well as in the British corpus. The five examples in category D in the Brown Corpus all occur in quotations from the Bible. There are no examples with a second person subject, which, according to traditional rules, requires *shall*, when *shall* is expected in the answer (Poutsma 1926:231). As pointed out by Fries (1925), *shall* is infrequent with second person subjects in British as well as in American 'contemporary' drama. *Will* is the auxiliary used. In the opinion of Evans and Evans (1957:447) *shall you?* sounds like a 'ridiculous affectation' in American English. In both corpora there are many examples of second person questions with WILL, in Brown 25, and in LOB as many as 48.

The number of questions with SHOULD is higher in the Brown Corpus than in the LOB Corpus. This agrees with the statements found in Myers (1959:421-22) and in Leech (1971:85) that in American English *should* is preferred to *shall* in questions. The difference between the use of SHALL and SHOULD in questions in British and American English is well illustrated by the figures found in category K (General fiction). Here the LOB Corpus has seven examples of SHALL with a first person subject, while the Brown Corpus has none at all. In the case of SHOULD we have the opposite situation, with seven examples in Brown and only two in LOB (see Tables 7-10).

#### *Subordinate clauses*

As pointed out above, there are very few examples of SHALL with a first person subject in reported speech, where, in accordance with Fries's results (1925), we might have expected to find the reason for the difference in frequency between our two corpora. In the LOB Corpus the majority of the examples of *I/we shall* in subordinate clauses are found in relative and adverbial clauses, not in *that*-clauses. The two corpora have close to the same number of instances in informative prose, with 27 examples in Brown and 25 in LOB. However, while the examples are evenly distributed among the various categories in the LOB Corpus, *I/we shall* in subordinate clauses is completely absent from several of the text categories in the Brown Corpus, the majority of the occurrences being concentrated to one genre, Belles lettres,

biography, essays (Category G) (see Tables 7 and 8). It is the examples in the fiction categories that account for the slight over-representation of *I/we shall* in subordinate clauses in the LOB Corpus. Eight of the fifteen examples occur in *that*-clauses. There are several examples after governing expressions such as 'I'm afraid that', 'I presume that', etc., where *shall*, according to traditional rules, is the auxiliary to be used with the first person, and *will* with the second and third persons (cf. Jespersen 1931:284, Taglicht 1970:203-204). There is, however, no consistent use of *shall* with the first person in the British material after verbs of doubt, fear and belief. After 'I hope' only *will*, '*ll* and *won't* occurred, with the exception of one example of *shall* after 'let us hope' (LOB E18:180). After 'I think' *will* and '*ll* are more frequent than *shall* in both corpora. With regard to the latter verb, Joos (1964:160) claims that only *I shall* can be used after 'I think', while Taglicht (1970:203) distinguishes between two meanings of 'I think', the one being 'To conceive or entertain the notion of doing something', taking *shall*, and the other 'To be of opinion...', taking *will*. Both corpora have only one example each of 'I think I shall', while *will* occurred once and '*ll* five times in Brown, and '*ll* twice in LOB after 'I think'. It is quite possible that *shall* is more common in British than in American English after verbs of hope, fear and belief, but the number of examples here is too small to warrant any definite conclusions in this respect.

SHALL with a second person subject is rare in subordinate clauses in both corpora. With a third person subject it survives mainly in certain types of formal language, especially after verbs of volition, expressing a demand, a request, etc. It is this use that is responsible for the high frequency of SHALL in the third person in subordinate clauses in category H. There is no difference between American and British usage on this point (see Tables 7 and 8).

Table 6 shows that there is a striking difference between the two corpora in the use of SHOULD in subordinate clauses, with the first as well as with the third person. In the second person examples are few and the distribution is fairly similar. While the over-representation in the first person is most obvious in two of the categories (G and K), there is a consistently higher frequency of SHOULD with the third person in the LOB Corpus in all the categories apart from one

(L). With regard to the higher frequency in the case of the first person, it is difficult to say whether this is due to the use of SHOULD as a stylistic variant of WOULD or whether the examples have 'normative' or 'putative' meaning (cf. Krogvig 1981). However, it is quite clear that the differences in the third person are partly due to a more extensive use of SHOULD in British English after verbs expressing a request or a demand, where American speakers are more likely to employ the mandative subjunctive (cf. Quirk et al. 1979:76).

Note, in conclusion, that while the over-representation of SHOULD in the first person is found in independent declarative clauses as well as in subordinate clauses, the difference in the third person is due to its use in subordinate clauses only (see Table 6).

#### CONCLUSION

Our investigation of the frequencies of SHALL, WILL and 'LL, SHOULD, WOULD and 'D in the Brown and LOB corpora has shown that the main difference between American and British English lies in the use of SHALL and SHOULD. This is in itself not surprising, and agrees with previous observations. It should be noted, however, that the difference is more marked with SHOULD than with SHALL. Furthermore, the over-representation of SHALL and SHOULD in the British material is not matched by a corresponding increase of WILL and WOULD in the American material. On the contrary, the British material has more examples of all the auxiliaries except WOULD, but the under-representation here is only fifty examples, or less than 2%.

The use of the auxiliaries is clearly genre-bound in American as well as in British English. There is a fairly close similarity between the two corpora in the distribution of the auxiliaries with regard to the two main divisions of informative and imaginative prose, allowing for differences with respect to the individual categories. The only exception to this is SHALL, which is almost five times more frequent in British fiction than in American fiction. The difference indicated here is due to the use of SHALL with a first person subject. In relative distribution SHALL makes up 22.2% of the auxiliaries (SHALL, WILL and 'LL) with a first person subject in the fiction categories, while the corresponding American categories have only 6.93% of SHALL. It is interesting to note that SHALL is evidently less frequent in

present-day American fiction than in the fiction material investigated by Luebke (1929) and in the American drama material examined by Fries (1925). The percentage of *shall* in the first person in Luebke's material is 28.55, and in that of Fries 16.28.<sup>8</sup>

In informative prose SHALL has a similar distribution in the two corpora, with regard to the first as well as to the third person. SHALL is apparently used to much the same extent in British and American texts characterized by a certain degree of formality, in particular in legal and religious language. A typical feature of legal language is the use of SHALL with a third person subject.

The most important difference between the two corpora is the much higher frequency of SHOULD in the British material. The over-representation is found in the first as well as in the third person. Although SHOULD may have been used as a stylistic variant of WOULD to a larger extent in the British than in the American corpus, the main reason for the discrepancy is probably the use of SHOULD in *that*-clauses, where American English frequently employs alternative expressions, such as the subjunctive or the *for - to* construction.

Needless to say, frequency counts like those presented here are not sufficient to describe British-American differences in the use of SHALL, WILL, SHOULD, and WOULD. We feel, however, that they form a good starting-point for further analysis. Krogvig (1981) includes a more detailed study of the uses and meanings of SHALL and SHOULD, primarily based on the LOB Corpus, but to give a further account is beyond the scope of this brief paper.

Table 1 Total frequencies of SHALL, WILL, 'LL, SHOULD, WOULD, 'D <sup>1</sup>

Types	Brown Corpus		LOB Corpus		Difference coefficient
	Absolute frequency	Relative frequency in %	Absolute frequency	Relative frequency in %	
shall	266		350		0.13
shan't	1		5		0.67
SHALL	267	0.03	355	0.04	0.14
will	2160		2208		0.00
other forms	1		0		- 1.00
won't	105		111		0.02
WILL	2266	0.23	2319	0.23	0.01
'LL	442	0.04	505	0.05	0.07
should	888		1276		0.17
other forms	1		1		0.00
shouldn't	22		25		0.06
SHOULD	911	0.09	1302	0.13	0.18
would	2716		2682		- 0.01
other forms	1		6		0.71
wouldn't	129		108		- 0.09
WOULD	2846	0.28	2796	0.28	- 0.01
'D	202	0.02	236	0.02	0.08

Table 2 Distribution of SHALL across text categories

Category	Absolute frequency		Relative frequency in %		Expected frequency	
	Brown	LOB	Brown	LOB	Brown	LOB
A	5	14	0.01	0.02	23.6	31.2
B	19	9	0.04	0.02	14.5	19.2
C	2	4	0.01	0.01	9.1	12.1
D	22	25	0.06	0.07	9.1	12.1
E	5	15	0.01	0.02	19.3	27.0
F	12	8	0.01	0.01	25.7	31.2
G	35	26	0.02	0.02	40.2	53.3
H	99	95	0.17	0.16	16.1	21.3
J	42	60	0.03	0.04	42.9	56.8
K	3	28	0.01	0.05	15.5	20.6
L	4	11	0.01	0.02	12.9	17.0
M	3	1	0.03	0.01	3.2	4.3
N	10	15	0.02	0.03	15.5	20.6
P	4	41	0.01	0.07	15.5	20.6
R	2	3	0.01	0.02	4.8	6.4
Total	267	355	0.03	0.04		

Table 3 Distribution of SHOULD across text categories

Category	Absolute frequency		Relative frequency in %		Expected frequency	
	Brown	LOB	Brown	LOB	Brown	LOB
A	64	120	0.07	0.14	80.2	114.6
B	93	146	0.17	0.27	49.2	70.3
C	18	19	0.05	0.06	31.0	44.3
D	45	41	0.13	0.12	31.0	44.3
E	74	146	0.10	0.19	65.6	99.0
F	78	98	0.08	0.11	87.5	114.6
G	105	187	0.07	0.12	136.7	200.5
H	113	126	0.19	0.21	54.7	78.1
J	179	204	0.11	0.13	145.8	208.3
K	38	64	0.07	0.11	52.8	75.5
L	30	35	0.06	0.07	43.7	62.5
M	4	11	0.03	0.09	10.9	15.6
N	20	41	0.03	0.07	52.8	75.5
P	43	49	0.07	0.08	52.8	75.5
R	7	15	0.04	0.08	16.4	23.4
Total	911	1302	0.09	0.13		



Table 4 Distribution of SHALL, WILL and 'LL in relation to person

Person	Tokens	Brown		LOB	
		No. of occurrences	%	No. of occurrences	%
1st	SHALL	114	22.5	204	32.9
	WILL	149	29.3	138	22.3
	'LL	245	48.2	278	44.8
	Total	508	100	620	100
2nd	SHALL	5	2.3	6	1.9
	WILL	122	56.2	191	59.9
	'LL	90	41.5	122	38.2
	Total	217	100	319	100
3rd	SHALL	148	6.6	145	6.5
	WILL	1995	88.6	1990	88.8
	'LL	107	4.8	105	4.7
	Total	2250	100	2240	100

Table 5 Distribution of SHOULD, WOULD and 'D in relation to person

Person	Tokens	Brown		LOB	
		No. of occurrences	%	No. of occurrences	%
1st	SHOULD	126	31.6	204	36.9
	WOULD	198	49.6	228	41.2
	'D	75	18.8	121	21.9
	Total	399	100	553	100
2nd	SHOULD	42	31.8	43	24.3
	WOULD	67	50.8	85	48.0
	'D	23	17.4	49	27.7
	Total	132	100	177	100
3rd	SHOULD	743	21.7	1055	29.3
	WOULD	2581	75.3	2483	68.9
	'D	104	3.0	66	1.8
	Total	3428	100	3604	100

Table 5 Distribution of SHOULD, WOULD and 'D in relation to person

Person	Tokens	Brown		LOB	
		No. of occurrences	%	No. of occurrences	%
1st	SHOULD	126	31.6	204	36.9
	WOULD	198	49.6	228	41.2
	'D	75	18.8	121	21.9
	Total	399	100	553	100
2nd	SHOULD	42	31.8	43	24.3
	WOULD	67	50.8	85	48.0
	'D	23	17.4	49	27.7
	Total	132	100	177	100
3rd	SHOULD	743	21.7	1055	29.3
	WOULD	2581	75.3	2483	68.9
	'D	104	3.0	66	1.8
	Total	3428	100	3604	100

Table 6 Frequency in relation to clause type: SHALL and SHOULD

Person	Independent declarative clauses				Questions				Subordinate clauses			
	SHALL		SHOULD		SHALL		SHOULD		SHALL		SHOULD	
	Brown	LOB	Brown	LOB	Brown	LOB	Brown	LOB	Brown	LOB	Brown	LOB
1st person	60	132	65	110	25	32	27	11	29	40	34	83
2nd person	5	3	31	23	0	0	1	5	0	3	10	15
3rd person	100	91	408	413	7	1	42	29	41	53	293	613
Total	165	226	504	546	32	33	70	45	70	96	337	711

Table 7 The distribution of SHALL with respect to person, clause type, and text category (Brown Corpus)

Category	1st person				2nd person				3rd person			
	I	Q	S	T	I	Q	S	T	I	Q	S	T
A	2	-	-	2	-	-	-	-	2	-	1	3
B	3	1	1	5	-	-	-	-	4	1	9	14
C	2	-	-	2	-	-	-	-	-	-	-	-
D	3	6	-	9	1	-	-	1	4	5	3	12
E	1	-	1	2	-	-	-	-	1	-	2	3
F	4	1	4	9	-	-	-	-	-	-	3	3
G	12	2	13	27	1	-	-	1	5	1	1	7
H	4	2	1	7	-	-	-	-	74	-	18	92
J	22	3	7	32	-	-	-	-	6	-	4	10
K	2	-	-	2	-	-	-	-	1	-	-	1
L	1	3	-	4	-	-	-	-	-	-	-	-
M	1	1	1	3	-	-	-	-	-	-	-	-
N	2	4	1	7	3	-	-	3	-	-	-	-
P	1	1	-	2	-	-	-	-	2	-	-	2
R	-	1	-	1	-	-	-	-	1	-	-	1
	60	25	29	114	5	0	0	5	100	7	41	148

I = independent declarative clauses

Q = questions

S = subordinate clauses

T = total number of occurrences

Table 8 The distribution of SHALL with respect to person, clause type, and text category (LOB Corpus)

Category	1st person				2nd person				3rd person			
	I	Q	S	T	I	Q	S	T	I	Q	S	T
A	9	-	3	12	-	-	-	-	1	-	1	2
B	4	1	1	6	-	-	-	-	-	-	3	3
C	3	-	1	4	-	-	-	-	-	-	-	-
D	2	-	3	5	2	-	1	3	13	-	4	17
E	11	2	2	15	-	-	-	-	-	-	-	-
F	2	1	2	5	-	-	-	-	1	-	2	3
G	19	-	4	23	-	-	-	-	-	1	2	3
H	4	-	2	6	-	-	-	-	69	-	20	89
J	26	1	7	34	-	-	-	-	6	-	20	26
K	13	7	4	24	1	-	2	3	-	-	1	1
L	7	3	1	11	-	-	-	-	-	-	-	-
M	1	-	-	1	-	-	-	-	-	-	-	-
N	7	7	1	15	-	-	-	-	-	-	-	-
P	23	8	9	40	-	-	-	-	1	-	-	1
R	1	2	-	3	-	-	-	-	-	-	-	-
	132	32	40	204	3	0	3	6	91	1	53	145

I = independent declarative clauses

Q = questions

S = subordinate clauses

T = total number of occurrences

Table 9 The distribution of SHOULD with respect to person, clause type, and text category (Brown Corpus)

Category	1st person				2nd person				3rd person			
	I	Q	S	T	I	Q	S	T	I	Q	S	T
A	-	-	4	4	-	-	-	-	23	3	34	60
B	5	2	3	10	-	-	-	-	53	2	28	83
C	-	-	2	2	-	-	-	-	10	-	6	16
D	2	4	-	6	-	-	1	1	18	-	20	38
E	1	-	-	1	7	-	1	8	51	1	13	65
F	1	2	1	4	1	-	2	3	46	3	22	71
G	12	1	7	20	-	-	-	-	31	6	48	85
H	8	-	3	11	5	-	2	7	70	4	21	95
J	21	4	5	30	-	-	-	-	80	14	55	149
K	5	7	2	14	4	-	-	4	3	3	14	20
L	1	1	-	2	4	-	2	6	6	2	14	22
M	-	-	-	-	2	-	-	2	1	-	1	2
N	1	2	2	5	2	-	1	3	3	2	7	12
P	7	4	4	15	5	1	1	7	9	2	10	21
R	1	-	1	2	1	-	-	1	4	-	-	4
	65	29	34	126	31	1	10	42	408	42	293	743

I = independent declarative clauses

Q = questions

S = subordinate clauses

T = total number of occurrences

Table 10 The distribution of SHOULD with respect to person, clause type, and text category (LOB Corpus)

Category	1st person				2nd person				3rd person			
	I	Q	S	T	I	Q	S	T	I	Q	S	T
A	9	1	4	14	1	-	-	1	47	-	58	105
B	3	-	6	9	-	-	-	-	46	7	84	137
C	4	-	-	4	-	-	-	-	3	2	10	15
D	5	-	2	7	-	-	-	-	10	-	24	34
E	7	3	5	15	4	-	2	6	75	3	47	125
F	-	-	3	3	3	-	2	5	53	-	37	90
G	35	1	23	59	-	-	1	1	28	3	96	127
H	5	-	4	9	2	-	-	2	33	-	82	115
J	5	-	6	11	-	-	-	-	91	4	98	193
K	12	2	13	27	1	1	2	4	6	3	24	33
L	8	-	5	13	1	-	5	6	6	1	9	16
M	2	-	2	4	-	1	-	1	-	-	6	6
N	3	2	3	8	6	3	-	9	5	1	18	24
P	11	2	6	19	4	-	2	6	4	5	15	24
R	1	-	1	2	1	-	1	2	6	-	5	11
	110	11	83	204	23	5	15	43	413	29	613	1055

I = independent declarative clauses

Q = questions

S = subordinate clauses

T = total number of occurrences



Figure 1 Relative deviation of absolute frequency from expected frequency: SHALL. The categories of the LOB Corpus have been ranked from negative to positive deviation.

O = LOB Corpus  
 Δ = Brown Corpus

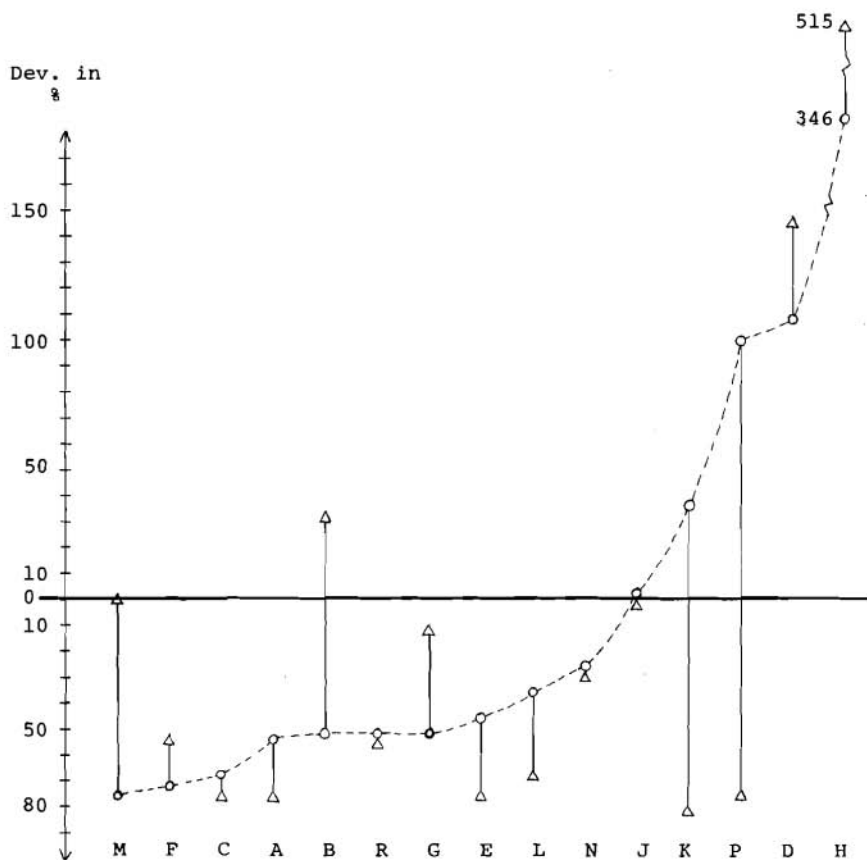


Table 10 The distribution of SHOULD with respect to person, clause type, and text category (LOB Corpus)

Category	1st person				2nd person				3rd person			
	I	Q	S	T	I	Q	S	T	I	Q	S	T
A	9	1	4	14	1	-	-	1	47	-	58	105
B	3	-	6	9	-	-	-	-	46	7	84	137
C	4	-	-	4	-	-	-	-	3	2	10	15
D	5	-	2	7	-	-	-	-	10	-	24	34
E	7	3	5	15	4	-	2	6	75	3	47	125
F	-	-	3	3	3	-	2	5	53	-	37	90
G	35	1	23	59	-	-	1	1	28	3	96	127
H	5	-	4	9	2	-	-	2	33	-	82	115
J	5	-	6	11	-	-	-	-	91	4	98	193
K	12	2	13	27	1	1	2	4	6	3	24	33
L	8	-	5	13	1	-	5	6	6	1	9	16
M	2	-	2	4	-	1	-	1	-	-	6	6
N	3	2	3	8	6	3	-	9	5	1	18	24
P	11	2	6	19	4	-	2	6	4	5	15	24
R	1	-	1	2	1	-	1	2	6	-	5	11
	110	11	83	204	23	5	15	43	413	29	613	1055

I = independent declarative clauses

Q = questions

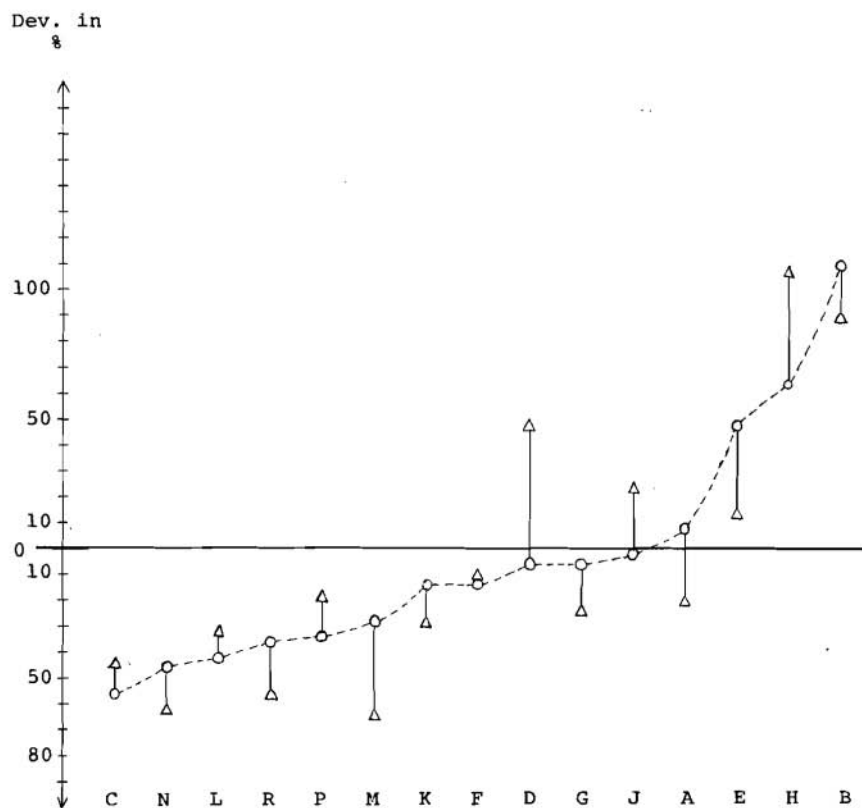
S = subordinate clauses

T = total number of occurrences

Figure 2 Relative deviation of absolute frequency from expected frequency: SHOULD. The categories of the LOB Corpus have been ranked from negative to positive deviation.

○ = LOB Corpus

△ = Brown Corpus



# NOTES

- 1 Under 'other forms' in Table 1 we have included occasional spellings like *shoulda*, *wouldya*, *wouldna* (found mainly in dialogue in imaginative prose).
- 2 In the rest of this paper we shall use capital letters to refer to the occurrences of each major type in our material, e.g. SHALL to include *shall* and *shan't*, SHOULD to include *should* and *shouldn't*. Italicized forms will be used in the general discussion of the auxiliaries.
- 3 The number of occurrences of uncontracted negative forms was:

	Brown	LOB
shall not	16	28
should not	64	95
will not	109	108
would not	172	188
- 4 The text categories of the LOB Corpus are listed on page 4 of this issue of *ICAME NEWS*. The category division is, with minor exceptions, the same as in the Brown Corpus (cf. Johansson et al. 1978). The expected frequency in Tables 2 and 3 is the number we would expect assuming an even distribution of the forms in the text categories of the corpora.
- 5 As Figure 1 does not say anything about the number of occurrences of SHALL and the size of each category, it is necessary to consult Table 2 in order not to overestimate the differences indicated by the diagram. Thus the difference between the two corpora in the case of category M is hardly more interesting than the similarity indicated for category R, since these categories are small and the examples found are very few.
- 6 The close resemblance in the shape of the curves for SHALL and SHOULD in Figures 1 and 2 does not reflect the degree of similarity with respect to the type of text. In both figures the categories of the LOB Corpus have been ranked from negative to positive deviation. Note that the ordering of the categories from left to right differs in the two figures.
- 7 It should be noted, however, that six of the examples in the Brown Corpus occur in quotations from the Bible, such as 'Whom shall I fear', in category D. This leaves us with only 19 examples, as against 32 in the LOB Corpus.
- 8 The percentages have been calculated on the basis of the figures for *shall* in the first person, indicated separately for independent declarative clauses, questions, and subordinate clauses in Luebke (1929: 454) and in Fries (1925:1012-15).

# REFERENCES

- Coates, J. and G. Leech. 1980. 'The Meanings of the Modals in Modern British and American English'. *York Papers in Linguistics* 8. 23-34.
- Evans, B. and C. Evans. 1957. *A Dictionary of Contemporary American Usage*. New York: Random.
- Forgue, G.J. and R.I. McDavid, Jr. 1972. *La Langue des Américains*. Paris: Aubier Montaigne.
- Fowler, H.W. 1965. *A Dictionary of Modern English Usage*, 2nd ed. Revised by E. Gowers. Oxford: Clarendon Press.
- Fries, C.C. 1925. 'The Periphrastic Future with *Shall* and *Will* in Modern English', *Publications of the Modern Language Association* 40. 963-1024.
- Fries, C.C. 1940. *American English Grammar*. New York: Appleton-Century-Crofts.
- Jacobsson, B. 1962a. 'A Note on the Use of *Will* in Questions in the First Person'. *Moderna Språk* 56. 17-21.
- Jacobsson, B. 1962b. 'A Final Remark on the Use of *Will* in Questions in the First Person'. *Moderna Språk* 56. 397-400.
- Jespersen, O. 1931. *A Modern English Grammar on Historical Principles*. Part IV. London: George Allen & Unwin Ltd.
- Johansson, S., Leech, G.N. and H. Goodluck. 1978. *Manual of Information to Accompany the Lancaster-Oslo/Bergen Corpus of British English, for Use with Digital Computers*. Department of English, University of Oslo.
- Joos, M. 1964. *The English Verb: Form and Meanings*. Madison: The University of Wisconsin Press.
- Krapp, G.P. 1925. *The English Language in America*. Vol. 2. New York: Century.
- Krovgig I. 1981. *Shall, Will, Should, and Would in Present-Day American and British English*. With Special Reference to *Shall* and *Should* in British English. Unpublished 'hovedfag' thesis, Department of English, University of Oslo.
- Leech, G.N. 1971. *Meaning and the English Verb*. London: Longman.
- Luebke, W.F. 1929. 'The Analytic Future in Contemporary American Fiction'. *Modern Philology* 26. 451-57.
- Mencken, H.L. 1936. *The American Language*. 4th ed. New York: Alfred A. Knopf.
- Myers, L.M. 1959. *Guide to American English*. 2nd ed. Englewood Cliffs, N.J.: Prentice-Hall.
- Poutsma, H. 1926. *A Grammar of Late Modern English*. Part II. Section II. Groningen: P. Noordhoff.
- Quirk, R., Greenbaum, S., Leech, G.N. and J. Svartvik. 1979. *A Grammar of Contemporary English*. 8th impr. (corrected). London: Longman.
- Švejc, A.D. 1978. *Standard English in the United States and England*. The Hague: Mouton.

- Taglicht, J. 1970. 'The Genesis of the Conventional Rules for the Use of *Shall* and *Will*'. *English Studies* 51. 193-213.
- Taubitz, R. 1978. 'British and American English: Some Differences and Their Implications for the EFL Teacher'. *Die Neueren Sprachen* 2. 159-64.
- Zandvoort, R.W. 1968. 'American English'. In A.N.J. den Hollander and S. Skard, eds., *American Civilization: An Introduction*. London: Longman. 375-88.
- Yule, G.U. 1944. *The Statistical Study of Literary Vocabulary*. Cambridge: Cambridge University Press.

## ICAME PROJECTS

### THE BROWN CORPUS

W. Nelson Francis and Henry Kučera are publishing a book called *Frequency Analysis of English Usage: Vocabulary and Grammar*, based on the grammatically tagged version of the Brown Corpus. Other publications using or commenting on the Brown Corpus (and not included in the bibliography in *ICAME NEWS 2*) are:

- Elsness, Johan. 1981. "On the Syntactic and Semantic Functions of *That*-Clauses". In *Papers from the First Nordic Conference for English Studies, Oslo, 17-19 September, 1980*, ed. Stig Johansson and Bjørn Tysdahl. Institute of English Studies, University of Oslo. 281-303.
- Francis, W. Nelson. 1980. "A Tagged Corpus--Problems and Prospects". In *Studies in English Linguistics for Randolph Quirk*, ed. Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. London: Longman. 192-209.
- Francis, W. Nelson and Henry Kučera. 1979. *Manual of Information to Accompany a Standard Sample of Present-Day Edited American English, for Use with Digital Computers*. Revised and augmented edition. Providence, R.I.: Department of Linguistics, Brown University.
- Kajita, M. 1968. *A Generative-Transformational Study of Semi-Auxiliaries in Present-Day English*. Tokyo.
- Kiyokawa, Hideo. 1978. A Statistical Analysis of American English (1). Tokyo: Shukutoku Daigaku Kenkyu Kiyo, No. 13. (in Japanese)
- Kjellmer, Göran. 1980. "Accustomed to swim: accustomed to swimming. On Verbal Form after TO". In *ALVAR. A Linguistically Varied Assortment of Readings. Studies Presented to Alvar Ellegård on the Occasion of His 60th Birthday*, ed. Jens Allwood and Magnus Ljung. Stockholm Studies in English Language and Literature 1. Department of English, University of Stockholm. 75-99.
- Kučera, Henry. 1980. "Computational Analysis of Predicational Structures in English". In *Proceedings of the 8th International Conference on Computational Linguistics, Sept. 30-Oct. 4, 1980*, Tokyo.
- Lynch, M.F. and S.D. Rawson. 1976. "Equiprevalent Character Strings --A Novel Text Characterization Method". In *The Computer in Literary and Linguistic Studies*, ed. A. Jones and R.F. Churchhouse. Cardiff: University of Wales Press. 47-58.
- Sahlin, Elisabeth. 1979. *Some and Any in Spoken and Written English*. Acta Universitatis Upsaliensis: Studia Anglistica Upsaliensia 38. Uppsala: Almqvist & Wiksell.
- Solso, R.L. and J.F. King. 1976. "Frequency and Versatility of Letters in the English Language", *Behavior Research Methods & Instrumentation* 8, 283-86.

- Solso, R.L., P.F. Barbuto, Jr., and C.L. Juel. 1979. "Bigram and Trigram Frequencies and Versatilities in the English Language", *Behavior Research Methods & Instrumentation* 11, 475-84.
- Solso, R.L. and C.L. Juel. 1980. "Positional Frequency and Versatility of Bigrams for Two- through Nine-Letter English Words", *Behavior Research Methods & Instrumentation* 12, 297-343.
- Yates, A.R. 1977. Text Compression in the Brown Corpus Using Variety-Generated Keysets, with a Review of the Literature on Computers in Shakespearean Studies. M.A. dissertation, University of Sheffield.
- Zettersten, Arne. 1978. *A Word Frequency List Based on American English Press Reportage*. Publications of the Department of English, University of Copenhagen, Vol. 6. Copenhagen: Akademisk Forlag.

Work in progress:

- Bäcklund, Ingegerd: English Non-Finite and Verbless Clauses. (Department of English, University of Uppsala)
- Fåhraeus, Ann-Mari: Degree Words in Absolute Use. (Department of English, University of Uppsala)

Some publications using both the Brown Corpus and the LOB Corpus are mentioned below.

#### THE LOB CORPUS

Grammatical tagging of the LOB Corpus is in progress, in co-operation between the University of Lancaster, the University of Oslo, and the Norwegian Computing Centre for the Humanities. The project is funded by the Social Science Research Council and the Norwegian Research Council for Science and the Humanities. A book by Knut Hofland and Stig Johansson on *Word Frequencies in British and American English* (based on the LOB Corpus and the Brown Corpus) is being printed and will appear shortly after the distribution of this newsletter; see the enclosed brochure. Other work using the LOB Corpus includes:

- Coates, Jennifer and Geoffrey Leech. 1980. "The Meanings of the Modals in Modern British and American English", *York Papers in Linguistics* 8, 23-34. (uses the LOB Corpus and the Brown Corpus)
- Engels, L.K., van Beckhoven, B., Leenders, Th., and I. Brasseur. 1981. *Leuven English Teaching Vocabulary-List Based on Objective Frequency Combined with Subjective Word-Selection*. Department of Linguistics, Catholic University of Leuven. (includes word frequency lists for the Leuven Drama Corpus, the LOB Corpus, and the Brown Corpus)
- Johansson, Stig. 1980. "The LOB Corpus of British English Texts:



Presentation and Comments", *ALLC Journal* 1:1, 25-36.

- Johansson, Stig. 1980. "Corpus-Based Studies of British and American English". In *Papers from the Scandinavian Symposium on Syntactic Variation, Stockholm, May 18-19, 1979*, ed. S. Jacobson. Stockholm Studies in English 52. Stockholm: Almqvist & Wiksell. 85-100. (uses the LOB Corpus and the Brown Corpus)
- Johansson, Stig. 1980. "Word Frequencies in British and American English: Some Preliminary Observations". In *ALVAR. A Linguistically Varied Assortment of Readings. Studies Presented to Alvar Ellegård on the Occasion of His 60th Birthday*, ed. Jens Allwood and Magnus Ljung. Stockholm Studies in English Language and Literature 1. Department of English, University of Stockholm. 56-74. (uses the LOB Corpus and the Brown Corpus)
- Johansson, Stig. 1980. *Plural Attributive Nouns in Present-Day English*. Lund Studies in English 59. Lund: C.W.K. Gleerup. (uses the LOB Corpus and the Brown Corpus)
- Krogvig, Inger. 1981. *Shall, Will, Should, and Would in Present-Day American and British English. With Special Reference to Shall and Should in British English*. Unpubl. "hovedfag" thesis, Department of English, University of Oslo. (uses the LOB Corpus and the Brown Corpus)
- Leech, Geoffrey and Jennifer Coates. 1980. "Semantic Indeterminacy and the Modals". In *Studies in English Linguistics for Randolph Quirk*, ed. Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. London: Longman. 79-90. (uses the LOB Corpus and the Brown Corpus)
- Sørheim, Mette-Cathrine Jahr. 1980. The s-Genitive in Present-Day English. Unpubl. "hovedfag" thesis, Department of English, University of Oslo. (uses the LOB Corpus and the Brown Corpus)
- Sørheim, Mette-Cathrine Jahr. 1981. "The Genitive in a Functional Sentence Perspective". In *Papers from the First Nordic Conference for English Studies, Oslo, 17-19 September, 1980*, ed. Stig Johansson and Bjørn Tysdahl. Institute of English Studies, University of Oslo. 405-23.

See further the articles included in this issue of *ICAME NEWS*.

An international symposium on "Computer Corpora in Research and Teaching" was held in Bergen on June 1-3 at the Norwegian Computing Centre for the Humanities. Papers from the symposium will be printed in a volume on *Computer Corpora in English Language Research*. Preliminary list of contents:

- John McH. Sinclair "Computer Corpora in English Language Research"
- Willem Meijs and Gert van der Steen "Exploring Brown with QUERY"
- Jan Aarts and Theo van den Heuvel "Current Work on the Dutch Computer Corpus Pilot Project"
- Jan Svartvik and Mats Eeg-Olofsson "Grammatical Tagging of the

## London-Lund Corpus"

Stig Johansson and Mette-Cathrine Jahr "Predicting Word Class from Word Endings"

The book will be published by the Norwegian Computing Centre for Science and the Humanities.

## THE LONDON-LUND CORPUS

The complete London-Lund Corpus, representing educated spoken British English, is now available from Bergen on magnetic computer tape. The Corpus was compiled and transcribed at University College London under the direction of Randolph Quirk and has been prepared for the computer by Jan Svartvik and his co-workers at the University of Lund. It consists of 87 'texts', each of some 5,000 running words, with detailed prosodic marking. The prosodic analysis includes such basic distinctions as tone unit, nucleus, booster, onset, and stress (see *ICAME NEWS* 3). The following main categories of texts are represented:

- spontaneous, surreptitiously recorded conversations
- non-surreptitious public conversations
- non-surreptitious private conversations
- telephone conversations
- spontaneous commentary (sports and non-sports)
- spontaneous oration (speeches in court, political speeches etc.)
- prepared but unscripted oration (sermons, lectures etc.)

Also available are KWIC concordances for the material:

London-Lund KWIC I: a complete concordance for the 34 texts representing spontaneous, surreptitiously recorded conversation (text categories 1-3), made available both in computerised and printed form (J. Svartvik and R. Quirk (eds.), *A Corpus of English Conversation*, 1980).

London-Lund KWIC II: a complete concordance for the remaining 53 texts of the London-Lund Corpus (text categories 4-12)

Some publications using the London-Lund Corpus are:

Oreström, Bengt, Svartvik, Jan, and Cecilia Thavenius. 1976. Manual for Terminal Input of Spoken English Material. SSE Report.

Oreström, Bengt. 1977. Why "/di/ book?". SSE Report.

- Oreström, Bengt. 1977. Supports in English. SSE Report.
- Oreström, Bengt. 1978. Turn-Taking: On Speaker-Shift in Face-to-Face Conversation. SSE Report.
- Oreström, Bengt. 1980. Turn Length in Face-to-Face Conversation. SSE Report.
- Oreström, Bengt and Cecilia Thavenius. 1978. Auditory and Acoustic Analysis: An Experiment. SSE Report.
- Quirk, Randolph and Jan Svartvik. 1978. "A Corpus of Modern English". *Empirische Textwissenschaft. Aufbau und Auswertung von Text-Corpora*, ed. H. Bergenholz & b. Schaefer. Königstein: Scriptor. 204-218.
- Svartvik, Jan. 1976. Projektet Engelskt talspråk. SSE Report.
- Svartvik, Jan. 1980. Computer-Aided Grammatical Tagging of Spoken English. SSE Report.
- Svartvik, Jan. 1980. "Well in Conversation". *Studies in English Linguistics for Randolph Quirk*, ed. S. Greenbaum, G. Leech and J. Svartvik. London: Longman. 167-177.
- Svartvik, Jan. 1980. "Interactive Parsing of Spoken English". In *Proceedings of the 8th International Conference on Computational Linguistics*, Sept. 30-Oct. 4, 1980, Tokyo.
- Svartvik, Jan. 1980. "Tagging Spoken English" In *ALVAR. A Linguistically Varied Assortment of Readings. Studies presented to Alvar Ellegård on the Occasion of His 60th Birthday*, ed. Jens Allwood and Magnus Ljung. Stockholm Studies in English Language and Literature 1. Department of English, University of Stockholm. 182-206.
- Svartvik, Jan and Randolph Quirk (eds.). 1980. *A Corpus of English Conversation*. Lund Studies in English 56. Lund: Liber.
- Thavenius, Cecilia. 1977. A Select Bibliography of Studies in Spoken English. SSE Report.
- Thavenius, Cecilia. 1978. A Pilot Study of Referential 'it' in Spoken English. SSE Report.
- Thavenius, Cecilia and Bengt Oreström (eds.). 1979. Konkordanser: Föredrag från 2:a svenska kollokviet i språklig databehandling i Lund 1979. SSE Report.
- Work in progress:
- Oreström, Bengt: Turn-Taking and Interruption in Face-to-Face Conversation.
- Stenström, Anna-Brita: Questions and Answers in English Conversation.
- Thavenius, Cecilia: Reference in English Conversation. The Pragmatics of Third Person Reference.

## MATERIAL AVAILABLE FROM BERGEN

Apart from the London-Lund texts and KWIC concordances just mentioned, the following material is available from Bergen:

Brown Corpus, text format I: Typographical information is preserved; the same line division is used as in the original version from Brown University except that words at the end of the line are never divided.

Brown Corpus, text format II: Typographical information is reduced; the line division is new.

LOB Corpus: text.

Also available are KWIC concordances for the LOB Corpus and the Brown Corpus (on tape and microfiche). The microfiche set for the Brown Corpus, but *not* for the LOB Corpus, includes the complete text of the corpus. A printed manual accompanies the tape of the LOB Corpus. Printed manuals for the Brown Corpus cannot be obtained from Bergen.

The material has been described in greater detail in previous issues of *ICAME NEWS*. Prices and technical specifications are given on the order forms which accompany this newsletter.

The grammatically tagged version of the Brown Corpus can only be ordered from: Henry Kučera, TEXT RESEARCH, 196 Bowen Street, Providence, R.I. 02906, U.S.A.

## EDITORIAL NOTE

Further ICAME newsletters will appear irregularly and will, for the time being, be distributed free of charge. The Editor is grateful for any information or documentation which is relevant to the field of concern of ICAME.



ICAME NEWS is published by the Norwegian Computing Centre  
for the Humanities.  
Address: Harald Hårfagresgate 31, P. O. 53, 5014 Bergen - University, Norway