

Modal verbs in TIME: Frequency changes 1923 - 2006

Neil Millar (2009)

Lancaster University, UK

International Journal of Corpus Linguistics 14:2

Background

- Many diachronic studies use corpora that span centuries (ARCHER, Helsinki)
- Coverage of recent changes (20th century) in English typically use Brown corpora
 - Brown, Frown, LOB, FLOB
 - Limited to 30 year time period (1961 and 1991)
- Study by Leech (2003) found modal verbs declined between 1961 and 1991

Brown study results

	LOB 1961	FLOB 1991	Log likhd	% diff. 1961-91	Brown 1961	Frown 1991	Log likhd	% diff. 1961-91
<i>would</i>	3,028	2,694	20.4	-11%	3,053	2,868	5.6	-6.1%
<i>will</i>	2,798	2,723	1.2	-2.7%	2,702	2,402	17.3	-11.1%
<i>can</i>	1,997	2,041	0.4	+2.2%	2,193	2,160	0.2	-1.5%
<i>could</i>	1,740	1,782	2.4	+2.4%	1,776	1,655	4.1	-6.8%
<i>may</i>	1,333	1,101	22.8	-17.4%	1,298	879	81.1	-32.4%
<i>should</i>	1,301	1,147	10.1	-11.8%	910	787	8.8	-13.5%
<i>must</i>	1,147	814	57.7	-29.0%	1,018	668	72.8	-34.4%
<i>might</i>	777	660	9.9	-15.1%	635	635	0.7	-4.5%
<i>shall</i>	355	200	44.3	-43.7%	267	150	33.1	-43.8%
<i>ought (to)</i>	104	58	13.4	-44.2%	70	49	3.7	-30.0%
<i>need(n't)</i>	87	52	9	-40.2%	40	35	0.3	-12.5%

Table 1. Frequencies of modals in the Brown corpora (Leech 2003: 228)

Millar's Motivation

- To ...
 - Determine how ‘the frequencies of modal verbs behaved across the 20th century as a whole’
 - Determine to what degree language change can be “extrapolated from two data points separated by 30 years” [methodological issues]
 - See if the trends observed by Leech hold true given a larger corpus, specifically, using the much larger TIME Corpus (Davies, 2008)

Different approaches

- Breadth and depth
- Single-genre vs. language-in-its-entirety
- American English only

	FROWN	Brown Corpus	TIME Magazine Corpus
Total words	1 million	1 million	100 million
Text description	15 different categories ⁵	15 different categories	1 single category 275,000 articles
Time period	1991	1961	1923-2006
	1 year	1 year	~83 years
Text words/year	1 million	1 million	1-1.5 million words

Modal verbs

- Auxiliary verbs that express:
 - Permission, ability, or obligation (deontic modality)
 - “And if I may be so bold as to offer a word of advice to you...”
 - “The dirigible can meet all emergencies except fire.”
 - “The purpose of education should be truth and not theories.”
 - Probability, possibility (epistemic modality)
 - “The X-ray may prove of more value in treatment of whooping-cough than any other remedy, including vaccine.”
- could, may, might, must, shall, should, will, would

Examples from the TIME corpus, 1923, simple search using [vm]

TIME Magazine Corpus

- 83 years of text (from March, 1923 to 2006)
- Each year represented by 1-1.5 million words
- Over 100,000,000 words
- 275,000 articles
- Marked-up using CLAWS¹⁷
- Searchable by POS tag
- <http://corpus.byu.edu/time/>

1. Constituent Likelihood Automatic Word-tagging System

Modal verb frequencies per decade, TIME magazine

	1920s	1930s	1940s	1950s	1960s	1970s	1980s	1990s	2000s	% diff. 1920s– 2000s
<i>will</i>	2,194.63	1,681.76	1,856.4	1,988.37	1,965.76	2,135.73	2,057.43	2,273.23	2,362.52	+7.7%
<i>would</i>	1,690.70	1,665.01	2,095.76	1,669.18	1,513.30	1,828.92	1,758.44	1,797.03	1,693.19	+0.1%
<i>can</i>	832.91	742.3	955.73	1,093.39	1,233.13	1,305.82	1,231.99	1,475.95	1,777.07	+113.4%
<i>could</i>	661.33	822.72	1,188.24	998.83	950.73	1,106.25	1,156.61	1,378.39	1,342.56	+103.0%
<i>may</i>	583.59	515.12	496.93	502.74	628.13	743.66	775.92	937.08	931.91	+59.7%
<i>should</i>	577.46	450.07	454.87	495.26	441.96	475.5	453.33	521.46	593.27	+2.7%
<i>must</i>	485.31	418.03	456.57	417.62	401.36	390.47	347.02	306.69	250.59	-48.4%
<i>might</i>	374.52	375.4	500.33	408.9	399.8	458.99	416.81	474.23	433.34	+15.7%
<i>shall</i>	212.19	120.79	96.42	70.52	50.48	35.65	25.93	16.09	9.26	-95.6%
<i>ought</i>	50.22	37.94	39.31	40.34	36.91	34.29	28.27	34.9	27.65	-44.9%
Total	7,662.86	6,829.14	8,140.56	7,685.15	7,621.56	8,515.28	8,251.75	9,215.05	9,421.36	+22.9%

Table 3. Modal verb frequencies per decade (words per million)

Frequency plots per decade

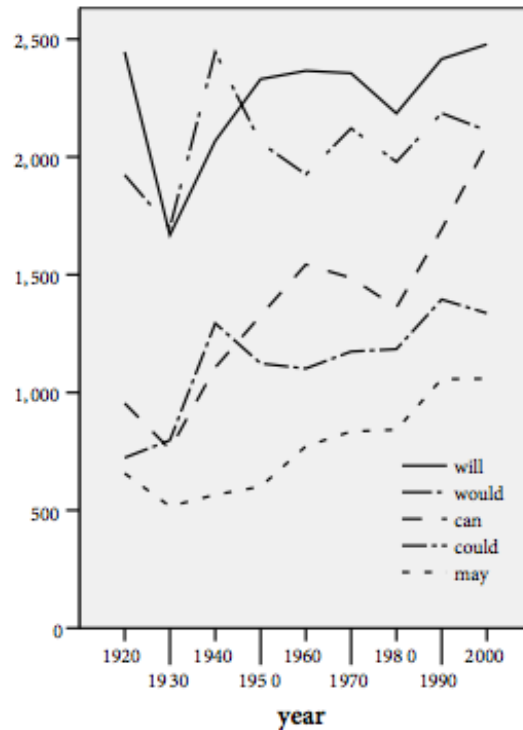


Figure 2a. High-mid frequency modals
(words per million)

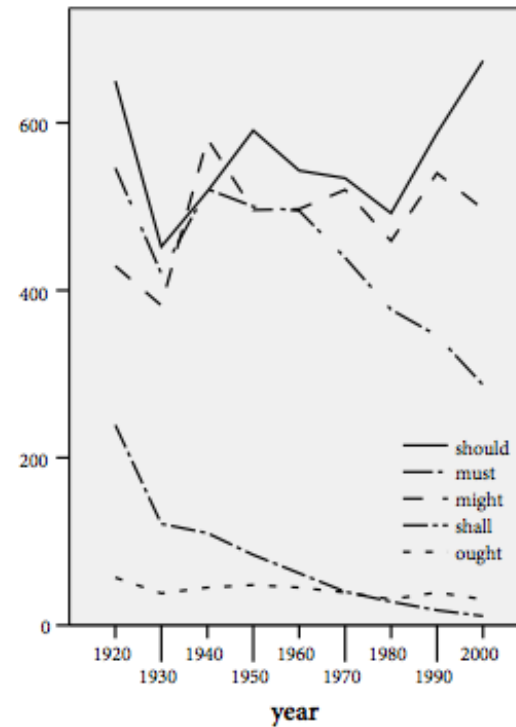
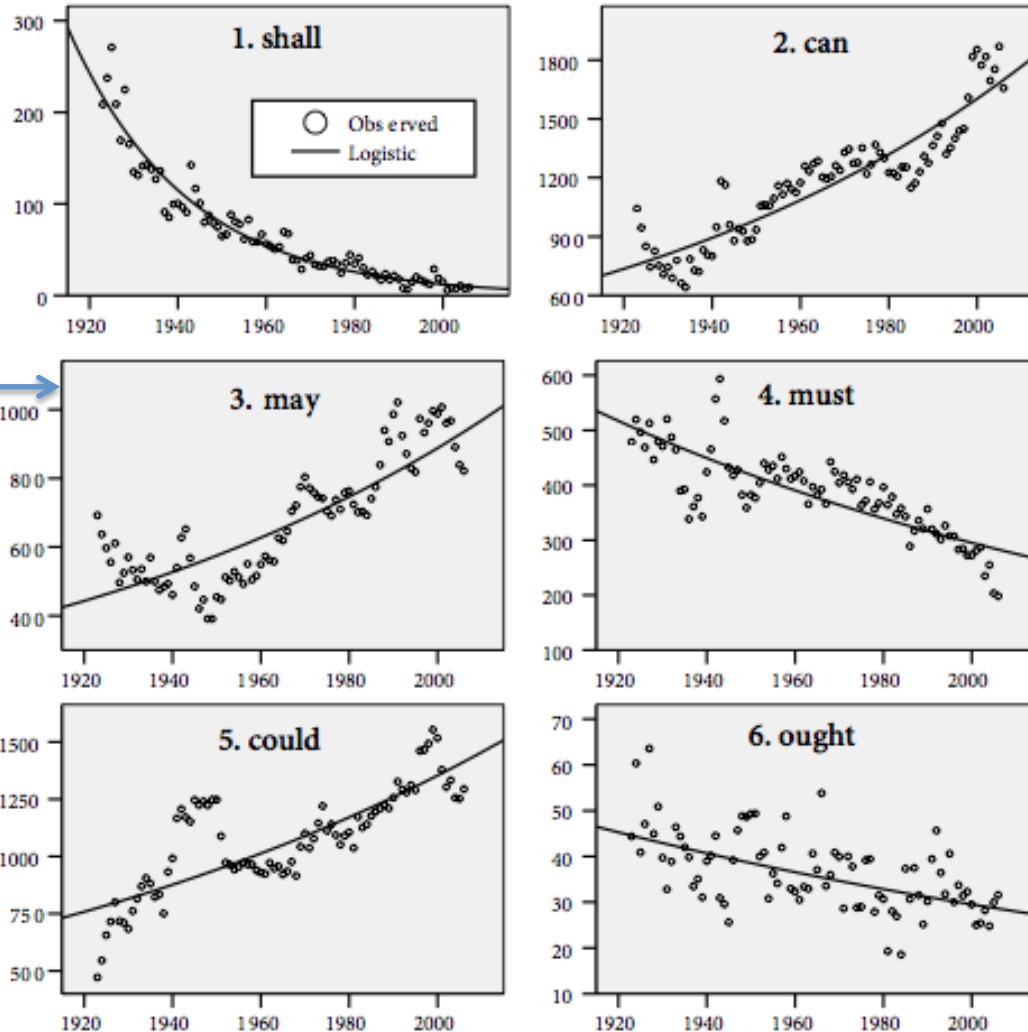


Figure 2b. Mid-low frequency modals
(words per million)

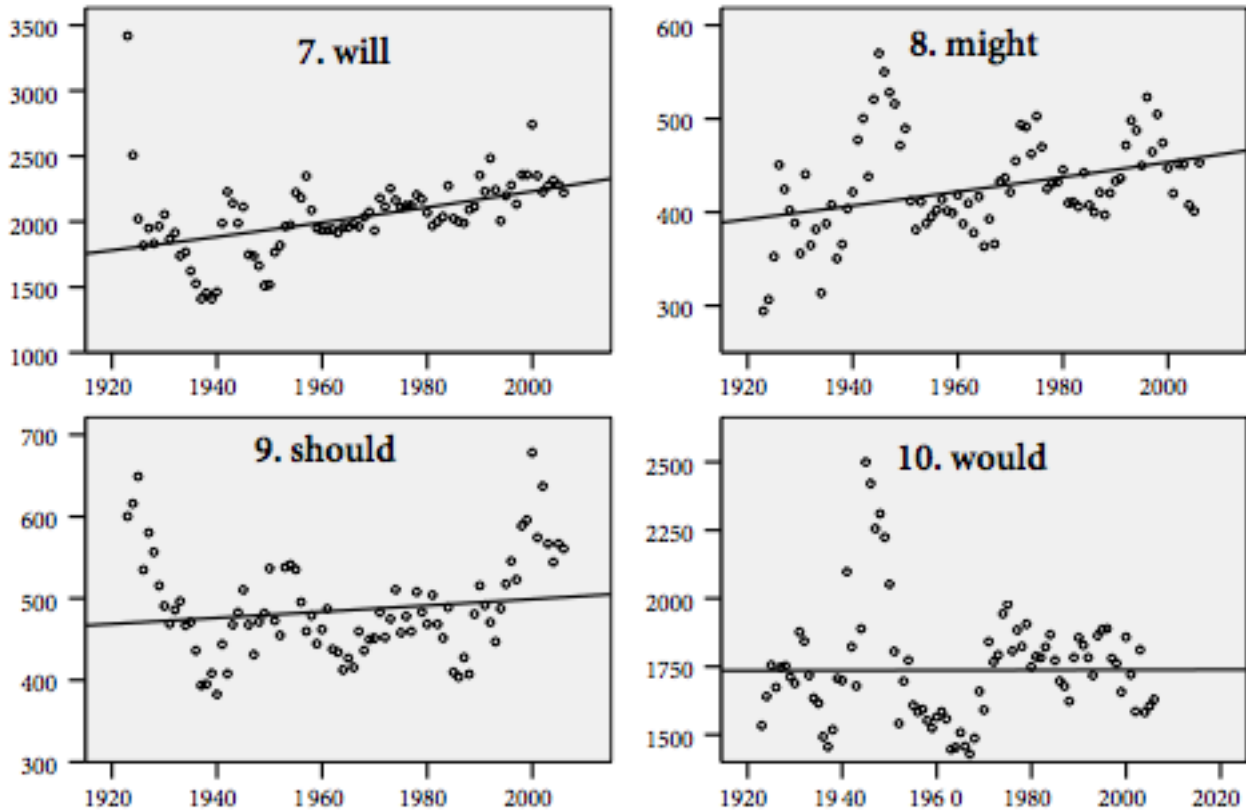
Pattern of change



*Opposite trend
cf. Leech study*

Note. x axis: year (1923-2006) ; y axis: frequency (words per million)

No discernible pattern



Note. x axis: year (1923-2006) ; y axis: frequency (words per million)

Observations

- Decline in frequency of “shall” and “must”
- Increase in frequency of “can,” “could”, and “may”
- supported by the high percentage of variation accounted for by the statistical model





Statistical Analysis

Table 4. Statistical analyses for modal auxiliary verbs

modal verb	<i>shall</i>	<i>can</i>	<i>may</i>	<i>must</i>	<i>could</i>	<i>ought</i>	<i>will</i>	<i>might</i>	<i>should</i>	<i>would</i>
Curve estimation regression statistics for a logistic model										
R^2	0.924	0.819	0.675	0.664	0.612	0.322	0.225	0.127	0.025	0.000
Significance	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.152	0.969
Binary logistic regression										
B	-0.036	0.010	0.090	-0.070	0.070	-0.050	0.030	0.020	0.010	0.000
<i>Nagelkerke R²</i>	0.031	0.102	0.030	0.010	0.020	0.010	0.000	0.000	0.000	0.000
Significance	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.380

Note. R^2 = coefficient of determination; B = regression coefficient

Comparison of Studies

Brown-Frown	1961-1991	TIME corpus	1920s-2000s
small decrease 	<i>can</i> -1.5%	large increase	<i>may</i> +59.7%
	<i>might</i> -4.5%		<i>can</i> +113.4%
	<i>would</i> -6.1%		<i>could</i> +103.0%
	<i>could</i> -6.8%	small increase	<i>will</i> +7.7%
	<i>will</i> -11.1%	fluctuation  	<i>might</i> +15.7%
	<i>should</i> -13.5%		<i>should</i> +2.7%
large decrease 	<i>ought</i> -30.0%	large decrease	<i>would</i> +0.1%
	<i>may</i> -32.4%		<i>ought</i> -44.9%
	<i>must</i> -34.4%		<i>must</i> -48.4%
	<i>shall</i> -43.8%		<i>shall</i> -95.6%

Results

- Modal verb “may” has increased considerably, not decreased (since 1923, in TIME)
- Examples of:

Selective semantic analysis

- Random sample of the data for modal verbs
 - may, should, must
- Over three temporal points
 - 1923, 1960, 2000
- Minimum one-third of all concordance lines semantically coded*
- * No further detail 😞

may

Table 5. Senses of *may*

<i>may</i>	1923	1960	2000
a. epistemic possibility	136 (56%)	255 (78%)	414 (94%)
b. root possibility	59 (24%)	28 (9%)	6 (1%)
c. permission	9 (4%)	25 (8%)	7 (2%)
d. quasi-subjunctive	19 (8%)	5 (2%)	5 (1%)
e. unclear	19 (8%)	14 (4%)	9 (2%)
Total	242	327	441

- E.g.:
- Forget to buy an alarm, and you may wake up without a couch...*
 - ...his death at this time may undoubtedly be traced to...*
 - ...it stipulates who may not marry whom...*
 - Whatever changes time may bring, it will never...*
 - There will be no material that may give offense...*

should

Table 6. Senses of *should*

<i>should</i>	1923	1960	2000
a. weak inference epistemic	32 (15%)	47 (17%)	37 (12%)
b. weak obligation (root/deontic)	105 (50%)	196 (72%)	242 (80%)
c. putative, quasi-subjunctive	60 (29%)	19 (7%)	11 (4%)
d. <i>should</i> = <i>would</i> (1st pers.)	3 (1%)	2 (1%)	0 (0%)
e. unclear	9 (4%)	9 (3%)	13 (4%)
Total	209	273	303

- E.g.:
- a. OS X's brand-new code *should* eliminate system crashes...
 - b. *Should* you pre-order one of the 1 million PlayStation 2s...?
 - c. "It's funny that you *should* ask."
 - d. *I should* like to describe the Tusitala.
 - e. O'Donnell *should* know: he's a former Senate staff member...

must

Table 7. Senses of *must*

<i>must</i>	1923	1960	2000
a. epistemic (necessity)	28 (14%)	44 (17%)	58 (29%)
b. root deontic	166 (83%)	207 (81%)	134 (67%)
c. unclear	6 (3%)	5 (2%)	8 (4%)
Total	200	256	200

- E.g.:
- “What a jerk that hacker must be.”*
 - A jury decided Disney must pay \$240 million to two plaintiffs...*
 - ...kids with dyslexia ... must work twice as hard...*

Readers letters, 1920s

- (1) *Sirs: I ought not to write at this time because I am angry and disappointed.*
(1925)
- (2) *Sirs: May I correct an inaccuracy in your otherwise letter-perfect issue of Aug. 3? (1925)*
- (3) *Sirs: Will you kindly advise me to what extent your typographic room is unionized, that I may judge how much your reports of the present British labor situation are influenced by such unionization? (1926)*

Readers' letters, 2007

- (4) *I don't think all the Republican candidates "had noisily rattled sabers about Iran." I can't even begin to imagine Ron Paul wanting to go to war with Iran. He didn't even want to go to war with Iraq. You did a great disservice to Paul by using such a blanket statement. (2007)*
- (5) *Why the excitement over the latest NIE report? Since Bush hasn't regarded intelligence in the past, why should he change now? (2007)*
- (6) *Morality is bunk. What separates man from the beasts is the capacity to reason. Yet we seldom do — we're too busy moralizing. Reason is the only frontier left. (2007)*

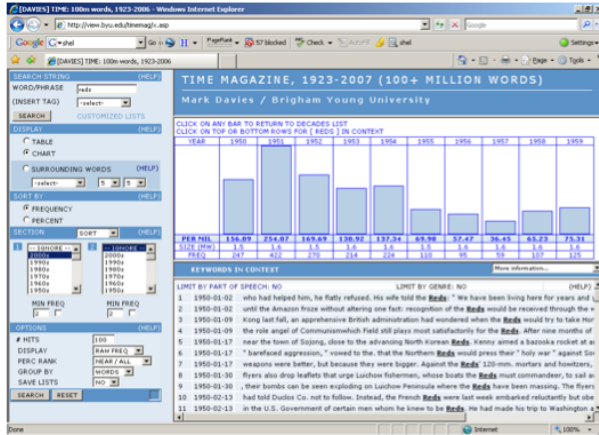
Conclusions

- Colloquialization
 - Stylistic drift of written language towards that of speech
 - Contractions as evidence--writers at TIME have shown an increasing preference over the past 83 years (Biber)
- Democratization of language
 - Speakers/writers suppressing overt claims to power and authority
 - Shift from formal -> informal
- Stylistic changes
 - Greater speculation in reporting
 - Frequency changes in *may* and *could* point to surge in epistemic modality

Methodological Issues

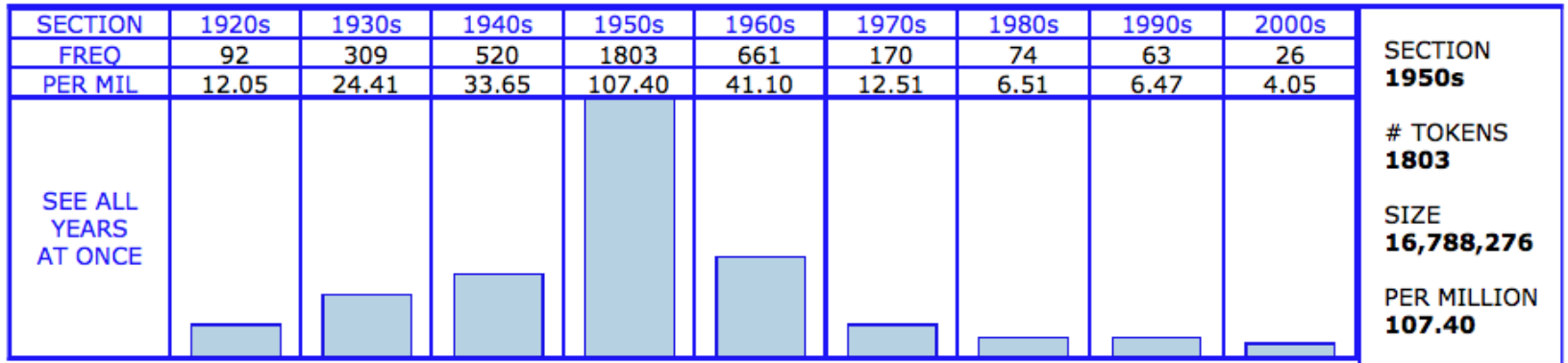
- Discrepancies due to different corpus type
 - Using only the “Press” sub-corpora from Brown and comparing to only 1961 and 1991 in TIME narrowed the differences in the numbers
 - But still, comparing two data points “produces a picture of change that contradicts the overall pattern
- Corpus size issues: the “Press” sub-corpora amounts to less than 180,000 words

Impressionist observations



Historical events reflected in the language in the corpus.

← Millar's search for "Reds"



↑ "McCarthyism"