

Has American English gotten any different?

Using the Brown family to track diachronic change in
American (and British) English

Paul Baker, Lancaster University

Research Questions

- 1. What words in American English are
 - a) increasing over time
 - b) decreasing over time
 - c) remaining stable over time
- 2. Are these patterns similar to British English?
- 3. To what extent American English leading change?
- 4. What do 1-3 tell us about changing language and culture?

Applications

- 1. Pedagogy – to keep learner language from looking “old-fashioned”. If certain words are declining in frequency, should they be given as much attention by teachers?
- 2. Cultural interest – how are societies changing in terms of what we write and how we write about them?

The Brown family

- Each corpus is 1 million words, 15 genres of writing.
- Similar sampling methods.
- The 3 American corpora: Brown (1961), Frown (1992), AmEo6 (2006)
- The 4 British corpora: BLOB (1931), LOB (1961), FLOB (1991), BEo6 (2006)

Sampling framework

Text category letter and description		Files	Sources
A	Press: Reportage	44	Nexis/Factiva
B	Press: Editorial	27	
C	Press: Reviews	17	
D	Religion	17	Religious and academic websites
E	Skills, Trades and Hobbies	36	Websites for magazines (Popular Mechanics, Elle, Family Circle etc)
F	Popular Lore	48	Simonandschuster.com, magazines
G	Belles Lettres, Biographies, Essays	75	Ebscohost.com, bookbrowse.com etc.
H	Miscellaneous: Government documents, industrial reports etc	30	Government (.gov) and company websites e.g. Ford Foundation
J	Academic prose in various disciplines	80	Metalib, Google scholar
K	General Fiction	29	Book publishers and authors websites e.g. Bookbrowse.com
L	Mystery and Detective Fiction	24	
M	Science Fiction	6	
N	Adventure and Western	29	
P	Romance and Love story	29	
R	Humour	9	

Issues in using the Brown family

- Pros

- Relatively easy to collect (BEo6 took 10 working days)
- Same sampling frame means comparisons have high validity

- Cons

- Small sample size means analysis better suited for high frequency phenomena
- Need to consider dispersion carefully (*hedonic*)
- ‘Snapshot’ corpora mean that the family has a low granularity (e.g. American Brown 3 sampling points, 45 year span, $\text{granularity}=0.066$)

Existing research on the British side

- Smith (2002) The progressive increasing between LOB '61 and FLOB '91
- Leech (2002) Most modal verbs decreasing between LOB and FLOB (except the weak *can* and *could*)
- deHaan (2002) Use of *whom* decreasing (but increasing for religious texts, press reportage, and editorials and skills, trades & hobbies)
- Baker (2009, 2010, 2011) BE – decreases in male pronouns, gradable adverbs, *on* replacing *upon*, *around* replacing *round*, more contractions and numerals

Cultural differences

- Leech and Fallon “a picture of US culture in 1961 – masculine to the point of machismo, militaristic, dynamic and actuated by high ideals, driven by technology, activity and enterprise – contrasting with one of British culture as more given to temporizing and talking, to benefiting from wealth rather than creating it, and to family and emotional life, less actuated by matters of substance than by status” (1992: 44-45)
- Oakes (2003) similar findings for 1990s FLOB/FROWN

The British lag

- Leech (2002) patterns of modal verbs in British English 1991 similar to American English in 1961
- McEnery and Xiao (2005) choice of the bare or full infinitive (*help/help to*) in British English 1991 similar to American English in 1961

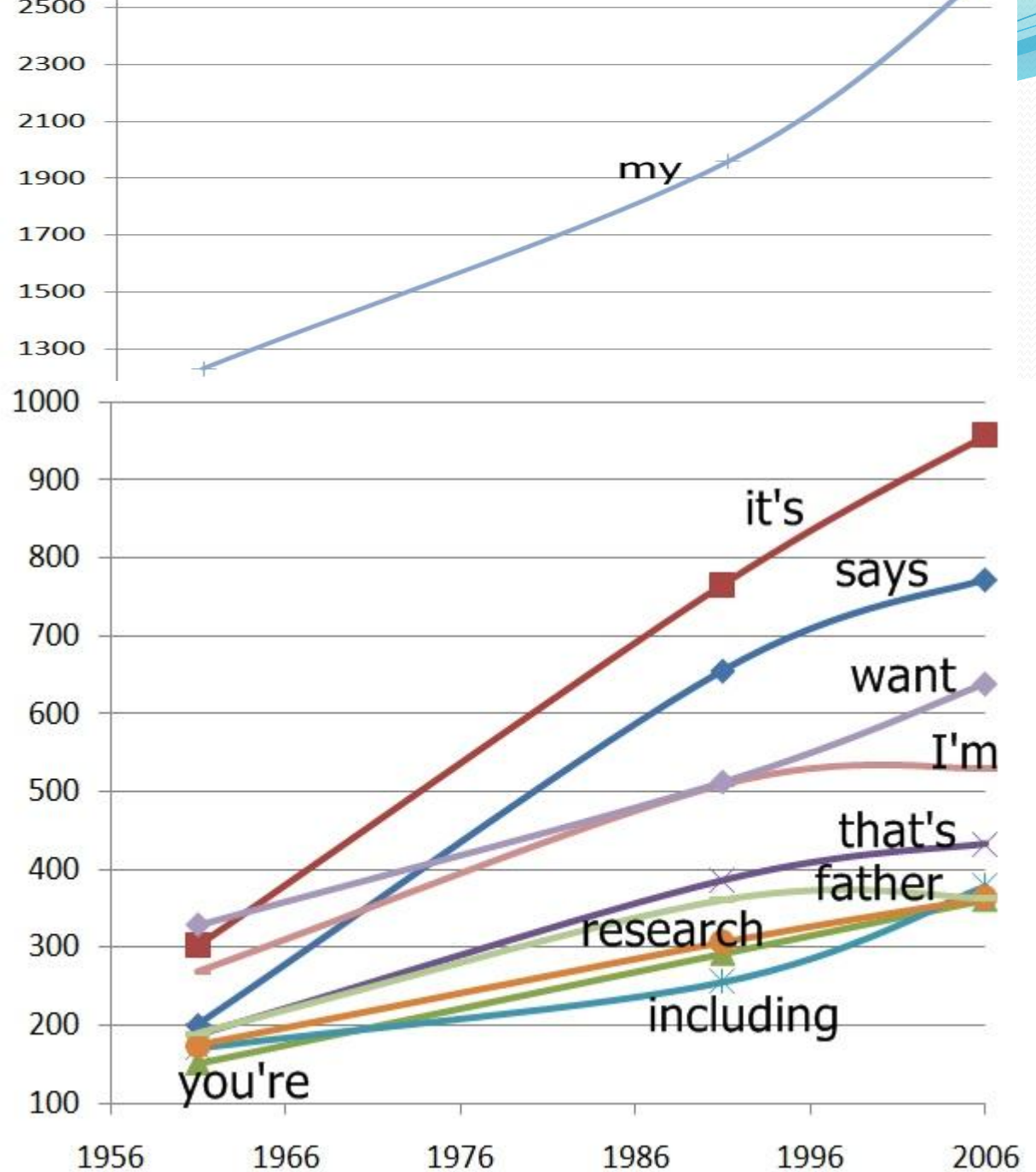
Identifying diachronic change/stability

File	Edit	View	Compute	Settings	Windows	Help
N	Word	Total	Brown_files	Frown_files	AmE06	
1	THE	195,518	70,002	62,418	63,098	
2	OF	100,664	36,472	32,300	31,892	
3	AND	87,519	28,935	28,026	30,558	
4	TO	79,659	26,238	26,193	27,228	
5	A	71,656	23,522	23,195	24,939	
6	IN	62,849	21,419	20,393	21,037	
7	#	43,365	10,898	13,740	18,727	
8	THAT	34,554	10,596	11,645	12,313	
9	IS	28,293	10,102	9,315	8,876	
10	FOR	28,052	9,499	9,235	9,318	
11	WAS	26,631	9,815	8,183	8,633	
12	HE	23,895	9,542	7,239	7,114	

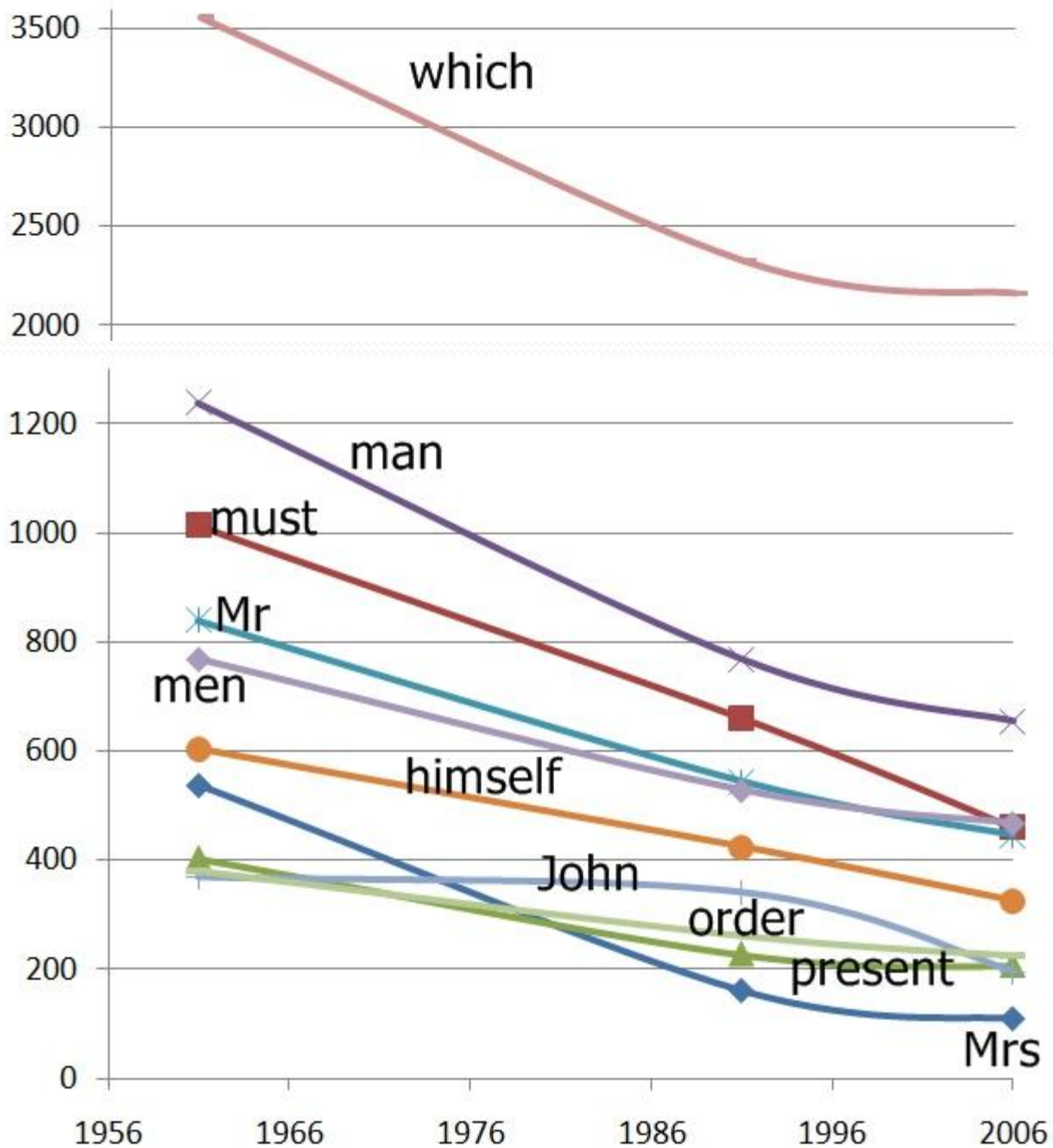
Method

- 1. Only frequent words considered - those with a mean of 250 or more across all three corpora. 386 words used (accounts for 61% of all tokens across the corpora)
- 2. Work out standard deviations for each word
E.g. *Mrs* (537, 160, 108) mean = 268, SD =191
having (279, 278, 260) mean=272, SD=9
- 3. Because SD increases with frequency, work out coefficient of variance (CV) for each word
 $CV = (\text{standard deviation}/\text{mean frequency}) * 100$
E.g. *the*: SD = 3426 CV=5
Mrs: SD = 191 CV= 71

Top 10 words with highest CVs (increasing)



Top 10 words with highest CVs (decreasing)



Must vs want to

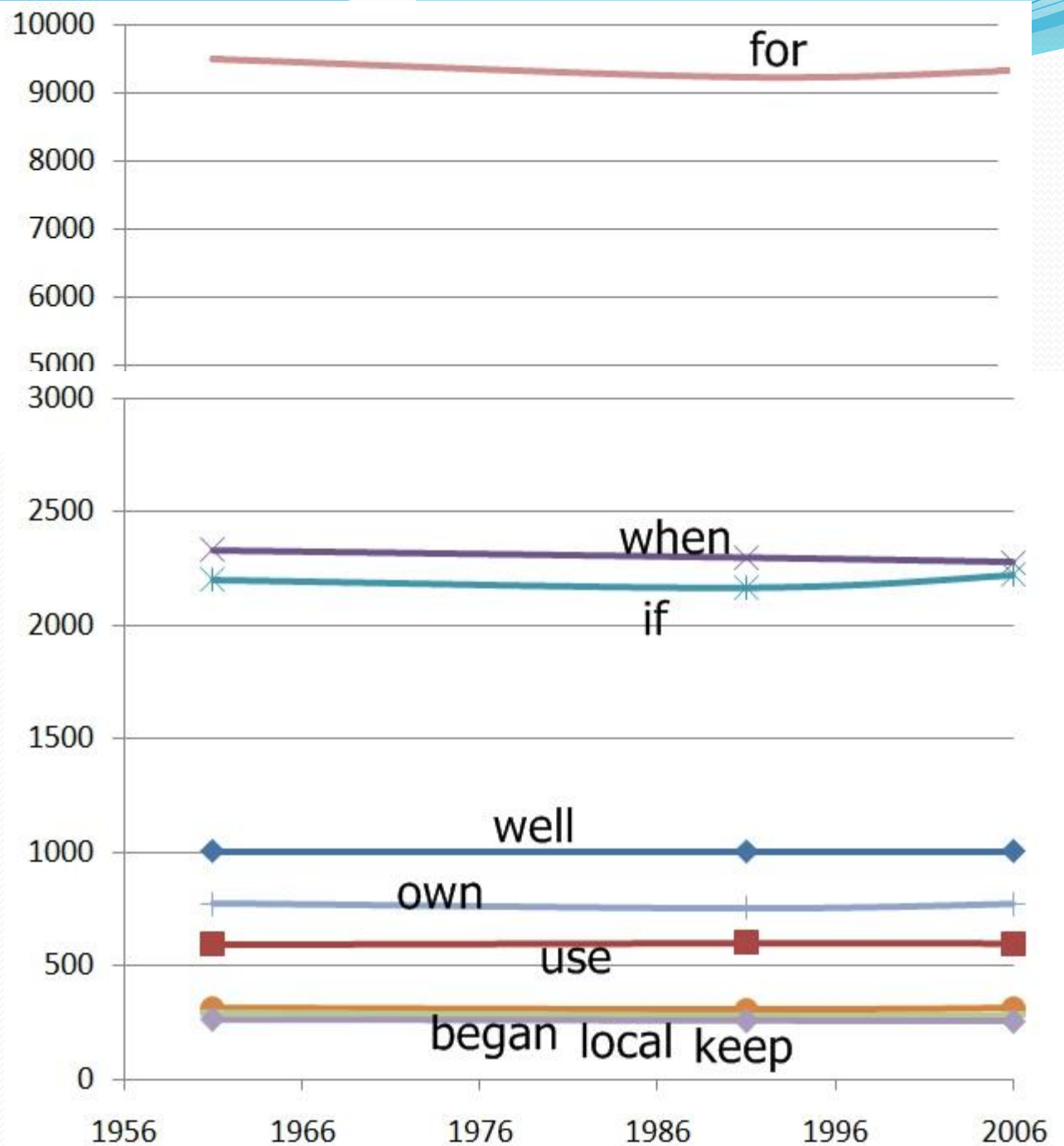
You **must** attach a statement to your return, if you take advantage of this automatic extension, showing that you were in Alaska or were outside the United States or Puerto Rico on April 15 or other due date.

Brown (1961)

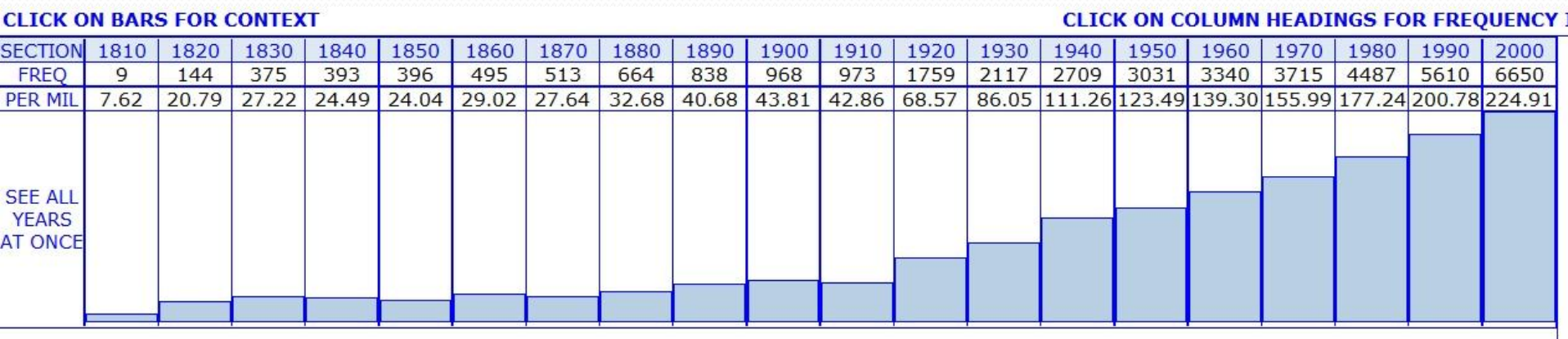
Maybe you **want to** stay here and help Meri get started

AEo6 (2006)

Top 10 words with lowest CVs (stable)



Checking against COHA



Standardised frequency of *including* (change from 1810 to 2010)

Corpus created by Mark Davies at BYU see: <http://corpus.byu.edu/coha/>

Checking against COHA

- Almost all increasing/decreasing words had same pattern when compared over the period 1810-2010.
- Three decreasing words: *man*, *men*, *Mrs* increased in the 19th century and fell in the 20th.
- Two increasing words: *my*, *father* decreased until about 1940 then increased.
- Three stable words showed decreases over the COHA (*well*, *if*, *own*).
- Two stable words showed increases over the COHA (*use*, *local*)
- Method – better at spotting long term increase/decrease. Stability should only be considered as short-term.

The overall picture

- 1. Divide most frequent 386 words into 3 equal-sized groups
- Highest CV = high change
 - Increasing words (*including*: 171, 256, 379)
 - Decreasing words (*must*: 1013, 661, 460)
 - No clear pattern (*major*: 256, 358, 232)
- Lowest CV = low change
(*well*: 1004, 1002, 1002)

Those in the middle → ignore for now

Categorising words

- Words put into a table based on their most common grammatical function (e.g. *mind* occurs collectively as 245 times as a verb and 1404 times as a noun).
- Grammatically tagged versions of the corpora used to help calculate these frequencies.
- Some categories based on CLAWS C7 tagset
<http://ucrel.lancs.ac.uk/claws7tags.html>
- Some categories added (e.g. abstract/concrete noun)

Grammatical words

	Increased (top third)	Decreased (top third)	Stayed stable (bottom third)
modality	need	must	might, could, can
pronouns	my, her, me, she, I, you, your	himself , he	they , its, their
terms of address		Mrs, Mr	
contractions/full forms	it's , you're, that's, I'm, didn't, don't		
prepositions	including, across		with, to, in, out, on, from, for, along , up, over, away, at , off, during, down, after , by
conjunctions		until	if, but, as , yet, than, or , since, and , not
determiners/articles/ numerals	# , million	thus, there, any, nothing	own , anything, some, many, two , least, an, a , first, less, enough, four, most , one, six, all
question words	how, why	which	when, who, where , whose
auxiliary verbs		be	are, have, having, do, were, has, being , does

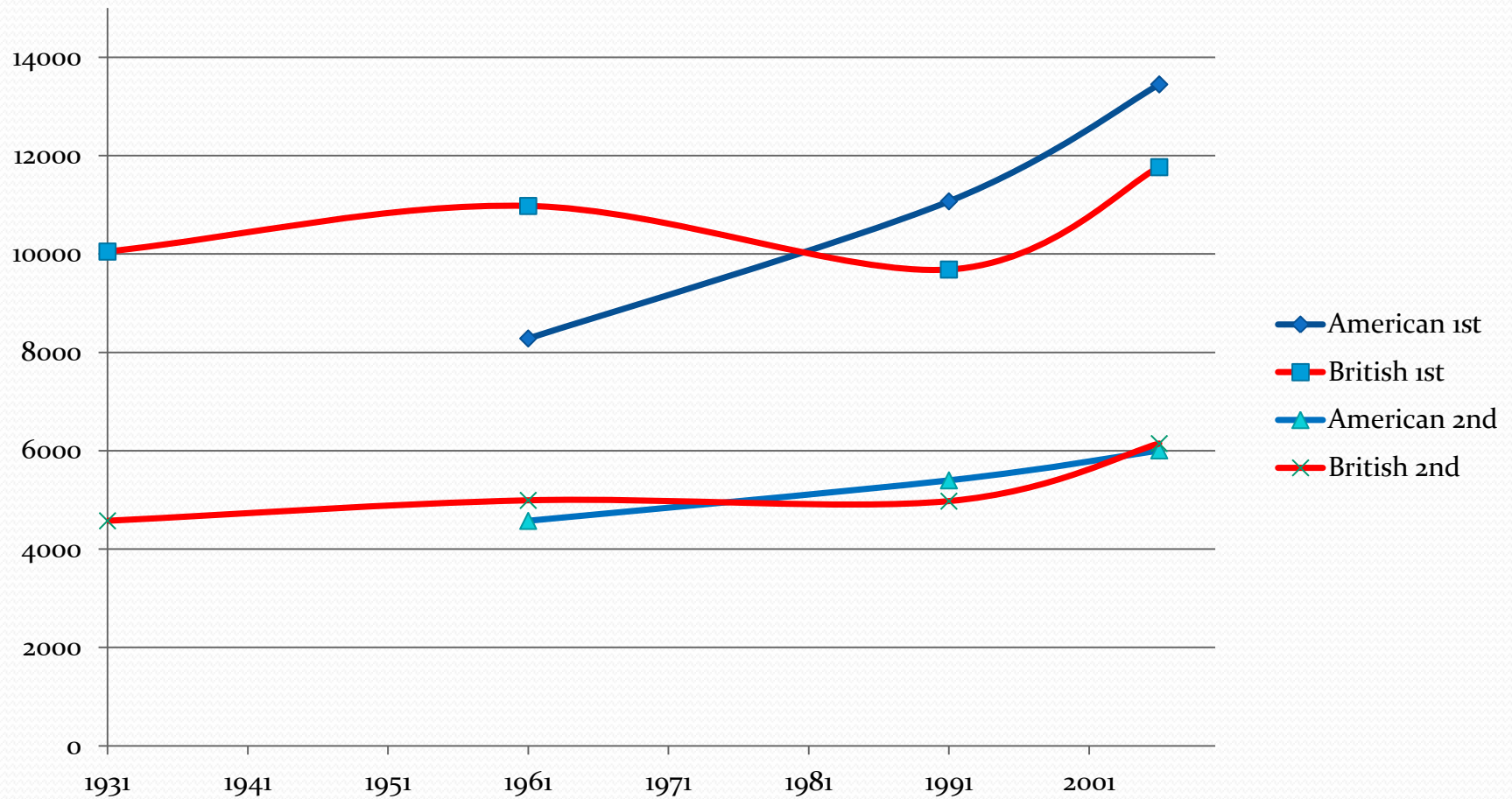
Lexical words

	Increased (top third)	Decreased (top third)	Stayed stable (bottom third)
lexical verbs	says , looking, going, love, want, wanted, think, said, looked, like	made , went, means , give	use, began, keep , seen, told, go , thought, act, make , help, used, turned, look, take , called , found, asked, saw, come, put , see , took , making
adverbials	just , really, about	almost	well , probably, ever, so, together, too , never , rather , here , perhaps, now , always , still , even
adjectives	hard, different, past	great , possible , whole , general , certain , present	local, best, another, other , left, good, kind, high, national, open, later, sure
lexical nouns (abstract)	research, story, order, support, education, times, war, information	area, interest , problem, matter , mind	sense , question, point, set, part , end, way, things, others
lexical nouns (concrete)	father, human, students, family , people	man, John, men,	door, children, city, house, country, company, room, world , line , office

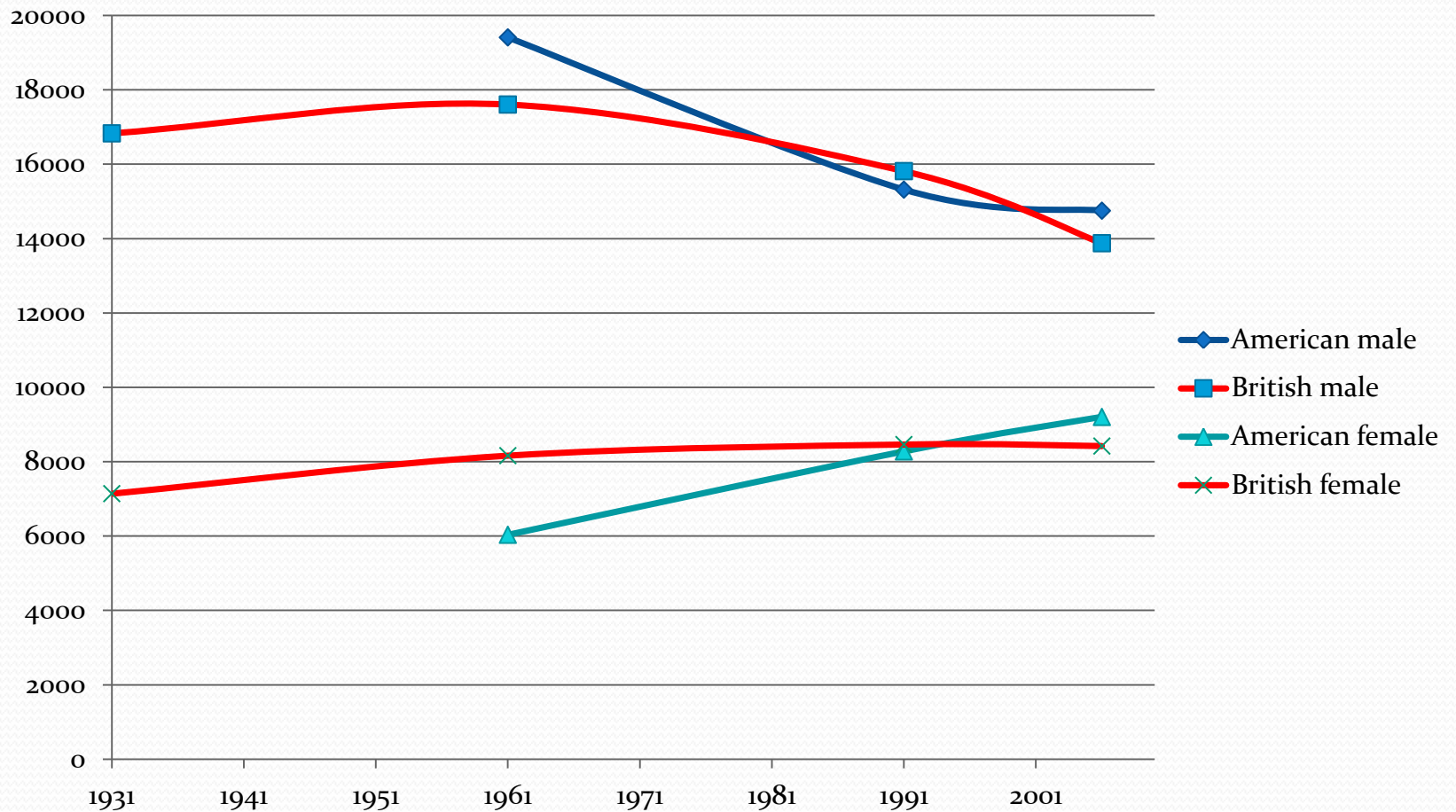
Notable points

- No cases where AE and BE go in opposite directions
- Both American and British English show decreases in the same 6 adjectives: *great, possible, whole, general, certain, present*. (Densification?)
- Contracted forms (*it's, you're* etc) becoming popular – influence of spoken English
- Democratisation/informalisation: decrease in strong modal *must* and terms of address (*Mr, Mrs*)
- Increase in semi-modal forms (*want, wanted, need*)
- Move towards first, second and female third pronouns. Move away from male pronouns (*he, himself*) and nouns (*man, men*)

1st and 2nd person pronouns



Male and female pronouns



Cultural differences - method

- Considered top 300 sets of keywords between American and British English in 2006 (spelling differences and geographical words removed).
- Remove words with low dispersion (more than 95% of cases in 1 or 2 texts)
- Words placed into categories based on Leech and Fallon's (1992) study of 1960s English
- New categories added if needed

Example: category 5 - military

	Key in British English	Key in American English
1960s (Leech and Fallon)	disarmament	armed, army, enemy, forces, missile(s), warfare, bullet(s), gun(s), rifle(s), shot
1990s (Oakes)	armed	gun, hawk, nuclear, shotgun, submarines, terrorist, terrorism, treaty, victim, violence
2000s		border, military, nuclear, operations, revolution, security, terrorist, torture, war

Summary of differences

#	Cultural Category	1960s (Leech and Fallon)	1990s (Oakes)	2000s (This study)
1	Sport	Am		
2	Travel and transport	Am	Am	
3	Admin and politics	Am	Am	Am
4	Social hierarchy	BE	BE	BE
5	Military	Am	Am	Am
6	Law and Crime	Am	Am	
7	Business	Am		
8	Mass media	Am	Am	
9	Science and Technology	Am	Am	Am
10	Education	Am	Am	Am
11	Arts	Am		Am
12	Religion	Am		
13	Personal Reference	BE		BE
14	If, Buts and Modality	BE	BE	BE
15	Be and have	Not included	BE	BE
16	Equality/rights	Not included	Not included	AE
17	Social welfare	Not included	Not included	BE
18	Health/illness	Not included	Not included	BE
19	Data analysis	Not included	Not included	BE

Notable points

- American focus on sport, business, transport, law, crime, mass media, religion – no longer holds in 2006
- AE more focus on admin/politics, military, science/technology, education, arts, equality, civil rights in 2006
- BE more focus on social hierarchy, personal reference, modality, be/have, welfare, health, data analysis in 2006.

Spelling differences

- Of the top 50 Am/Br keywords 20% involved spelling differences (*program, toward, center, defense* etc).
- Are spellings differences ‘holding’?
- E.g. color/colour

American	<i>or</i>	<i>our</i>	ratio	British	<i>or</i>	<i>our</i>	ratio
1961	807	13	62	1961	33	916	27
1992	866	11	78	1991	8	968	121
2006	699	8	87	2006	41	850	20

AE – preference for “or”, getting stronger over time

BE – preference for “our” – no clear pattern of change over time

6 Spelling differences examined

- 1. cent**re**/center
- 2. col**our**/color
- 3. real**ise**/realize
- 4. analy**se**/analyse
- 5a. encyclop**a**edia/encyclop**e**dia
- 5b. fo**e**tal/fetal
- 6. counsell**l**or/counsel**l**or

- Peters, Pam (2004). *The Cambridge Guide to English Usage*. Cambridge: Cambridge University Press.
- Howard, Philip (1984). *The State of the Language—English Observed*. London: Hamish Hamilton

Summary of spelling differences

Trend	American English	British English
No perceivable trend	center realize analyze	centre colour analyse foetal realise
Preference getting stronger	color fetal counselor	
Preference getting weaker		counsellor

Main patterns

- For half the spelling differences, American English is moving away from the British English spelling
- No cases where American English is moving towards the British English spelling
- For 5/6 of the spelling differences, British English is holding its spelling (no clear trend in any direction).
- But in 1 case, the “ll” preference is getting weaker in British English

Overall Conclusions

- American English – becoming more personalised, feminised, democratic, densified and informal (as is BE)
- American English - still driven by technology, enterprise, still militaristic, emphasis on identity equality rather than economic equality and personal health
- American English – appears more confident in its spelling choices. Not so much for BE.

Reflections on the study

- Sampling frame is based on 1960s model – so produces a very conservative view of change (published writing, 1960s categories)
- Different statistical techniques likely to give different results
- Other language varieties are available (e.g. Australian English)
- The two 2006 corpora are available from my website.
<http://www.ling.lancs.ac.uk/profiles/Paul-Baker/>

Thank you

- Baker, P. (2011) 'Times may change but we'll always have money: a corpus driven examination of vocabulary change in four diachronic corpora.' *Journal of English Linguistics* 39: 65-88.
- Baker, P. (2010) 'Will *Ms* ever be as frequent as *Mr*? A corpus-based comparison of gendered terms across four diachronic corpora of British English.' *Gender and Language* 4.1: 125-129.
- Baker, P. (2009) 'The BE06 Corpus of British English and recent language change.' *International Journal of Corpus Linguistics*. 14:3 312-337.
- de Haan, P. (2002) 'Whom is not dead?' In P. Peters, P. Collins and A. Smith (eds) *New Frontiers of Corpus Research*. Amsterdam: Rodopi, pp. 215-228.
- Leech, G. (2002) 'Recent grammatical change in English: data, description, theory.' In K. Aijmer and B. Altenberg (eds) *Proceedings of the 2002 ICAME Conference*, Gothenburg, pp. 61-81.
- Leech, G. and Fallon, R. (1992) 'Computer corpora – what do they tell us about culture?' *ICAME Journal* 16, 29-50.
- McEnery, T. and Xiao, Z. (2005) '*Help* or *Help to*: what do corpora have to say?' *English Studies* 86: 2, 161-187.
- Oakes, M. (2003) 'Contrasts between US and British English in the 1990s.' In E. H. Oleksy and B. Lewandowska-Tomaszczyk (eds) *Research and Scholarship in Integration Processes*. Łódź: Łódź University Press, pp. 213-222.
- Smith, N. (2002) 'Ever moving on? The progressive in recent British English.' In P. Peters, P. Collins and A. Smith (eds) *New Frontiers of Corpus Research*. Amsterdam: Rodopi, pp. 317-330.