

Is academic British English becoming more colloquial? Evidence from the Written BNC2014

Abi Hawtin

 @AbiHawtin

 @BNC_2014

 #BNC2014

Outline:

- What is colloquialisation?
- Research questions
- Methodology
- Results

What is colloquialisation?

- “a tendency for features of the conversational spoken language to infiltrate and spread in the written language” (Leech, 2002: 72)
- a form of “stylistic drift” wherein the style of written language moves toward that of spoken language (Miller, 2009)
- Baker (2017: 243) suggests that colloquialisation of written language can make messages “more accessible to wider audiences”

What is colloquialisation?

- Leech (2002: 72) observes that there are two ways in which colloquialisation can be demonstrated quantitatively: “(a) by an increasing frequency of phenomena associated with spoken language, and (b) by a decreasing frequency of phenomena associated with the written language”.
- Thus, in order to research the phenomenon of colloquialisation it is first necessary to have an understanding of the typical features of spoken and written language

Features associated with colloquialisation:

- Nouns
- Relative pronouns
- Passives
- Verb contractions
- Negative contractions
- Present tense verbs
- WH questions
- First and second person pronouns
- Articles
- and-coordinated adjectives
- Prepositional phrases as post-modifiers
- Verbs
- Semi-modal verbs
- Present progressive
- Progressive passive
- 's genitives

Features associated with colloquialisation:

- Nouns
- Relative pronouns
- Passives
- Verb contractions
- Negative contractions
- Present tense verbs
- WH questions
- First and second person pronouns
- Articles
- and-coordinated adjectives
- Prepositional phrases as post-modifiers
- Verbs
- Semi-modal verbs
- Present progressive
- Progressive passive
- 's genitives

Research questions:

- **RQ1:** Have features of language associated with colloquialisation become more or less frequent in academic writing since 1994?
- **RQ2:** Do the results of RQ1 differ between academic books and academic journal articles?
- **RQ3:** Do the results of RQ1 differ across different genres of academic writing?

Why focus on academic English?



- The literature shows that academic writing seems to be the least ‘speech-like’, and thus least colloquial, type of writing.
- Biber et al. (1999) find that, often, features which are most common in speech are least common in academic writing.
- Academic language is constrained by strong conventions and traditions regarding expected level of formality.
- Colloquialisation in academic language could point to even greater colloquialisation in other types of language which are not constrained by these conventions.

Methodology:

- Search for each feature in each corpus (or sub-corpus)
- Record relative frequencies for each feature in each text
- Carry out a Bootstrap test to confirm the statistical significance of any differences observed

The data (BNC1994):

Sub-genre	Word count (tokens)	Total academic books and total academic journals word count (tokens)	Total word count (tokens)
W_ac_humanities_arts (books)	3,506,992	14,153,936	17,233,631
W_ac_medicine (books)	139,933		
W_ac_nat_science (books)	1,036,307		
W_ac_polit_law_edu (books)	4,425,905		
W_ac_soc_science (books)	4,513,798		
W_ac_tech_engin (books)	531,001		
W_ac_humanities_arts (journals)	190,532	3,079,695	
W_ac_medicine (journals)	1,497,792		
W_ac_nat_science (journals)	242,960		
W_ac_polit_law_edu (journals)	833,327		
W_ac_soc_science (journals)	282,622		
W_ac_tech_engin (journals)	32,462		

The data (BNC2014):

Sub-genre	Word count (tokens)	Total academic books and total academic journals word count (tokens)	Total word count (tokens)
W_ac_book_humanities_arts (books)	741,762	4,132,820	6,395,903
W_ac_book_medicine (books)	270,688		
W_ac_book_nat_science (books)	482,340		
W_ac_book_polit_law_edu (books)	845,165		
W_ac_book_soc_science (books)	900,990		
W_ac_book_tech_engine (books)	891,875		
W_ac_journal_humanities_arts (journals)	631,738	2,263,083	
W_ac_journal_medicine (journals)	231,428		
W_ac_journal_nat_science (journals)	646,114		
W_ac_journal_polit_law_edu (journals)	304,139		
W_ac_journal_soc_science (journals)	189,874		
W_ac_journal_tech_engin (journals)	259,790		

The search terms:

Linguistic feature	Positively or negatively associated with colloquialisation	Search
First and second person pronouns	+	[word = "I me we us you"]
Present tense verbs	+	[tag="VBP VBZ VHP VHZ VVP VVZ"]
Verb contractions	+	[word = "'s" & tag = "V.*" word = "'ve" word = "'re" word = "'ll" word = "'d" word = "'m"]
Negative contractions	+	n't
Questions (all)	+	\?
Verb frequency	+	[tag = "V.*"]
's Genitives	+	[tag = "NNSZ" tag = "NNZ" tag = "NPSZ" tag = "NPZ"]

The search terms:

Linguistic feature	Positively or negatively associated with colloquialisation	Search
Semi-modals	+	([word = "going"] [tag="TO"] [word = "gonna"] [lemma = "be"] [tag="PP" tag="N.*" tag="DT"]? [tag = "RB.*"]{0,2} [tag="TO"] [lemma = "have"] [tag="PP" tag="N.*" tag="DT"]? [tag = "RB.*"]{0,2} [word="better"] [tag = "RB.*"]{0,2} [tag = "V.*"] [word="got"] [tag = "TO"] [word = "gotta"] [lemma = "have"] [tag = "RB.*"]{0,2} [tag = "TO"] [lemma = "need"] [tag = "TO"] [lemma = "want"] [tag = "TO"] [word = "wanna"] [word = "used"] [tag = "TO"])
Passive forms (all)	-	[lemma = "be" & tag = "VB.*"] [tag = "R.*"] {0,2} [tag = "V.N"]
Relative pronouns	-	[word = "who which whose whom what"]
Noun frequency	-	[tag = "N.*"]

Have features of language associated with colloquialisation become more or less frequent in academic writing since 1994?



Research question 1

	BNC1994 (freq per mill)	BNC2014 (freq per mill)	Difference (+/- %)	Statistically Significant? (p<0.05)
First and second person pronouns	3,976.12	5,154.20	+19.627	NO
Present tense verbs	38,018.11	37,957.90	-8.244	NO
Verb contractions	346.68	713.9	+90.094	YES
Negative contractions	189.75	384.03	+86.775	YES
Questions (all)	780.8	805.65	-4.808	NO
Verb frequency	134,352.71	122,456.80	-16.154	YES
's Genitives	3,664.52	3,675.00	-7.531	NO
Semi-modals	2,033.06	1,659.6	-24.674	YES (p<0.001)
Passive forms (all)	15,441.50	11,181.40	-33.176	YES (p<0.001)
Relative pronouns	8,125.45	5,723.80	-34.993	YES (p<0.001)
Noun frequency	254,176.85	273,095.00	-1.131	NO

Have features of language associated with colloquialisation become more or less frequent in academic writing since 1994?



Research question 1

	BNC1994 (freq per mill)	BNC2014 (freq per mill)	Difference (+/- %)	Statistically Significant? (p<0.05)
First and second person pronouns	3,976.12	5,154.20	+19.627	NO
Present tense verbs	38,018.11	37,957.90	-8.244	NO
Verb contractions	346.68	713.9	+90.094	YES
Negative contractions	189.75	384.03	+86.775	YES
Questions (all)	780.8	805.65	-4.808	NO
Verb frequency	134,352.71	122,456.80	-16.154	YES
's Genitives	3,664.52	3,675.00	-7.531	NO
Semi-modals	2,033.06	1,659.6	-24.674	YES (p<0.001)
Passive forms (all)	15,441.50	11,181.40	-33.176	YES (p<0.001)
Relative pronouns	8,125.45	5,723.80	-34.993	YES (p<0.001)
Noun frequency	254,176.85	273,095.00	-1.131	NO

Have features of language associated with colloquialisation become more or less frequent in academic writing since 1994?



Research question 1

	BNC1994 (freq per mill)	BNC2014 (freq per mill)	Difference (+/- %)	Statistically Significant? (p<0.05)
First and second person pronouns	3,976.12	5,154.20	+19.627	NO
Present tense verbs	38,018.11	37,957.90	-8.244	NO
Verb contractions	346.68	713.9	+90.094	YES
Negative contractions	189.75	384.03	+86.775	YES
Questions (all)	780.8	805.65	-4.808	NO
Verb frequency	134,352.71	122,456.80	-16.154	YES
's Genitives	3,664.52	3,675.00	-7.531	NO
Semi-modals	2,033.06	1,659.6	-24.674	YES (p<0.001)
Passive forms (all)	15,441.50	11,181.40	-33.176	YES (p<0.001)
Relative pronouns	8,125.45	5,723.80	-34.993	YES (p<0.001)
Noun frequency	254,176.85	273,095.00	-1.131	NO

Have features of language associated with colloquialisation become more or less frequent in academic writing since 1994?



Research question 1

- Not straightforward...
- Four features associated with colloquialisation have shown statistically significant changes in frequency (in a direction predicted by colloquialisation) between 1994 and 2014.
- Two features have shown statistically significant changes contrary to what colloquialisation theory would predict.
- When looked at together these results seem to indicate that academic writing is certainly not becoming *less* colloquial, and in some aspects is becoming markedly more colloquial than in the 1990s.

Do the results of RQ1 differ between academic books and academic journal articles?



Research question 2

	Frequency change in academic books (+/- %)	Statistically Significant? (p<0.05)	Frequency change in academic journals (+/- %)	Statistically Significant? (p<0.05)
First and second person pronouns	+195.335	YES (p<0.001)	-53.886	YES (p<0.001)
Present tense verbs	+104.141	YES (p<0.001)	-51.595	YES (p<0.001)
Verb contractions	+334.906	YES (p<0.001)	+75.067	NO
Negative contractions	+353.998	YES (p<0.001)	+24.73	NO
Questions (all)	+125.878	YES (p<0.001)	-49.277	YES (p<0.001)
Verb frequency	+85.753	YES (p<0.001)	-60.944	YES (p<0.001)
's Genitives	+98.243	YES (p<0.001)	-52.644	YES (p<0.001)
Semi-modals	+71.788	YES (p<0.001)	-59.343	YES (p<0.001)
Passive forms (all)	+40.504	YES (p<0.001)	-70.048	YES (p<0.001)
Relative pronouns	+50.155	YES (p<0.001)	-68.237	YES (p<0.001)
Noun frequency	+109.371	YES (p<0.001)	-54.546	YES (p<0.001)

Do the results of RQ1 differ between academic books and academic journal articles?



Research question 2

	Frequency change in academic books (+/- %)	Statistically Significant? (p<0.05)	Frequency change in academic journals (+/- %)	Statistically Significant? (p<0.05)
First and second person pronouns	+195.335	YES (p<0.001)	-53.886	YES (p<0.001)
Present tense verbs	+104.141	YES (p<0.001)	-51.595	YES (p<0.001)
Verb contractions	+334.906	YES (p<0.001)	+75.067	NO
Negative contractions	+353.998	YES (p<0.001)	+24.73	NO
Questions (all)	+125.878	YES (p<0.001)	-49.277	YES (p<0.001)
Verb frequency	+85.753	YES (p<0.001)	-60.944	YES (p<0.001)
's Genitives	+98.243	YES (p<0.001)	-52.644	YES (p<0.001)
Semi-modals	+71.788	YES (p<0.001)	-59.343	YES (p<0.001)
Passive forms (all)	+40.504	YES (p<0.001)	-70.048	YES (p<0.001)
Relative pronouns	+50.155	YES (p<0.001)	-68.237	YES (p<0.001)
Noun frequency	+109.371	YES (p<0.001)	-54.546	YES (p<0.001)

Do the results of RQ1 differ between academic books and academic journal articles?



Research question 2

	Frequency change in academic books (+/- %)	Statistically Significant? (p<0.05)	Frequency change in academic journals (+/- %)	Statistically Significant? (p<0.05)
First and second person pronouns	+195.335	YES (p<0.001)	-53.886	YES (p<0.001)
Present tense verbs	+104.141	YES (p<0.001)	-51.595	YES (p<0.001)
Verb contractions	+334.906	YES (p<0.001)	+75.067	NO
Negative contractions	+353.998	YES (p<0.001)	+24.73	NO
Questions (all)	+125.878	YES (p<0.001)	-49.277	YES (p<0.001)
Verb frequency	+85.753	YES (p<0.001)	-60.944	YES (p<0.001)
's Genitives	+98.243	YES (p<0.001)	-52.644	YES (p<0.001)
Semi-modals	+71.788	YES (p<0.001)	-59.343	YES (p<0.001)
Passive forms (all)	+40.504	YES (p<0.001)	-70.048	YES (p<0.001)
Relative pronouns	+50.155	YES (p<0.001)	-68.237	YES (p<0.001)
Noun frequency	+109.371	YES (p<0.001)	-54.546	YES (p<0.001)

Do the results of RQ1 differ between academic books and academic journal articles?



Research question 2

	Frequency change in academic books (+/- %)	Statistically Significant? (p<0.05)	Frequency change in academic journals (+/- %)	Statistically Significant? (p<0.05)
First and second person pronouns	+195.335	YES (p<0.001)	-53.886	YES (p<0.001)
Present tense verbs	+104.141	YES (p<0.001)	-51.595	YES (p<0.001)
Verb contractions	+334.906	YES (p<0.001)	+75.067	NO
Negative contractions	+353.998	YES (p<0.001)	+24.73	NO
Questions (all)	+125.878	YES (p<0.001)	-49.277	YES (p<0.001)
Verb frequency	+85.753	YES (p<0.001)	-60.944	YES (p<0.001)
's Genitives	+98.243	YES (p<0.001)	-52.644	YES (p<0.001)
Semi-modals	+71.788	YES (p<0.001)	-59.343	YES (p<0.001)
Passive forms (all)	+40.504	YES (p<0.001)	-70.048	YES (p<0.001)
Relative pronouns	+50.155	YES (p<0.001)	-68.237	YES (p<0.001)
Noun frequency	+109.371	YES (p<0.001)	-54.546	YES (p<0.001)

Do the results of RQ1 differ between academic books and academic journal articles?



Research question 2

	Frequency change in academic books (+/- %)	Statistically Significant? (p<0.05)	Frequency change in academic journals (+/- %)	Statistically Significant? (p<0.05)
First and second person pronouns	+195.335	YES (p<0.001)	-53.886	YES (p<0.001)
Present tense verbs	+104.141	YES (p<0.001)	-51.595	YES (p<0.001)
Verb contractions	+334.906	YES (p<0.001)	+75.067	NO
Negative contractions	+353.998	YES (p<0.001)	+24.73	NO
Questions (all)	+125.878	YES (p<0.001)	-49.277	YES (p<0.001)
Verb frequency	+85.753	YES (p<0.001)	-60.944	YES (p<0.001)
's Genitives	+98.243	YES (p<0.001)	-52.644	YES (p<0.001)
Semi-modals	+71.788	YES (p<0.001)	-59.343	YES (p<0.001)
Passive forms (all)	+40.504	YES (p<0.001)	-70.048	YES (p<0.001)
Relative pronouns	+50.155	YES (p<0.001)	-68.237	YES (p<0.001)
Noun frequency	+109.371	YES (p<0.001)	-54.546	YES (p<0.001)

Do the results of RQ1 differ between academic books and academic journal articles?



Research question 2

- YES!
- Books are definitely changing more in the direction of colloquialisation than journal articles.
- Journal articles are actually becoming less colloquial in a lot of ways.
- Could this be because books were less colloquial than journal articles in the BNC1994, so a greater amount of change is expected?
- No – the relative frequencies of the linguistic features in the 1994 books and journals corpora are actually very similar.

Do the results of RQ1 differ across different genres of academic writing?



Research
question 3

	Humanities and arts (+/- %)	Medicine (+/- %)	Natural science (+/- %)	Politics, law and education (+/- %)	Social science (+/- %)	Technology and engineering (+/- %)
First and second person pronouns	-42.74*	+34.864	-14.571*	+222.418*	+134.469*	-77.082*
Present tense verbs	-49.359*	-42.042	-53.631*	+130.014*	+85.463*	-76.446*
Verb contractions	+32.131	+956.599	+391.451	+221.94*	+161.907	+448.786
Negative contractions	-15.761	+602.846	+16.604	+267.677*	+231.571*	+107.216
Questions (all)	-39.752*	+32.385	-12.479	+88.937*	+81.246*	-68.214
Verb frequency	-59.904*	-55.523*	-53.631*	+111.584*	+87.733*	-73.53*
's Genitives	-46.539*	-50.08	-23.626	+148.509*	+83.48*	-32.745
Semi-modals	-70.737*	-34.171	-33.825	+86.384*	+78.593*	-77.17*
Passive forms (all)	-67.047*	-68.851*	-66.421*	+51.743*	+49.484*	-81.956*
Relative pronouns	-68.342*	-47.518	-55.644*	+70.001*	+33.37	-73.479*
Noun frequency	-51.567*	-54.495*	-47.378*	+142.484*	+111.191*	-66.948*

	Humanities and arts (+/- %)	Medicine (+/- %)	Natural science (+/- %)	Politics, law and education (+/- %)	Social science (+/- %)	Technology and engineering (+/- %)
First and second person pronouns	-42.74*	+34.864	-14.571*	+222.418*	+134.469*	-77.082*
Present tense verbs	-49.359*	-42.042	-53.631*	+130.014*	+85.463*	-76.446*
Verb contractions	+32.131	+956.599	+391.451	+221.94*	+161.907	+448.786
Negative contractions	-15.761	+602.846	+16.604	+267.677*	+231.571*	+107.216
Questions (all)	-39.752*	+32.385	-12.479	+88.937*	+81.246*	-68.214
Verb frequency	-59.904*	-55.523*	-53.631*	+111.584*	+87.733*	-73.53*
's Genitives	-46.539*	-50.08	-23.626	+148.509*	+83.48*	-32.745
Semi-modals	-70.737*	-34.171	-33.825	+86.384*	+78.593*	-77.17*
Passive forms (all)	-67.047*	-68.851*	-66.421*	+51.743*	+49.484*	-81.956*
Relative pronouns	-68.342*	-47.518	-55.644*	+70.001*	+33.37	-73.479*
Noun frequency	-51.567*	-54.495*	-47.378*	+142.484*	+111.191*	-66.948*

	Humanities and arts (+/- %)	Medicine (+/- %)	Natural science (+/- %)	Politics, law and education (+/- %)	Social science (+/- %)	Technology and engineering (+/- %)
First and second person pronouns	-42.74*	+34.864	-14.571*	+222.418*	+134.469*	-77.082*
Present tense verbs	-49.359*	-42.042	-53.631*	+130.014*	+85.463*	-76.446*
Verb contractions	+32.131	+956.599	+391.451	+221.94*	+161.907	+448.786
Negative contractions	-15.761	+602.846	+16.604	+267.677*	+231.571*	+107.216
Questions (all)	-39.752*	+32.385	-12.479	+88.937*	+81.246*	-68.214
Verb frequency	-59.904*	-55.523*	-53.631*	+111.584*	+87.733*	-73.53*
's Genitives	-46.539*	-50.08	-23.626	+148.509*	+83.48*	-32.745
Semi-modals	-70.737*	-34.171	-33.825	+86.384*	+78.593*	-77.17*
Passive forms (all)	-67.047*	-68.851*	-66.421*	+51.743*	+49.484*	-81.956*
Relative pronouns	-68.342*	-47.518	-55.644*	+70.001*	+33.37	-73.479*
Noun frequency	-51.567*	-54.495*	-47.378*	+142.484*	+111.191*	-66.948*

	Humanities and arts (+/- %)	Medicine (+/- %)	Natural science (+/- %)	Politics, law and education (+/- %)	Social science (+/- %)	Technology and engineering (+/- %)
First and second person pronouns	-42.74*	+34.864	-14.571*	+222.418*	+134.469*	-77.082*
Present tense verbs	-49.359*	-42.042	-53.631*	+130.014*	+85.463*	-76.446*
Verb contractions	+32.131	+956.599	+391.451	+221.94*	+161.907	+448.786
Negative contractions	-15.761	+602.846	+16.604	+267.677*	+231.571*	+107.216
Questions (all)	-39.752*	+32.385	-12.479	+88.937*	+81.246*	-68.214
Verb frequency	-59.904*	-55.523*	-53.631*	+111.584*	+87.733*	-73.53*
's Genitives	-46.539*	-50.08	-23.626	+148.509*	+83.48*	-32.745
Semi-modals	-70.737*	-34.171	-33.825	+86.384*	+78.593*	-77.17*
Passive forms (all)	-67.047*	-68.851*	-66.421*	+51.743*	+49.484*	-81.956*
Relative pronouns	-68.342*	-47.518	-55.644*	+70.001*	+33.37	-73.479*
Noun frequency	-51.567*	-54.495*	-47.378*	+142.484*	+111.191*	-66.948*

	Humanities and arts (+/- %)	Medicine (+/- %)	Natural science (+/- %)	Politics, law and education (+/- %)	Social science (+/- %)	Technology and engineering (+/- %)
First and second person pronouns	-42.74*	+34.864	-14.571*	+222.418*	+134.469*	-77.082*
Present tense verbs	-49.359*	-42.042	-53.631*	+130.014*	+85.463*	-76.446*
Verb contractions	+32.131	+956.599	+391.451	+221.94*	+161.907	+448.786
Negative contractions	-15.761	+602.846	+16.604	+267.677*	+231.571*	+107.216
Questions (all)	-39.752*	+32.385	-12.479	+88.937*	+81.246*	-68.214
Verb frequency	-59.904*	-55.523*	-53.631*	+111.584*	+87.733*	-73.53*
's Genitives	-46.539*	-50.08	-23.626	+148.509*	+83.48*	-32.745
Semi-modals	-70.737*	-34.171	-33.825	+86.384*	+78.593*	-77.17*
Passive forms (all)	-67.047*	-68.851*	-66.421*	+51.743*	+49.484*	-81.956*
Relative pronouns	-68.342*	-47.518	-55.644*	+70.001*	+33.37	-73.479*
Noun frequency	-51.567*	-54.495*	-47.378*	+142.484*	+111.191*	-66.948*

	Humanities and arts (+/- %)	Medicine (+/- %)	Natural science (+/- %)	Politics, law and education (+/- %)	Social science (+/- %)	Technology and engineering (+/- %)
First and second person pronouns	-42.74*	+34.864	-14.571*	+222.418*	+134.469*	-77.082*
Present tense verbs	-49.359*	-42.042	-53.631*	+130.014*	+85.463*	-76.446*
Verb contractions	+32.131	+956.599	+391.451	+221.94*	+161.907	+448.786
Negative contractions	-15.761	+602.846	+16.604	+267.677*	+231.571*	+107.216
Questions (all)	-39.752*	+32.385	-12.479	+88.937*	+81.246*	-68.214
Verb frequency	-59.904*	-55.523*	-53.631*	+111.584*	+87.733*	-73.53*
's Genitives	-46.539*	-50.08	-23.626	+148.509*	+83.48*	-32.745
Semi-modals	-70.737*	-34.171	-33.825	+86.384*	+78.593*	-77.17*
Passive forms (all)	-67.047*	-68.851*	-66.421*	+51.743*	+49.484*	-81.956*
Relative pronouns	-68.342*	-47.518	-55.644*	+70.001*	+33.37	-73.479*
Noun frequency	-51.567*	-54.495*	-47.378*	+142.484*	+111.191*	-66.948*

Do the results of RQ1 differ across different genres of academic writing?



Research question 3

- YES!
- The changes seen in the politics, law & education genre mirror the changes seen in the academic books corpora.
- Perhaps genres with a 'social' aspect are changing the most?
- Medicine, natural science, and technology & engineering show the least statistically significant changes.
- Perhaps 'hard' science genres are the most stable? Or were they already very colloquial to begin with?

Summary

- Some features associated with colloquialisation have certainly become more frequent in academic writing since the 1990s
- These changes vary depending on medium and genre
- Books show many more changes in line with colloquialisation than journals
- Genres with a ‘social’ aspect showed more changes in line with colloquialisation than the ‘hard’ science genres
- Can we conclude anything about the colloquialisation of language in general?

Future work:

- Repeat the study with the full academic data set from the Written BNC2014
- Repeat the study on other genres of writing
- Use other corpora to provide more data points

Thank you for listening!
Any questions?

References:

- Baker, P. (2017). *American and British English: Divided by a common language?* Cambridge: Cambridge University Press.
- Leech, G. (2002). Recent grammatical change in English: Data, description, theory. In K. Aijmer & B. Altenberg (Eds.), *Advances in Corpus Linguistics. Papers from the 23rd International Conference on English Language Research on Computerized Corpora (ICAME 23)*, Göteborg 22–26 May 2002 (pp. 61-81), Amsterdam: Rodopi.
- Millar, N. (2009). Modal verbs in TIME: Frequency changes 1923–2006. *International Journal of Corpus Linguistics*, 14 (2), 191–220.