Corpus Research on the Development of Children's Writing in L1 English

Phil Durrant
University of Exeter

That morning a meerkat mob was snoring there heds of Suddenly a snaka slithed into the brow. The snaka saw the baby. The baby ran to the mum, the mum froo the snak. Next a jackal ran in to Sunny. The jackal sede can I be your frend? And they play together tag. Sunny a vitid him for tea time the end.

one luge time ago there was a king colld king james the first and the cathlixs did not like him. and there was a bad man called Guy Fawkes he wantied to bow the houses of Parliament he wantid to cill the king to as well as the cathlixs he had 36 barols of gunpowder and he hid it. Robert Catesby sent a leter to the king.

Dear Sir, I am writing to express my views on the article you recently printed, detailing a scheme by the Divert Trust to help difficult students. At first I was unsure if this scheme could ever work, and was indignant, like so many others, that many good students remained unrewarded. However, after researching this scheme I have come to realise that it is rather a brilliant idea. Research shows that around 88% of schools admit to not being able to cope with difficult students.

Year 2: Detail of content to be introduced (statutory requirement)	
Word	Formation of **nouns** using **suffixes** such as *–ness*, *–er* and by compounding [for example, *whiteboard*, *superman*] Formation of **adjectives** using **suffixes** such as *–ful*, *–less* (A fuller list of **suffixes** can be found on page 57 in the year 2 spelling section in English Appendix 1) Use of the **suffixes** *–er*, *–est* in **adjectives** and the use of –ly in Standard English to turn adjectives into **adverbs**
Sentence	**Subordination** (using *when*, *if*, *that*, *because)* and **co-ordination** (using *or*, *and*, *but*) Expanded **noun phrases** for description and specification [for example, *the blue butterfly*, *plain flour*, *the man in the moon*] **How the grammatical patterns in a sentence indicate its function as a** statement, question, exclamation or command

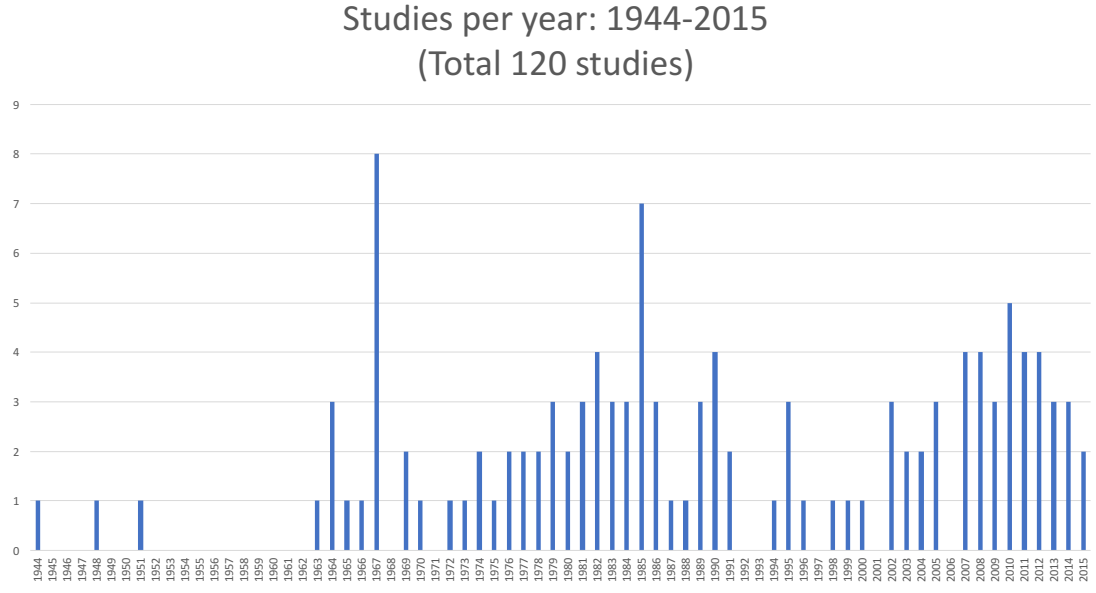
Department of Education, 2014

Year 6: Detail of content to be introduced (statutory requirement)	
Word	The difference between vocabulary typical of informal speech and vocabulary appropriate for formal speech and writing [for example, *find out – discover; ask for – request; go in – enter*] How words are related by meaning as synonyms and antonyms [for example, *big, large, little*].
Sentence	Use of the **passive** to affect the presentation of information in a **sentence** [for example, *I broke the window in the greenhouse* versus *The window in the greenhouse was broken (by me)*]. The difference between structures typical of informal speech and structures appropriate for formal speech and writing [for example, the use of question tags: *He's your friend*, *isn't he?*, or the use of **subjunctive** forms such as *If I were* or *Were they to come* in some very formal writing and speech]

Department of Education, 2014

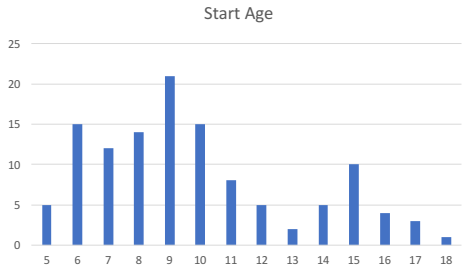
Historical Overview

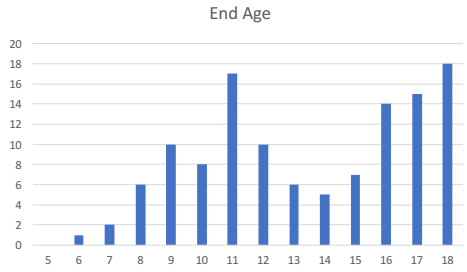
Quantitative comparisons of linguistic features in mainstream children's writing

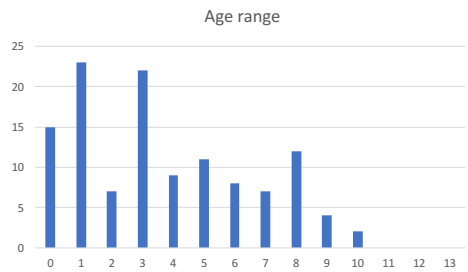


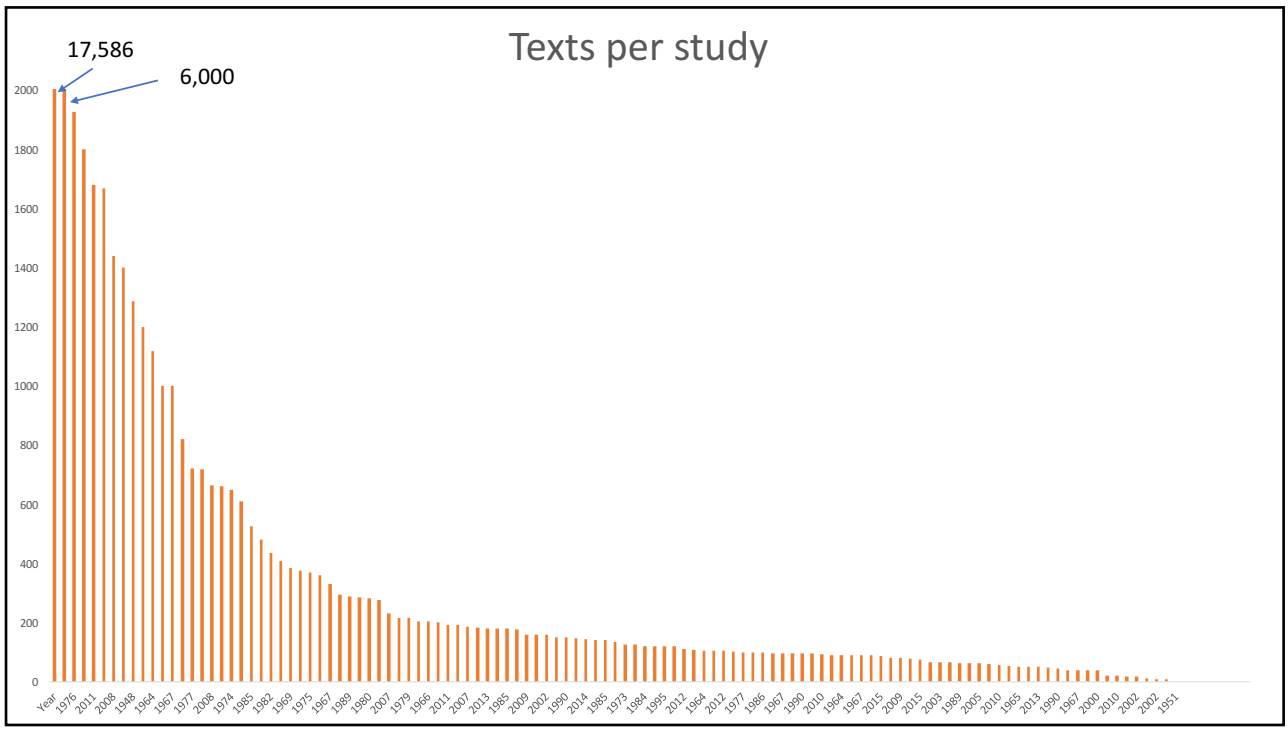
Provenance

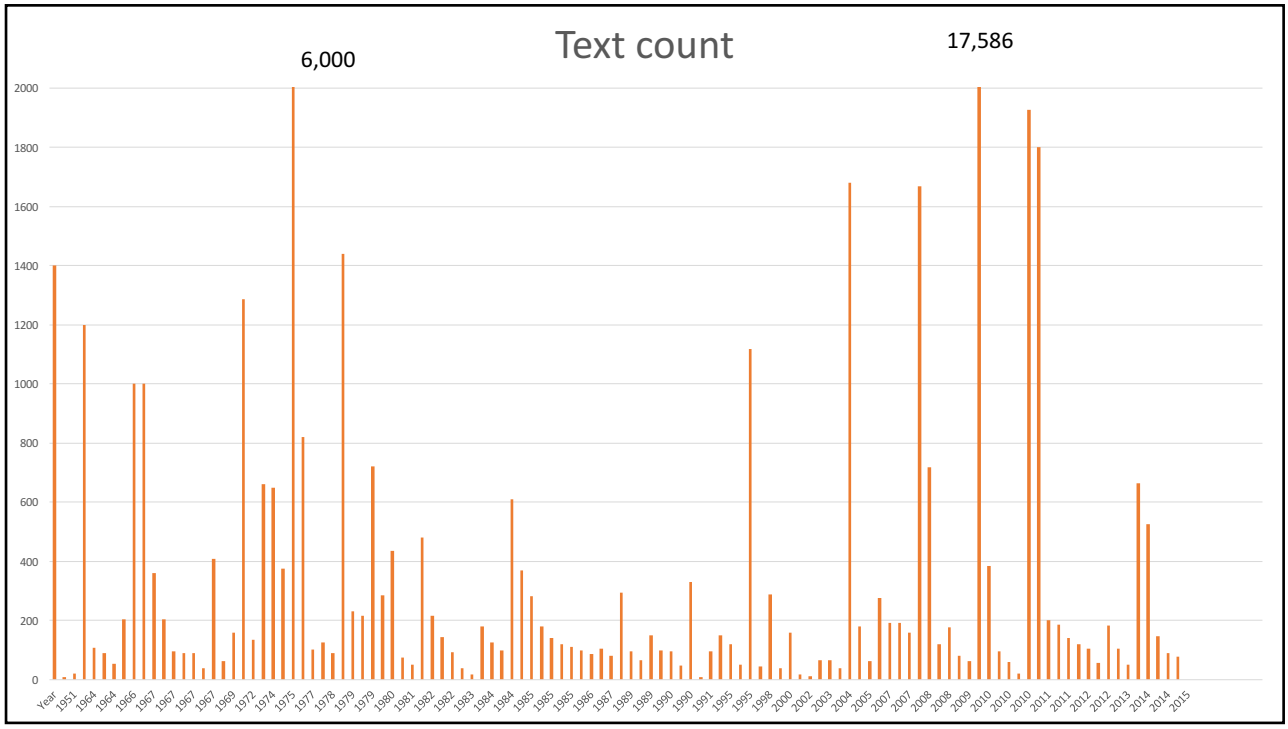
Country	Number of Studies
USA	77
UK	22
Canada	16
Australia	2
USA, UK & New Zealand	2
USA & New Zealand	1

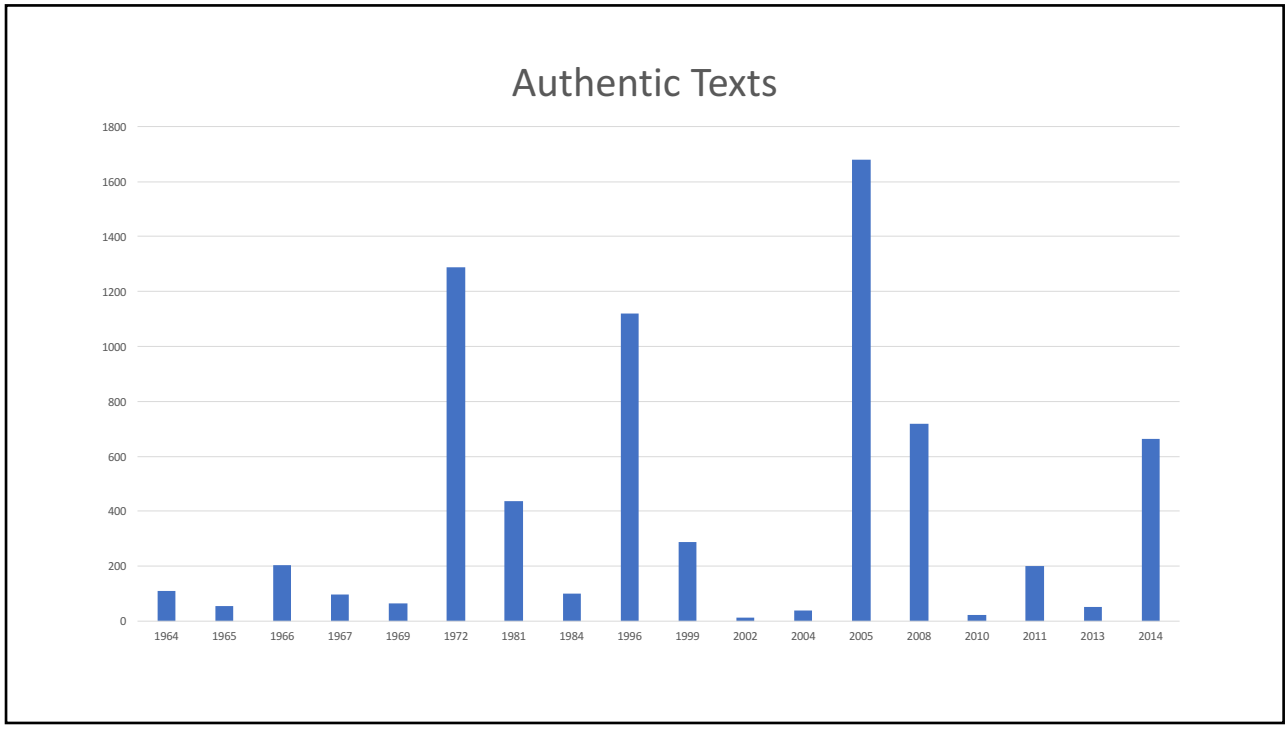


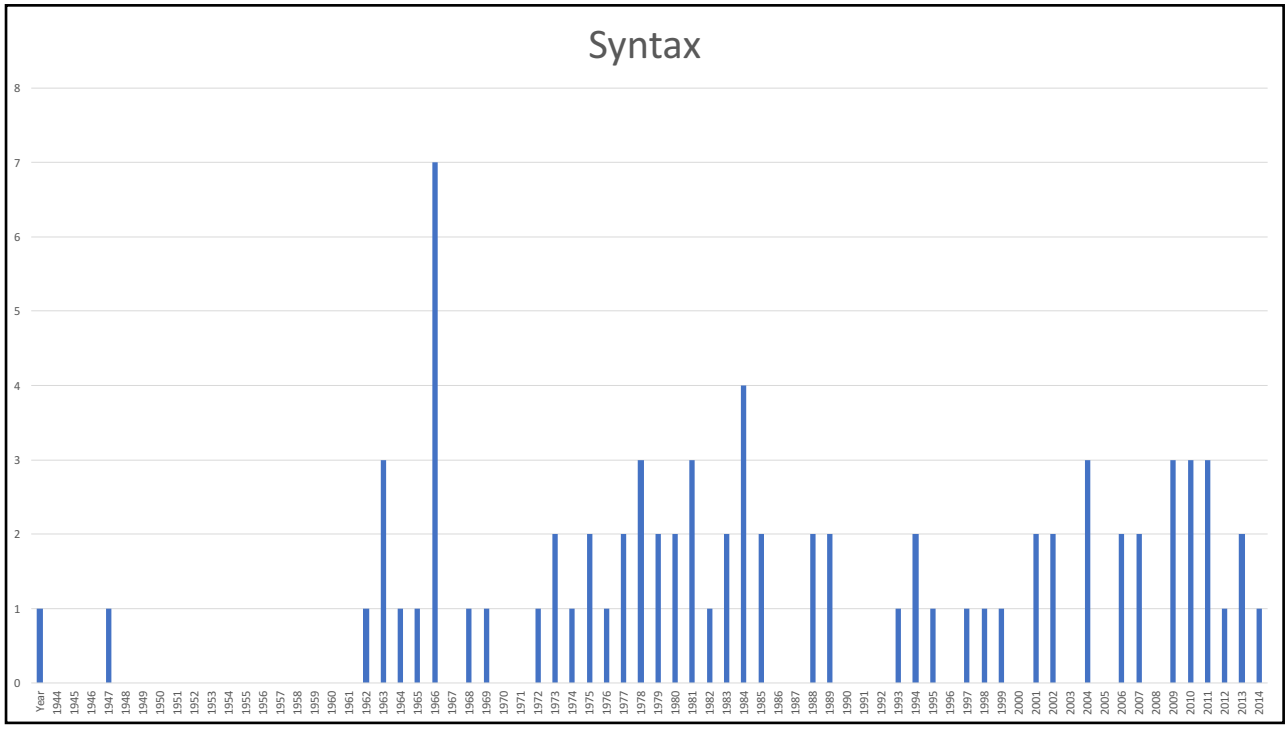


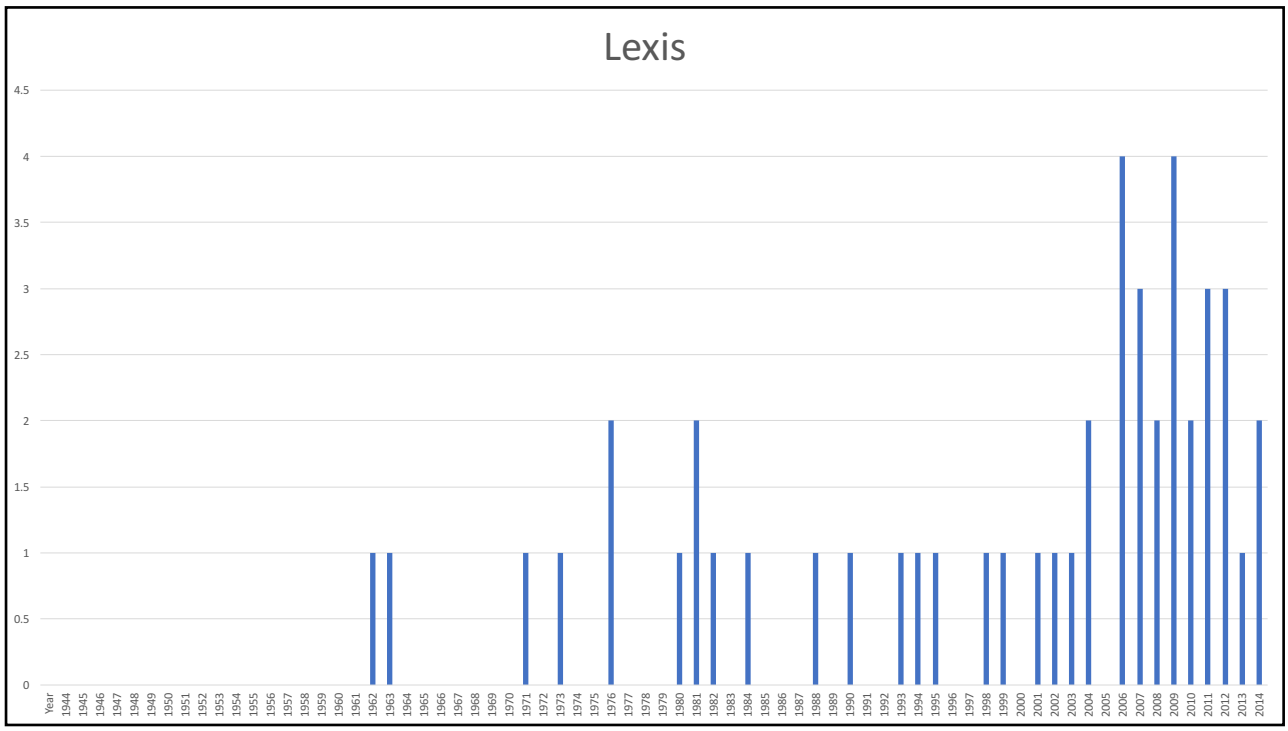


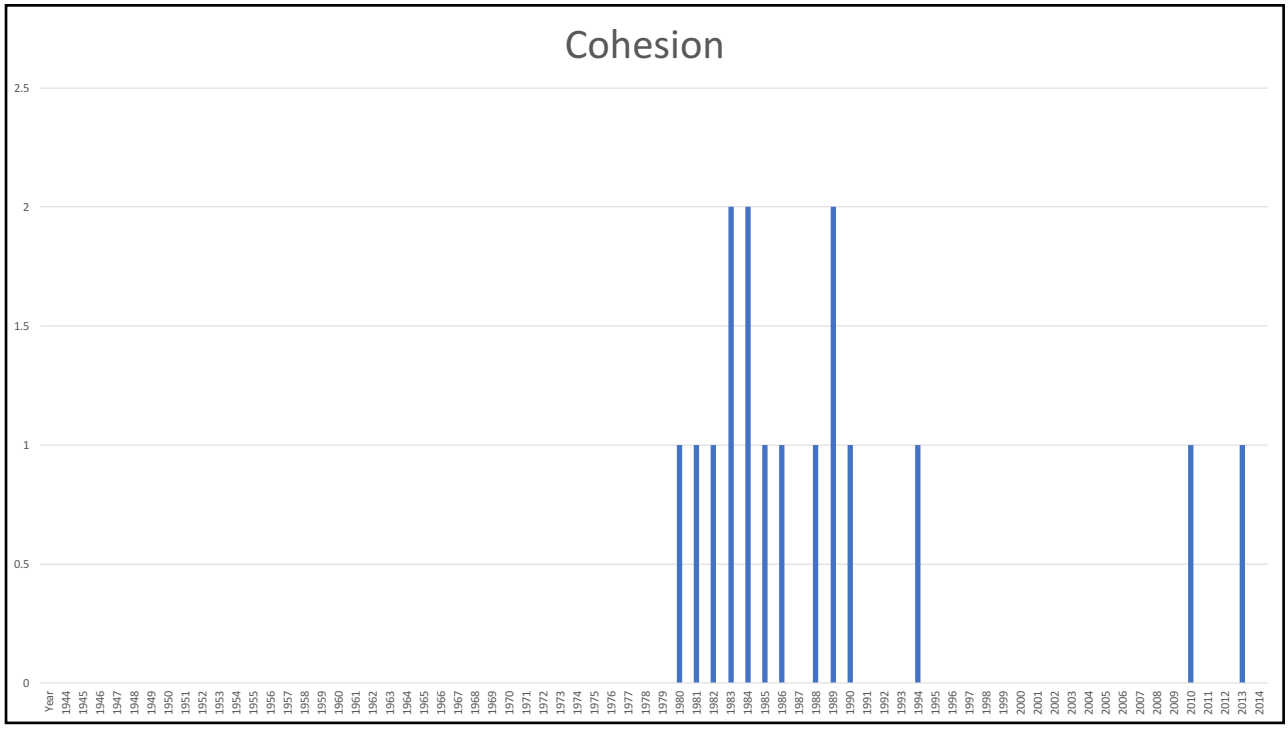

Texts per study
17,586
6,000
2000
1800
1600
1400
1200
1000
800
600
400
200
0


Text count
6,000
17,586
2000
1800
1600
1400
1200
1000
800
600
400
200
0


Authentic Texts
1800
1600
1400
1200
1000
800
600
400
200
0
1964
1965
1966
1967
1969
1972
1981
1984
1996
1999
2002
2004
2005
2008
2010
2011
2013
2014


Syntax
8
7
6
5
4
3
2
1
0
Year
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014


Lexis
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Year
1944 1945 1946 1947 1948 1949 1950 1951 1952 1953 1954 1955 1956 1957 1958 1959 1960 1961 1962 1963 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014


Cohesion
2.5
2
1.5
1
0.5
0
Year
1944 1945 1946 1947 1948 1949 1950 1951 1952 1953 1954 1955 1956 1957 1958 1959 1960 1961 1962 1963 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014

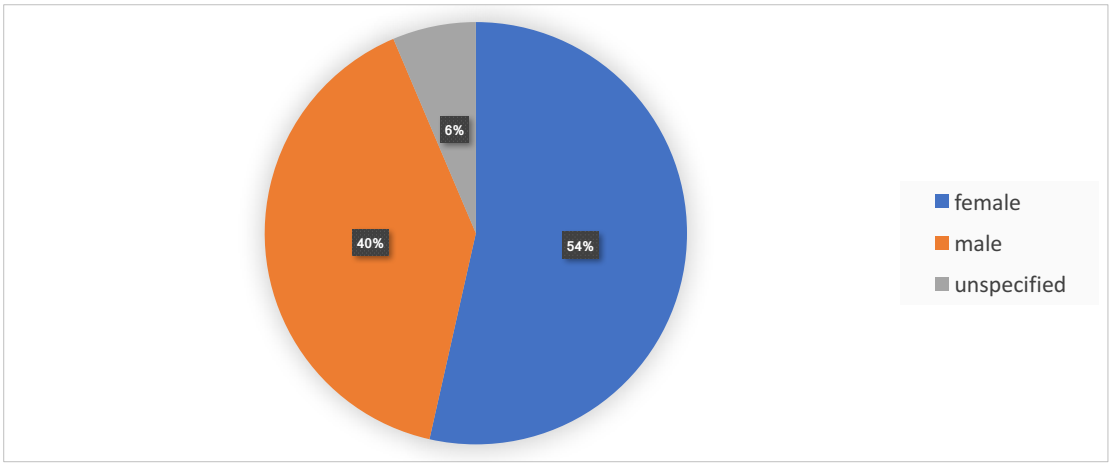
The Growth in Grammar Project

- Creating a corpus of educationally authentic writing from children in schools across England at Key Stages 1-4
- To be analysed for changes at the levels of lexis, phrase and clause.
  - NB: Analysis of structures used, not of accuracy.
- Corpus to be made publicly available from August 2018
- Combining with APU Corpus (we hope!)

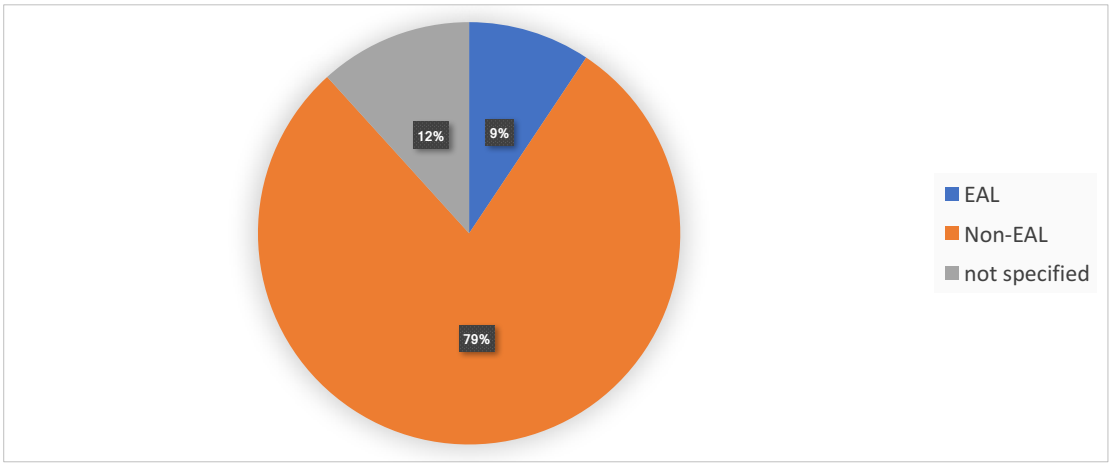
Corpus to date

	Total texts	Total schools	Total children	Discipline		
				English	Science	Other
Year 2	543	5	138	452	7	84
Year 4	49	2	10	25	2	22
Year 6	868	7	185	548	171	149
Year 9	761	12	465	483	166	112
Year 11	191	6	85	137	54	0
All years	2,412	21	883	1,645	400	367

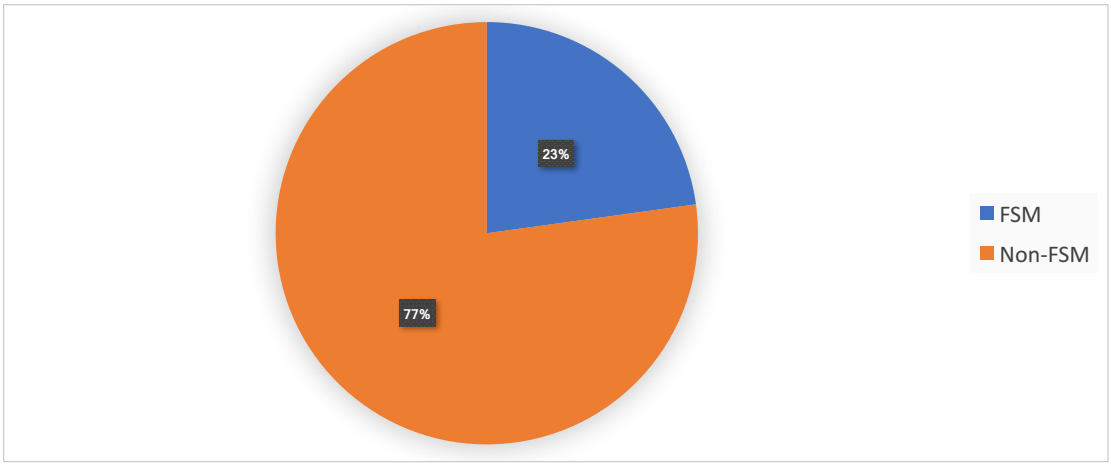
Corpus to date: Gender



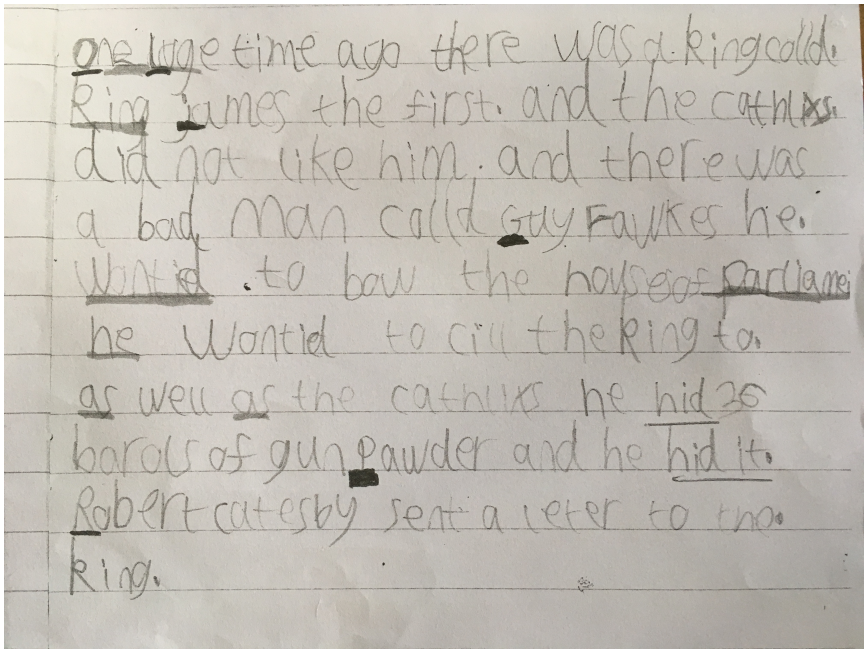
Corpus to date: EAL



Corpus to date: SES



Issues 1: Transcription



one huge time ago there was a king coldd.
King James the first. and the cathlxs.
did not like him. and there was
a bad man calld Guy Fawkes he.
Wantid to bow the houses of parliament
he Wantid to cill the King to.
as well as the cathlxs he hid 36
barols of gun powder and he hid it.
Robert catesby sent a leter to the.
King.

Transcription issues

- Legibility – especially with irregular features of spelling/punctuation
- Creating a 'parser-ready' version: regularization of spelling, punctuation; syntax.
- Anonymisation

Issues 2: Genre

Issues 3:
Coding Grammatical Features

Coding grammatical features

The captain ordered me to douse the fire in the galley in case the fire spread and burnt down the ship

Coding grammatical features

The captain ordered **me** to douse **the fire in the galley** in case **the fire** spread and burnt down **the ship**

Coding grammatical features

The captain ordered **me** to douse **the fire in the galley in case the fire spread and burnt down the ship**

Coding grammatical features

the **fire** in the **galley** in **case** the **fire spread** and burnt down the **ship**

Coding grammatical features

the **fire** in the **galley** in **case** the **fire spread** and burnt down the **ship**

noun modifier

Coding grammatical features

the **fire** in the **galley** in **case** the **fire spread** and burnt down the **ship**

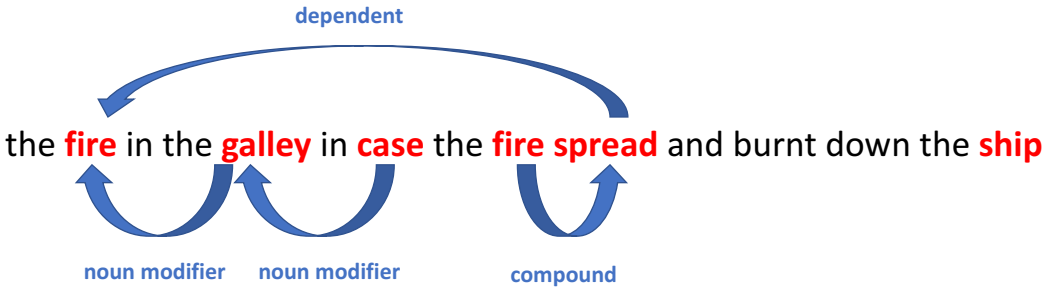
noun modifier noun modifier

Coding grammatical features

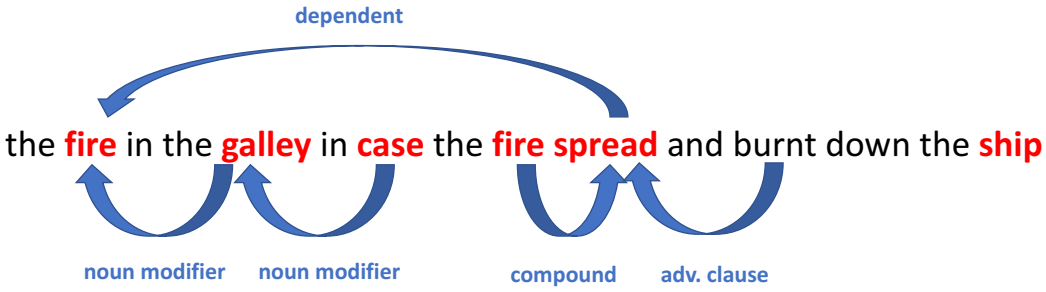
the **fire** in the **galley** in **case** the **fire spread** and burnt down the **ship**

noun modifier noun modifier compound

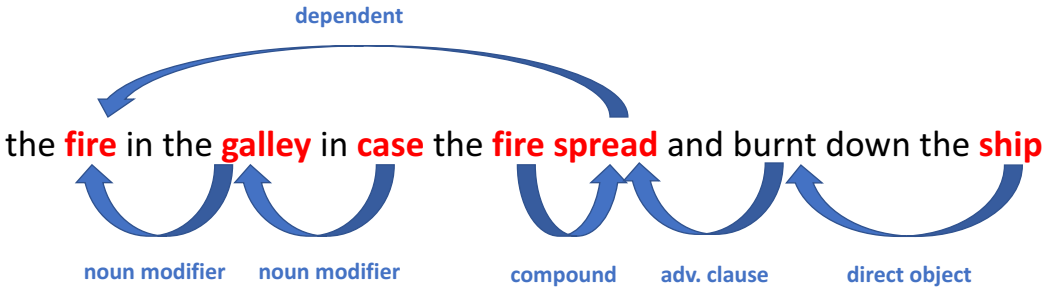
Coding grammatical features



Coding grammatical features

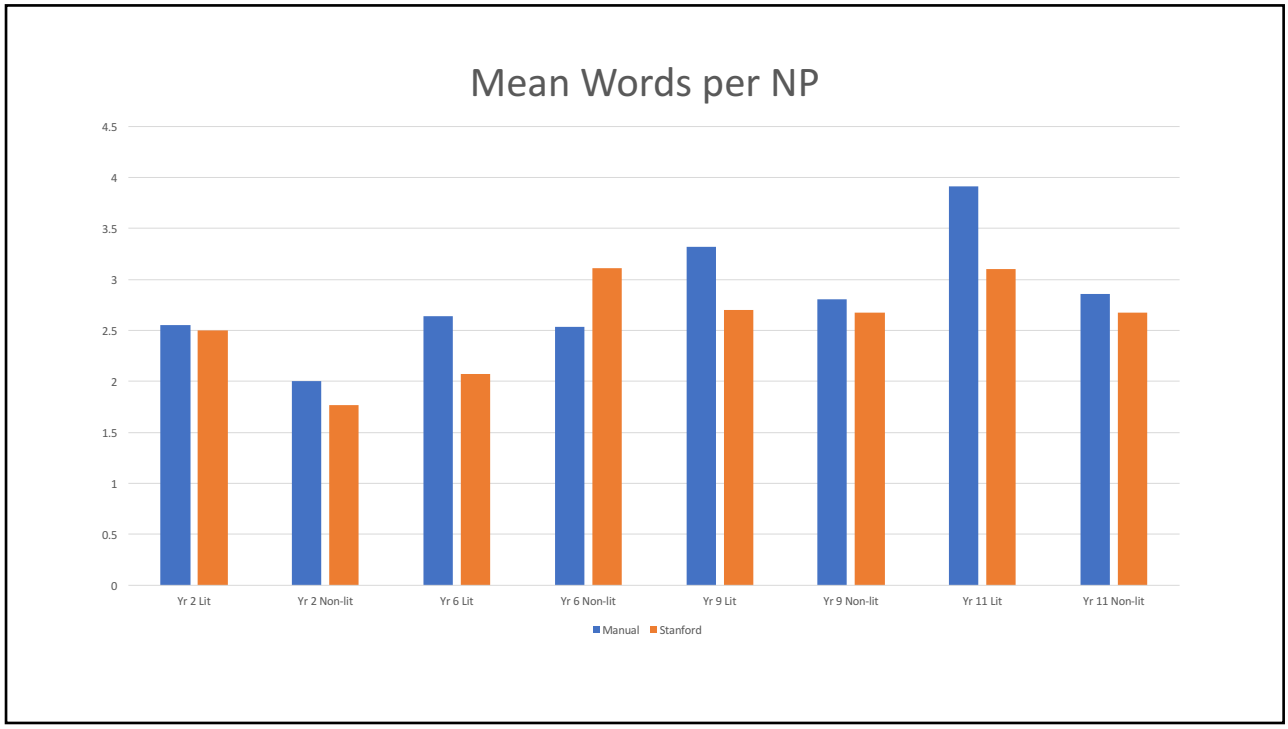


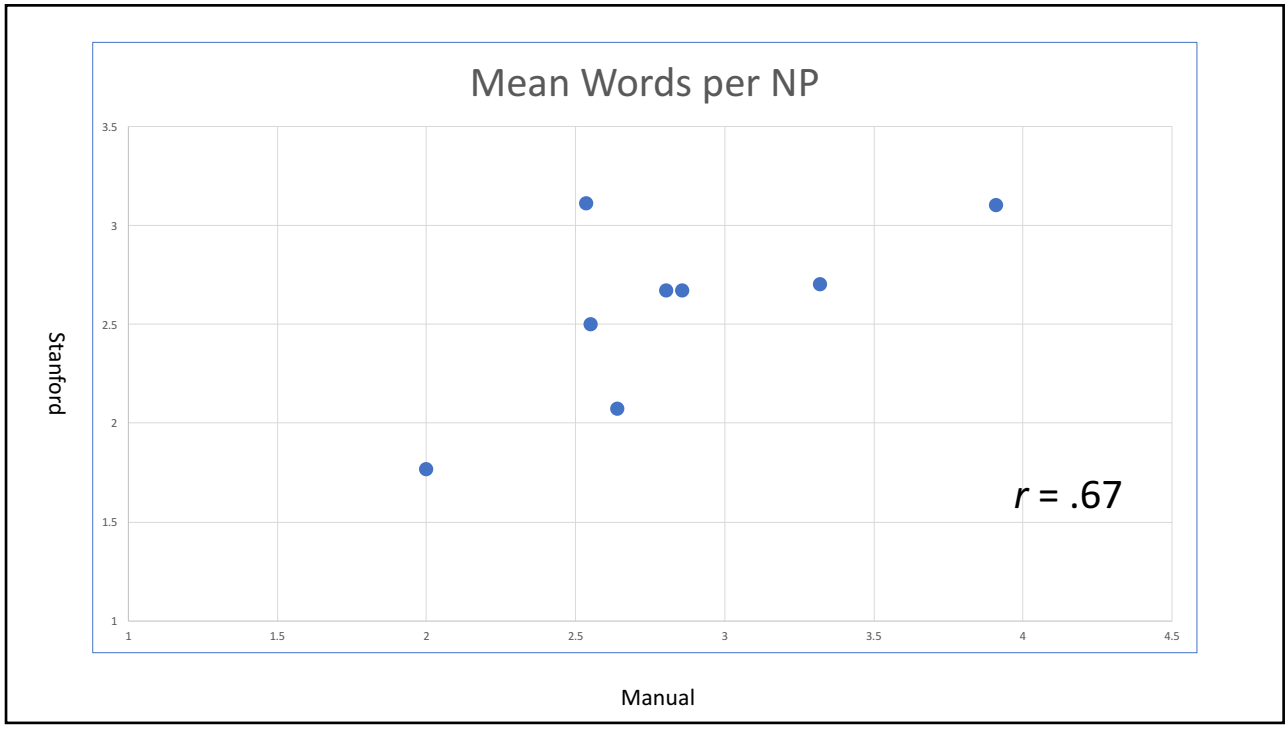
Coding grammatical features

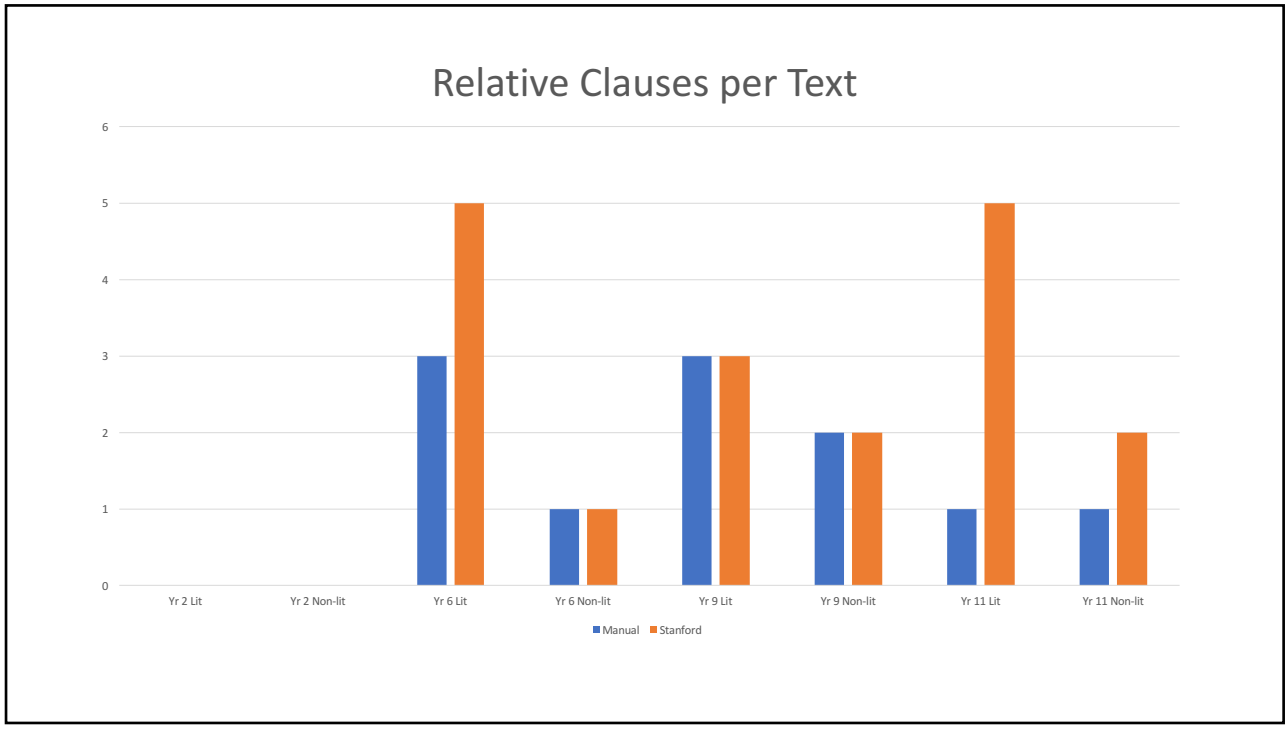


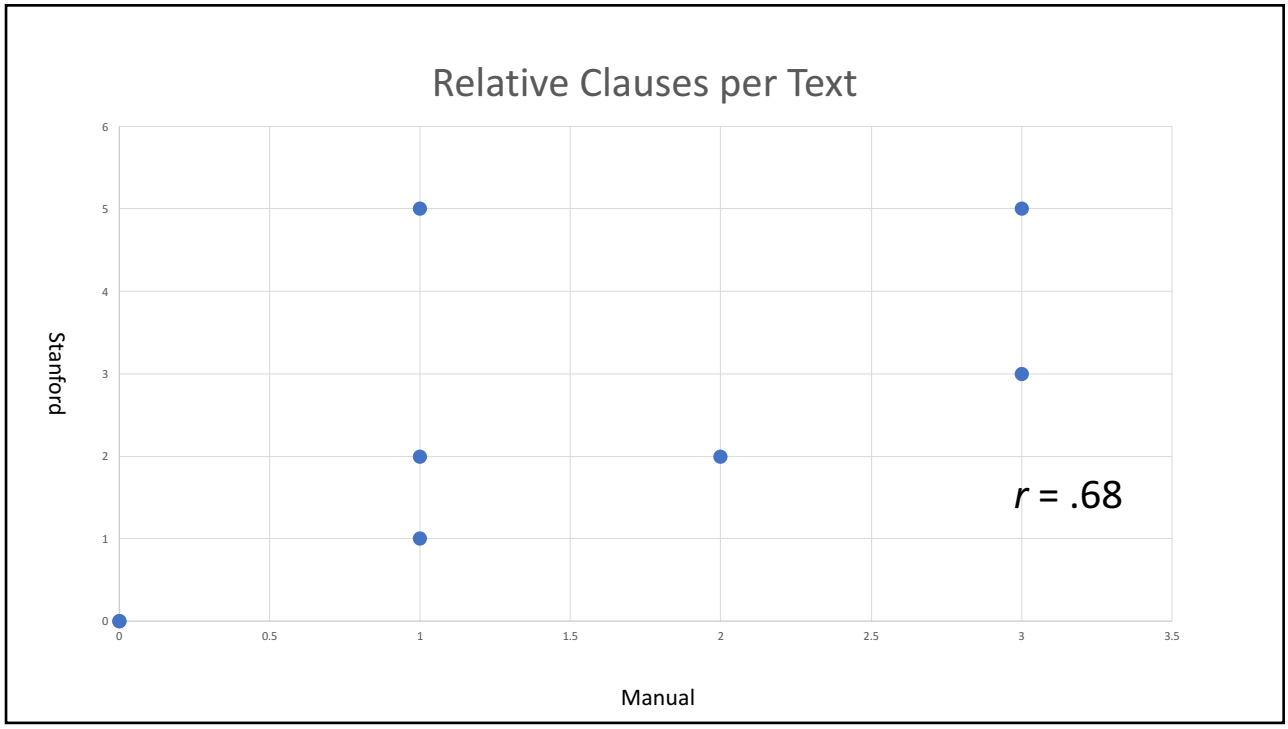
Coding grammatical features

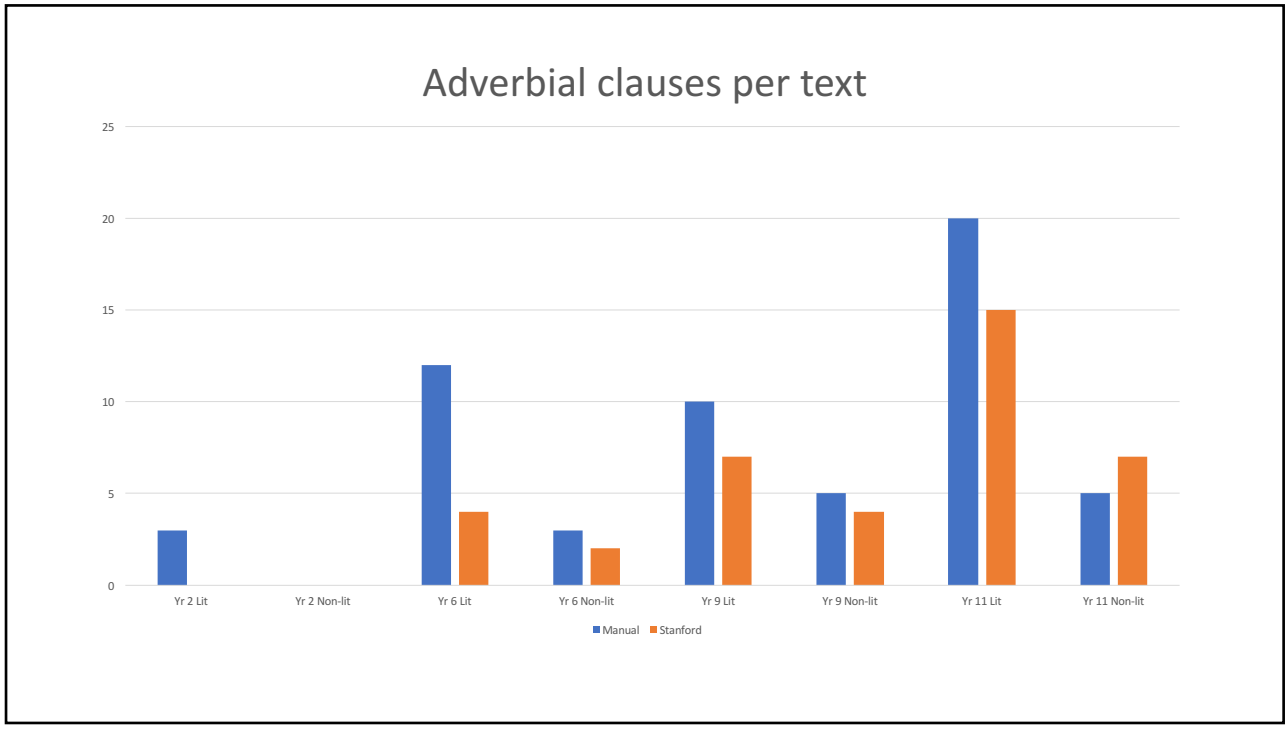
- Target result:
  - 1 x 1 word; 3 x 2 word; 1 x 5 word
  - Mean: 2.4 words per noun phrase
- Stanford:
  - 2 x 1 word; 2 x 2 word; 1 x 4 word; 1 x 8 word; 1 x 15 word
  - Mean: 4.7 words per noun phrase

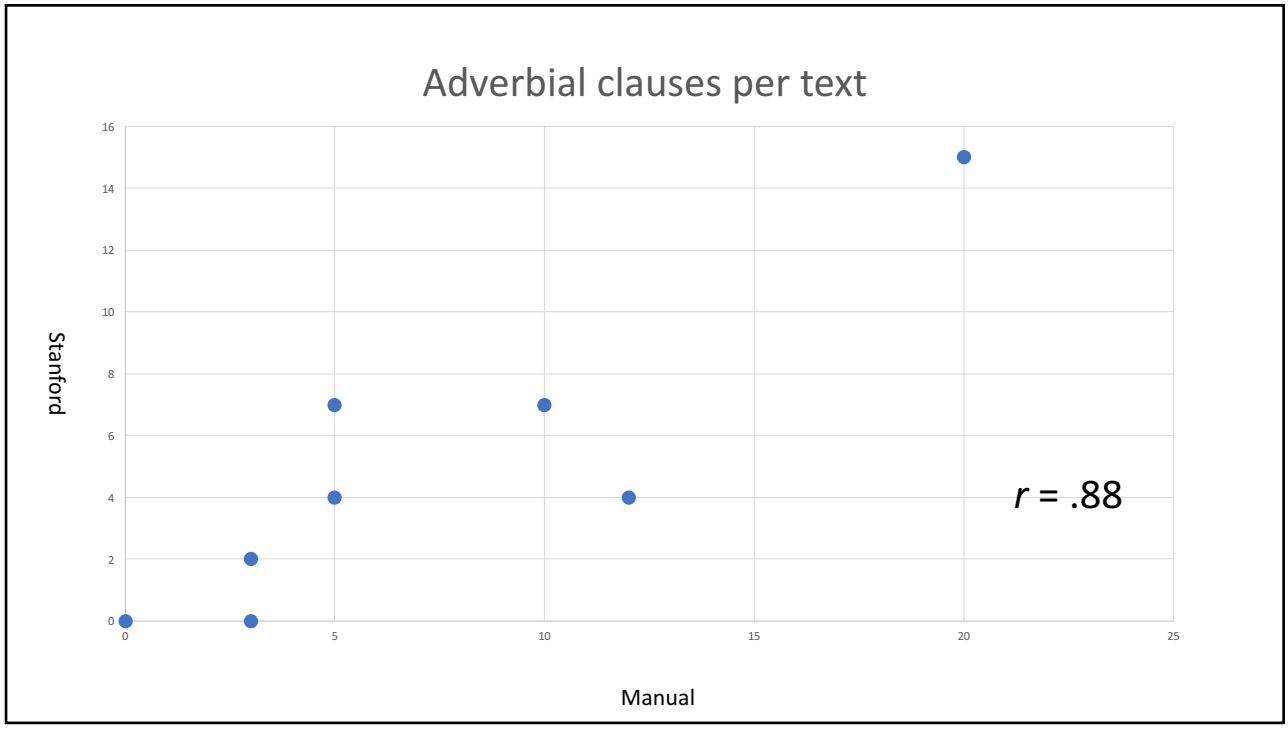

Mean Words per NP
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Yr 2 Lit
Yr 2 Non-lit
Yr 6 Lit
Yr 6 Non-lit
Yr 9 Lit
Yr 9 Non-lit
Yr 11 Lit
Yr 11 Non-lit
Manual
Stanford

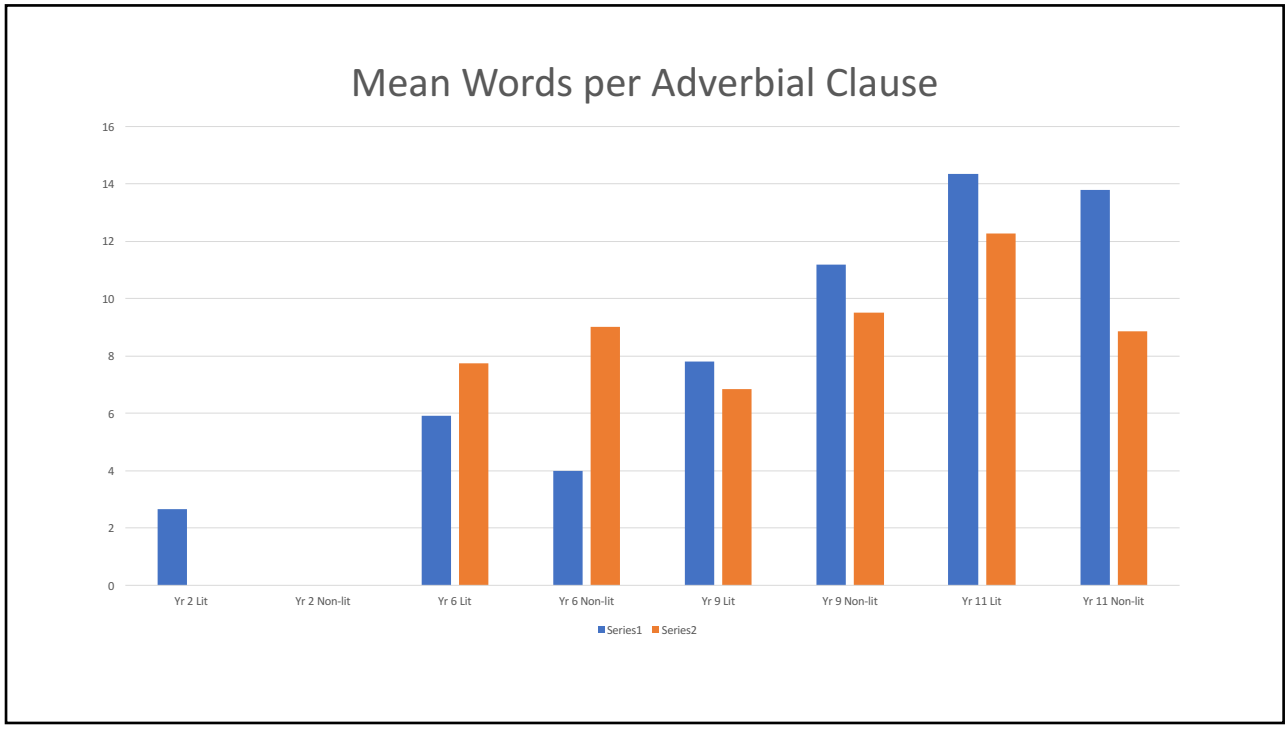

Mean Words per NP
Stanford
Manual
r = .67
3.5
3
2.5
2
1.5
1
1
1.5
2
2.5
3
3.5
4
4.5

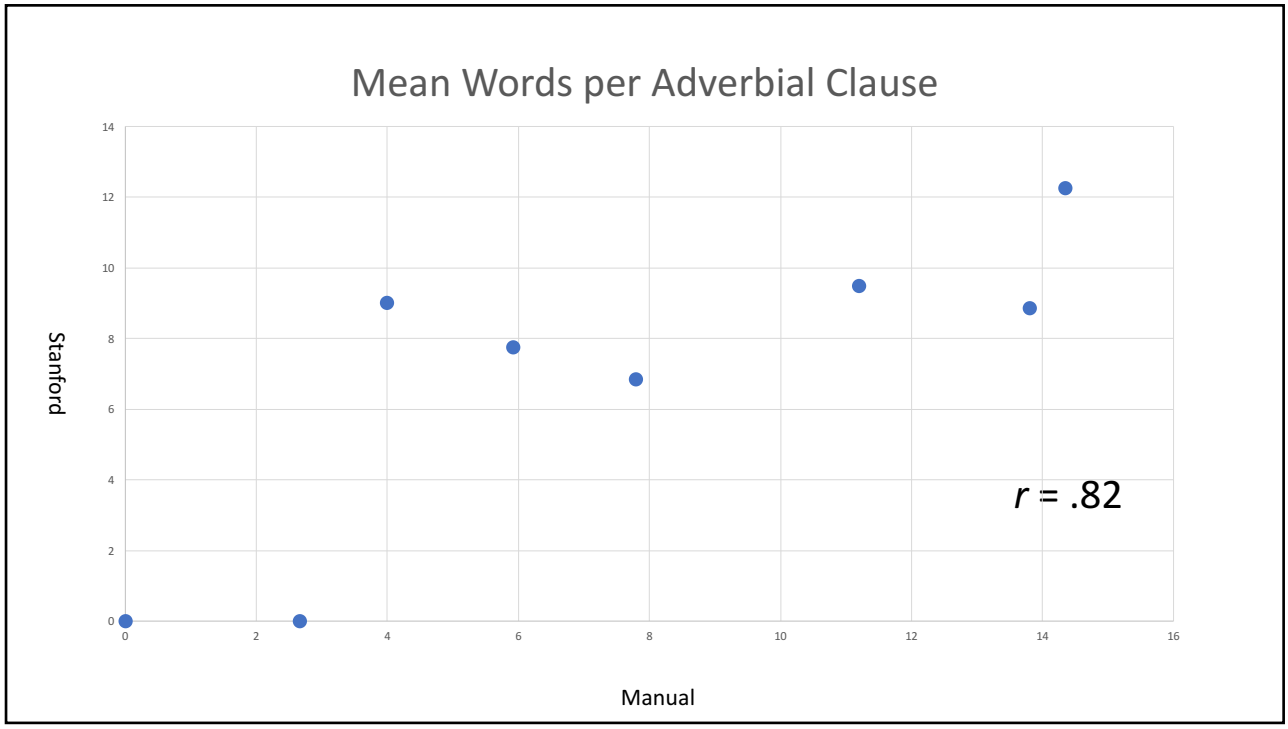

Relative Clauses per Text
6
5
4
3
2
1
0
Yr 2 Lit
Yr 2 Non-lit
Yr 6 Lit
Yr 6 Non-lit
Yr 9 Lit
Yr 9 Non-lit
Yr 11 Lit
Yr 11 Non-lit
Manual
Stanford


Relative Clauses per Text
Stanford
Manual
0
0.5
1
1.5
2
2.5
3
3.5
r = .68


Adverbial clauses per text
25
20
15
10
5
0
Yr 2 Lit
Yr 2 Non-lit
Yr 6 Lit
Yr 6 Non-lit
Yr 9 Lit
Yr 9 Non-lit
Yr 11 Lit
Yr 11 Non-lit
Manual
Stanford


Adverbial clauses per text
Stanford
Manual
16
14
12
10
8
6
4
2
0
0
5
10
15
20
25
r = .88


Mean Words per Adverbial Clause
16
14
12
10
8
6
4
2
0
Yr 2 Lit
Yr 2 Non-lit
Yr 6 Lit
Yr 6 Non-lit
Yr 9 Lit
Yr 9 Non-lit
Yr 11 Lit
Yr 11 Non-lit
Series1
Series2


Mean Words per Adverbial Clause
Stanford
Manual
r = .82
0
2
4
6
8
10
12
14
16

Issues 4:
Understanding attainment

Some Initial Findings:
Vocabulary development

Vocabulary development

Density

Effect	Source	Ages
No age effects	Golub & Frederick 1979	9-10 vs. 11-12
	Berman & Nir 2010	9-10 vs. 12-13 vs. 16-17 vs. adult
	Hall-Mills & Apel 2015	7-8 vs. 8-9 vs 9-10
No correlation with quality	Uccelli et al 2013	Mean age 18;4

Diversity: total types

Effect	Source	Ages
Increases with age	Thomas 1972	7-8 vs. 10-11
	Finn 1977	9-10 vs. 13-14 vs. 16-17
	Wagner et al 2011	6 vs. 9
	Nelson & Van Meter 2007	6-7 vs. 7-8 vs. 8-9 vs. 9-10
No age effects	Hall-Mills & Apel 2015	7-8 vs. 8-9 vs 9-10
Increases with quality	Berse 1974	10-11
	Blatchford 1991	Mean 7;6
	Grobe 1981	10-11; 13-14; 16-17
	Bourke & Adams 2003	Mean 6;11
	Bourke & Adams 2011	Mean 4;9
	Roessingh et al 2015	8-9

Diversity: types controlled for text length

- Types in N words
- Type:token ratio
- Corrected type:token ratio (types/2*tokens^2)
- D
- MTLD

Diversity: Adjusted measures

Effect	Measure	Source	Ages
Increases with age	Types in N words	Clendon & Erikson 2008	Mean 5;7 vs. 6;7 vs. 7;6 vs. 8;5 years
	TTR	Holdich et al 2002	11 years
	Corrected TTR	Moore 1977	14 vs. 17 years
		Olinghouse & Graham 2009	Mean 7.6 vs. 9.6 years
		Olinghouse & Leaird 2009	7-8 vs. 9-10 years
	D	Malvern et al 2004	7 vs. 11 vs. 14 years
		Berman & Nir 2010	9-10 vs. 12-13 vs. 16-17 vs. adult
		Crossley et al 2011	14-15 vs. 16-17 vs. 18-19 years
No age effects	TTR	Hall-Mills & Apel 2015	7-8 vs. 8-9 vs 9-10

Diversity: Adjusted measures

Effect	Measure	Source	Ages
Increases with quality	Types in N words	Koutsoftas & Gray 2012	10 years
	TTR	Grobe 1981	10-11 vs. 13-14 vs. 16-17
	Corrected TTR	Olinghouse & Graham 2009	Mean 7.6 vs. 9.6
		Olinghouse & Leaird 2009	7-8 vs. 9-10
	D	Malvern et al 2004	7 vs. 11 vs. 14
		Ucelli et al 2013	Mean age 18;4
	MTLD	Olinhouse & Wilson 2013	10-11
No correlation with quality	TTR	Cameron et al 1995	9;10

Sophistication

- Word length
- Word Frequency
- Register
- Semantics

Sophistication: Word length

Effect	Measure	Source	Ages
Increases with age	Words >= 7 letters	Houck & Billingsley 1989	10 vs. 14. vs. 17
	Syllables per word & Words > 3 syllables	Olinghouse & Leaird 2009	7-8 < 9-10
Ambiguous age effect	Words > 3 syllables	Berman & Nir-Sagiv 2007	9-10 = 12-13 < 16-17 = adult
No age effect	Letters per word	Malvern et al 2004	7 = 11 = 14
Increases with quality	Letters per word	Massey et al 1996 & 2005	7 & 11 & 14
		Malvern et al 2004	16
		Vanderberg & Swanson 2007	15
		Myhill 2009	16
Ambiguous quality effect	Letters per word	Grobe 1981	✅ 10-11 & 16-17 ❌ 13-14
	Syllables per word; Words > 3 syllables	Olinghouse & Leaird 2009	Task/age/measure dependent weak correlations

Sophistication: Word frequency (age)

Effect	Measure	Source	Ages
Decreases with age	% words < 1/100K	Finn 1977	9-10 < 16-17
	% words not on high-frequency list	Olinghouse & Leaird 2009	7-8 > 9-10
	% 1K words	Sun, Zhang & Scardamalia 2010	8 > 10
	% 2K words	Sun, Zhang & Scardamalia 2010	8 < 10
	% off-list words	Sun, Zhang & Scardamalia 2010	8 < 10
Ambiguous age effects	% words not on high-frequency list	Lawton 1963	12 > 14 for working-class children, not middle class
	Mean frequency from reference corpus	Crossley et al 2011	14-15 < 18-19 14-15=16-17 16-17=18-19
No age effects	P-Lex	Malvern et al 2004	7 = 11 = 14

Sophistication: Word frequency (quality)

Effect	Measure	Source	Ages
Decreases with quality	% words on low-frequency list	Massey et al 1996, 2005	16
	% words on low-frequency list	Olinghouse & Leaird 2009	7-8 & 9-10
Ambiguous effect of quality	% words on low-frequency list	Olinghouse & Wilson 2009	10-11 (effect in 1/3 genres)

Sophistication: Register

Effect	Measure	Source	Ages
Age effect	Greco-Latin vocabulary	Corson 1985	12 < 15
	AWL words	Sun, Zhang & Scardamalia 2010	8 < 10
Ambiguous age effect	Latinate:Germanic ratio	Berman & Nir-Sagiv 2007	9-10 = 12-13 < 16-17 < adult

Sophistication: Semantics

Effect	Measure	Source	Ages
Age effect	Abstract nouns	Fox 1982	15-16 < 17-18
	Word concreteness	Crossley et al 2011	14-15 < 16-17 < 18-19
Ambiguous age effects	Abstract nouns	Nippold, Ward-Londergan & Fanning 2005	7=8 < 11=12
		Berman & Nir-Sagiv 2007	4=7 < 11=adults
	Word polysemy	Crossley et al 2011	14-15 < 18-19 14-15=16-17; 16-17=18-19
Increases with quality	Abstract nouns	Fox 1982	15-16 & 17-18

To summarise…

Lexical Density	No effect (not a vocabulary measure?)
Lexical Diversity	Increases with both age and quality (of course!)
Word Length	Appears to increase with age and quality, but variable results. But what is the construct here?
Word Frequency	Appears to increase with age and quality, but variable results.
Register	Greek/Latin vocabulary and academic vocabulary appear to increase with age, but few studies
Semantics	Contradictory findings re. abstraction. Initial evidence of effect of polysemy
Phraseology	???????

Vocabulary Development in the GiG baby corpus

Sample for the current study

- Years 2, 6 and 9 only
- English/Humanities classes only
- Exclude texts with > 100 illegible words per 1,000
- Exclude poems
- Exclude samples more than 1SD from mean word length
- Randomly select texts to give equal numbers in each year group

Study Corpus

	Schools	Writers	Texts	Text Length				Genre		
				Mean	Median	Min	Max	Story	Exposition	Persuasion
Year 2	3	78	219	66.6	62	27	131	116	99	4
Year 6	4	90	219	284.2	261	120	521	114	82	23
Year 9	6	189	219	343.3	330	181	560	130	59	30

TAALES* Indices

- Frequency
- Ngram frequency/association
- Concreteness ratings

*Kyle & Crossley 2014

Frequency/Range: 72 indices each

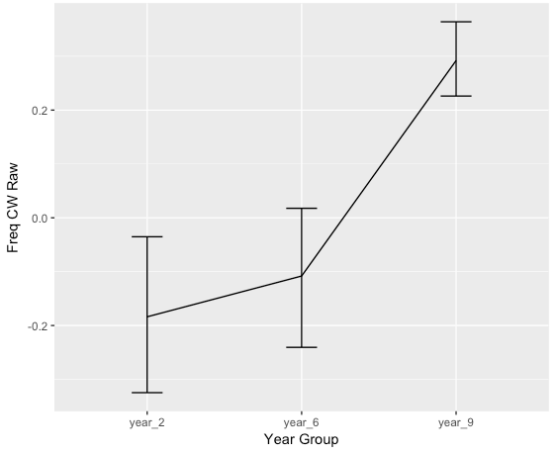
- Range of reference corpora
- Separate indices for:
  - all words vs. content words vs. function words
  - raw frequency vs. log frequency

Combining results from different corpora: frequency

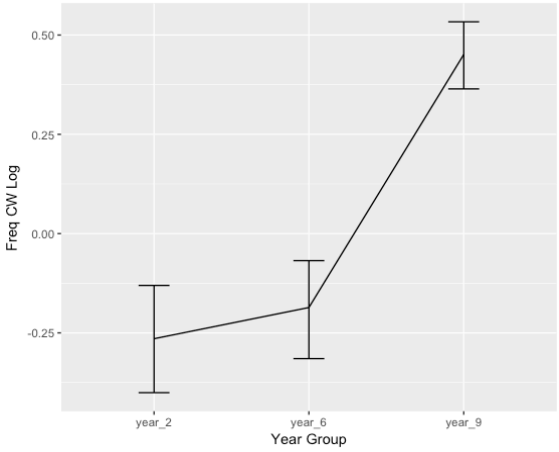
Sub-category (1)	Sub-category (2)	Cronbach's alpha	Deleted
All words	Raw	.99	SUBTLEXus (*r* = .41)
	Log	.98	
Content words	Raw	.98	
	Log	.98	
Function words	Raw	.99	SUBTLEXus (*r* = .41)
	Log	.98	

		All words		Content words		Function words	
		Raw	Log	Raw	Log	Raw	Log
All words	Raw	1.00					
	Log	0.18	1.00				
Content words	Raw	-0.15	0.61	1.00			
	Log	-0.27	0.80	0.76	1.00		
Function words	Raw	0.83	-0.17	-0.30	-0.42	1.00	
	Log	0.68	-0.08	-0.29	-0.35	0.86	1.00

Content words



F(2, 654)=18.38, p < .001, ω^2=.06
post-hoc (FDR correction): year 2 = year 6 < year 9

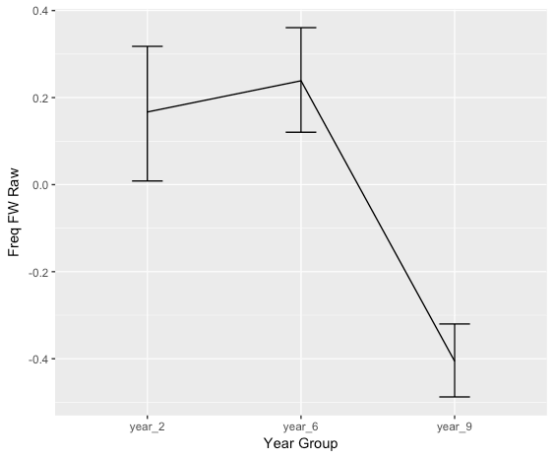


F(2, 654)=43.99, p < .001, ω^2=.11
post-hoc (FDR correction): year 2 = year 6 < year 9

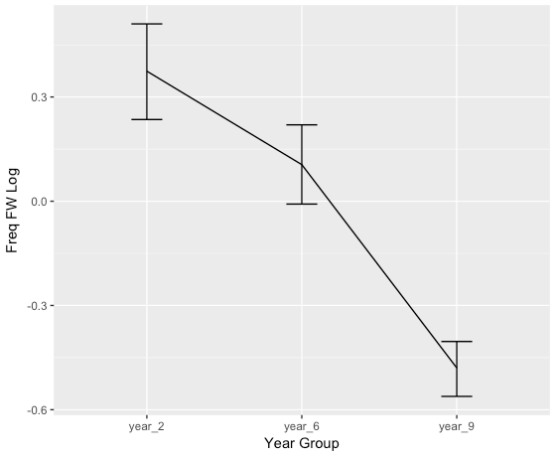
For example

One morning, five meerkats finished eating crunchy scorpians but there preditors like slithering snakes, lions and falcons. But one woke up and went meerkats, meerkats and woke up the others. One stepped on the snake and it went hiss hiss! And the other predators ran away. But a fennec fox was coming to take two pups. The scary fox wishes to eat them even the nice scorpians.

Function words



F(2, 654)=32.47, p < .001, ω^2=.09
post-hoc (FDR correction): year 2 = year 6 > year 9



F(2, 654)=57.45, p < .001, ω^2=.15
post-hoc (FDR correction): year 2 > year 6 > year 9
*Kyle & Crossley 2014

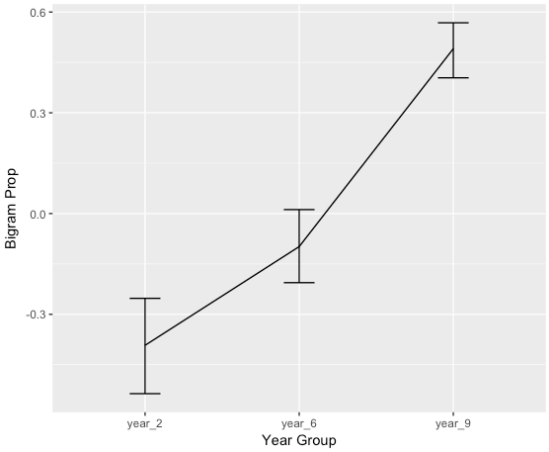
Ngram measures: Frequency & Proportion

- Frequency: 7 x reference corpora; Association: 5 x reference corpora
- Frequency: Raw vs. Log vs. Proportion
  - Proportion: 10K; 20K; 30K…100K)

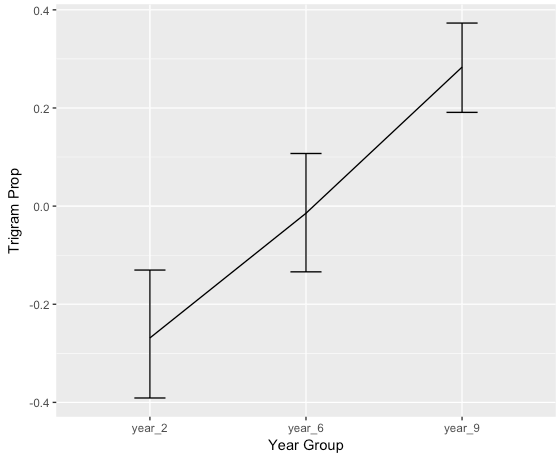
Combining results from different corpora: Ngram proportions/frequencies

Category	Sub-category	Cronbach's alpha	Deleted
Frequency	Bigram Proportion	1	
	Bigram Log	.95	
	Trigram Proportion	1	
	Trigram Log	.93	BNC Spoken: .14 BNC Written: .20 COCA Academic: .65

Proportions: Bigram & Trigram

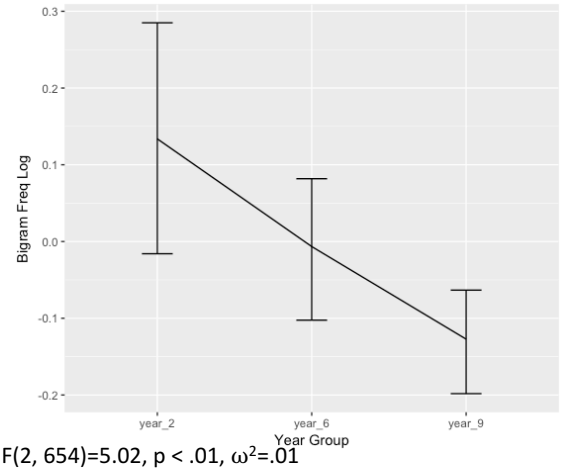


F(2, 654)=58.43, p < .001, ω^2=.15
post-hoc (FDR correction): year 2 < year 6 < year 9

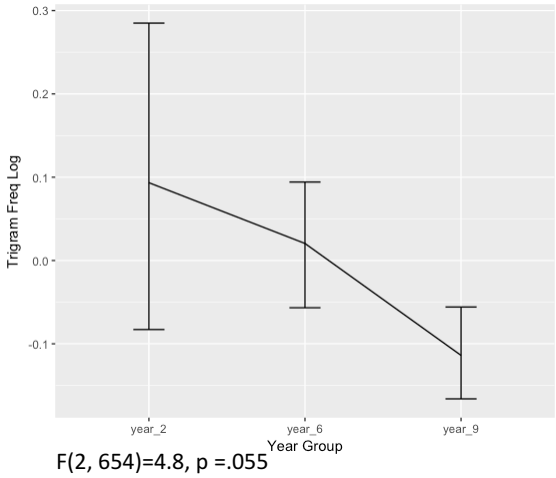


F(2, 654)=21.38, p < .001, ω^2=.06
post-hoc (FDR correction): year 2 < year 6 < year 9

Frequency: Bigram & Trigram (Logs)



F(2, 654)=5.02, p < .01, ω^2=.01
post-hoc (FDR correction): year 2 < year 9;
year 2 = year 6; year 6 = year 9



F(2, 654)=4.8, p =.055

Ngram measures: Association

- Association:
  - MI, MI2; t-score; Delta-P; Collexeme
  - Bigram vs. Trigram
    - Trigram 1 (*double* – *espresso please*)
    - Trigram 2 (*double espresso* – *please*)

Combining results from different corpora

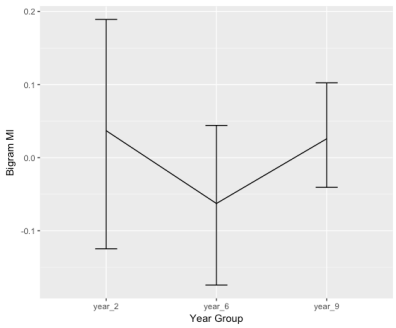
Sub-category	Cronbach's alpha	Deleted
Bigram / MI	.95	
Bigram / MI2	.95	
Bigram t-score	.97	
Bigram Delta-P	.97	
Bigram Collexeme	.99	

Combining results from different corpora

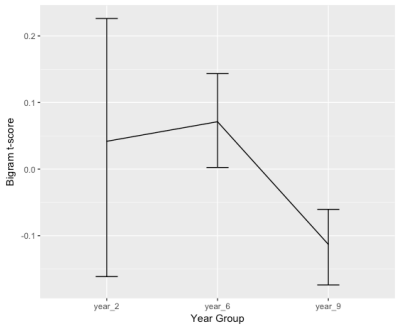
	Sub-category	Cronbach's alpha	Deleted
Trigram (1)	MI	.93	COCA Academic: .59
	MI2	.91	COCA Academic: .64
	t-score	.94	
	Delta-P	.94	
	Collexeme	.93	
Trigram (2)	MI	.95	COCA Academic: .65
	MI2	.93	COCA Academic: .62
	t-score	.94	COCA Fiction: .67
	Delta-P	.94	
	Collexeme	.94	COCA Academic: .68

	2G MI	2G MI2	2G T	2G DP	2G Clex	3G1 MI	3G1 MI2	3G1 T	3G1 DP	3G1 Clex	3G2 MI	3G2 MI2	3G2 T	3G2 DP	3G2 Clex
2G MI	1.00														
2G MI2	.61	1.00													
2G T	.30	.65	1.00												
2G DP	.30	.37	.44	1.00											
2G Clex	-.06	.32	.76	.43	1.00										
3G(1 MI	.45	.05	-.12	.13	-.24	1.00									
3G1 MI2	.39	.36	.23	.22	.08	.67	1.00								
3G1 T	.23	.42	.43	.20	.29	.19	.72	1.00							
3G1 DP	.06	.12	.29	.38	.43	.19	.34	.38	1.00						
3G1 Clex	.13	.36	.39	.21	.35	.09	.60	.91	.43	1.00					
3G2 MI	.46	.07	-.09	.16	-.19	.65	.47	.14	.13	.09	1.00				
3G2 MI2	.43	.33	.20	.24	.07	.49	.80	.61	.26	.53	.73	1.00			
3G2 T	.22	.42	.44	.28	.32	.12	.64	.84	.38	.84	.22	.67	1.00		
3G2 DP	.15	-.01	.11	.49	.19	.34	.35	.24	.49	.24	.44	.43	.31	1.00	
3G2 Clex	.12	.36	.39	.22	.36	.07	.59	.87	.43	.98	.11	.56	.89	.27	1.00

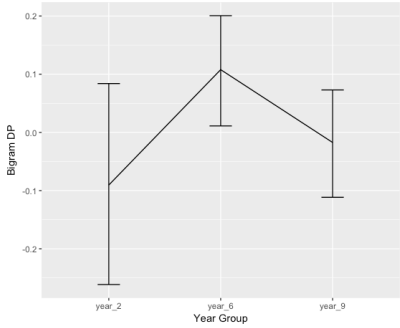
Bigram associations: MI; t-score; DP



F(2, 654)=0.79, p >.05

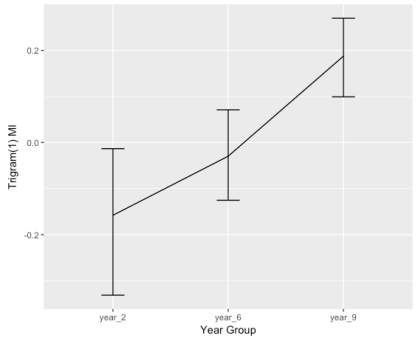


F(2, 654)=2.37, p >.05

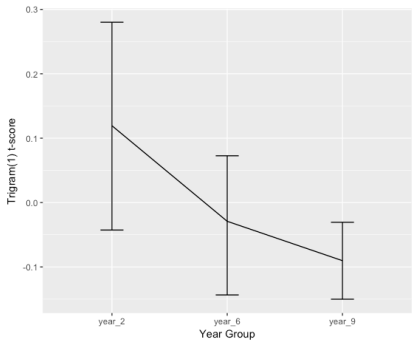


F(2, 654)=2.44, p >.05

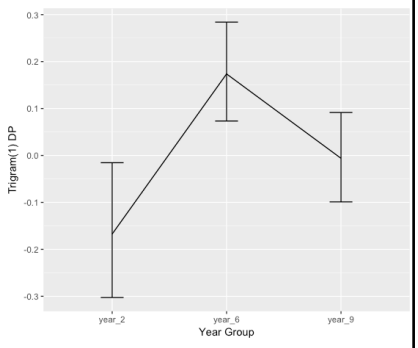
Trigram (1) associations: MI; t-score; DP



F(2, 654)=8.30, p < .001, ω^2=.02
post-hoc (FDR correction):
year 2 = year 6 < year 9

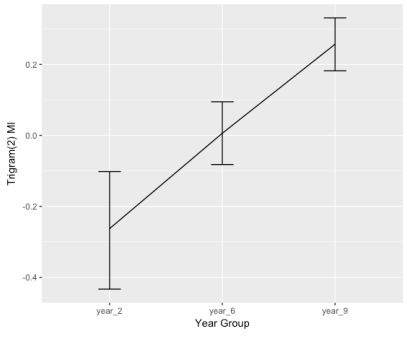


F(2, 654)=3.17, p < .05, ω^2=.01
post-hoc (FDR correction):
year 2 = year 6; year 6 = year 9
Year 2 < year 9

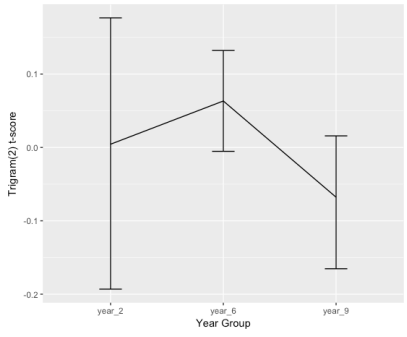


F(2, 654)=7.97, p < .001, ω^2=.02
post-hoc (FDR correction):
year 2 = year 9; year 6 = year 9
year 2 < year 6

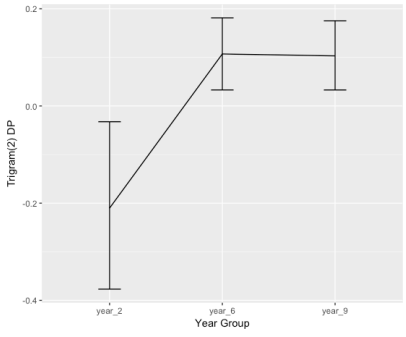
Trigram (2) associations: MI; t-score; DP



F(2, 654)=17.79, p < .001, ω^2=.05
post-hoc (FDR correction):
year 2 < year 6 < year 9



F(2, 654)=1.11, p > .05

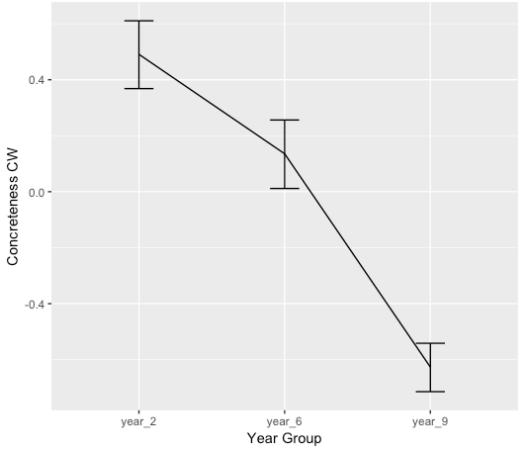


F(2, 654)=9.32, p < .001, ω^2=.03
post-hoc (FDR correction):
year 2 < year 6; year 2 < year 9
Year 6 = year 9

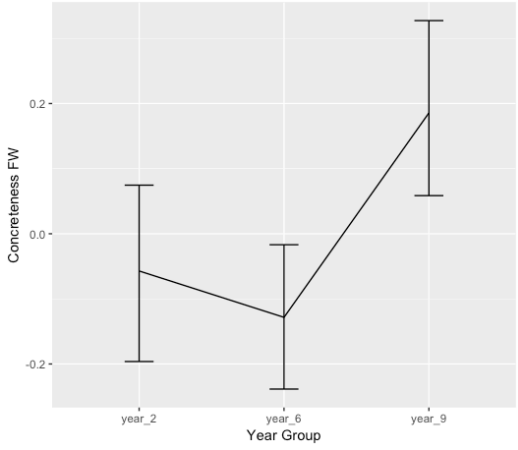
Concreteness

Sub-category	Cronbach's alpha
Content words	.95
Function words	.94

Concreteness



F(2, 654)=96.49, p < .001, ω^2=.23
post-hoc (FDR correction): year 2 > year 6 > year 9



F(2, 654)=11.9, p < .005, ω^2=.02
post-hoc (FDR correction): year 2 = year 6 < year 9

Conclusions - methodological

- Counts from different reference corpora mostly consistent
- Log frequencies enable patterns to emerge more clearly

Conclusions: frequency

- Mean content word frequency increases with age
- Mean function word frequency decreases with age

Conclusions: n-grams

- Percentage of ngrams attested in corpora increases with age
- Frequency of attested ngrams may decrease with age
- MI & DP of attested trigrams increase with age
- T-scores of attested trigrams decrease with age(?)

Conclusions - concreteness

- Older children use more abstract content words
- Older children use more concrete function words(??)

So…

- Corpus research under-exploited in study of L1 English writing development.
- Our corpus to be completed early 2018
- Online (with APU) late 2018
- Full analyses of vocabulary, NP-expansion, subordination soon.
- And the book of the lit. review…
- Future prospects:
  - Historical oral/written corpus
  - Studies of attainment

Keep in touch!

- Twitter: @growing grammar
- Facebook: facebook.com/growthingrammar
- Email: p.l.durrant@exeter.ac.uk

References

- Berman, R. A., & Nir, B. (2010). The lexicon in writing-speech-differentiation. *Written Language and Literacy, 13*(2), 183-205.
- Berman, R. A., & Nir-Sagiv, B. (2007). Comparing narrative and expository text construction across adolescence: A developmental paradox. *Discourse Processes, 43*(2), 79-120.
- Berse, P. (1974). Criteria for the assessment of pupils' compositions. *Educational Research, 17*(1), 54-61.
- Blatchford, P. (1991). Children's writing at 7 years: Associations with handwriting on school entry and pre-school factors. *British Journal of Edcucational Psychology, 61*, 73-84.
- Bourke, L., & Adams, A.-M. (2003). The relationship between working memory and early writing assessed at the word, sentence and text level. *Educational and Child Psychology, 20*(3), 19-36.
- Bourke, L., & Adams, A.-M. (2011). Is it differences in language skills and working memory that account for girls being better at writing than boys. *Journal of Writing Research, 3*(3), 249-277.
- Cameron, C. A., Lee, K., Webster, S., Munro, K., Hunt, A. K., & Linton, M. J. (1995). Text cohesion in children's narrative writing. *Applied Psycholinguistics, 16*, 257-269.
- Clendon, S. A., & Erikson, K. A. (2008). The vocabulary of beginning writers: Implications for children with complex communication needs. *Augmentative and alternative communication, 24*(4), 281-293.
- Corson, D. (1985). *The lexical bar*. Oxford: Pergamon.
- Crossley, S. A., Weston, J. L., Sullivan, S. T. M., & McNamara, D. (2011). The development of writing proficiency as a function of grade level: A linguistic analysis. *Written Communication, 28*, 282-311.

References

- Department of Education (2014). *The national curriculum in England: Framework document. December 2014.*
- Finn, P. J. (1977). Computer-aided description of mature word choices in writing. In C. R. Cooper & L. Odell (Eds.), *Evaluating writing: Describing, Measuring, Judging* (pp. 69-89). Urbana, Illinois: National Council of Teachers of English.
- Fox, J., & Mahood, J. (1982). Survey review: Lexicons and the ELT materials writer. *ELT Journal, 36*(2), 125-129. doi:10.1093/elt/36.2.125
- Golub, L. S., & Frederick, W., C. (1979). *Linguistic Structures in the discourse of fourth and sixth graders*. Madison, Wisconsin: Center for Cognitive Learning, The University of Wisconsin.
- Grobe, C. (1981). Syntactic Maturity, Mechanics, and Vocabulary as Predictors of Quality Ratings. *Research in the Teaching of English, 15*(1), 75-85.
- Hall-Mills, S., & Apel, K. (2015). Linguistic feature development across grades and genre in elementary writing. *Language, speech and hearing services in schools, 36*, 242-255.
- Holdich, C. E., Holdich, R. G., & Chung, P. W. H. (2002). Assessing aspects of children's written grammar: automating the process. *Computers and Education, 39*, 27-50.
- Houck, C. K., & Billingsley, B. S. (1989). Written expression of students with and without learning disabilities: Differences across the grades. *Journal of Learning Disabilities, 22*(9), 561-572.
- Koutsoftas, A. D., & Gray, S. (2012). Comparison of narrative and expository writing in students with and without language-learning disabilities. *Language, speech and hearing services in schools, 43*, 395-409.
- Lawton, D. (1963). Social class differences in language development: A study of some samples of written work. *Language and Speech, 6*(3), 120-143.

References

- Malvern, D., Richards, B. J., Chipere, N., & Duran, P. (2004). *Lexical diversity and language development*. Basingstoke: Palgrave Macmillan.
- Massey, A. J., & Elliott, G. L. (1996). *Aspects of Writing in 16+ English examinations between 1980 & 1994*. Cambridge: University of Cambridge Local Examinations Syndicate.
- Massey, A. J., Elliott, G. L., & Johnson, N. K. (2005). *Variations in aspects of writing in 16+ English examinations between 1980 and 2004: Vocabulary, spelling, punctuation, sentence structure, non-standard English*. Cambridge: Cambridge Assessment.
- Moore, E. M. (1977). Lexical characteristics of adolexcent prose: writer and task effects. *British Journal of Edcucational Psychology, 47*, 285-295.
- Myhill, D. (2009). From talking to writing: Linguistic development in writing. *BJEP Monograph Series II, Number 6 - Teaching and Learning Writing, 27*(44), 27-44.
- Nelson, N. W., & Van Meter, A. M. (2007). Measuring written language ability in narrative samples. *Reading and Writing Quarterly, 23*(3), 287-309.
- Nippold, M. A., Ward-Lonergan, J. M., & Fanning, J. L. (2005). Persuasive writing in children, adolescents, and adults: A study of syntactic, semantic, and pragmatic development. *Language, speech and hearing services in schools, 36*(2), 125-138.
- Olinghouse, N., G., & Graham, S. (2009). The relationship between the discourse knowledge and the writing performance of elementary-grade students. *Journal of Educational Psychology, 101*, 37-50.

References

- Olinghouse, N., G., & Leaird, J. T. (2009). The relationship between measures of vocabulary and narrarive writing quality in second- and fourth-grade students. *Reading and Writing, 22*, 545-565.
- Olinghouse, N., G., & Wilson, J. (2013). The relationship between vocabulary and writing quality in three genres. *Reading and Writing, 26*, 45-65.
- Roessingh, H., Elgie, S., & Kover, P. (2015). Using lexical profiling tools to investigage children's written vocabulary in grade 3: An exploratory study. *Language Assessment Quarterly, 12*(1), 67-86.
- Sun, Y., Zhang, J., & Scardamalia, M. (2010). Knowledge building and vocabulary growth over two years, Grades 3 and 4. *Instructional Science, 38*, 147-171.
- Thomas, V. (1972). The basic writing vocabulary of elementary school children. *Alberta Journal of Educational Research, 18*(4), 243-248.
- Uccelli, P., Dobbs, C. L., & Scott, J. (2013). Mastering academic language: Organizatoin and stance in the persuasive writing of high school students. *Written Communication, 30*(1), 36-62.
- Vanderberg, R., & Swanson, H. L. (2007). Which components of working memory are important in the writing process? *Reading and Writing, 20*(721-752).
- Wagner, R. K., Puranik, C. S., Foorman, B., Foster, E., WIlson, L. G., Tschinkel, E., & Kantor, P. T. (2011). Modeling the development of written language. *Reading and Writing, 24*, 203-220.