

Growth in Grammar

Phil Durrant

Debra Myhill

Mark Brenchley

A Year 2 Text (Age 6-7)

one time ago there was a king called king james the first. and the catholics did not like him. and there was a bad man called Guy Fawkes he wanted to blow the houses of parliament he wanted to kill the king too . as well as the catholics he hid 36 barrels of gun powder and he hid it. Robert catesby sent a letter to the king.

A Year 11 Text (Age 15-16)

Dear Editor, I am writing to express my opposition to the article regarding “teenage tearaways” which was recently published in your newspaper. I could not help but notice the considerable bias in the article but I read on, only to witness your reporter put words into your readers’ mouths sentence, after sentence, after sentence.

Big Questions

- How does children's written language change as they get older?
- What clues does this give us about their cognitive/social development?
- How is writing which receives high grades different from writing which receives low grades?
- How do these interact with genre, topic etc.?

Our study: The short version...

- **Our focus:** development in the written language of school children in England.
- **Our approach:** quantitative analysis of features in a large sample of texts
- **Our data:** a corpus of school writing at Years 2, 5, 9 and 11 (ages 5-16)
- **Our analysis:** what sets of linguistic features occur at each level? How do these distinguish ages/levels of attainment?

THE LONG VERSION...

OUR FOCUS



What are we studying?

- Development in written language
- Development in written language

What are we studying?

- Grammar as a resource for making meaning
- NOT grammatical accuracy

OUR APPROACH

Studying texts

- An important aspect of understanding writing development is understanding how use of written language develops
- This is best done by studying authentic texts
- Much can be learned through intensive analysis of small numbers of texts. But...
- Studying large numbers of texts offers an important additional perspective

Advantages of studying large numbers of texts

- Enables emergence of patterns which may not be obvious to analysts working on smaller samples:
 - Subtle patterns, such as semantic prosody
 - Frequency variations across texts/writers
- Enables robust generalizations

Disadvantages of studying large numbers of texts

- Requires transformation to electronic format, so loss of some original features
- Limited to features which can be (semi-) automatically identified
- Decontextualizes features

WHAT DO WE ALREADY (THINK WE) KNOW?

Age-related differences found in

- How clauses are linked together
- Use/structure of phrases
- Use of vocabulary
- Information structure

How clauses are linked together

- From co-ordination:
 - *I went to the zoo yesterday and I saw large fat rinos. I also saw some girafes and they were very very very long. I saw lions aswel but they were asleep* (Year 2; Age 6-7)

How clauses are linked together

- To subordination:
 - *In fact, I would suggest that it was irresponsible for you to publish such a letter without allowing people to form their own opinion*
 - *Perhaps this article was so busy condemning the organisation that you were not made aware that it is actually a charity (Year 11; Age 15-16)*

Types of subordination: Relative clauses

- Increase with age
- Greater % of non-defining relatives with age
 - *I am writing to express my opposition to the article regarding “teenage tearaways” which was recently published in your newspaper.*
 - *The charity has worked hard to raise the money to send these students (who are probably quite troubled or misled) to an inspiring place to learn.*
 - *These children are likely to need support and advice, which is what the Divert Trust offers.*

(Year 11; Age 15-16)

Types of subordination: Noun clauses

- E.g.:
 - *It is rightly stated that children in Jamaica value their education even though they are impoverished.*
 - *This will enable the “tearaways” to see what opportunities are offered in the UK*
 - *That opportunities like this are not offered to other children may seem unfair*
- Increase with age
- Move from direct object only to other uses

Types of subordination: Adverbial clauses

- Decrease with age
- Increased sophistication in meaning:
 - From **time**:
 - *When the meerkat woke up* and he said to himself *I am going to be in trouble*
 - To **concession**:
 - *Although I understand some of the points made in the article*, *I am appalled at the negative way your journalist portrayed...*

The structure of phrases: noun phrases

- Increase in length
 - Greater/more complex/diverse modification
 - Move from packaging information across clauses to modification within phrase:
 - E.g.: Janet (Harpin, 1976: 72-3):
 - Aged 7.5: *My little dog Kim is brown and he barks a lot and he has a furry coat and a red collar with his name on*
 - Aged 9: *those old, tangled, mossy, twisted branches creaking in the wind*

The structure of phrases: noun phrases

- More frequent nominalizations
 - Gerunds: **Talking in front of a large audience** can be intimidating
 - Appositives: **Mr Jones, the banker**, was...
 - Infinitivals: the aim was **to hit the other team's supply of ammunition**
- Fewer personal pronouns

The structure of phrases: verb phrases

- Auxiliary verbs: increase
 - modals; perfect/progressive aspect; passive...
- Passive voice: increases
 - *get* passive: decreases
- Non-finite verbs: increase
- Prepositional verbs increase; phrasal verbs decrease

The structure of phrases: non-clausal adverbs

- Decrease overall
- Time/place/motion decrease
- Manner increases

Vocabulary

- Greater variety
- Lower frequency words
- Greater 'lexical density'
- More Latinate words
- More abstract words

Information structure

- Greater use of ‘marked themes’:
 - Fewer sentence-initial subjects
 - More adverbial sentence openings
- Wider range of discourse connectors
- Greater use of substitution/ellipsis

A key caveat

- Text genre, topic and audience also have a strong influence on the above features

OUR DATA



The corpus

- Around 6,000 texts which students have already produced as part of their school work in English, History or Science
- From schools across England:
 - Northern : Southern
 - Urban : Rural
 - Range of Socio-economic status

The corpus

- Writing produced towards the end of:
 - Key Stage 1 (Year 2; approx. age 7)
 - Key Stage 2 (Year 6; approx. age 11)
 - Key Stage 3 (Year 9; approx. age 14)
 - Key Stage 4 (Year 11; approx. age 16)
- At a full range of attainment levels

Transcription

- Transcribed by a team of assistants
- Formatted according to TEI-guidelines to ensure compatibility

Identifying linguistic features

- POS tagging
 - CLAWS
 - Stanford tagger
- Syntactic parsing
 - Stanford CFPG parser
- Custom programs

Identifying linguistic features

Part of Speech tagging

- CLAWS

030 I
040 am
050 writing
060 to
070 express
080 my
090 opposition
100 to
110 the
120 article
130 regarding

93 [PPIS1/100] MC1@/0 ZZ1@/0
93 [VBM/100] RA@/0
93 [VVG/100] JJ%/0 NN1/0
97 T0
97 VVI
93 APPGE
03 NN1
93 [II/100] T0/0
93 AT
03 NN1
03 [II/80] VVG/20

Identifying linguistic features

Part of Speech tagging

- Stanford

```
<token id="4">
  <word>I</word>
  <lemma>I</lemma>
  <POS>PRP</POS>
</token>
<token id="5">
  <word>am</word>
  <lemma>be</lemma>
  <POS>VBP</POS>
</token>
<token id="6">
  <word>writing</word>
  <lemma>write</lemma>
  <POS>VBG</POS>
</token>
```

```
<token id="7">
  <word>to</word>
  <lemma>to</lemma>
  <POS>TO</POS>
</token>
<token id="8">
  <word>express</word>
  <lemma>express</lemma>
  <POS>VB</POS>
</token>
<token id="9">
  <word>my</word>
  <lemma>my</lemma>
  <POS>PRP$</POS>
</token>
```

Identifying linguistic features

Syntactic Parsing: Stanford Parser

(S
 (NP
 (PRP I)
)
 (VP
 (VBP am)
 (VP
 (VBG writing)
 (S
 (VP
 (TO to)
 (VP
 (VB express)
 (NP
 (PRP\$ my)
 (NN opposition)
)
)
)
)
)
)
)

Analysis

- Linguistic features of possible interest identified through systematic lit. review
- Features automatically counted through programmes, in combination with taggers/parser
- Generates a very large set of data!

Analysis: MDA

- Linguistic features often occur in combination, e.g.:
 - Nominalization, low-frequency words, passives
 - Conversation verbs, 1st person pronoun, NOT attributive adjectives

Analysis: MDA¹

¹ Biber, 1988

- Derive components summarizing underlying dimensions in the data
- Interpret communicative functions
- How do texts from different age levels, attainment levels and genres differ in their use of dimensions?

Archiving the Corpus

- Corpus will be freely available to teachers and researchers in 'clean' and tagged forms
- Online interactive tool to be developed for easy searching of corpus

Future directions

- Making the corpus accessible to teachers
- Parallel studies in other contexts/languages

References

- Berman, R. A., & Nir, B. (2010). The lexicon in writing-speech-differentiation: Developmental perspectives. *Written Language & 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.
- Biber, D. (1988). *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
- Blount, N. S., Johnson, S. L., & Frederick, W. C. (1969). *A comparison of the writing of eighth- and twelfth-grade students*. Madison: University of Wisconsin, Research and Development Center for Cognitive Learning.
- Chipere, N., Malvern, D., Richards, B., & Duran, P. (2001). Using a corpus of school children's writing to investigate the development of vocabulary diversity. *Technical papers, Volume 13. Special Issue: Proceedings of the Corpus Linguistics 2001 Conference*, 126-133.
- Christie, F., & Derewianka, B. (2008). *School discourse: Learning to write across the years of schooling*. New York: Continuum.
- Crowhurst, M., & Piche, G. L. (1979). Audience and mode of discourse effects on syntactic complexity in writing at two grade levels. *Research in the teaching of English*, 13(2), 101-109.

References

- Harpin, W. (1976). *Second 'R': Writing Development in the Junior School*. London: Allen & Unwin.
- Hunt, K. W. (1965). *Grammatical Structures Written at Three Grade Levels*. Urbana, IL: NCTE.
- Loban, W. (1976). *Language development: Kindergarten through grade twelve*. MCTE Committee on Research Report No. 18. Urbana, Ill: National Council of Teachers of English.
- Myhill, D. (2008). Towards a linguistic model of sentence development in writing. *Language and Education*, 22(5), 271-288.
- Olinghouse, N. G., & Leaird, J. T. (2009). The relationship between measures of vocabulary and narrative writing quality in second- and fourth-grade students. *Reading and Writing*, 22(5), 545-565.
- Ravid, D., & Berman, R. A. (2010). Developing noun phrase complexity at school age: A text-embedded cross-linguistic analysis. *First language*, 30(1), 3-26.
- Reilly, J., Zamora, A., & McGivern, R. F. (2005). Acquiring perspective in English: the development of stance. *Journal of Pragmatics*, 37, 185-208.