The Growth in Grammar Corpus:
On Working with Children
(but not Animals)

Mark Brenchley
Phil Durrant
Debra Myhill
Growth in Grammar Project

CONTEXT

- Increasing curricular emphasis on grammatical development
- Not much contemporary evidence to guide this emphasis
- Especially re: quality
- No interest in grammatical “accuracy” (at least for now)
- Grammar approached as a resource for “meaning-making”
Growth in Grammar Project

RESEARCH QUESTIONS

1) How does children’s written language change as they get older?
2) How does it change depending on the quality of the writing?
3) How does it change according to the kinds of writing asked of students?
4) Are there differences between how children’s texts group “grammatically” and how they are grouped conceptually?
Growth in Grammar Project

THE CORPUS

- Build & (MD-)analyse corpus of “school writing”
- Naturalistic: already produced as part of normal classwork
- England-wide: north/south, rural/urban, PP/\¬PP
- 3 curricular areas: English, Science, Humanities (History)
- 4 ages: 6-7, 10-11, 13-14, 15-16
- 3 attainment levels: “weak”, “average”, “good”
- Not “EAL” (i.e. not not-native speakers!)
Where We Are Now

- 1,200-1,500 texts (out of 6,000)
- Mix of primary and secondary
- Good mix of primary genres
- OK mix of secondary genres
- Begun transcription and annotation process
Challenges

- “Authentic” vs. “Relevant” vs. “Ethical”
Challenges

Their’s no hope

There’s no hope  Their’s no hope

Exeter School

Anon  Anon Anon

The School of Exeter

Anon  The Anon of Anon
Challenges

- “Want” vs. “Can”
Challenges

- Subordinate Clause Types
- Modifiers per NP
- Adverbial Placement
- Appositional Structures
- Finiteness
- Subject-Verb Inversions
- Depth of Embedding
- Ellipsis
- Cohesive Ties
- Relative Clause Gap Position
Challenges

- “Raw” material (i.e. ¬adult, ¬published, ¬typed)
Challenges

It wasn't just the support of the Vietnamese people which was important, America and France in particular needed support from their citizens. However, as people soon realised the Vietnam War was a lost cause and this support gradually slipped away. Source 1 shows a girl running naked after having napalm dropped on her. At the time of the war this was a very famous image to reach the American public and it helped show that they were suffering too. What was going on in Vietnam and did not think it right to support such brutality. Source 2 shows an Anti-war demonstration in Washington. Demonstrations like these were not uncommon as death rates rose.

As well as taxes, people weren't willing to support a war which they couldn't see. This source could be seen as subjective because it only appeared to be white women and men of a similar age typically protesting. Therefore we don't know if the general feel of the American population called for Source 3 also shows support for the French against the Vietnamese, though the 1954 defeat of 1952 weakened increasing criticism in France. This shows that the French weren't happy about continuing a war when they were just seeing a lot of men dying.
Challenges

- Material that varies in kind
Challenges

Rose & Martin (2012)

Learning to Write, Reading to Learn
Challenges

- Material that varies in quality
Challenges

It wasn't just the support of the Vietnamese people which was important; America and France, in particular, needed support from their citizens. However, as people soon realised the Vietnam War was a lost cause and thus support gradually slipped away.

Source 1 shows a girl running naked after having Napalm dropped on her. At the time of the war this was a very famous image to reach the American public and they saw this as they were horrified by what was going on in Vietnam and did not think it right to support such brutality.

Source 2 shows an Anti-war demonstration in Washington. Demonstrations like these were not uncommon as both sides rose.

As well as taxes, people weren't willing to support a war which couldn't be won. This source could be seen as subjective because there only appears to be white women and men of a similar age. Also, women are not often pictured when discussing the Vietnam War. Women were not killed in the war and were not often mentioned in discussions of the Vietnam War.

Source 3 shows support for the French, but the French were killing civilians in their own towns and villages. Their own casualties were over 10,000 dead and wounded by the time the CEZ2 were causing increasing criticism in France.

This shows that the French were not happy about continuing a war when they were just seeing a lot of men dying.
Challenges

The Swimming Carnival.

I went in the race twice.

And fun Thursday.

A fish was swimming in the worto. He heard a splash and a car fell in the water. "Help! Help!" said the man. The fish help the man swim to the beach.
Some “Simple” Solutions?

- Anonymise the whole name:
  - *Exeter School* → Institution_name
  - *The School of Exeter* → Institution_name

- Mark up spelling “errors”:
  - *worta* → `<sp worta>water</sp>`
Some Interim Solutions

- Mark up grammatically “incoherent” stretches:
  - *They went into to town*
    - → <garble>They went into to town</garble>

- Mark up verb “errors”:
  - *He does loves her*
    - → He does <gram>loves</gram> her
Some Less Simple Problems?

PUNCTUATION

I lost. She won. → ROOT; ROOT
I lost, she won. → ccomp(won, lost)
I lost she won. → ccomp(lost, won)

I lost. But she won. → ROOT; ROOT
I lost, but she won. → conj(lost, won)
I lost but she won. → ccomp(lost, won)

I lied. Then I died. → ROOT; ROOT
I lied, then I died. → parataxis(lied, died)
I lied then I died. → ccomp(lied, died)
Some Less Simple Problems?

PUNCTUATION

- Piloted full stop insertion after “independent” clauses
  - *I lost she won* → *I lost*<sent>.<sent> *She won.*

- Definite differences between two versions
- Mostly not critical
- But some are…
Some Less Simple Problems?

- *This isn't coming* from taxpayers' money either, it is entirely fundraised.
  - ccomp(fund-raised, coming)
Some Less Simple Problems?

PUNCTUATION

- "Interpretive" Issues
  - I hope your readers remember that travel broadens the mind and that trips like these have been proven to work in the past.
  - I think he’s great and she’s great.
  - I think that’s terrible and we should do better.

- Make our peace with it?
Some Less Simple Problems?

Children who watch too much television are turning into Couch Potatoes.

- Limited
- B
- C
- What we do
- Fill the bottle A.
- Empty the index of C.
- Refill C.
- Empty C in B.
- Refill C from A.
- Refill B from C.
- Empty the index of B.
- Empty the index of C in B.
- Refill C from A.
- Empty C in B.
Some Less Simple Problems?

- Laborious to transcribe
- Error-prone
<table>
<thead>
<tr>
<th></th>
<th>Main Colour</th>
<th>Has wings</th>
<th>Can fly</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Blackbird</strong></td>
<td>black</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td><strong>Parrot</strong></td>
<td>loads!</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td><strong>Penguin</strong></td>
<td>black and white</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td><strong>Seagull</strong></td>
<td>white</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>
Some Less Simple Problems?

- Laborious to transcribe
- Error-prone
- Grammatically Awkward
Some Less Simple Problems?

- *folded secondary feathers*  
  root(ROOT, folded-VBN)  
  dobj(folded, feathers)

- *twitching ears*  
  root(ROOT, twitching-VBG)  
  dobj(twitching, ears)

- *lower beak*  
  root(ROOT, lower-JJR)  
  dep(lower, beak)

- *lower beak*  
  nsubj(beak, lower)  
  root(ROOT, beak)
Some Less Simple Problems?

“CHARACTERISTIC” FEATURES

- Need to track the grammatical bases of writing development
- Requires identifying two characteristic features:
  1) Different types of discourse – the “genres” of school writing
  2) General later language development, especially re: “quality”
- Many such structures not all that straightforward
Some Less Simple Problems?

FORMULAIC MARKERS

- Dear X,

- Yours sincerely/faithfully, Y.
Some Less Simple Problems?

- Yours sincerely/faithfully
  root(ROOT, yours-PRP$)
  advmod(yours, sincerely/faithfully)

- yours sincerely/faithfully
Some Less Simple Problems?

- Dear Editor,
  compound(Editor-NNP, Dear-NNP)
  nsubj(MAIN CLAUSE, Editor)

- Dear Sir,
  compound(Sir-NNP, Dear-NNP)
  nsubj(COMP. CLAUSE, Sir)

- Dear Sir or madam,
  compound(Sir-NNP, Dear-NNP)
  nmod:tmod(MAIN CLAUSE, Sir)

- dear editor/sir/sir or madam,
  amod(sir-NN, dear-RB)
  ccomp(MAIN CLAUSE, editor/sir)
Some Less Simple Problems?

**ISOLATED NPs**

- folded secondary feathers
- twitching ears
- lower beak

- Clouds of dust as blinding as fog and the sound of animal roars dancing around the arena.

- The sound of two strong, sturdy, swords clashing together.

- The sound of the gladiators, declaring war on each other.
Some Less Simple Problems?

**ISOLATED NPs**

- Potentially characteristic of:
  - narrative fiction
  - scientific descriptions
  - poetry
  - ?
  - “sophisticated” writers
Some Less Simple Problems?

- *folded secondary feathers*  
  root(ROOT, folded-VBN)  
  dobj(folded, feathers)

- *twitching ears*  
  root(ROOT, twitching-VBG)  
  dobj(twitching, ears)

- *lower beak*  
  root(ROOT, lower-JJR)  
  dep(lower, beak)
Some Less Simple Problems?

- Clouds of dust as blinding as fog and the sound of animal roars dancing around the arena.
  - nsubj(roars-VBZ, clouds)
  - root(ROOT, roars-VBZ)
  - xcomp(roars-VBZ, dancing)

- The sound of two strong, sturdy, swords clashing together.
  - amod(sturdy, strong)
  - amod(sound, sturdy)
  - dep(sound, clashing)

- The sound of the gladiators, declaring war on each other.
  - nsubj(declaring, sound)
  - root(ROOT, declaring)
  - root(ROOT, sound)
  - acl(gladiators, declaring)
Some Less Simple Problems?

“DISPLACED” AdjPs

- The beast, monstrous, ravenous, roamed the house.
- He’s a great student, dedicated, hardworking and ambitious.
- He is a terrible student – nasty, lazy, stupid.

- Monstrous, ravenous, the beast roamed the house.
- The beast roamed the house, monstrous, ravenous.
- Green, bronze, and golden, it flowed through the weeds.
- John chuckled, highly amused.
Some Less Simple Problems?

“DISPLACED” AdjPs

- Potentially characteristic of:
  - Fiction
  - ?
  - “sophisticated” writers
Some Less Simple Problems?

- The beast, monstrous, ravenous, roamed the house.
  
  appos(beast, monstrous)
  appos(monstrous, ravenous)

- Monstrous, ravenous, the beast roamed the house.
  
  nsubj(roamed, monstrous—JJ)
  appos(monstrous, ravenous)
  appos(ravenous, beast)

- The beast roamed the house, monstrous, ravenous.
  
  nsubj(ravenous, house)
  appos(house, monstrous)
  xcomp(roamed, ravenous)
Some Less Simple Problems?

- Green, bronze, and golden, it flowed through the weeds
  - dep(flowed, Green-NNP)
  - conj(Green-NNP, bronze-NN)
  - conj(Green-NNP, golden)

- John chuckled, highly amused.
  - xcomp(chuckled, amused-VBN)

- He’s a great student, dedicated, hard-working and ambitious.
  - acl(student, dedicated-VBN)
  - xcomp(dedicated, hardworking-VBG)
  - conj(hardworking-VBG, ambitious)

- He is a terrible student, nasty, lazy, stupid.
  - amod(stupid, nasty-JJ)
  - amod(stupid, lazy-JJ)
  - amod(student, stupid-JJ)
Some Less Simple Problems?

- Maybe not all that much of a surprise – what you’d expect when working with a highly variable, even “deviant” corpus

- And maybe we can’t “count” these more problematic features

- And maybe that’s not a major problem -

- Perhaps too sparse for substantive, reliable counts anyway

- BUT
Some Less Simple Problems?

- This isn’t something we yet know, which raises two issues:
  - First, even if they aren’t pervasive across the corpus generally, they might be pervasive for certain kinds of texts
    - Science reports
    - High level science reports
  - In which case, we will lose our capacity to pick up on some core developmental differences
  - Perhaps even be the core differences
Some Less Simple Problems?

- Second, the annotations marking these more obviously problematic features also implicated in other features that we might like to measure.

- But suppose we can’t reliably separate out these annotations when they mark the problematic features from when they mark these other features.

- Then these other features also become compromised.

- In which case, we’ll lose even more of our capacity to pick out those grammatical features that really count in the development of school writing.
http://socialsciences.exeter.ac.uk/education/researchCENTRES/centreforresearchinwriting/projects/growthingrammar/