Using corpus linguistics to examine the linguistic challenge of starting high school

Alice Deignan

University of Leeds

Early ideas

Previous project on the language of science (2014- 2016), which included interviewing school students aged 11- 16 about their understandings of climate change.

Evidence of students not being able to articulate ideas using academic language (both examples students aged 14-15):

- if we're recycling stuff like the landfills, I don't know, it *releases* something like, you know, less landfills and less pollution and stuff like that.
- It's getting thicker because erm, there's more pollutants and they're like carbon dioxide, so cos it's getting thicker, less oxygen, over less gases, like bounce back off. So they're getting less *released* so there's holes in there, which makes it more warmer.

The language of school

‘Response’ used in textbooks/ educational websites

1. *[T]emperatures* warm in *response* to increasing amounts of greenhouse gases in the air (climatecentral.org)
2. the *Earth's* temperature is rising in *response* to emissions of greenhouse gases from the burning of oil, coal and gas (geographical.co.uk)
3. Shaped by orbital variations, *responses* such as the rise and fall of continental ice sheets and significant sea-level changes helped create the climate. (Wikipedia)

Everyday language

Use of *response* is mostly literal, with human agency. All examples from spoken BNC2014.

1. he says it's quite (.) easy (.) to see whether people are (.) er in their *response* he says it's quite easy to see whether people are bringing stuff through or not
2. erm well I I was ah relatively positive with my *response*
3. there was no *response* on Facebook
4. I bet you any money at all if you emailed the suppliers direct you'd get a you'd get a better *response* and a quicker *response*

So, is there a language problem for school students?

If so, is this for any particular groups of students, and for particular stages in schooling?

Teachers think so:

Children are able to think but they can't articulate their thoughts because of the lack of language […] it is not the concepts they are finding difficult at Key Stage 3, it is the ability to access material given to them.

Discussion with a history teacher

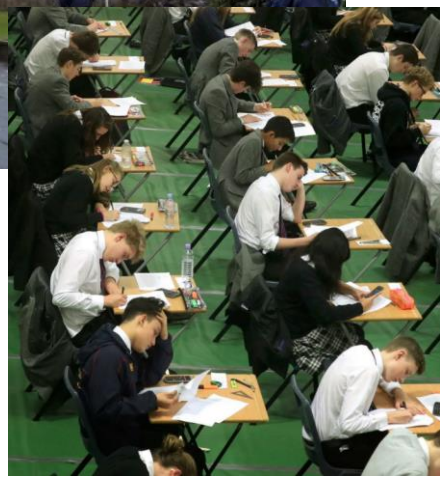
Schools in England and Wales

Key Stage	Year	Age	
Early years	Nursery	3-4	
	Reception	4-5	
Key Stage 1	Years 1 & 2	5-7	Primary school
Key Stage 2	Years 3-6	7-11	
Key Stage 3	Years 7-9	11-14	Secondary school
Key Stage 4	Years 10 & 11	14-16	
Key Stage 5	Years 12 & 13	16-18	

The transition: different environments, new challenges

Primary school

Secondary school



The transition dip

“There is a large dip in mathematical attainment and attitudes towards maths as children move from primary to secondary school.”

Educational Endowment Foundation, Nov 2017

“There is evidence across the UK that a drop in attainment takes place during the transition. Characteristics of pupils particularly affected by the drop in attainment include: pupils receiving free school meals, those with poor prior attainment, those with low self-esteem and those from minority ethnic backgrounds.”

A rapid evidence assessment: Investigating the drop in attainment during the transition phase Wilson, P., 2011 for the Welsh Assembly Government

Language and the transition

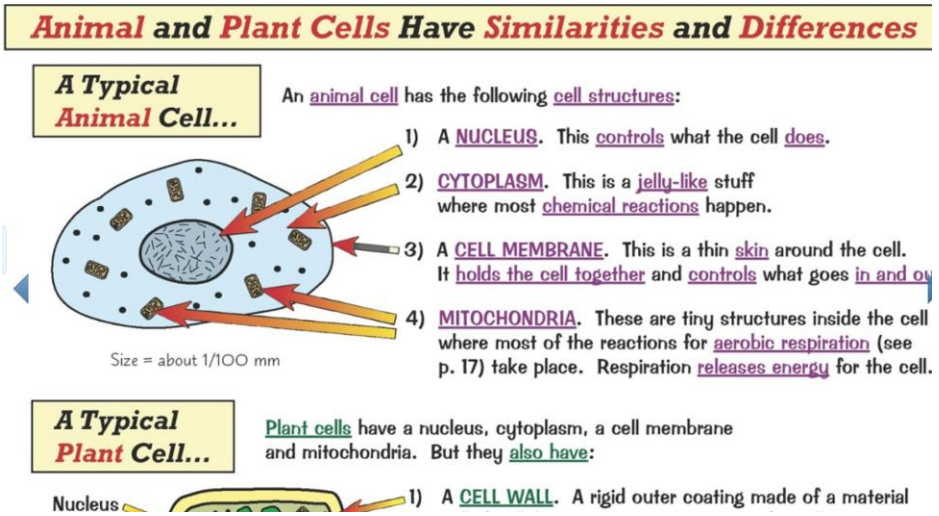
teaching environments […] and teachers' language are very different in secondary schools from primary schools

(Braund & Driver, 2005, p. 78)

Fossils

In the eighteenth and nineteenth centuries people began to carry out a closer study of the strange animal and plant shapes embedded in rocks. They did not know what they were or how they came to be there. Some people said that they were nothing but patterns in the rocks that just happened to look like animals. Nowadays we call them fossils and know that they show us that the animals and plants that lived millions of years ago were very different to those alive today.

These early geologists could not have fully understood this because they had no idea of how old the Earth actually was. One way of calculating a possible age was by adding up all the ages of the people mentioned in the Book of Genesis of the Christian Bible.



Investigating the problem

'The linguistic challenges of the transition from primary school to secondary school'

2018-2021, funded by Economic and Social Research Council (ESRC), UK

Principal investigator: **Alice Deignan**, (University of Leeds)

Co-investigators: **Gary Chambers**, **Michael Inglis** (University of Leeds), **Elena Semino**, **Vaclav Brezina** (Lancaster University)

Research Fellows: **Duygu Candarli** (now Dundee University, formerly Leeds), **Dogus Oksuz** (University of Leeds)

Research Assistants: **Robbie Love** (now Aston University, formerly Leeds); **Florence Oxley** (University of Leeds)

Consultant: **Marcus Jones**, Huntington School, York

Data

- 13 schools have contributed data, across the North of England: 5 secondary schools, 8 primary schools;
- Of these, 5 of the primary schools directly ‘feed’ 3 of the secondary schools;
- Teachers were requested to upload written teaching materials via OneDrive, but this was endlessly problematic. In the end, project staff physically collected material on USB sticks from schools;
- Some textbooks scanned– a lot of cleaning needed due to messy format; The school materials comprise a huge dataset, 1000s of files, several million words– a lot of sorting, cleaning, deleting duplicate files;
- Spoken data recorded via a lanyard microphone worn by teachers, collected by project staff.

Project data

Written data (Key Stage 2 and Key Stage 3)

- Worksheets
- Textbooks
- Exams and assessment tasks
- Lesson presentations
- Vocabulary/glossary booklets

Spoken data (Key Stage 2 and Key Stage 3)

- Audio recordings of lessons- teacher utterances only

Interviews with pupils and teachers for qualitative analysis

Subjects: English, maths, science, history, geography

Corpora

Key Stage 2 (Primary school) corpus

Data from the last two years at primary school, when children are aged 9- 11 years

Key Stage 3 (Secondary/ high school) corpus

Data from the first two years at secondary school, when children are aged 11-13 years

Reference corpus

To be drawn from existing corpora, aiming to proxy the language experience outside school. (Sections of the BNC2014, Oxford Children's Corpus etc)

Interview data

30 school students aged 10- 12 years at time of interviews

1. March 2019
We interviewed the children in 5 groups of 6 students. They were in the middle of their final year at 5 different primary schools, average age 11y.

2. June 2019
Interviewed in same 5 groups of 6 students, towards the end of their final year at their primary schools. Their final exams had finished, they had visited the secondary schools they will move to, average age 11y3m.

3. October 2019
The 30 students had now progressed to 3 different secondary schools in September. Interviewed in pairs (timetables made it too complicated to reconstruct the original groups), average age 11y7m.

4. February 2020
As above, average age 11y11m.

Schools in UK closed on March 20th 2020, final interviews planned for May 2020 not possible. Teacher interviews have continued online.

Examples from interviews

Q. Which subjects do you think you will find difficult? Why?

Maths: Algebra because of the use of letters, and word problems.

Cathy: *I don't like word problems because you have to work out what the word problem means as well as actually doing the sums...When you go to secondary school, it gets more complicated.*

Science: Scientific terms can be hard to understand. In primary school, they do not learn as many 'hard' words as they will learn at secondary school.

James: *physics and biology I think I'll be not very good at that because we haven't done it before.*

The corpora: current size

Main divisions:
Key Stage 2/ Key Stage 3
Written/ Spoken

Written corpus: 1.5 million tokens
Key Stage 2: over 800,000 tokens; Key Stage 3: over 600,000 tokens

Spoken: 400,000 tokens, split roughly equally between Key Stage 2 and Key Stage 3

Subdivided by school subject: Maths, English, Science, Geography History.
Questions: how to deal with literature in the English sub-corpus? How to weight the different subjects, given they take very unequal amounts of class time, especially in Key Stage 2?

Analysis and some early findings

Corpus data is being analysed using LancsBox, early work done with AntConc

Comparisons made between:

Primary all subjects & Secondary all subjects

Primary and Secondary for specific subjects

Separate years.. How words and structures the students are exposed to change from Y5 through to Y8 (aged 9 through to aged 13)

Primary/ Secondary language and language outside the school

Polysemous words in classroom data – ‘mean’

Key Stage 2

one habitat. What Does Adapted Mean?| ‘Adapted’ means to adjust to
measure angles What do we mean by angle? What are they
MISCONCEPTION ALERT! While offspring does mean child, it does not mean
your partner. What does percentage (%) mean? Converting fractions to percentages _93 100 34 5
of invention9 What does this mean? Do you agree? Can you
invention’ What does this quote mean? Invention research: Using Google Slides,
using partitioning. What does this mean? Partitioning: 4,379 - 243 Step 1: Partition 24

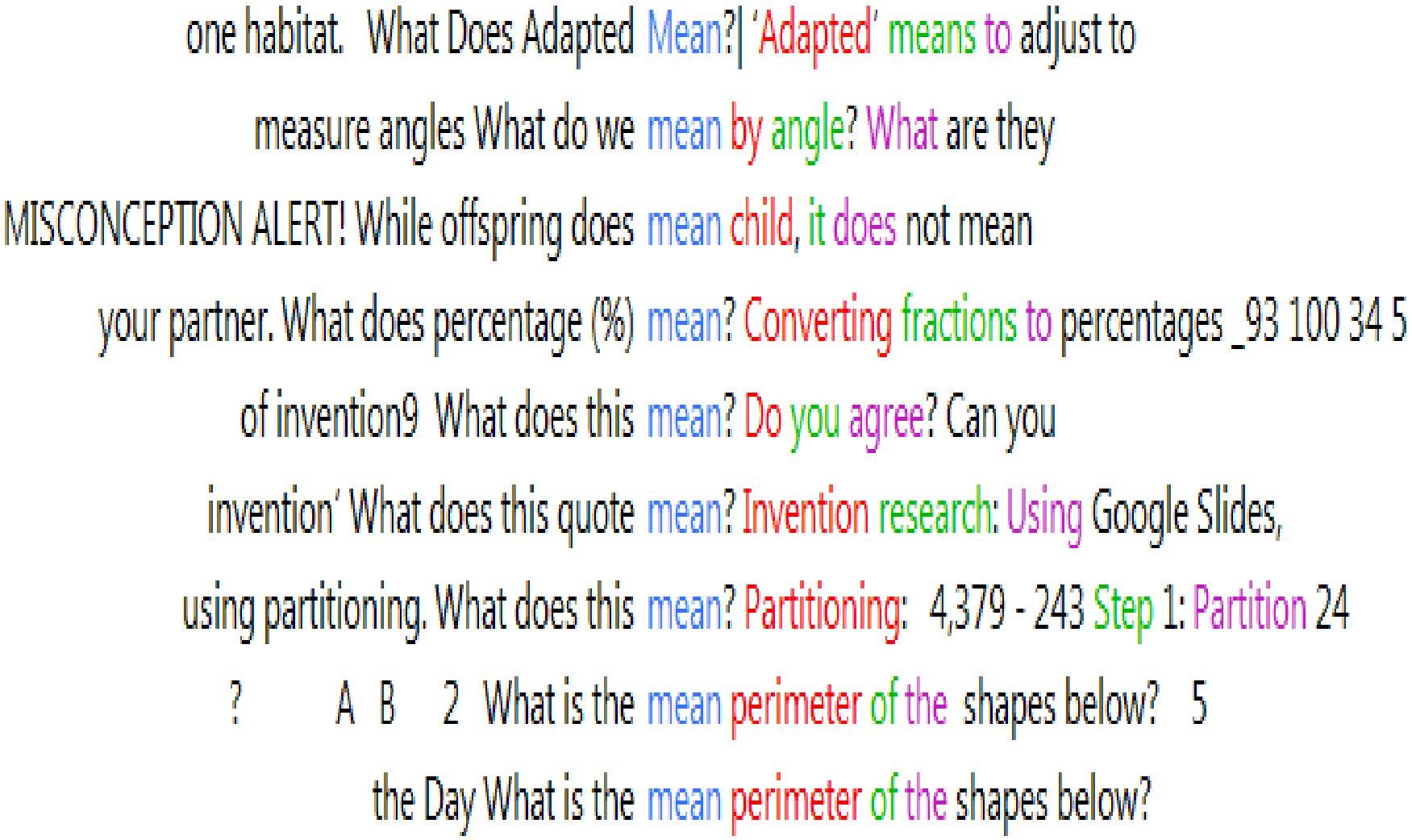
? A B 2 What is the mean perimeter of the shapes below? 5
the Day What is the mean perimeter of the shapes below?

use BIDMAS What does BIDMAS mean? Santa has some presents in
same characteristics, it does not mean that they are identical. There
mean child, it does not mean that you are only offspring
talk about inheritance, we often mean things that are passed on
similar words and what they mean to work out the meaning
traits. Variation What does variation mean? What causes variation? Inheritance These
genome. Variation What does variation mean? What causes variation? Inheritance These
percentages to decimals What does % mean? What is 0.7 as a percentage?
a partner. What does ‘percent’ (%) mean? What is 25% as a fraction?
decimal place. What does this mean? Which two tenths will 7.65 be

Key Stage 3

the same way. Key Words outlier, mean, line graph, bar chart, pie chart
granite from the quarry. Calculate the mean mass of granite dug out of
players is 425kg and the average (mean) mass of ten ballet dancers is 40
the meaning of ☒ average Find the mean, median and mode from a set
account the following: ☒ What are the mean, median and mode? ☒ Which is the
mean 7 and mode 10 3. Three numbers with mean 8, median 10 and range 8 4. Four numbers with
7.5, mode 6 and median 7 5. Four numbers with mean 6, median 6.5 and range 11 6. Five numbers wi
following properties? Mode < Median < Mean Mode < Mean < Median Mean < Mode < Median Mean < Median <
< Median Mean < Median < Mode Median < Mode < Mean Median < Mean < Mode M, M and
following properties? Mode < Median < Mean Mode < Mean < Median Mean < Mode < Median Mean < Median <
< Median Mean < Median < Mode Median < Mode < Mean Median < Mean < Mode M, M and
following properties? Mode < Median < Mean Mode < Mean < Median Mean < Mode < Median Mean < Median <
< Median Mean < Median < Mode Median < Mode < Mean Median < Mean < Mode Not all of
Girls Frequency Boys Range: Mean: Median: Mode: 1 2 3 4 5 6 7
whole numbers with the following properties: ☒ Mean = 4 ☒ Median = 3 ☒ Mode = 3 Can you find al
whole numbers with the following properties: • Mean = 4 • Median = 3 • Mode = 3 Can you find all
Girls Range: Mean: Median: Mode: Do girls have different
Mode < Mean < Median Mean < Mode < Median Mean < Median < Mode Median < Mode < Mean Median <
Mode < Mean < Median Mean < Mode < Median Mean < Median < Mode Median < Mode < Mean Median <

A closer look at 'mean' in classroom data (KS2)



Emma: *We learned 'mean' in maths. It was in the SATs practice test.*
I couldn't guess the meaning. We haven't heard 'mean' in maths before.

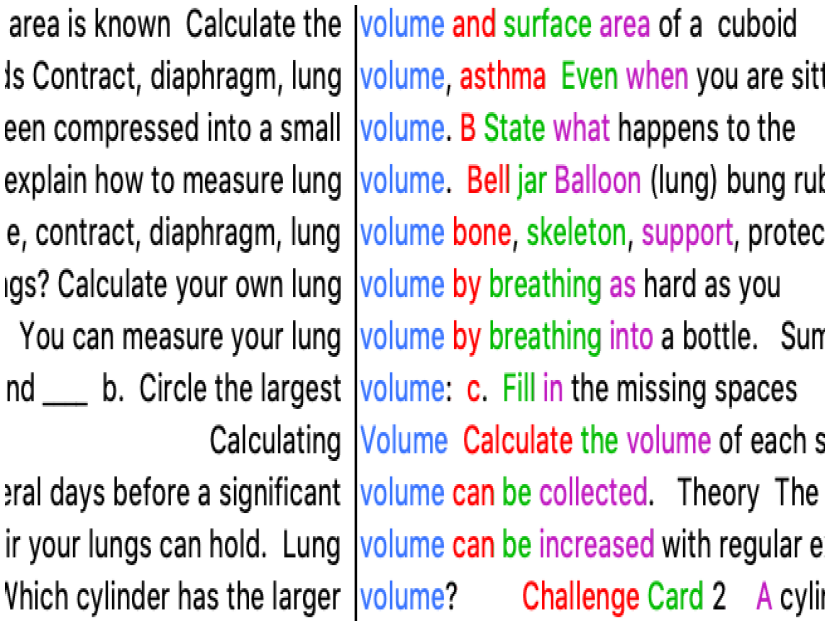
Polysemous words in classroom data – ‘volume’

Key Stage 2

As the volume of music at the school disco became unbearable…

(1 occurrence in the existing classroom data set)

Key Stage 3



Jacob: *You have volume on TV, but then you come across it in maths. Difficult to understand.*

Polysemous words in classroom data – ‘volume’ (KS3)

area is known Calculate the	volume and surface area of a cuboid
ls Contract, diaphragm, lung	volume, asthma Even when you are sitt
een compressed into a small	volume. B State what happens to the
explain how to measure lung	volume. Bell jar Balloon (lung) bung rub
e, contract, diaphragm, lung	volume bone, skeleton, support, protec
igs? Calculate your own lung	volume by breathing as hard as you
You can measure your lung	volume by breathing into a bottle. Sum
nd ___ b. Circle the largest	volume: c. Fill in the missing spaces
Calculating	Volume Calculate the volume of each s
eral days before a significant	volume can be collected. Theory The
ir your lungs can hold. Lung	volume can be increased with regular e
Vhich cylinder has the larger	volume? Challenge Card 2 A cylin

Polysemous words in classroom data – ‘concentration’

Concentration – example explanations from pupils during the interviews

to focus

you’re not distracted

concentration camp

pointing towards one thing

when stuff is really high

Polysemous words in classroom data – ‘concentrat*’

Key Stage 2

Harry wished they wouldn’t, because he was trying to concentrate to find his way to classes.

(1 occurrence in the existing classroom data set)

Key Stage 3

r unit volume (litre or cubic metre).	concentration A measure of the number of
the air particles are close together.	concentrated A solution is concentrated if it
loride per litre than acid B. Acid A is	concentrated. Acid B is dilute. The concentra
igh concentration to an area of low	concentration, across a partially permeable
chloric acid cotton wool soaked in	concentrated ammonia solution Particles o
e is called OSMOSIS! Title: Osmosis	Concentration: Amount of substance in a pa
arkable amount of resistance in the	concentration and death camps. To resist he
talis • describe differences between	concentrated and dilute solutions of an acid
ed with a book but found it hard to	concentrate, and just then the Hopeless Cas
a high-concentration area to a low-	concentration area, for example, water and
movement of particles from a high-	concentration area to a low-concentration a
she so effectively have brought the	concentrated attention of millions to bear u
18 huts was built at Sachsenhausen	concentration camp. According to the officia
Hierarchy - Stereotype - Holocaust	Concentration camp Back of your book. No
Hierarchy - Stereotype - Holocaust	Concentration camp Back of your books W
to death after a trial at Flossenburg	concentration camp. He was hanged at daw
eeches. He was sent to Buchenwald	concentration camp in 1937, but still manag
minorities. Niemoller was sent to a	concentration camp in 1938. He continued t

Polysemous words in classroom data – ‘concentrat*’ (KS3)

chloric acid cotton wool soaked in concentrated ammonia solution Particles o
e is called OSMOSIS! Title: Osmosis Concentration: Amount of substance in a pa
arkable amount of resistance in the concentration and death camps. To resist he
talis • describe differences between concentrated and dilute solutions of an acid
ed with a book but found it hard to concentrate, and just then the Hopeless Cas
a high-concentration area to a low- concentration area, for example, water and
movement of particles from a high- concentration area to a low-concentration a
she so effectively have brought the concentrated attention of millions to bear u
18 huts was built at Sachsenhausen concentration camp. According to the officia
Hierarchy - Stereotype - Holocaust Concentration camp Back of your book. No
Hierarchy - Stereotype - Holocaust Concentration camp Back of your books W
to death after a trial at Flossenburg concentration camp. He was hanged at daw
eeches. He was sent to Buchenwald concentration camp in 1937, but still manag
minorities. Niemoller was sent to a concentration camp in 1938. He continued t

How interview data informed our corpus findings

- Student also find technical words hard.

  Examples: fertilization, Australopithecus, parentheses

- The students tend to evaluate their reading as successful if they can pronounce all the words, even if they do not understand the meaning of some of those words

- Over-confidence about understanding certain technical words that have everyday use

  Example: energy

  “the force when something’s happening”

  “Energy can be used as a synonym for force”

Comparing lesson powerpoints across Years 5-7

- There is a dip in Year 6 lessons presentations in terms of subject content;
- Assessing students' knowledge is most frequent in Year 6;
- Lesson presentations tend to be more task-oriented in Key Stage 3 than in Key Stage 2;
- Organisational text, including 'presenting learning objectives' and 'restating learning objectives' decrease across key stages, except for 'referring to a website'. It seems that students are expected to infer learning objectives at Key Stage 3.

Early conclusions

How does the language of Key Stage 3 differ from the language of Key Stage 2?

There's a lot more of it. Densely written textbooks and powerpoints, compared to KS2;

A higher frequency of academic words at KS3 than at KS2;

Polysemous words tend to occur more with their domain-specific meanings at KS3 than at KS2;

Technical/subject-specific words that students find hard occur very rarely at both KS2 and KS3;

Information presented differently, for example in teacher powerpoint presentations.

Changes in the language of maths from Primary to Secondary school

From personal to impersonal

Primary School
More personal/involved
Problems clearly related to simple, real-world scenarios
More familiar, everyday lexis
Simpler grammar?

Secondary School
Less personal/involved – more ‘clinical’
Tasks detached from the real-world (e.g. algebra)
More scientific notation
Variation in how learning objectives are expressed

What should we do about this?

As the project progresses, we'll have an increasingly clear idea of where the difficulties are exactly.

The project will generate a list of the vocabulary of Secondary school that is likely to be new- or at least, not encountered previously in Primary, and other new features of Secondary school language at the sentence and discourse level.

In our view, the solution is not to dumb down in any way, but to put in language support.

Ultimately our goal is to improve social justice by enabling corpus-informed language interventions for children from lower socio-economic status groups, who have less access to the language of school.

Looking ahead

Further development of the classroom data set as we finish transcribing and cleaning spoken data from lessons (teacher talk)

More comparative corpus analysis across the full datasets

Subject comparisons: eg is there much in common between the language of History and Geography?

Analysis of the literature that students read in English classes

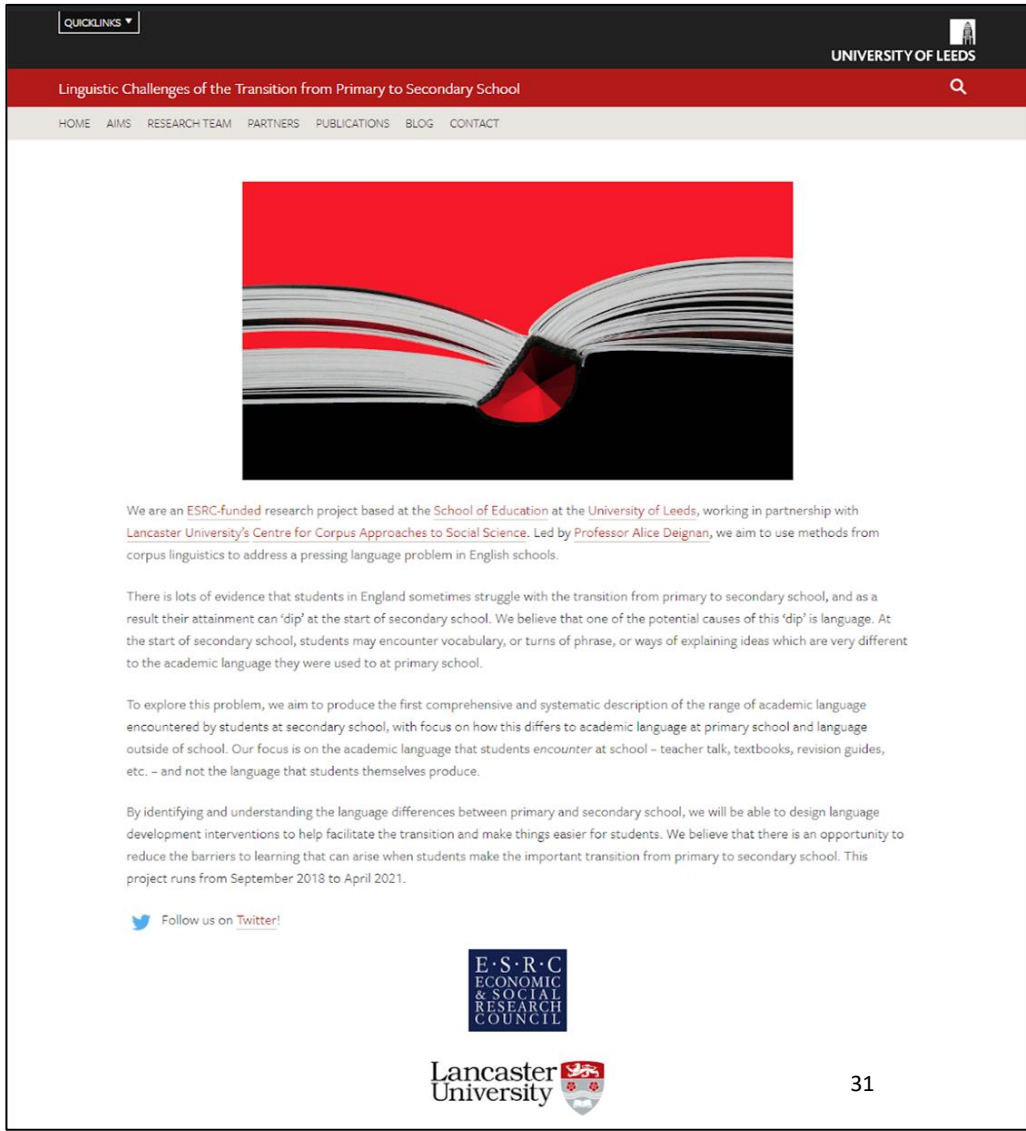
Covid crisis permitting: revisiting schools

Dissemination to education professionals. So far, 4 presentations to non-academic audiences, most recently March 7th 2020, ResearchEd Birmingham.

Linguistic challenges of the transition from primary to secondary school

https://linguistictransition.leeds.ac.uk/
@LeedsTransition

a.h.deignan@education.leeds.ac.uk
@alicedeignan



UNIVERSITY OF LEEDS

Project publications

Various conference presentations on project website; various work in progress

Work on polysemy:

Deignan A, Love R. in press. Using corpus methods to identify subject specific uses of polysemous words in English secondary school science materials. *Corpora.*

(link to pre-print copy from my Twitter feed/ http://eprints.whiterose.ac.uk/154115/)

References

Braund, M. & Driver, M. (2005). Pupils' perceptions of practical science in primary and secondary school: implications for improving progression and continuity of learning. *Educational Research*, 47, 77-91.

Brooks, G. (2016). *What works for children and young people with literacy difficulties? The effectiveness of intervention schemes*. The Dyslexia/ SpLD Trust.

Christie, F. (2002). *Classroom discourse analysis: A functional perspective*. London: Continuum.

Christie, F. & Derewianka, B. (2008). *School discourse: Learning to write across the school years*. London: Continuum.

Department for Education. (2011). How do pupils progress between Key Stages 2 and 3 Research Report.

Howe, A., & Richards, V. (2011). *Bridging the transition from primary to secondary school*. Abingdon: Routledge.

Jindal-Snape, D., & Cantali, D. (2019). A four-stage longitudinal study exploring pupils' experiences, preparation and support systems during primary–secondary school transitions. *British Educational Research Journal,* 1-24.

Leung, C. (2014). Researching language and communication in schooling. *Linguistics in Education*, 26, 136-144.

Schleppegrell, M. (2012). Academic language in teaching and learning. *The Elementary School Journal*, *112*(3), 409–418.