



Using corpora to support UK school students

Alice Deignan, Duygu Candarli



A.H.Deignan@education.leeds.ac.uk; dcandarli001@dundeee.ac.uk



[@alicedeignan](https://twitter.com/alicedeignan); [@duygucandarli](https://twitter.com/duygucandarli)

The transition: different environments, new challenges

Primary school



Secondary school



A poor transition is associated with negative outcomes.

West et al (2010), research in Scotland:

Longitudinal study of 2000 school students from age 11 to 19. A poor school transition predicted lower attainment and well-being than peers at age 15, and the effect was still detectable, albeit reduced, at age 18/19, after the participants had left school.

Who struggles?

Higgins et al (2014), research in England:

Students whose reading is below expected levels at the end of primary school are statistically likely to fall further behind. They have a less than 1 in 10 chance of getting 5 good GCSEs (including maths and English).

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.”
Wilson, 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.

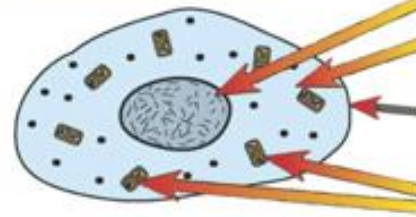
These early **geologists** could not have fully understood this because they had no idea of how old the Earth actually was. One way of calculating a possible age was by adding up all the ages of the people mentioned in the Book of Genesis of the Christian Bible.

Animal and Plant Cells Have Similarities and Differences

A Typical Animal Cell...

An **animal cell** has the following **cell structures**:

- 1) A **NUCLEUS**. This **controls** what the cell **does**.
- 2) **CYTOPLASM**. This is a **jelly-like** stuff where most **chemical reactions** happen.
- 3) A **CELL MEMBRANE**. This is a thin **skin** around the cell. It **holds the cell together** and **controls** what goes **in and out**.
- 4) **MITOCHONDRIA**. These are tiny structures inside the cell where most of the reactions for **aerobic respiration** (see p. 17) take place. Respiration **releases energy** for the cell.



Size = about 1/100 mm

A Typical Plant Cell...

Plant cells have a nucleus, cytoplasm, a cell membrane and mitochondria. But they **also have**:

- 1) A **CELL WALL**. A rigid outer coating made of a material



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

Research from 2016 onwards; from 2018-2021, funded by Economic and Social Research Council (ESRC), UK; no-cost extension until 31st Dec 2022.

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

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

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

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

Consultant: **Marcus Jones**, Literacy lead, Huntington School, York.

Data collection

- 13 schools 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, that is, the children from the primary schools all go to the same secondary school;
- Some textbooks were scanned. The school materials comprise a huge dataset, 1000s of files, several million words– a lot of sorting, cleaning, deleting duplicate files;
- Spoken data was recorded via a lanyard microphone worn by teachers, collected by project staff.



**Northallerton School
& Sixth Form College**



Project data

1. Written data (Key Stage 2 and Key Stage 3)

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

2. Spoken data (Key Stage 2 and Key Stage 3)

Audio recordings of lessons: teacher utterances only

3. Interviews with pupils and teachers for qualitative analysis

Subjects: English, maths, science, history, geography

Corpora

Main divisions

Key Stage 2/ Key Stage 3

Written/ Spoken

Written corpus: 1.9 million tokens

Key Stage 2: approx. 800,000 tokens;

Key Stage 3: approx. 1,100,000 tokens

Spoken corpus: 600,000 tokens

split roughly equally between Key Stage 2 and Key Stage 3

Subdivided by school subject: Maths, English, Science, Geography History.

Case study 1: the English corpora

	Key Stage 2		Key Stage 3	
	Number of texts	Tokens	Number of texts	Tokens
Written	600	303,257	334	258,869
Spoken	15	72,475	8	47,595
Total	615	375,732	306	306,464

Some of our ways into the corpora

Frequency (Baker et al., 2013) #LancsBox 6.0 We identified the most frequent types in each corpus, and eliminated 'general' words using the top 200 lemmas from the 'New General Service List (Brezina & Gablasova, 2015);

Keywords (Gabrielatos, 2018): KS3/ KS2, and studies using BNC2014 baby+ and Spoken BNC2014 as reference corpora;

Leading to:

Collocation analysis;

Detailed concordance analysis.

Using frequency to find 'aboutness' (Baker et al., 2013)

'Most frequent topic-specific types' was taken as types that occur 100 times or more in the KS3 English corpus.

Corresponding to 32.6 occurrences per 100,000 words.

KS2: 126 types; KS3: 129 types.

Most frequent topic-specific words (types) in KS2 English

1	word	11	clause	21	down	31	describe
2	sentence	12	right	22	evidence	32	main
3	words	13	paragraph	23	punctuation	33	commas
4	mark	14	answer	24	marks	34	able
5	write	15	writing	25	add	35	test
6	text	16	noun	26	bit	36	section
7	sentences	17	question	27	clauses	37	example
8	read	18	information	28	box	38	book
9	verb	19	spelling	29	sure	39	speech
10	correct	20	relative	30	explain	40	past

Most frequent topic-specific words (types) in KS3 English

1	write	11	poem	21	range	31	questions
2	words	12	create	22	example	32	characters
3	writing	13	language	23	important	33	describe
4	down	14	text	24	question	34	vocabulary
5	sentence	15	explain	25	paragraph	35	identity
6	word	16	effect	26	evidence	36	explore
7	gothic	17	features	27	person	37	remember
8	key	18	reader	28	extract	38	quote
9	story	19	understand	29	setting	39	love
10	read	20	ideas	30	character	40	chapter

KS2 frequent word types

Three semantic groups, in order of frequency of words in group:

- Assessment: *correct, answer, explain, circle, match*
- Grammatical description: *noun, verb, adjective, past (tense), clause*
- Reading and creative writing: *meaning, author, reader*

KS3 Frequent word types

Five overlapping semantic groups. The majority are associated with the analysis of text.

- central vocabulary of text analysis: *text, character(s), set, setting*.
- major themes or strong emotions in literature: *war, death, fear, identity, love, horror, theme, heart*
- a small group associated with the head: *head, eyes, face*, often as indicators of characters' emotions.
- used in genre analysis discussions; *features, poems, poetry, purpose, language, story, gothic*.
- used in discussion of the effect that writing is intended to have on the reader, and how this is achieved: *create, tension, reader, effect, audience, effective, technique(s), language*.

Qualitative differences in word use

	Language
KS3	<p>'Consider the effect that this language would have.'</p> <p>'How does the writer use language to ...?'</p> <p>'By the end of the extract, Wells' use of language becomes more intense.'</p>
KS2	<p>'Formal language often uses longer words'</p>
ref	<p>The French language is beautiful..</p> <p>... accused him of using foul language.</p>

	Explore
KS3	<p>'Today we are going to explore the context of this novel.'</p> <p>'How did the poems explore the theme of identity?'</p>
KS2	<p>'[He] explored the empty room in the farmhouse'</p>
ref	<p>'We hoped to hire a boat to explore all those tiny beaches'</p>

Case study 2: Lexical sophistication

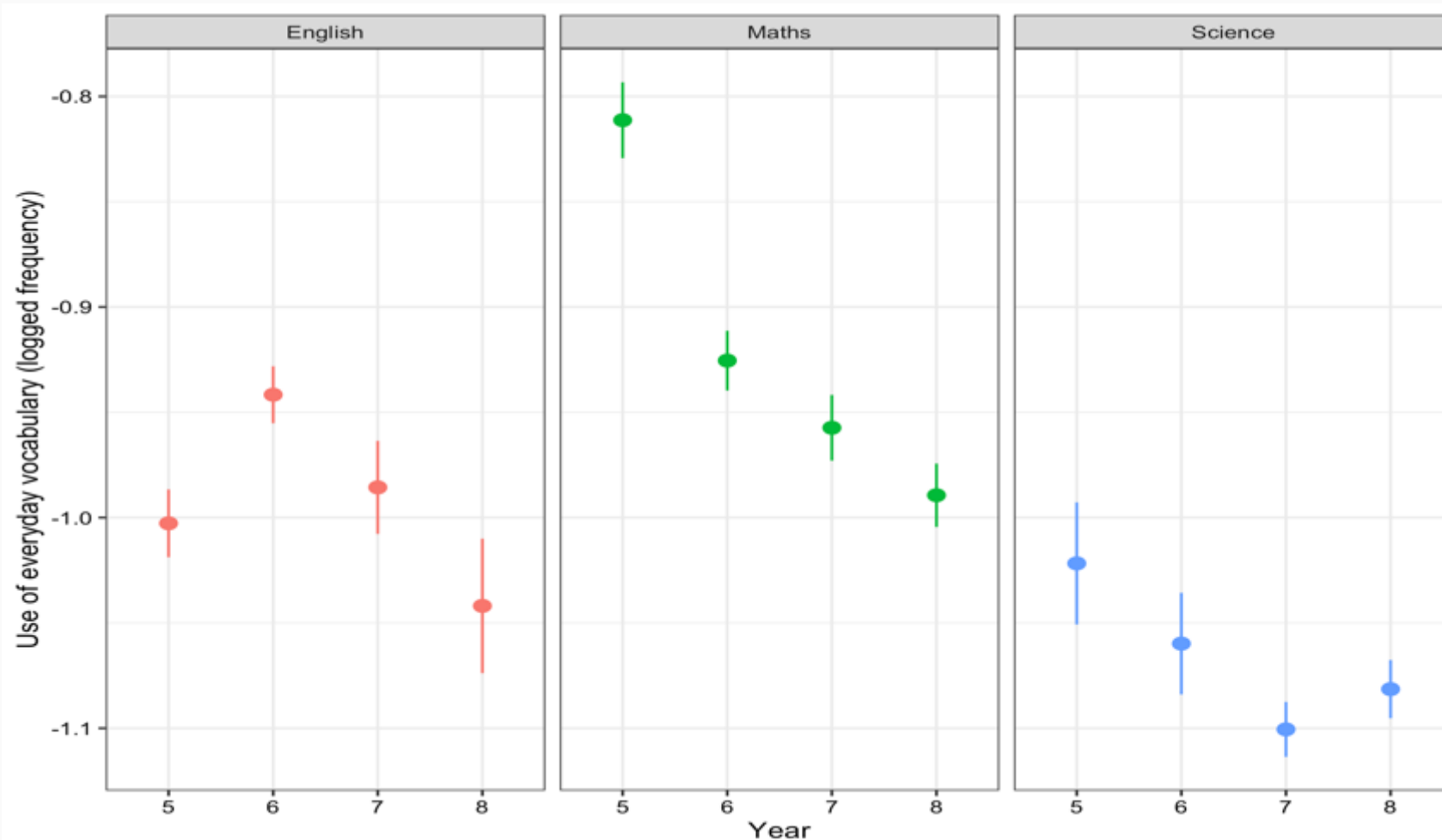
“Selection of **low frequency words** that are appropriate to the topic and style of writing, rather than just everyday vocabulary” (Read, 2020, p. 200)

Our operationalisation (see Kyle & Crossley, 2015):

- The less use of everyday vocabulary (British National Spoken Corpus)
- The use of academic vocabulary (Coxhead, 2000)
- The use of polysemous words (more senses and thus more ambiguous)

A freely available tool called the Tool for the Automatic Analysis of Lexical Sophistication 2.2 (TAALES 2.2; Kyle & Crossley, 2015) was used for the analysis.

Use of everyday vocabulary (spoken British National Corpus)



As the use of everyday vocabulary decreases, the written resources become lexically more sophisticated.

Example text extracts

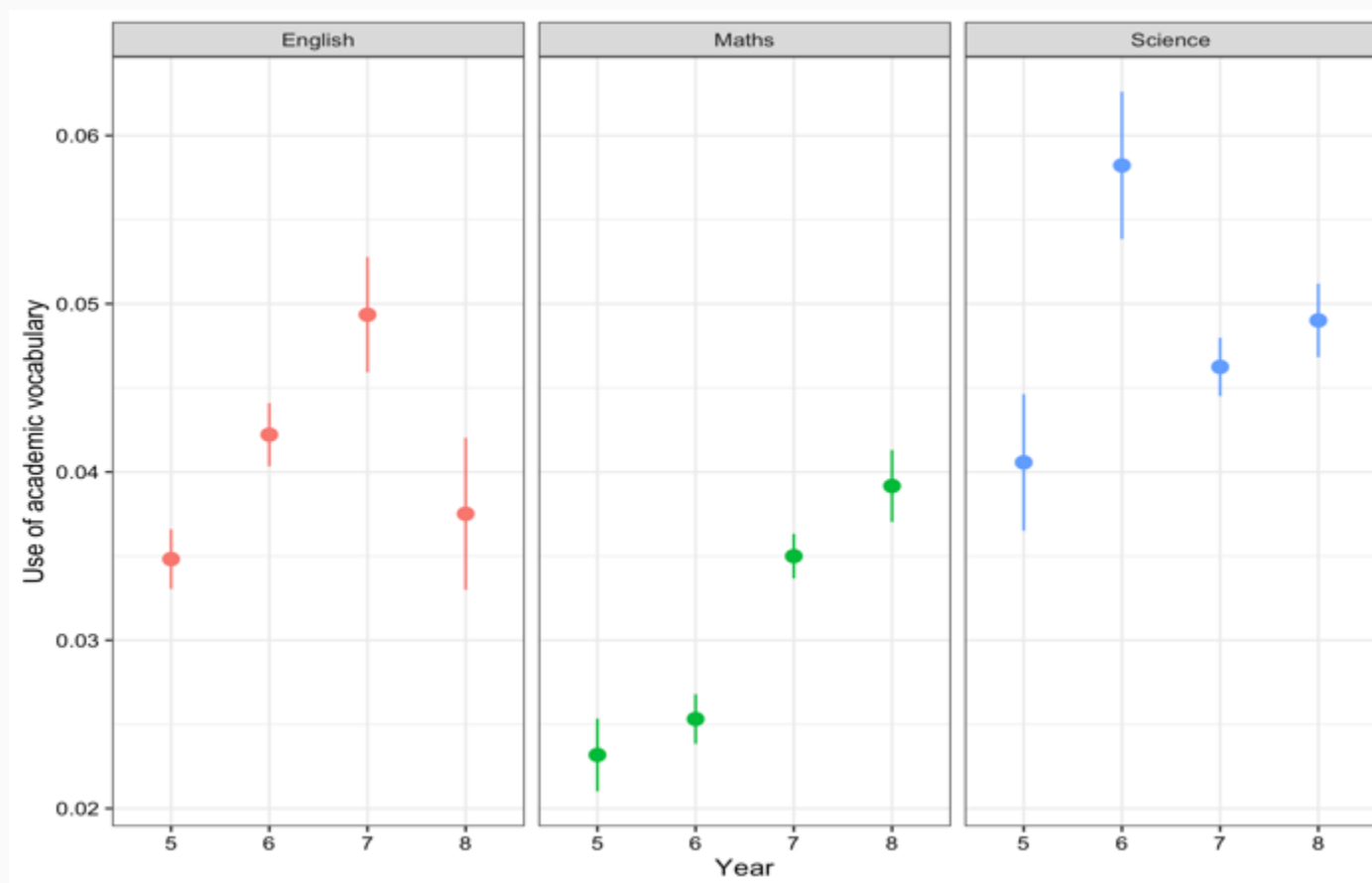
Least lexically sophisticated (everyday vocabulary)

A television programme starts at 9:53 and lasts for 18 minutes. What time did it end? (Year 5_maths)

The most lexically sophisticated (technical/specialised vocabulary)

Describe the difference between a **trapezium** and a **parallelogram**. (Year 7_maths)

Use of academic vocabulary (Coxhead, 2000)



As the use of academic vocabulary increases, the written resources become lexically more sophisticated.

Example text extracts

Least lexically sophisticated (no academic vocabulary)

What was the story about? Why did the children laugh? (Year 5_English)

The most lexically sophisticated (use of academic vocabulary)

In this essay, you are aiming to understand, **identify**, explain, **analyse** & explore. (NB-think about how they make the reader feel) (Year 8_English)

Year 6 science – Use of academic vocabulary in self-assessment sheets

I can explain the scientific **concept** of inheritance.

I can **identify** inherited characteristic that are passed on from parent to offspring.

I can explain how inherited characteristics can lead to **variation**.

I can **demonstrate** understanding of the scientific meaning of **adaptation**.

I can understand that **adaptations** are mutations.

I can **identify adaptive** traits.

I can **identify** the key ideas of the **theory of evolution**.

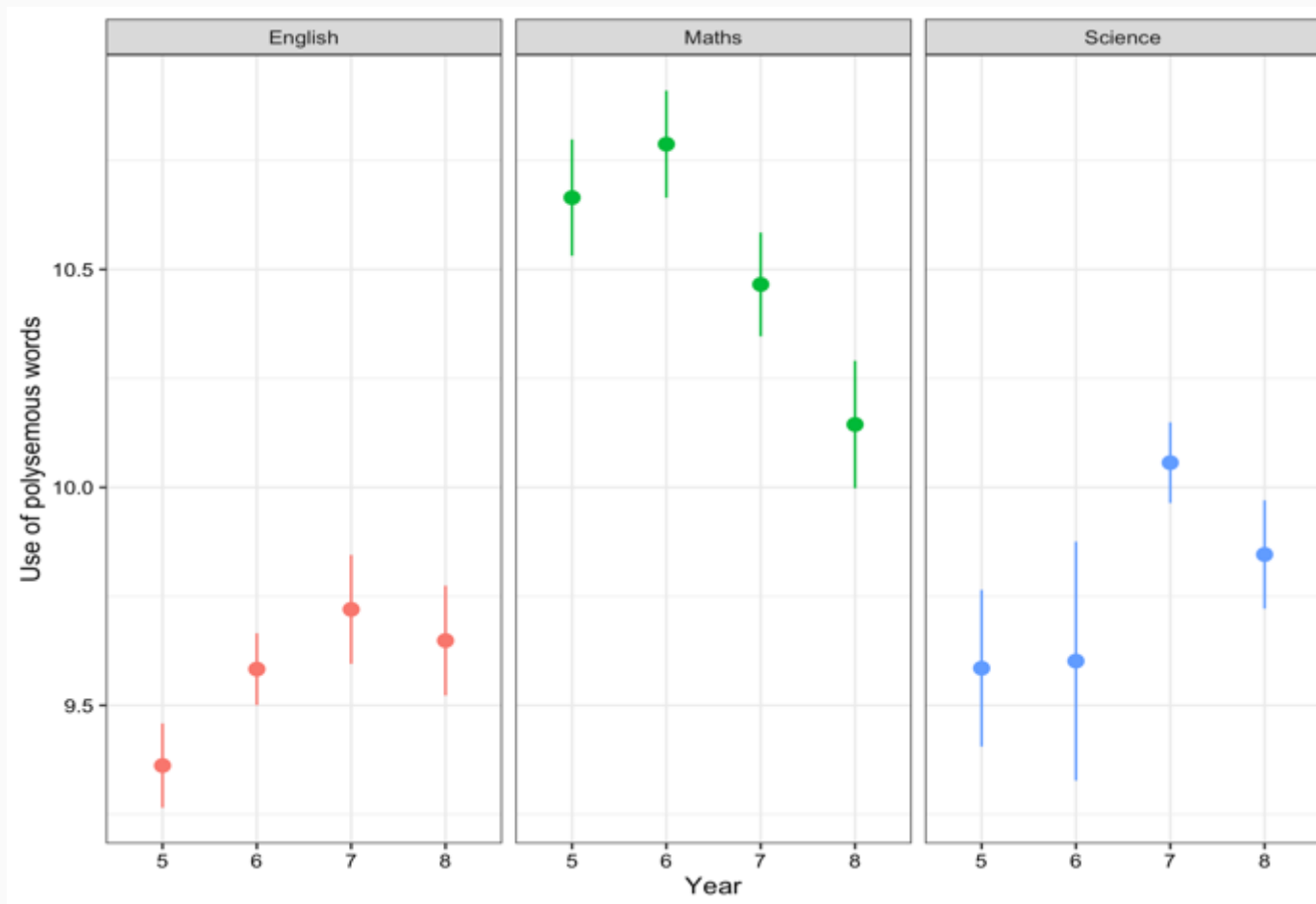
I can **demonstrate** understanding of how ideas about **evolution** developed over time.

I can explain the terms **adaptation, evolution** and natural **selection**.

I can **identify evidence** for **evolution** from fossil records.

I can examine fossil **evidence**.

Use of polysemous words



As the use of polysemous words increases, the written resources become lexically more sophisticated.

Example text extracts

Least lexically sophisticated (fewer polysemous words)

There are 30 children in the class. On Monday one tenth of the class was absent. What percentage is that? (Year 5_maths)

The most lexically sophisticated (use of polysemous words)

Using coloured filters

Filters let certain colours of light pass through, but absorb all other colours. (Year 8_science)

Conclusion

- The different components of lexical sophistication give valuable information on the construct of lexical sophistication.
- Subjects differ across the Key Stages in terms of lexical sophistication.
- The use of everyday vocabulary overall decreases from Key Stage 2 (KS2) to Key Stage 3 (KS3) for all subjects, especially for English and science subjects.
- The use of academic vocabulary increases from KS2 to KS3 for English and maths subjects.
- The use of polysemous words increases from KS2 to KS3 for science and English subjects.

Implications

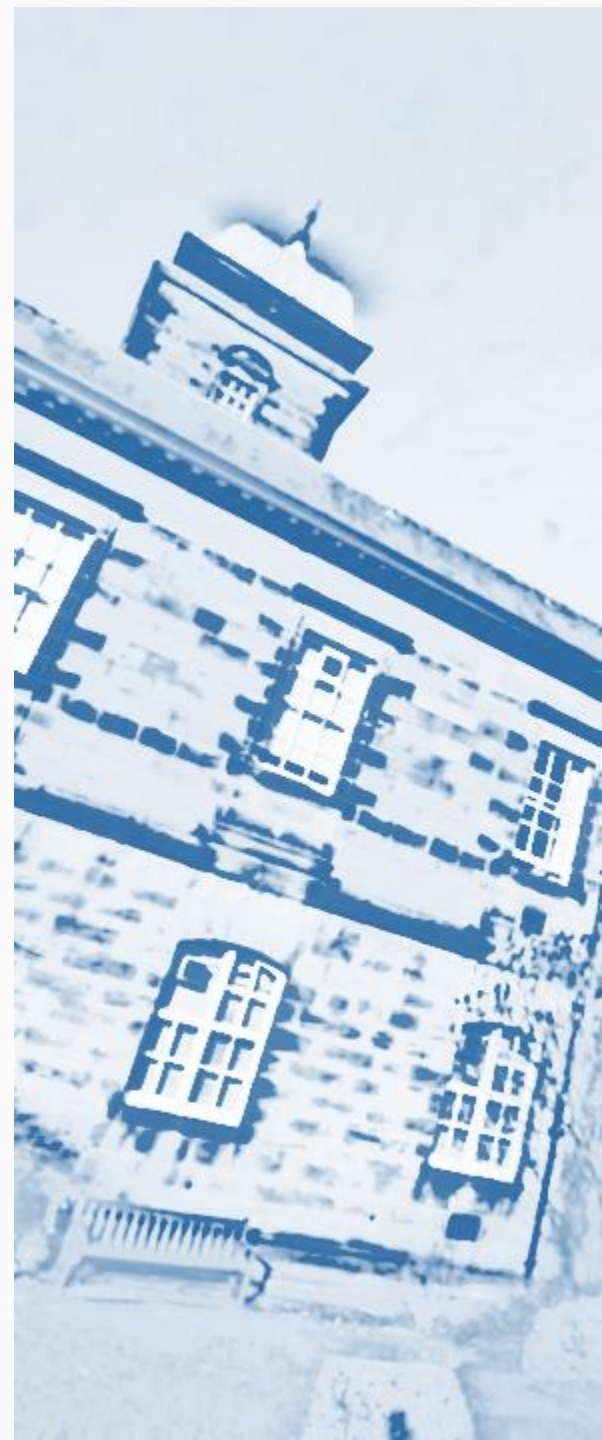
- The science subject has the heaviest load of technical vocabulary (opposite of everyday vocabulary) overall, and this load increases at KS3 in all three subjects. Increased literacy of technical vocabulary/low-frequency words is necessary at KS3 (see Daborn, Zacharias, & Crichton, 2020).
- The use of academic word lists and subject-specific words lists may be useful for English and maths subjects.
- Understanding of the meaning of different word senses is very important for the science subject at KS3 and somewhat important for the English subject at KS3.

Thank you for listening!

<https://linguistictransition.leeds.ac.uk/>

A.H.Deignan@education.leeds.ac.uk

dcandarli001@dundee.ac.uk



References

- Baker, P., Gabrielatos, C., McEnery, T. (2013) *Discourse analysis and media attitudes: The representation of Islam in the British press*. Cambridge: Cambridge University Press.
- Brezina, V., Gablasova, D. (2015) Is there a core general vocabulary? Introducing the *New General Service List*, *Applied Linguistics*
- 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.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34, 213–238.
- Daborn, E., Zacharias, S., & Crichton, H. (2020). *Subject Literacy in Culturally Diverse Secondary Schools: Supporting EAL Learners*. London: Bloomsbury Publishing.
- Gabrielatos, C. (2018) Keyness analysis: nature, metrics and techniques. In Taylor, C., Marchi, A. (eds.) *Corpus Approaches to Discourse: A critical review*. Routledge.
- Higgins, S., Katsipataki, M., & Coleman, R. (2014). Reading at the transition: Interim evidence brief. Education Endowment Foundation: https://educationendowmentfoundation.org.uk/public/files/Publications/EEF_Interim_Evidence_Brief_ReadingAtTheTransition.pdf
- Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*, 49, 757–786.
- Read, J. (2000). *Assessing Vocabulary*. Cambridge: Cambridge University Press.
- West, P., Sweeting, H., Young, R. (2010). Transition matters: Pupils' experiences of the primary- secondary transition in the West of Scotland and consequences for well-being and attainment. *Research Papers in Education*, 25(1), 21-50.