



# The linguistic challenges of the transition from primary to secondary school: challenges in corpus design

Dr Duygu Candarli, Dr Robbie Love & Professor Alice Deignan

School of Education  
University of Leeds

[d.candarli@leeds.ac.uk](mailto:d.candarli@leeds.ac.uk)

[r.love@leeds.ac.uk](mailto:r.love@leeds.ac.uk)

[a.h.deignan@education.leeds.ac.uk](mailto:a.h.deignan@education.leeds.ac.uk)

 @duygucandarli

 @lovermob

 @alicedeignan



Lancaster  
University



CAMBRIDGE  
UNIVERSITY PRESS

# Plan

---



UNIVERSITY OF LEEDS

- England's school system
- Background to the project
- Research questions
- Challenges in corpus design and data collection
- Discussion

# Background: English state schools

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	
<hr style="border-top: 1px dashed red;"/>			
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	

# Background: the project

---



UNIVERSITY OF LEEDS

- Many students in England find the transition from primary to secondary school difficult (DfE 2011; Howe & Richards 2011; Evangelou et al. 2008; Brooks 2016)
  - social reasons
  - larger school
  - change in friendship groups
  - being the youngest again
  - tougher academic demands
  - **tougher linguistic demands?**

# Academic school language

---



UNIVERSITY OF LEEDS

- The language of school comprises numerous registers, not all of them academic
- Our focus is on academic registers = ‘school language’ i.e. the **instructional / regulative** registers of academic learning
- Comprises overlapping subject-specific registers (Christie 2002; Christie & Derewianka 2008)

*academic activities are associated with a kind of language that is different from that used in everyday activities*

(Leung 2014: 137)

# Academic school language

---



UNIVERSITY OF LEEDS

- Students' writing has been extensively researched in the UK education contexts (e.g. Durrant & Brenchley 2018; Nesi & Gardner 2012)
- We are interested in the language that students **receive** in academic contexts at school
- What are they expected to **understand** and **respond to** in order to access the curriculum?

# School language & the transition

---



UNIVERSITY OF LEEDS

- Dip in attainment at start of Key Stage 3
- Difference in language as a barrier to understanding the curriculum?

*[t]eaching environments [...] and teachers' language are very different in secondary schools from primary schools*

(Braund & Driver 2005: 78)

*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.*

Interview with history teacher

# What might be unfamiliar at KS3?

---



A few ideas...

- academic and semi-technical words
- multiword verbs
- grammatical metaphor
- tendency to compress information into complex noun groups
- use of passive voice
- subordination and other complex embedding
- unfamiliar discourse structures in both speaking and writing



# Research questions

---



UNIVERSITY OF LEEDS

RQ1: What are the linguistic characteristics of texts that students are required to understand and respond to at **Key Stage 2**, in terms of lexis, grammar and discourse?

RQ2: What are the linguistic characteristics of texts that students are required to understand and respond to at **Key Stage 3**, in terms of lexis, grammar and discourse?

RQ3: How does the language of Key Stage 3 vary according to **subject area**?

RQ4: How does the language of Key Stage 3 differ from the **language students have previously encountered**, at the levels of lexis, grammar and discourse?

RQ5: How do teachers and students **perceive** the linguistic challenges of the transition from primary to secondary school?

# The project

---



UNIVERSITY OF LEEDS

- Linguistic challenges of the transition from primary to secondary school
- ESRC-funded, 2018-2021

## University of Leeds

- Alice Deignan (PI)
- Gary Chambers (Co-I), Michael Inglis (Co-I)
- Duygu Candarli (RF), Robbie Love (RA)

## Lancaster University

- Elena Semino (Co-I), Vaclav Brezina (Co-I)

## Advisors/consultants

- Niall Curry (CUP), Marcus Jones (Huntington School), Constant Leung (KCL)

# Project plans

---



UNIVERSITY OF LEEDS

- the first comprehensive and systematic description of the academic registers of secondary school
- with focus on how they differ from primary school and non-specialist language outside the school
- engaging with a range of schools in England
- building corpora of spoken and written academic language that students encounter at end of KS2 and start of KS3
- corpus analysis to compare to each other, & across subjects, & to reference corpora of British English, including the BNC2014
- plus, interviews with students / teachers about their views on the transition



# Challenges in design and data collection

# School recruitment

---



UNIVERSITY OF LEEDS

So far...11 schools

## **Yorkshire**

- 3 secondary schools, 4 primary schools

## **Newcastle**

- 1 secondary school, 3 primary schools

Things at schools change far quicker than in the academy...

# Corpus design - background

---



UNIVERSITY OF LEEDS

- Interviews with school teachers
- Curriculum (timetable and distribution of lesson time)
- Literature review
  
- Representativeness - what are we trying to sample?

## Written corpora (Key Stage 2 and Key Stage 3)

- Teacher-designed worksheets
- Textbooks
- Exams, rubrics
- PowerPoint presentations
- Vocabulary/glossary booklets
- Web-based exercises (maths)

## Spoken corpora (Key Stage 2 and Key Stage 3)

- Audio recordings of lessons

Subjects: English, maths, sciences, history, geography

# Representativeness

---



UNIVERSITY OF LEEDS

## **Science lessons in secondary schools**

Textbooks 20% - sampling the textbooks

Worksheets 60%

Marking rubrics 20%

## **English lessons in secondary schools**

Fiction – Of Mice and Men, Private Peaceful



# Representativeness

---



- Text length
- Size of the corpus – Key Stage 2 and Key Stage 3
- Short texts (< 100 words at Key Stage 2)
- Written vs spoken
- Other variables that we cannot control

# Example texts



## Year 5 (156 words)

OLO- To identify features of a News report.	Underline feature an example of each.
Headline.	
Hook. (Who, What, Where, Why, When)	
Factual Language	
Quotes- reported speech in " "	
Descriptive vocabulary- carefully chosen; adjectives, verbs and adverbs.	
Time conjunctions.	

OLO- To identify features of a News report.	Underline feature an example of each.
Headline.	
Hook. (Who, What, Where, Why, When)	
Factual Language	
Quotes- reported speech in " "	
Descriptive vocabulary- carefully chosen; adjectives, verbs and adverbs.	
Time conjunctions.	

OLO- To identify features of a News report.	Underline feature an example of each.
Headline.	
Hook. (Who, What, Where, Why, When)	
Factual Language	
Quotes- reported speech in " "	
Descriptive vocabulary- carefully chosen; adjectives, verbs and adverbs.	
Time conjunctions.	

OLO- To identify features of a News report.	Underline feature an example of each.
Headline.	
Hook. (Who, What, Where, Why, When)	
Factual Language	
Quotes- reported speech in " "	
Descriptive vocabulary- carefully chosen; adjectives, verbs and adverbs.	
Time conjunctions.	

## Year 6 (355 words)

<b>Tuesday 8<sup>th</sup> January 2019</b>		
<b>O.L.O. To write a successful description</b>		
Please use the image of the Ancient Child to inspire you to write a descriptive paragraph based on this mysterious figure.		
Who is he? What does he look/sound/act like? Where is he? Why is he here?		
<u>I have successfully used:</u>	Writers	Checkers
6. Consistent tense and subject-verb agreement		
7. Conjunctions and fronted adverbials to show cohesion and chronology		
8. Passive voice		

<b>Tuesday 8<sup>th</sup> January 2019</b>		
<b>O.L.O. To write a successful description</b>		
Please use the image of the Ancient Child to inspire you to write a descriptive paragraph based on this mysterious figure.		
Who is he? What does he look/sound/act like? Where is he? Why is he here?		
<u>I have successfully used:</u>	Writers	Checkers
6. Consistent tense and subject-verb agreement		
7. Conjunctions and fronted adverbials to show cohesion and chronology		
8. Passive voice		

<b>Tuesday 8<sup>th</sup> January 2019</b>		
<b>O.L.O. To write a successful description</b>		
Please use the image of the Ancient Child to inspire you to write a descriptive paragraph based on this mysterious figure.		
Who is he? What does he look/sound/act like? Where is he? Why is he here?		
<u>I have successfully used:</u>	Writers	Checkers
6. Consistent tense and subject-verb agreement		
7. Conjunctions and fronted adverbials to show cohesion and chronology		
8. Passive voice		

# Example texts



## Year 7 ( > 3000 words)

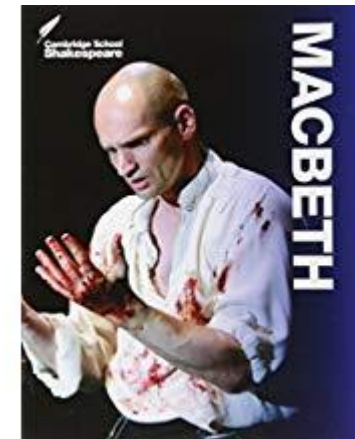
### Macbeth: Key Scenes – How Does Macbeth Change Throughout the Play?

- 1) **Act 1 Scene 2:** Macbeth established as a hero
- 2) **Act 1 Scene 3:** Macbeth's reaction to the witches' prophecy
- 3) **Act 1 Scene 7:** Macbeth reconsiders the plot to kill Duncan; ultimately he is persuaded by his wife
- 4) **Act 3 Scene 4:** Macbeth imagines he sees the ghost of murdered Banquo
- 5) **Act 4 Scene 1:** Macbeth receives three apparitions from the witches
- 6) **Act 5 Scene 8:** Macbeth fights Macduff to the death

### Act 1

*These quotations about Macbeth come early on in the play. For each, unpick what you think they mean and what impression of Macbeth we get from the outset.*

1. For brave Macbeth – well he deserves that name –  
Disdaining fortune, with his brandish'd steel,  
Which smoked with bloody execution,  
Like valour's minion carved out his passage  
Till he faced the slave;  
...And fix'd his head upon our battlements.
2. O valiant cousin! worthy gentleman!
3. 'eagle(s)' 'lion'
4. Doubly redoubled strokes upon the foe
5. No more that thane of Cawdor shall deceive  
Our bosom interest: go pronounce his death,  
And with his former title greet Macbeth
6. What he hath lost noble Macbeth hath won.
7. every one did bear/Thy praises in his kingdom's great defense (1.3)
8. most worthy thane! (1.3)



# Example texts



## Year 8 ( > 1000 words)

**Year 8**  
**'Of Mice and Men'**  
*End of Unit Quiz*

### Lennie

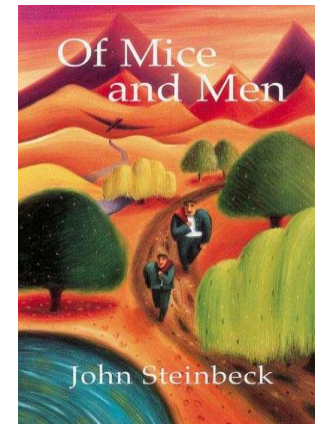
**1. Complete the following key quotation:**

"An' live off the fatta \_\_\_\_\_," Lennie shouted. 'An' have \_\_\_\_\_. Go on, George! Tell about what we're gonna have in the garden and the \_\_\_\_\_ in the cages and about the rain in the winter and the stove . . .' p.

**2. What were migrant workers and why was their life a hard one?**

**3. What does this key quotation tell us about Lennie's attitude to George?**

"An' I got you. We got each other, that's what, that gives a hoot in hell about us,' Lennie cried in triumph. P.103



# Early findings - 1

## The keywords in Maths written sub-corpus at Key Stage 3 in comparison to Key Stage 2

Single-words	Keyness	Effect size
mean	+71.96	0.01
data	+67.8	0.01
median	+56.02	0.009
mode	+52.56	0.008
pie	+43.55	0.007
factors	+39.4	0.006
range	+34.56	0.005
multiple	+31.79	0.005
prime	+29.02	0.005
average	+28.33	0.004

# Keywords at Key Stage 3

---

## Data

Each group has a statement, a set of **data** and one extra piece of data to make the statement true. (KS3\_M8\_3)

Match these types of **data** to their meanings.

Primary **data**

Secondary **data**.... (KS3\_M8\_8)

# Early findings - 2

## The key multi-words in Maths written sub-corpus at Key Stage 3 in comparison to Key Stage 2

Multi-words	Score
pie chart	+1299.83
prime number	+547.88
bar chart	+513.7
same time	+342.8
total number	+308.62
ascending order	+240.26
perfect number	+171.9
square root	+171.9
frequency table	+171.9
high degree	+137.72

# Keywords at Key Stage 3

---

## The same time

Challenge! Two Formula 1 cars race around a track. The first car takes 54 seconds to complete a lap. The second car is slower and takes 63 seconds. After how many seconds will they be at the starting place at exactly **the same time**? (KS3\_M7\_12)

A green light flashes every 8 seconds, a red light flashes every 15 seconds. After how many seconds will they both flash at **the same time**? (KS3\_M7\_9)



- Does the transition from KS2 and KS3 present language barriers which prevent students from accessing the curriculum?
- We believe that the transition to secondary school seems to involve a step change in academic language, which is likely to be especially difficult for children from lower SES backgrounds
- We aim to investigate this in order to inform the design of more accessible curricula for all students



- Representativeness, representativeness, representativeness...
  - Designing a sampling frame is much more complicated than we first expected
- Responsiveness to changes at schools
  - Schools are very keen, but we have learned to expect the unexpected
- Communication & avoiding intrusion

# Thank you



UNIVERSITY OF LEEDS

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

[d.candarli@leeds.ac.uk](mailto:d.candarli@leeds.ac.uk)  
@duygucandarli

[r.love@leeds.ac.uk](mailto:r.love@leeds.ac.uk)  
@lovermob

The screenshot shows a website page with a dark header containing the University of Leeds logo and a search icon. Below the header is a red navigation bar with the title 'Linguistic Challenges of the Transition from Primary to Secondary School' and a search icon. A light grey navigation bar below that contains links for HOME, AIMS, RESEARCH TEAM, PARTNERS, PUBLICATIONS, BLOG, and CONTACT. The main content area features a large image of an open book with a red cover and white pages. Below the image is a paragraph of text: 'We are an ESRC-funded research project based at the School of Education at the University of Leeds, working in partnership with Lancaster University's Centre for Corpus Approaches to Social Science. Led by Professor Alice Deignan, we aim to use methods from corpus linguistics to address a pressing language problem in English schools.' This is followed by two more paragraphs of text. At the bottom of the page, there is a Twitter icon with the text 'Follow us on Twitter!' and two logos: the E·S·R·C Economic & Social Research Council logo and the Lancaster University logo.

# References



UNIVERSITY OF LEEDS

- 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
- 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.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34, 213–238.
- Cummins, J. (1980). The cross-lingual dimensions of language proficiency: implications for bilingual education and the optimal age issue. *TESOL Quarterly*, 14(2), 175-187.
- Cummins, J. (2008). BICS and CALP: Empirical and theoretical status of the distinction. In Street, B. & Hornberger, N. H. (Eds.). (2008). *Encyclopedia of Language and Education*, 2nd Edition, Volume 2: Literacy. (pp. 71-83). New York: Springer Science & Business Media LLC.
- Cummins, J. (2014). Beyond Language: Academic communication and student success. *Linguistics and Education*, 26, 145-154
- Deignan, A., Semino, E., & Paul, S. (2017). Metaphors of climate science in three genres : research articles, educational texts, and secondary school student talk. *Applied Linguistics*, 1–26.
- Department for Education. (2011). *How do pupils progress between Key Stages 3 and Research Report*.
- Durrant, P., & Brenchley, M. (2018). Development of vocabulary sophistication across genres in English children's writing. *Reading and Writing*.
- Howe, A., & Richards, V. (2011). *Bridging the transition from primary to secondary school*. Abingdon: Routledge.
- Kyle, K., Crossley, S. A., & Berger, C. (2018). The tool for the analysis of lexical sophistication (TAALES): Version 2.0. *Behavior Research Methods*, 50, 1030-1046.
- Leung, C. (2014). Researching language and communication in schooling. *Linguistics in Education*, 26, 136-144.
- Love, R., Dembry, C., Hardie, A., Brezina, V., & McEnery, T. (2017). The Spoken BNC2014: designing and building a spoken corpus of everyday conversations. In *International Journal of Corpus Linguistics*, 22(3), 319-344.
- Nesi, H., & Gardner, S. (2012). *Genres across the disciplines: Student writing in higher education*. Cambridge: Cambridge University Press.
- Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31, 487–512.